



HAL
open science

Structuration automatique de flux télévisuels

Jean-Philippe Poli

► **To cite this version:**

Jean-Philippe Poli. Structuration automatique de flux télévisuels. Interface homme-machine [cs.HC]. Université Paul Cézanne - Aix-Marseille III, 2007. Français. NNT: . tel-00207960

HAL Id: tel-00207960

<https://theses.hal.science/tel-00207960>

Submitted on 18 Jan 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Laboratoire des Sciences de l'Information
et des Systèmes



Institut National de l'Audiovisuel

THÈSE

Structuration automatique de flux télévisuels

Présentée et soutenue publiquement le :

29 Mai 2007

Par

Jean-Philippe Poli

Pour l'obtention du :

Doctorat de l'Université Paul Cézanne – Aix-Marseille III
(spécialité informatique)

Composition du jury

Président du jury : Bernard Merialdo

Rapporteurs : Patrick Gros
Bernard Merialdo

Examineurs : Philippe Joly
François Pachet

Directeur de thèse : Philippe Jégou

Encadrant INA : Jean Carrive

Mis en page avec la classe thloria.

Remerciements

Je tiens à remercier l'ensemble des membres du jury pour leur attention, leurs disponibilités respectives et leurs remarques, et particulièrement les rapporteurs pour leur travail qui m'a permis d'améliorer le manuscrit.

La particularité d'une thèse CIFRE est de côtoyer deux mondes différents - dans mon cas, l'INA et le LSIS - et par conséquent les remerciements sont nombreux.

Je commence avec l'INA, en remerciant Jean pour ses précieux conseils pendant ces trois ans, dont j'espère avoir su apprécier toute la valeur, et pour avoir été une oreille attentive aux divers problèmes de différent ordre que j'ai pu rencontrer pendant cette période. Merci de m'avoir laissé toute l'autonomie dont j'avais besoin tout en m'encourageant dans mes choix. Je remercie Vincent de m'avoir offert l'opportunité d'une part de mon stage de DEA puis de ma thèse, sans compter son aide précieuse dans les tâches les plus pénibles. J'aimerais remercier Valérie à qui je dois toutes mes connaissances sur la télévision et qui m'a ouvert les yeux sur un monde fascinant ; tes cours ont été une véritable valeur ajoutée à ma thèse. Dans le même ordre, je remercie Régine Chaniac pour sa connaissance sur Médiamétrie et les grilles de programmes. Je remercie Laurent pour tous ses développements dans l'urgence : j'espère ne pas avoir été trop insistant et pressant parfois. Je délivre une mention particulière à Odile, ma confidente, pour son assistance dans bien des tâches, parfois même d'une manière un peu maternelle. Merci à Marie-Luce pour avoir été une oreille attentive et avoir, entre autres, relu ma thèse. Je remercie la bande du midi, incontournable, dont j'ai apprécié, surtout ces derniers mois, les conversations légères et agréables permettant une vraie coupure dans la journée : Thomas, Raphaël, Antoine, Véronique, Rémi, Jérôme, Marc, Nicolas, Ludovic, Hervé, Jérémy, Quentin et Jean-Pascal. Joyeux drilles, je ne désespère pas vous traîner un jour à Disneyland. Au détour des couloirs, d'un bureau ou d'une salle de réunion, j'ai pu rencontrer beaucoup de monde à l'INA, qui par une réflexion, dans les différents sens du terme, ont apporté de près ou de loin quelque chose à mon travail : Olivier, Bruno, Jean-Hugues, Louis C., Louis L., Steffen et Fabrice.

Je poursuis par le LSIS en remerciant Philippe pour ses précieux conseils qu'il m'a prodigué depuis bientôt huit ans et pour si bien gérer cette équipe dans laquelle il est agréable de venir passer des séjours, même courts. Merci aux doctorants de l'équipe que j'ai pu rencontrer lors de mes visites ainsi qu'à Gwladys. Une mention toute particulière pour Mathieu, qui s'est fait un défi de m'intégrer malgré mon immense timidité : merci Mathieu pour avoir transformé mes visites à Marseille. Comment ne pas remercier Geoffroy avec qui j'ai eu la chance d'assister à ICTAI et de visiter Hong-Kong ; dommage pour Washington.

Cependant, les collègues ne sont pas les seuls garants de l'équilibre d'un doctorant. Je remercie ma famille de m'avoir supporté pendant toutes ces années et d'avoir accepté tout ce que j'avais à leur dire, et même après mon départ à Paris : il est tellement plus facile de se rendre compte que l'on tient aux gens lorsqu'ils sont loin. Je remercie tout particulièrement ceux qui ont pu faire le déplacement pour assister à la soutenance : ma mère, Francine, Danielle et Cyril. Au fil des années, il arrive qu'on se découvre une nouvelle famille ; je remercie mes drôles de dames : Annabel, Hélène, Laurence et Mélanie, pour être là dans toutes les situations. Je remercie énormément Gaby et Jean, Franck et Marie, Pascal, Laurent pour tous ses souvenirs que nous continuons à construire. Merci à toi Alexandre T. pour être là, toujours, quoiqu'il arrive, sauf à la salle de sport. Merci à vous Alexandre R., Isabelle, Jean-Guy, Christian, Laurent,

Eric, Sébastien pour pimenter ma vie parisienne. Stéphane, merci d'avoir été ce patron ouvert d'esprit qui m'a permis de financer une partie de mes études : nos discussions ont toujours été enrichissantes et je t'en suis reconnaissant.

Table des matières

Table des figures	vii
Liste des tableaux	ix
Liste des algorithmes	xi
Introduction générale	1

Partie I Positionnement

Chapitre 1 Contexte industriel	7
1 L'Institut national de l'audiovisuel	7
2 L'Inathèque de France et le dépôt légal	8
3 La direction de la recherche et de l'expérimentation	10
Chapitre 2 Structuration et classification de vidéos	13
1 Structuration de documents vidéos	14
2 Détection de coupures publicitaires	19
3 Classification en genres	20
4 Travaux du projet Demi-Ton	22

Chapitre 3 Proposition d'une approche	23
1 Flux et documents télévisuels	23
2 Motivation des travaux	24
3 Objectifs de l'approche	24
4 Inadéquation des méthodes de structuration de documents vidéo	25
5 Présentation du système SAFARI	26

Partie II La programmation d'une chaîne de télévision

Introduction	35
Chapitre 4 Évolution de la télévision française	37
1 La télévision de l'offre	37
2 La télévision de la demande	38
3 La télévision aujourd'hui	40
Chapitre 5 La programmation d'une chaîne	43
1 Programmation et grilles de programmes	43
2 Construction d'une grille de programmes	45
3 Principes et règles de la programmation	48
4 Impacts de différents facteurs sur les grilles de programmes	50
5 Les différentes grilles de programmes	52
6 Grilles prévisionnelles et guides de programmes	52
7 Stabilité des grilles de programmes	56
Chapitre 6 Taxinomie des programmes	57
1 Définitions des genres télévisuels	58
2 Taxinomie des genres	66

Partie III Approche markovienne pour la modélisation des grilles de programmes

Introduction	77
Chapitre 7 Modélisation des grilles de programmes	79
1 Modèles de Markov	79
2 Modèles de Markov cachés contextuels	85
3 Application à la modélisation des grilles de programmes	99
4 Conclusion	104
Chapitre 8 Régression des durées des émissions	105
1 Régression par des méthodes mathématiques simples	106
2 Régression par des méthodes informatiques	108
3 Application des arbres de régression à la prédiction des durées des émissions	112
4 Dédution de la probabilité des durées	118
5 Conclusion	118
Chapitre 9 Prédiction de grilles de programmes	121
1 Fusion des prédictions et du guide des programmes	122
2 Réduction du graphe des possibilités	125
3 Zones d'incertitude	126
4 Étiquetage des émissions	127
5 Conclusion	127
Chapitre 10 Évaluation du modèle et des prédictions	129
1 Environnements d'apprentissage	130
2 Évaluation des arbres de régression	131
3 Évaluation de la prédiction des grilles	135
4 Conclusion	149

Partie IV Expérimentation du système de structuration

Introduction	155
Chapitre 11 Alignement des grilles prédites sur le flux	159
1 Détection des transitions	159
2 Identification d'un programme	168
3 Conclusion	172
Chapitre 12 Structuration d'émissions et de tranches horaires	173
1 Structuration d'une tranche horaire	173
2 Isolation des plateaux dans un journal télévisé	182
3 Conclusion	184

Conclusion

Conclusion	191
-------------------	------------

Annexes

Annexe A Nomenclature de Médiamétrie	197
Annexe B Environnement d'apprentissage utilisé	201
Annexe C Résultat de la structuration du mardi 25 octobre sur France 2	203
Annexe D Résultat de la structuration d'un <i>Morning Café</i>	205
Annexe E Résultat de la structuration d'un journal de 20H sur France 2	207
Glossaire	209
Bibliographie	211

Table des figures

1	Heures de programmation conservées par les archives professionnelles et l'In-athèque au 31 décembre 2004	8
2	Nombre de chaînes de télévision et de radios concernées par le dépôt légal	9
3	Volume d'heures captées par an dans le cadre du dépôt légal	10
4	Représentation hiérarchique d'un document vidéo	14
5	Exemples de transitions entre deux plans	15
6	Structure d'une coupure publicitaire pendant un film en Allemagne	20
7	Représentation de la structure hiérarchique d'un flux télévisuel	25
8	Schéma général du système SAFARI	26
9	Extrait d'un arbre représentant les séquences possibles de types de programmes	29
10	Extrait de la grille de programmes de France 2 de janvier à juin 2006, fournie à ses annonceurs	45
11	Évolution du coût de la grille de programmes de TF1 entre 1995 et 2005	49
12	Décalages entre le guide de programmes et la diffusion	55
13	Répartition des genres en fonction des chaînes de télévision en 2000	58
14	L'émission de <i>call-TV Tubissimo</i> , sur M6	65
15	Extrait de l'ontologie de l'audiovisuel	68
16	Extrait de la nomenclature de Médiamétrie	69
17	Représentation graphique d'une chaîne de Markov	81
18	Représentation graphique d'un modèle de Markov caché à trois états	82
19	Hiérarchie des extensions des modèles de Markov cachés	84
20	Représentation de champs de Markov en 4 et 8-connexité	85
21	Production ou classification d'une séquence d'observations avec un CHMM	88
22	Comparaison des réseaux bayésiens d'un CHMM et d'un HMM	89
23	Treillis représentant les calculs de la procédure <i>forward</i> pour un HMM	90
24	Exemple jouet des mêmes transitions avec un HMM et un CHMM	100
25	Durées (en secondes) des magazines en fonction de leur heure de diffusion, sur France 2 en 2004	107
26	Structure d'un neurone formel	108
27	Structure d'un réseau de neurones multicouche	109
28	Exemple de réseau de neurones récurrent	110
29	Exemple d'arbre de régression	111
30	Extrait de l'arbre de régression de TF1 entraîné sur les émissions de 2004	116

31	Extrait de l'arbre de régression de France 2 entraîné sur les émissions de 2004 . . .	116
32	Distribution des durées sur une feuille de l'arbre de décision	119
33	Schéma représentatif d'un arbre des grilles possibles avant et après l'application de la deuxième heuristique. La fusion des différents nœuds cibles diminue le temps de construction de l'arbre.	125
34	Illustration d'une zone d'incertitude	126
35	Taux d'erreur en fonction du paramètre σ pour les chaînes hertziennes	134
36	Nombre de nœuds dans l'arbre de régression pour les chaînes hertziennes en fonction du paramètre σ	134
37	Extrait du graphe acyclique dirigé représentant les grilles de programmes possibles du 10 Mai 2005 sur TF1	139
38	Première étape : construction du graphe entre les deux premières émissions du guide de programmes	140
39	Deuxième étape : application de la deuxième heuristique sur le graphe	140
40	Troisième étape : développement du graphe jusqu'à l'obtention des nœuds représentant l'émission de télé-achat	141
41	Quatrième étape : élimination des chemins impossibles qui n'atteignent pas l'émission de télé-achat au bout de 30 minutes	142
42	Cinquième étape : application de la première heuristique sur le graphe	143
43	Sixième étape : application de la deuxième heuristique sur le graphe	144
44	Septième étape : poursuite du développement du graphe jusqu'à la série	145
45	Huitième étape : application de la première heuristique et élimination des chemins impossibles	146
46	Dernière étape : application de la deuxième heuristique sur le graphe	147
47	Structuration du flux télévisuel par suivi d'un chemin dans le graphe prédit . . .	155
48	Détection des images monochromes	161
49	Effet visuel de transition utilisé par France 2	162
50	Exemples de jingles publicitaires	166
51	Architecture du réseau à convolution (Duffner et Garcia, 2006)	170
52	Exemples de logos mal détectés par les réseaux à convolution	172
53	Structure hiérarchique du <i>Morning Café</i>	174
54	Extrait du graphe représentant le début du <i>Morning Café</i>	175
55	Système de reconnaissance des jingles	176
56	Analyse spectrale	176
57	Comparaison du jingle de référence et du flux audio	176
58	Exemple de distances euclidiennes issues de la comparaison d'un jingle avec trois minutes de flux (Pinquier, 2004)	177
59	Distance entre le jingle de référence et le flux audio (Pinquier, 2004)	177
60	Module de classification en clip musical ou non clip	178
61	Apprentissage du système de classification	178
62	Différentes vues du plateau de <i>Morning Café</i>	179
63	Plateaux de différents journaux télévisés	183

Liste des tableaux

1	Comparaison des recettes des chaînes hertziennes de 1980 à 2005	39
2	Évolution des parts de marché annuelles de 1989 à 2005	41
3	Coûts de la grille de programmes de plusieurs chaînes	47
4	Définitions américaines du découpage de la télévision	47
5	Estimation de l'incomplétude des guides de programmes	54
6	Comparaison des volumes de publicité autorisés sur les principales chaînes hertziennes	63
7	Liste des genres possibles pour la télévision	66
8	Liste des thèmes possibles pour la télévision	67
9	Comparaison entre la typologie de l'INA et la nomenclature de Médiamétrie sur quelques programmes	70
10	Exemple de grille de programmes	103
11	Coefficients de corrélation de Pearson des tests et des durées	114
12	Exemples de durées prédites pour quelques émissions de TF1 en 2005	117
13	Extraits de la grille prévisionnelle et de la grille de programmes du 21 Septembre 2005 sur France 2	122
14	Taux d'erreur et envergure des arbres de régression des chaînes hertziennes	131
15	Classement des genres télévisuels en fonction de leur prédictibilité	132
16	Précision temporelle des arbres de régression sur les chaînes hertziennes	133
17	Nombre de journées entières de 2005 exprimables à l'aide du CHMM entraîné sur 2004	135
18	Nombre de tranches horaires entières de 2005 exprimables à l'aide du CHMM entraîné sur 2004	137
19	Extrait du guide de programmes du 10 mai 2005 sur TF1	138
20	Évaluation de la détection des « cuts » pour la localisation de programmes	160
21	Évaluation de la détection des images monochromes pour la localisation de programmes	161
22	Évaluation de la détection des silences pour la localisation de programmes	162
23	Évaluation de la détection simultanée des silences et des images monochromes pour la localisation de programmes	163
24	Évaluation de la détection de « cuts » précédés de silence pour la localisation de programmes	163

25	Nombre de fausses alarmes pour les détections locales	164
26	Règles utilisées pour la détection de France 2	165
27	Correction et précision temporelle des structures produites automatiquement . .	167
28	Récapitulatif des résultats obtenus sur les quatre logos	171
29	Règles utilisées pour détecter les transitions	180
30	Taux de recouvrement de segments détectés automatiquement sur les segments annotés manuellement en fonction du nombre de gaussiennes et du nombre de composantes utilisées	180
31	Évaluation de la reconnaissance des jingles	181
32	Correction et précision temporelle des structures produites automatiquement . .	181
33	Correction, fausses alarmes et précision temporelle des structures obtenues	184

Liste des algorithmes

1	Algorithme d'évaluation des probabilités	92
2	Algorithme de Viterbi contextuel	96
3	Algorithme d'estimation des paramètres	103
4	Algorithme d'induction d'un arbre de régression	113
5	Algorithme de développement complet du graphe	123

Introduction générale

Contexte

La recherche présentée dans ce manuscrit s'est faite dans le cadre d'une convention CIFRE (convention industrielle de formation par la recherche). Elle s'est principalement déroulée à la direction de la recherche et de l'expérimentation de l'Institut national de l'audiovisuel¹ (INA), dans le thème « description des contenus audiovisuels », en collaboration avec le laboratoire des sciences de l'information et des systèmes² (UMR CNRS 6168). Ces travaux qui ont été menés de janvier 2004 à décembre 2006 interviennent dans un contexte applicatif fort : le dépôt légal de la télévision géré par l'Inathèque de France.

Le dépôt légal de la télévision est une mission patrimoniale qui consiste à archiver toutes les émissions françaises diffusées à la télévision. Ces émissions doivent ensuite être indexées de façon à faciliter une recherche ultérieure. Une étape préalable à leur indexation est la structuration des flux télévisuels. Cette tâche consiste à cataloguer les différentes émissions qui composent le flux.

Les travaux que nous présentons dans ce manuscrit ont pour but de décrire une approche au problème de la structuration automatique des flux télévisuels.

Problématique de recherche

La structuration automatique d'un flux télévisuel est un sujet qui, à notre connaissance, a peu été abordé par la communauté d'indexation automatique de contenus audiovisuels. En revanche, de nombreuses recherches concernent la structuration de documents audiovisuels. La structuration d'un document ou d'un flux est le processus qui consiste à déterminer un horodatage précis de ses composantes. Par exemple, les composantes d'un magazine télévisé sont les génériques, les plateaux, les reportages alors que les composantes d'un flux télévisuel sont les émissions et les publicités. La structuration d'un flux est hiérarchique puisque le flux peut être divisé en émissions et les émissions en composantes plus petites, comme les plateaux.

Nos travaux ont été guidés par le procédé actuel de structuration des flux télévisuels de l'Institut national de l'audiovisuel. Nous proposons ainsi une méthode reposant en grande partie sur des connaissances sur la télévision, comme les grilles de programmes passées d'une chaîne, les génériques et les logos des émissions. Ces connaissances nous donnent une structure hypothétique du flux qui doit être confirmée par des détections locales dans le flux afin d'éviter d'une part les calculs inutiles et d'autre part les fausses alarmes. Ces détections peuvent être, par exemple, des détections de logos ou des outils de reconnaissance des génériques.

¹www.ina.fr

²www.lsis.org

Nous insistons dans ce manuscrit sur la modélisation des grilles de programmes déjà diffusées afin de prédire les prochaines grilles de programmes en les confrontant aux *guides* de programmes – comme les grilles prévisionnelles de la presse spécialisées ou les guides de programmes électroniques. Pour cela, nous proposons une extension contextuelle des modèles de Markov, les modèles de Markov cachés contextuels (CHMM). Cette extension est particulièrement adaptée à la production d’observations et donc en particulier à la prédiction de grilles de programmes. Les durées des émissions sont encadrées à l’aide d’un arbre de régression, créant ainsi de véritables fenêtres temporelles qui nous permettent d’analyser le flux de façon locale. Plus généralement, l’utilisation des CHMM combinés à un arbre de régression est très efficace pour prédire n’importe quelle grille de programmes ou conducteur d’émissions.

Les grilles de programmes prédites sont générées sous la forme d’un graphe acyclique dirigé dont les nœuds représentent une émission par son genre (par exemple, magazine ou journal télévisé), son heure de début et sa durée et chaque arc de ce graphe représente une transition d’un programme à l’autre. Les arcs sont étiquetés par la probabilité de la transition fournie par le CHMM. Ainsi, pour structurer automatiquement un flux, le système que nous proposons parcourt le graphe en commençant par les transitions les plus probables et essaye de trouver dans le flux l’élément caractéristique de ce type de transition. Ces caractéristiques sont fournies par un ensemble de règles. Par exemple, pour détecter la transition d’une émission suivie d’un bulletin météo, la règle consiste à trouver le générique du bulletin.

Cette approche nous permet d’utiliser des détecteurs spécifiques pour chaque type de transition. Puisque le système les exécute localement autour de la fin prédite de l’émission, le nombre de fausses alarmes et le coût de l’algorithme utilisé sont moins pénalisants.

Notre contribution est multiple. Dans un premier temps, nous fournissons des outils formels et des pistes d’algorithmes – dont un algorithme d’entraînement EM (espérance-maximisation) – pour une modélisation statistique contextuelle avec un modèle dérivé des modèles de Markov. Le modèle ainsi décrit, le CHMM, est plus général que certaines extensions contextuelles des modèles de Markov, comme les IOHMM (input-output hidden Markov models). Dans un second temps, nous fournissons une approche originale à la structuration des flux et des documents audiovisuels. L’intérêt de cette approche est d’utiliser des connaissances plus précises sur la structure du document audiovisuel ou du flux afin de pouvoir employer localement des détecteurs spécifiques.

Plan du document

Nous introduisons dans la première partie le contexte industriel de la thèse en présentant l’Institut national de l’audiovisuel et le dépôt légal de la télévision conduit par l’Inathèque de France. Nous présentons ensuite une bibliographie sur la structuration et la classification de documents vidéos avant de présenter l’approche que nous proposons.

La seconde partie est consacrée à l’évolution de la télévision afin d’expliquer les hypothèses faites pendant nos travaux et afin que le lecteur comprenne au mieux les raisonnements que nous avons mis en évidence dans ce manuscrit. Il s’agit d’une partie moins technique, mais qui nous paraît incontournable dans une thèse qui traite du flux télévisuel. C’est notamment dans cette partie que nous distinguerons les différences entre les grilles de programmes, les guides de programmes et les grilles prévisionnelles. Toutefois, ces définitions sont rappelées dans le glossaire (page 209). Nous évoquerons les différents genres télévisuels dans la mesure où nous avons eu besoin de typer les segments obtenus par la structuration. Nous terminerons la seconde partie par la mise en évidence et l’explication de la stabilité des grilles de programmes.

Nous avons consacré la troisième partie à la prédiction des grilles de programmes et à la prédiction des durées des émissions. Nous présenterons différentes extensions des modèles de Markov cachés afin de comparer leur adéquation à la télévision, puis nous décrirons notre extension, appelée modèle de Markov contextuel. Dans ces travaux, les durées des émissions sont encadrées à partir des feuilles d'un arbre de régression. La probabilité de ces durées est estimée à l'aide de lois gaussiennes asymétriques paramétrées par les feuilles de cet arbre de régression. À partir de l'arbre de régression et de l'extension des modèles markoviens que nous proposons, entraînés sur les grilles de programmes passées, nous décrirons comment prédire les prochaines grilles de programmes. Nous verrons aussi comment les guides de programmes permettent de réduire le nombre de grilles prédites. Nous évaluerons ensuite les capacités de ces modèles sur l'ensemble des chaînes hertziennes françaises.

La quatrième partie exposera dans un premier temps les différentes détections (détections de silences et d'images monochromes, reconnaissance de logos) qui rentrent en jeu dans la phase d'alignement des grilles possibles, produites par le modèle décrit dans la troisième partie : ces détections assez simples ne nécessitent pas un temps de calcul élevé. Nous verrons qu'il peut s'avérer utile de vérifier que le programme annoncé dans le guide de programmes est bien celui qui est diffusé : en effet, si une erreur se produit lors de la structuration, cela permet de s'en rendre compte rapidement. Nous mettrons en évidence l'utilité de détecteurs simples et génériques comme un détecteur de silence ou d'images monochromes ainsi que l'utilité de certains détecteurs plus spécifiques comme un détecteur de logo ou une signature des génériques. Après avoir retrouvé le début des émissions isolées, des interprogrammes et des tranches horaires, nous parlerons de la structuration des tranches horaires et des émissions de plateaux. Pour cela, nous prendrons l'exemple d'une tranche horaire matinale, le *Morning café*, et des journaux télévisés. Toutefois, les méthodes employées pour structurer ces deux types d'émission sont différentes. Pour le journal télévisé, nous nous contenterons d'extraire les plateaux avec une méthode très simple de classification des plans. En revanche, pour le *Morning café*, nous montrerons que la méthode utilisée pour le flux peut être appliquée à ce genre de tranches horaires et que de nombreux détecteurs peuvent être guidés pour la structurer.

Enfin, nous terminerons ce manuscrit par une évaluation du travail qui a été effectué et en précisant les tâches qui restent à accomplir afin que le système puisse être utilisé à l'Inathèque de France.

Première partie

Positionnement

Chapitre 1

Contexte industriel

Sommaire

1	L'Institut national de l'audiovisuel	7
2	L'Inathèque de France et le dépôt légal	8
3	La direction de la recherche et de l'expérimentation	10

1 L'Institut national de l'audiovisuel

L'INA est né en 1974 de l'éclatement de l'ORTF (Office de radiodiffusion-télévision française, 1964-1974) en six sociétés nouvelles et indépendantes : trois sociétés chargées de la conception et de la programmation d'émissions télévisées (TF1, Antenne 2 et FR3), une société pour la radio (Radio-France), une pour la production (Société française de production) et la dernière chargée du réseau de télécommunication (Télédiffusion de France). Ainsi le démantèlement de l'ORTF permettait de remédier à son dysfonctionnement en conciliant deux objectifs : le maintien du monopole et l'instauration d'un esprit de compétition. Certains services de l'ORTF ont, à ce moment là, été oubliés. Il s'agit des services des archives, de production de création, et de recherche et de formation. Un amendement est alors voté, établissant la création de l'INA (Hoog, 2006) le 6 janvier 1975.

L'INA est un établissement à caractère industriel et commercial (EPIC) dont l'État fixe le cadre général, législatif, réglementaire et financier. De par les conditions de sa création, des activités concurrentielles et des missions de service public doivent cohabiter au sein de l'Institut, épouser au mieux un secteur audiovisuel en perpétuelle mutation, et faire face au temps qui menace ses archives. Les missions de l'INA peuvent être regroupées en trois catégories :

– **la conservation du patrimoine audiovisuel national**

L'INA, à sa création, a hérité des archives de l'ORTF et doit en assurer la restauration et la préservation. En charge du dépôt légal audiovisuel, l'Institut doit aussi assurer la collecte des programmes audiovisuels français. Toutes ces archives doivent être documentées de façon efficace afin d'en faciliter l'exploitation ;

– **L’exploitation et la mise à disposition de ce patrimoine**

L’Institut doit à la fois valoriser et développer l’exploitation commerciale de ces fonds. Il s’agit, entre autres, des archives de l’ORTF (elles-mêmes constituées des archives de la RTF, Radio-télévision française, de 1949 à 1964), des archives de TF1 jusqu’en 1982, d’Antenne 2 et FR3 jusqu’en 1992, de fonds radiophoniques, de photos et de films. Le fond est encore enrichi par certaines émissions de France 2 et France 3 depuis 1992.

Dans le cadre de sa mission patrimoniale, l’INA doit aussi mettre à disposition ses archives à des fins scientifiques, éducatives, et culturelles. En avril 2006, l’Institut a mis en ligne sur son site³ plus de 10000 heures de radio et de télévision ;

– **L’accompagnement des évolutions du secteur audiovisuel**

Il se fait au travers des activités de recherche, de production et de formation de l’institut. C’est dans le cadre de cette dernière mission que la formation professionnelle s’oriente vers les technologies numériques et que la recherche converge vers les missions patrimoniales de l’INA.

Depuis sa création, l’INA a dû composer avec l’évolution de ses missions et les évolutions technologiques, en particulier l’avènement de l’audiovisuel numérique. Ainsi depuis 1999, l’institut a entrepris la numérisation de ses archives afin de les préserver de la destruction de leurs supports (figure 1) ; en 2015, la totalité du fond sera numérisé.

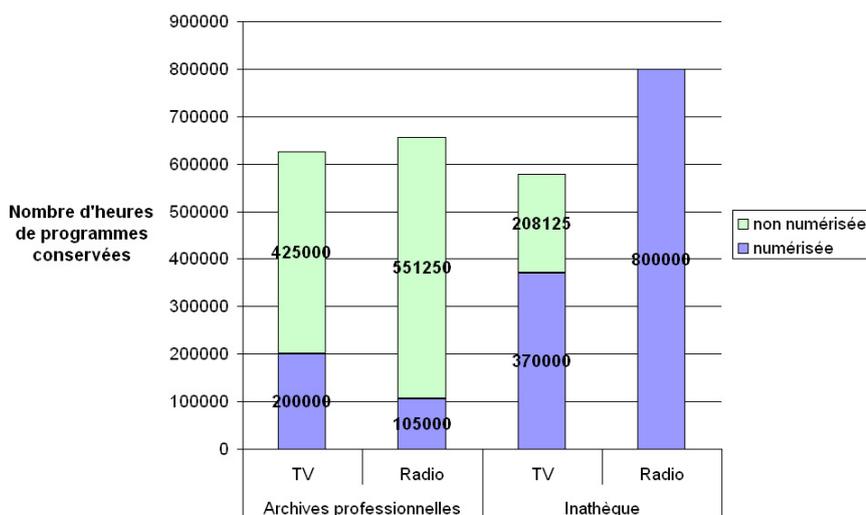


FIG. 1: Heures de programmation conservées par les archives professionnelles et l’Inatèque au 31 décembre 2004

Les travaux présentés dans ce manuscrit concernent en particulier une des extensions des missions de l’INA : le dépôt légal de la télévision effectué par l’Inatèque.

2 L’Inatèque de France et le dépôt légal

En 1988, suite à l’annonce par François Mitterrand de la création d’une grande bibliothèque, le gouvernement entreprend une refonte du dépôt légal en France. Pionnière pour le dépôt légal

³www.ina.fr/archivespourtous/

de l'imprimé introduit par François I^{er} en 1537, précurseur avec l'instauration en 1925 du dépôt légal du cinéma, des photographies et des phonogrammes, la France a souhaité poursuivre dans cette voie avec les œuvres diffusées par la radio et la télévision. C'est ainsi qu'en 1992, la loi révisant le dépôt légal est adoptée.

Dans sa première forme, le dépôt légal est une obligation pour les chaînes hertziennes (TF1, Antenne 2, FR3, la Cinquième, Arte, M6 et Canal plus) qui sont tenues de déposer leurs productions à l'INA. Du côté de la radio, seuls les programmes nationaux de Radio-France ont l'obligation d'être déposés. Les programmes concernés par le dépôt légal doivent être d'origine française et avoir été diffusé au moins une fois à l'antenne. L'INA ne peut en aucun cas procéder à une exploitation commerciale de ces documents.

Le dépôt légal débuta dès le 1^{er} janvier 1994 pour la radio et le 1^{er} janvier 1995 pour la télévision. L'INA crée une nouvelle direction afin d'assurer une parfaite étanchéité entre cette nouvelle mission patrimoniale et sa mission commerciale : l'Inathèque de France.

Dans un premier temps, l'Inathèque doit donner aux chaînes hertziennes deux semaines avant leur diffusion la liste des émissions entrant dans le cadre du dépôt légal. Les chaînes les transmettaient ensuite à l'Inathèque sur des supports analogiques (c'est pour cela que les archives de la télévision de l'Inathèque ne sont pas toutes numérisées, contrairement à celles de la radio - voir figure 1).

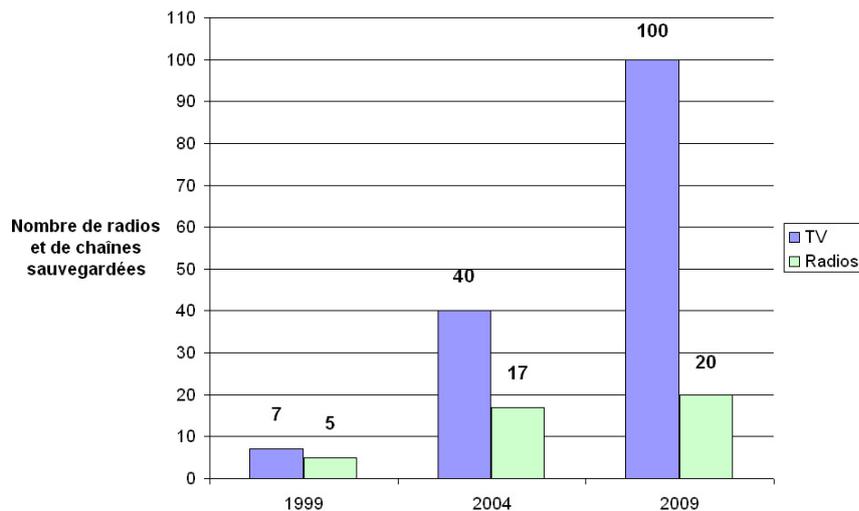


FIG. 2: Nombre de chaînes de télévision et de radios concernées par le dépôt légal

Une extension du dépôt légal est à nouveau votée : le 1^{er} janvier 2002, 12 chaînes du câble et du satellite y sont ajoutées. Il s'agit de Canal J, Canal Jimmy, Euronews, Eurosport, LCI, MCM, Paris Première, Planète, RTL 9, Série Club, TMC et TV5. À ce moment là, les coûts des fournitures et les technologies d'acquisition permettent à l'Inathèque une captation continue des chaînes qui les libère de l'envoi des émissions. Cette extension sera suivie en septembre 2003 d'un ajout de 22 chaînes à la collecte numérique : 13^e rue, AB Moteurs, Animaux, La Chaîne Météo, Chasse et Pêche, Demain!, Encyclopédia, Equidia, Escales, Festival, Fun TV, Histoire, La Chaîne Parlementaire et Public Sénat, Mangas, Mezzo, RFM TV, Seasons, Sport+, Télétoon, Téva, Toute l'Histoire. Enfin, en 2005, s'ajoutèrent les chaînes inédites de la TNT⁴ diffusées

⁴Télévision Numérique Terrestre

en clair (Direct 8, NT1, NRJ 12, France 4, W9, et TMC), et suivirent en fin d'année I-Télé, Europe 2 TV, Gulli et BFM TV. Une prochaine extension fera monter à 100 le nombre de chaînes archivées dans le cadre du dépôt légal.

Parallèlement, le dépôt légal de la radio, qui depuis 1995 archive intégralement les programmes proposés par Radio France (France-Inter, France-Musique, France-Culture, France-Info, France-Bleue), est étendu en 2001 avec RFI, Europe 1, RTL, RMC, Europe 2, Fun Radio, RTL2, NRJ, Chérie FM, Nostalgie, RFM et Skyrock.

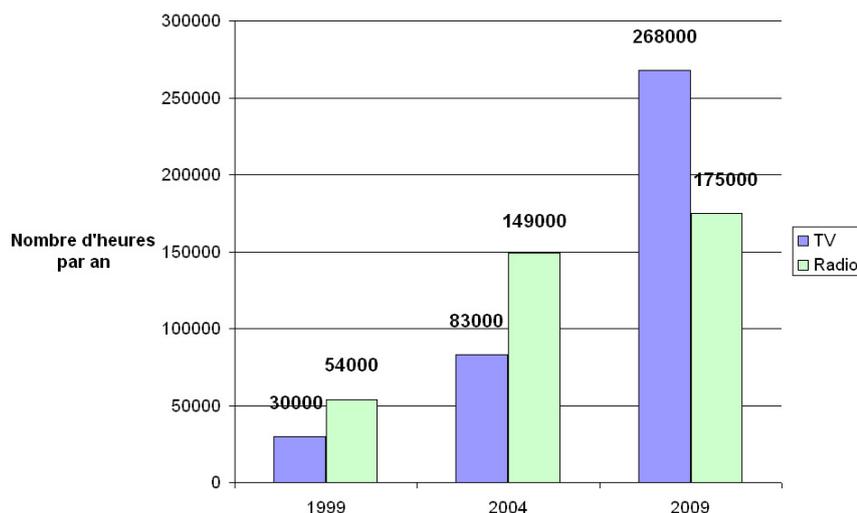


FIG. 3: Volume d'heures captées par an dans le cadre du dépôt légal

L'Inathèque doit faire face à un nombre croissant de documents et à un nombre d'heures archivées vertigineux (voir figure 2) lié à l'augmentation du nombre de chaînes et de radios (voir figure 2). Cependant, la tâche qui incombe l'Inathèque ne se limite pas à l'archivage et à la conservation de ces heures. Il faut aussi les mettre à disposition des chercheurs. Pour faciliter leurs recherches, les émissions sont donc décrites par des notices indiquant la nature, le sujet, le producteur et tout autre information qui pourrait être utilisée pour une recherche ultérieure. L'Inathèque possède ainsi plus de 2 500 000 notices sur les émissions issues du dépôt légal. À cela s'ajoute la numérisation de documents joints : grilles de programmes, conducteurs d'antenne, données de Médiamétrie ou des agences de presse.

Actuellement, la consultation des bases de données de l'Inathèque se fait exclusivement à la bibliothèque nationale de France (BNF) moyennant une accréditation. La consultation permet aux chercheurs de constituer leur corpus d'étude, de visionner les vidéos ou les sons et de consulter les documents connexes et les notices.

Prochainement, l'Inathèque sera chargée d'une partie du dépôt légal du Web, avec la BNF, en archivant les sites web audiovisuels (environ 10 000 sites qui représentent 50% du volume global du web français).

3 La direction de la recherche et de l'expérimentation

Dès janvier 1960, au sein de la RTF, la création d'un service de recherche est confiée à Pierre Schaeffer. Le groupe de recherche musicale (GRM) s'intéresse aux musiques expérimentales, le Groupe de Recherche-Images à la télévision et au cinéma, le groupe de recherches technologiques

à de nouveaux moyens de production (dont les *Shadocks* en sont le fruit) et le groupe d'études critiques analyse l'ensemble des problématiques de la recherche (Hoog, 2006). Puis, dans les années 80, la direction de la recherche tente de donner un nouvel élan à la production française de dessins-animés avec des logiciels d'animation 2D (Psyché en 1983, DAN en 1987 puis Toon Box en 1990) et 3D (Explore en 1987). En 1988, le projet synthetic-TV mélange, pour la première fois en temps réel, des images vidéos réelles et des images générées par un ordinateur.

Aujourd'hui, la recherche s'est recentrée sur les besoins des secteurs opérationnels de l'INA. Elle est découpée en cinq thèmes :

- **groupe de recherches musicales**, dont l'intérêt n'a pas changé depuis sa création ;
- l'équipe **description des contenus audiovisuels** s'intéresse à l'extraction et la gestion de métadonnées depuis un flux télévisuel ou une vidéo ainsi qu'à la publication automatique de collections de documents audiovisuels ;
- l'équipe **traitement technique de l'audiovisuel**, dont la problématique de recherche s'étend de la restauration à la détection de copie de contenus audiovisuels ;
- l'équipe **visualisation, interfaces, expérimentation** s'intéresse au résumé vidéo et à la représentation graphique de données en très grand nombre ;
- l'équipe **interprétation sémiotique de l'audiovisuel** a pour objectif d'étudier les formes visuelles.

La direction de la recherche et de l'expérimentation participe aussi à divers projets de diverses natures et de diverses ampleurs, dont, pour ne citer que ceux auxquels nous avons eu l'occasion de participer :

- **FERIA, framework pour l'expérimentation et la réalisation industrielle d'applications multimédias**⁵ est un projet RIAM qui regroupe des industriels (Communication et Systèmes, Vecsys, NDS Technologies⁶), des laboratoires de recherche (INA, IRISA, IRIT) et des utilisateurs (Arte). Le premier objectif du projet était d'obtenir un framework permettant le développement rapide d'applications multimédias en proposant des techniques de représentation et de manipulation des métadonnées. Le second objectif était d'utiliser le framework comme plate-forme d'expérimentation de nouvelles technologies d'indexation et de manipulation de contenus audiovisuels. Le projet FERIA est terminé mais a été repris en interne sous le nom de FERIA 2 ;
- **K-Space**⁷ est un réseau d'excellence européen réunissant plusieurs équipes de recherche intéressées par l'indexation, l'annotation et la recherche de contenus audiovisuels ;
- **Demi-Ton**⁸ est un projet retenu lors de l'appel à projets de l'action concertée incitative masse de données, réunissant les équipes TEXMEX et METISS de l'IRISA et l'équipe DCA de l'INA. Ce projet a pour thème la description multimodale des flux de télévision afin de les structurer automatiquement. Nous verrons les contributions de ce projet à la communauté dans le chapitre 2 de cette partie (page 13). La problématique de la présente thèse est au coeur du projet.

⁵www.ina.fr/recherche/projets/finis/feria/index.fr.html

⁶www.nds.com

⁷<http://www.kspace-noe.net>

⁸www.irisa.fr/textmex/demi-ton/

Chapitre 2

Structuration et classification de vidéos

Sommaire

1	Structuration de documents vidéos	14
1.1	Découpage en plans	14
1.2	Découpage en scènes	16
1.3	Détection d'événements	18
2	Détection de coupures publicitaires	19
3	Classification en genres	20
4	Travaux du projet Demi-Ton	22

L'indexation est le processus qui consiste à associer des index à un document pour permettre une recherche documentaire ultérieure. La création de ces index repose sur l'analyse ou l'interprétation du contenu du document de manière à en dégager les caractéristiques les plus représentatives (Bachimont, 2000). L'indexation automatique ne se distingue que par l'aspect automatique ou semi-automatique des techniques employées à la création des index. L'exploitation de ces index peut consister à rechercher un document, à parcourir rapidement une collection de documents, à résumer une collection, à décrire automatiquement des documents pour les archiver ou les publier sous une autre forme (site Web, DVD) (Gros, 2007). Li et Kuo définissent l'indexation automatique des documents audiovisuels comme étant l'association de descriptions à des segments temporels du document (Li et Kuo, 2003) : cela permet de définir la structuration de ces documents comme une étape préalable à un processus d'indexation automatique. La structuration est donc l'étape qui permet de déduire d'un document sa table des matières : à titre d'exemple, la table des matières d'une émission peut consister en un horodatage de ses génériques, de ses plateaux et de ses reportages.

La numérisation massive des documents a considérablement augmenté le nombre de collections concernées par ces techniques d'indexation automatique. Le cas de la télévision en est un très bon exemple qui se généralise facilement aux autres médias (presse et radio). Cependant, les collections personnelles ou professionnelles (telles que des documents médicaux, des vidéos de télésurveillance, ou encore des photographies satellites) produisent aussi de grandes quantités de documents à indexer en vue d'une consultation ultérieure.

Le problème de la structuration des flux télévisuels n'a été, à notre connaissance, que très peu

traité. Liang (Liang *et al.*, 2005) propose un algorithme de structuration des flux télévisuels en Chine, mais semble sous-estimer le problème : cette méthode ne tient pas compte des publicités et propose de structurer le flux en étudiant la similarité des plans du flux avec les génériques de toutes les émissions contenus dans une base. Cette méthode considère que toutes les émissions commencent et terminent par un générique identique à chaque fois, ce qui n'est pas toujours le cas : certaines séries, par exemple, ont un générique de fin qui change à chaque épisode (*un Gars, une fille, Kaamelott*). De plus, le problème de la mise à jour de la base des génériques n'a pas été abordé. Cependant, cet article montre que la communauté commence à s'intéresser au problème de la structuration des flux télévisuels.

Nous présentons dans ce chapitre quelques méthodes de structuration de documents vidéos, de détection de publicité et de classification de vidéos selon leur genre télévisuel apparentées à notre problématique de recherche. Nous décrivons ensuite des travaux menés dans le cadre de Demi-Ton (voir section 3, page 11), un projet dont la problématique de recherche est la structuration des flux télévisuels.

1 Structuration de documents vidéos

Cette partie présente différents travaux de structuration automatique de vidéos, aussi appelée macro-segmentation. Généralement, trois niveaux sont considérés : le plan, la scène et l'événement (voir figure 4). Nous verrons que certaines méthodes font appel à des modèles mathématiques et statistiques des structures des vidéos tandis que d'autres font appel à des règles de production. Vasconcelos indique que la structuration des documents vidéos permet ensuite de sélectionner un sous-ensemble d'éléments caractéristiques adéquats à l'indexation de chacune de ses parties (Vasconcelos et Lippman, 2000).

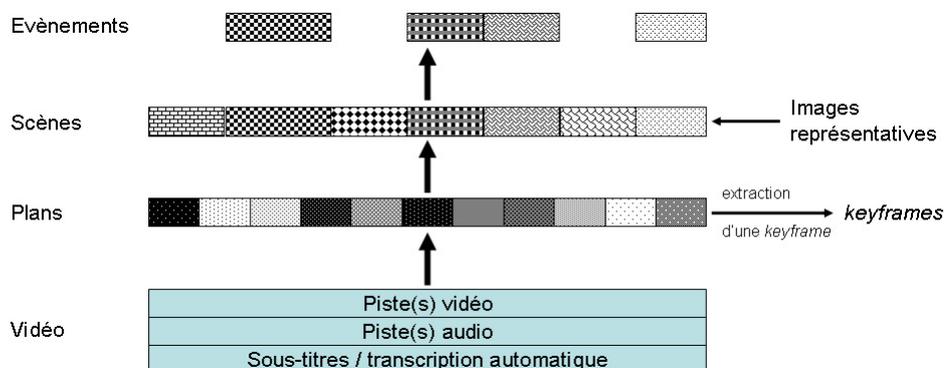


FIG. 4: Représentation hiérarchique d'un document vidéo selon (Li et Kuo, 2003)

1.1 Découpage en plans

Le découpage en plans est très souvent la première étape dans un système d'indexation ou de résumé automatique, ce qui explique que les premiers travaux sur l'analyse de vidéos se sont intéressés à cette structuration (Snoek et Worring, 2005). Le plan est une séquence d'images capturées sans interruption par une caméra. On distingue deux types de transition entre deux plans.

Le type le plus simple est une transition brutale entre deux plans sans recouvrement d'images (on parle alors de *cut*, voir figure 5a). Boreczky (Boreczky et Rowe, 1996) liste les diverses méthodes de détection de *cuts*, basées sur une comparaison pixel à pixel de l'image, une comparaison des propriétés statistiques des images (par exemple, la moyenne et l'écart-type de l'intensité lumineuse), une comparaison d'histogrammes ou de contours, une comparaison de la compression ou des vecteurs de mouvement. De nombreux articles (Boreczky et Rowe, 1996; Lienhart, 1998; Brunelli *et al.*, 1999; Browne *et al.*, 2000) comparent plusieurs de ces méthodes, présentant des taux de bonnes détections allant de 95% à 99%, avec des taux de fausses alarmes variables. Il ressort de ces quatre articles que les techniques les plus simples basées sur la comparaison d'histogrammes sont les plus performantes.



FIG. 5: Exemples de transitions entre deux plans

L'autre type de transition est dit progressif et relève d'effets d'édition. Il s'agit de transitions dans lesquelles le premier plan disparaît progressivement alors que le second apparaît. On distingue dans cette catégorie les fondus (fondu à l'ouverture, ou *fade-in*, fondu à la fermeture, ou *fade-out*, et fondu enchaîné, ou *dissolve* comme dans la figure 5b), jouant sur la transparence des deux plans, et les différentes sortes de volets, que ce soit le passage d'une ligne horizontale ou verticale faisant apparaître le nouveau plan (appelé *wipe*, voir figure 5c) ou tout autre effet ne faisant pas appel à la transparence des deux plans (voir figure 5d). Petersohn (Petersohn, 2004) propose une méthode de détection de ces trois types de transition. Chaque transition est détectée séparément à partir d'une combinaison d'informations sur la couleur, le mouvement et les contours. Pour l'ensemble de ces transitions, sa méthode obtient une précision de 85% et un

rappel de 70%. La méthode utilisée dans (Volkmer et Tahaghoghi, 2005) se démarque des autres par l'utilisation d'une fenêtre glissante pour détecter les transitions progressives par comparaison des couleurs sur chacune de ses images. L'avantage de la fenêtre glissante permet de tenir compte des dégradations des images sur un long interval temporel, plutôt que frame par frame. Cette méthode donne une précision et un rappel de 80%. Les chiffres parlent d'eux-mêmes : les transitions progressives sont beaucoup plus difficiles à détecter. De plus, la plupart des travaux font l'hypothèse de la stationnarité de l'image pendant la transition.

Une fois le document découpé en plans, il est d'usage de le représenter par une ou plusieurs images représentatives (appelées *keyframes*) pour le décrire de façon condensée. Il s'agit le plus souvent d'une image au début, à la fin ou au milieu du plan. Fauvet (Fauvet *et al.*, 2004) propose une façon plus originale qui permet d'adapter le nombre d'images représentatives en fonction du mouvement dominant afin de couvrir au mieux la séquence.

1.2 Découpage en scènes

Le plan est un objet vidéo parfaitement défini et la détermination de ses bornes temporelles est donc un problème bien posé. En revanche, la définition d'une scène est plus subjective : il s'agit en général de plusieurs plans adjacents partageant une même unité de sens. Dans la littérature internationale, de nombreux synonymes au mot anglais *scene* peuvent être employés : *sequence* (Aigrain *et al.*, 1995), *story unit* (Yeung *et al.*, 1996), *logical unit* (Vendrig et Worring, 2002), *video paragraph* (Watclar *et al.*, 1996) ou encore *act* (Aoki *et al.*, 1996). Le découpage en scènes se justifie par l'insuffisance du découpage en plans : la segmentation en plans donne en effet de trop nombreux éléments et qui sont de surcroît faiblement porteurs de sens. On peut séparer les méthodes de segmentation en scènes en deux grandes familles selon l'existence d'un modèle ou non.

D'un côté, certaines méthodes nécessitent la construction de modèles de structure (temporels et spatiaux) en s'appuyant sur des éléments caractéristiques de la scène comme par exemple l'affichage d'un logo ou l'apparition d'un texte. Par exemple, dans le cadre d'un journal télévisé, différents auteurs (Swanberg *et al.*, 1993; Zhang *et al.*, 1995) ont proposé une structure temporelle constituée de plateaux, de reportages, de coupures publicitaires et de bulletins météo et une structure spatiale composée de plusieurs modèles d'images avec un présentateur, un présentateur et un logo en haut à droite, etc. Des approches faisant appel aussi à la détection de texte et à la composante audio sont proposées dans (Maybury *et al.*, 1996; Merlino *et al.*, 1997). Gonsel (Gonsel *et al.*, 1998) propose d'utiliser le suivi d'objet (par exemple les logos et les visages) et une méthode de clustering non supervisé pour trouver la structure d'un journal télévisé. Les retransmissions sportives présentent aussi l'avantage d'être très formatées (faible nombre de vues, éléments caractéristiques forts, structure temporelle bien définie, éléments récurrents). Ces méthodes sont donc basées sur l'intégration de connaissances sur le domaine et sur l'extraction d'éléments caractéristiques (Zhong et Chang, 2001). Dans le cadre du projet européen DiVAN, auquel l'INA a collaboré, Carrive (Carrive *et al.*, 2000) décrit CLAVIS (classification of audiovisual sequences), un système de raisonnement temporel et terminologique pour la classification de séquences audiovisuelles, capable d'identifier dans un document toute sorte de structures à partir d'un modèle (*template*). Ces modèles sont définis à partir d'un réseau de contraintes portant sur des objets élémentaires (plans, logos, jingles ...) et doivent être écrits par un opérateur. Dans (Kijak *et al.*, 2003a,b), les auteurs s'intéressent à des retransmissions de tennis et utilisent des modèles de Markov cachés pour fusionner des informations provenant à la fois de l'image et de la bande son. Même si ces méthodes donnent des résultats satisfaisants, elles ont des domaines d'application très limités : le modèle développé pour une application n'est

pas généralisable ou applicable à une autre. De plus, pour des genres un peu plus complexes que le sport et les journaux télévisés, la modélisation peut être longue et nécessite des connaissances solides sur le domaine (Li et Kuo, 2003).

D'un autre côté, les méthodes qui ne reposent pas sur des modèles peuvent s'appliquer à des domaines moins restreints en utilisant soit les images, soit la bande son, soit les deux afin de regrouper les plans similaires et temporellement adjacents. Dans le cas d'une segmentation basée sur les aspects visuels, des informations sur la couleur et le mouvement sont le plus souvent utilisées. Par exemple, Yeung (Yeung *et al.*, 1996) regroupe les plans similaires au niveau des couleurs en construisant un Graphe des Transitions de Scènes (*scene transition graph* ou STG). Cette méthode consiste en un regroupement hiérarchique des plans ayant été tournés depuis la même position de caméra, en faisant l'hypothèse qu'à chaque position de caméra correspond des caractéristiques visuelles propres. Rui (Rui *et al.*, 1998a,b) propose une mesure de similarité basée sur l'histogramme des couleurs des images-clés et une mesure d'activité au sein d'un plan. Dans (Veneau *et al.*, 2000), un critère cophénétique est utilisé afin de mesurer la cohérence entre des plans successifs. Dans un premier temps, chaque plan est considéré comme une classe ; puis les classes sont rassemblées deux à deux de façon itérative en fonction de leur distance cophénétique. La distance cophénétique entre deux objets x_1 et x_2 est définie comme la distance à partir de laquelle les objets x_1 et x_2 appartiennent au même groupe lors d'un groupement.

Les méthodes basées sur des caractéristiques audio uniquement sont beaucoup plus rares. Dans (Sundaram et Chang, 2000), une scène audio est définie comme une collection de sources sonores dominantes pouvant être caractérisées par quelques éléments caractéristiques. Un changement de scène audio est donc détecté quand la majorité des sources sonores dominantes changent. Zhang et Kuo (Zhang et Kuo, 1999) obtiennent de bons résultats en segmentant la bande audio d'une vidéo en parole, silence, musique, parole et musique, chanson et ambiance en n'utilisant que des descripteurs audio bas niveau.

Enfin, il arrive que plusieurs modalités interviennent dans la segmentation en scène. À titre d'exemple, Huang (Huang *et al.*, 1998) utilise à la fois des indices visuels et auditifs ainsi que des informations sur le mouvements pour détecter les différentes scènes. La multimodalité dans la détection des scènes a souvent été illustrée par des travaux sur les journaux télévisés dans lesquels le discours du présentateur et les silences permettent de découper un journal télévisé en sujets et reportages à partir d'une segmentation en plans (Huang *et al.*, 1999; Nam *et al.*, 1997). L'application PACE (publication automatique de collections d'émissions) décrite dans (Caillet *et al.*, 2005) utilise pleinement la multimodalité afin de découper une émission en fonction de l'apparition d'un personnage ou de la diffusion d'une musique pour permettre une recherche efficace d'extraits dans une collection.

La méthode proposée dans (Aigrain *et al.*, 1995) est assez différente des autres. Il s'agit d'une méthode de segmentation par application des règles de production. Même si elle fait appel à des connaissances *a priori* sur la structure des documents, ses règles sont moins spécifiques à une collection ou à un genre donné. Dans (Rui *et al.*, 1998a) comme dans (Li et Kuo, 2003), cette méthode est de fait classée dans les méthodes générales qui ne se basent pas sur la structure du document. Le principe développé dans (Aigrain *et al.*, 1995) est d'exprimer sous la forme de règles de production des formes typiques permettant de repérer les limites des scènes, les plans importants et les groupes de plans faisant partie d'une même séquence. L'avantage est que les neuf règles proposées reposent sur un nombre limité de primitives : segmentation en plans, similarité globale des images, détection des transitions graduelles et détection de segments musicaux.

1.3 Détection d'événements

Le concept d'événement est encore plus subjectif que celui de scène, ce qui explique que des travaux assez différents sont regroupés dans ce domaine. La détection d'événements a été appliquée jusqu'à ce jour au sport et aux films. Dans les vidéos de sport, il s'agit de trouver les moments forts (*highlights*) d'un match ; la détection d'événements est alors capitale car elle permet de résumer une retransmission sportive de façon efficace. Par exemple, les coups-francs, les penaltys et les corners d'un match de football peuvent être détectés à l'aide d'un détecteur de mouvements de caméra et d'un modèle de Markov caché (Assfalg *et al.*, 2002). Yow (Yow *et al.*, 1995) s'intéresse à l'extraction d'événements dans un match de football et proposent de reconstruire une vue panoramique de l'action. Pour cela, ils détectent les cages sur le terrain, procèdent à un suivi du ballon et le mouvement est utilisé pour déterminer l'intensité de l'action. Zhao (Zhao *et al.*, 2006) considère que les ralentis des actions comme étant les événements à détecter dans la vidéo. Cabasson et Divakaran (Cabasson et Divakaran, 2002) combinent des règles basées sur des informations sonores et visuelles pour sélectionner les passages importants d'un match ; ils utilisent le mouvement pour détecter les actions importantes et l'énergie sonore pour détecter les réactions du public et les commentaires. Des travaux similaires ont été appliqués au basketball (Saur *et al.*, 1997; Nepal *et al.*, 2001), au cricket (Lazarescu *et al.*, 2002), au football américain (Li et Sezan, 2002; Babaguchi *et al.*, 2002), aux sports mécaniques (Petkovic *et al.*, 2002) et au baseball (Chang *et al.*, 2002). Même si ces techniques donnent des résultats appréciables, elles restent spécifiques à un seul sport ; d'autres chercheurs se sont intéressés à des méthodes fonctionnant sur différents sports. Li et Sezan (Li et Sezan, 2002) proposent une approche générale, à la fois déterministe et stochastique pour détecter les actions (*plays*) importantes d'une vidéo de sport ; leurs expérimentations portent sur des vidéos de football, de baseball et de combat de sumos. Sadlier et O'Connor (Sadlier et O'Connor, 2005) proposent une détection générique d'événements dans les vidéos de sport en utilisant des caractéristiques communes à toutes les grandes actions de match : la mesure de l'activité sonore, la détection de gros plans et de la foule et d'autres informations visuelles ou sonores sont extraites de la vidéo. Une machine à vecteurs de support est ensuite utilisée pour détecter les actions importantes parmi les actions candidates.

La détection d'événements dans les films de fiction consiste généralement à détecter une catégorie d'événements. L'exemple le plus simple est sans doute le dialogue. Lienhart (Lienhart *et al.*, 1999) recherche les dialogues d'une vidéo en se basant sur la détection d'une suite de champs et contrechamps. En effet, lors d'un dialogue entre deux personnes A et B, la caméra filme dans un premier temps la personne A, puis la B, puis encore la A, etc. Si une répétition de tels plans est détectée, alors un dialogue doit avoir lieu. Cependant, cette méthode ne s'applique qu'à des dialogues de deux personnes puisque lorsque le nombre de personnes dialoguant est supérieur, il est impossible de savoir dans quel ordre les personnes seront filmées. Li et Kuo (Li et Kuo, 2003) propose une extension de ce travail en détectant trois types d'événements : les dialogues entre deux personnes, les dialogues entre plusieurs personnes et tout ce qui ne constitue pas un dialogue. Seuls les dialogues sont traités comme des événements intéressants. Le cas des dialogues entre plusieurs personnes (A, B et C) est ramené à un cas de dialogue entre deux personnes en considérant qu'il y aura soit un plan avec les trois personnages, soit une alternance de plans présentant A et B, ou A et C, ou B et C. Alatan (Alatan *et al.*, 2001) utilise un modèle de Markov caché pour détecter les dialogues. Des critères visuels (couleurs et visages) et des critères audio sont utilisés par le modèle de Markov caché pour classer les scènes en dialogue ou non-dialogue. Contrairement à ces travaux qui sélectionnent des événements auxquels il faut prêter attention, Nam (Nam *et al.*, 1998) s'intéresse aux événements qui doivent être cachés (aux

enfants, par exemple) en détectant des scènes violentes. Pour cela, ils s'intéressent à l'activité globale d'un plan, à la détection de flammes dans les plans montrant une explosion ou une arme à feu, à la reconnaissance de scènes montrant du sang ainsi qu'à des sons ressemblant à des cris ou des explosions. Dans (Moncrieff *et al.*, 2001), la bande son est utilisée pour caractériser des scènes telles que des poursuites de voitures, des scènes violentes, à partir de la détection de sons caractéristiques : klaxons, sirènes, accidents, crissements de pneus, éclats de verre, explosions et tirs à l'arme à feu.

2 Détection de coupures publicitaires

La détection de publicités est un cas particulier de la reconnaissance de genres qui a été abondamment traité pour les avantages pratiques qu'elle représentait, comme le saut des publicités pendant une émission enregistrée avec un magnétoscope ou un terminal évolué (*set-top-box*). Le but est de différencier les publicités des programmes usuels dans un flux télévisuel ou dans une vidéo. La détection des publicités a un rôle capital dans un système de structuration des flux. Même si deux émissions ne sont pas forcément séparées par une coupure publicitaire (voir la partie II, page 33), la détection des publicités fournit une première structuration à raffiner par la suite.

Trois genres de méthodes peuvent être envisagés : les méthodes basées sur l'extraction de caractéristiques, les méthodes basées sur la reconnaissance de plans identiques et les méthodes exploitant les deux précédentes.

De nombreuses méthodes de classification peuvent être utilisées pour détecter les publicités. Par exemple, Albiol (Albiol *et al.*, 2004a,b) utilise la longueur des plans et un détecteur de logos (opaques ou transparents) pour déterminer à l'aide d'un modèle de Markov caché si un plan appartient à une coupure publicitaire ou pas, affichant 99% de plans correctement étiquetés. Haupmann et Witbrock (Haupmann et Witbrock, 1998) propose une détection des publicités dans les journaux télévisés en se basant sur des caractéristiques uniquement visuels. Hua (Hua *et al.*, 2005) adopte des descripteurs audiovisuels et l'entraînement d'une machine à vecteurs de support pour classifier les plans d'une vidéo en publicité ou programmes. Dans (Mizutani *et al.*, 2005), le flux télévisuel est modélisé par une chaîne de Markov de premier ordre possédant deux états (publicité et programme); Mizutani (Mizutani *et al.*, 2005) utilise un classifieur par descripteur audiovisuel et fusionne leurs réponses avec une machine à vecteurs de support. Les descripteurs utilisés dans (Divakaran *et al.*, 2004) sont un peu plus complexes que dans la méthode précédente : le son est segmenté en parole, musique, parole et musique, et bruits de foule (applaudissements, rires) et le descripteur vidéo consiste en une évaluation du mouvement en cinq niveaux allant de très faible à très élevé.

D'autres méthodes proposent une reconnaissance des publicités dans un document ou un flux télévisuel. Sanchez (Sanchez *et al.*, 1999) propose une segmentation en plans et la représentation des plans par leur image clé : dans leurs expérimentations, l'image clé est la première image du plan. Une base de données des signatures des images clés est ensuite constituée. La phase de reconnaissance consiste alors à comparer la signature des images clés de chacun des plans de la vidéo à traiter avec les signatures de la base. Une approche plus récente consiste à considérer une publicité comme une séquence répétée plusieurs fois dans le flux télévisuel (Pua *et al.*, 2004; Gauch et Shivadas, 2005; Shivadas et Gauch, 2006; Gauch et Shivadas, 2006; Naturel et Gros, 2006b). Les méthodes varient en fonction des signatures proposées mais l'idée et les problèmes à résoudre restent toujours les mêmes. Une base de signatures est constituée dans un premier temps manuellement, le problème étant la mise à jour de la base et l'action à entreprendre quand une

publicité inconnue est présentée. Gauch et Shivadas (Gauch et Shivadas, 2006) proposent une solution efficace : la vidéo est segmentée en plans avec une méthode qui privilégie un nombre élevé de plans et les plans qui se répètent sous vingt-quatre heures sont considérés comme candidats à une analyse plus profonde basée sur l'extraction de caractéristiques. Leur méthode obtient une précision de 100% et un taux de rappel de 96% pour une reconnaissance en temps réel. Leur système utilisé est dans leur laboratoire de façon continue sur plusieurs chaînes de télévision.

Lienhart (Lienhart *et al.*, 1997) propose une méthode évolutive de détection des publicités dans les films. Leur méthode s'appuie sur l'étude de la publicité à la télévision allemande : structure générale d'une coupure publicitaire (voir figure 6), réglementation en vigueur (durée maximale de la coupure) et caractéristiques visuelles (frames noires) et sonores (différence de volume). Pour la détection des publicités, leur système procède de façon itérative. Dans un premier temps, des zones de la vidéo sont sélectionnées comme étant de bonnes candidates pour être des publicités grâce à la détection de frames noires et du changement de volume. Ensuite, d'autres caractéristiques sont utilisées pour rechercher plus finement le début et la fin de la coupure publicitaire. Le jingle introductif est considéré comme la scène précédant la première publicité. Afin d'augmenter les performances, le système permet de sauvegarder la signature d'une publicité dans une base afin de la retrouver plus facilement et les nouvelles publicités, voire même les jingles, sont automatiquement ajoutés. Pour appliquer cette méthode, il faut être capable de regrouper les plans du jingle en une seule scène ce qui peut être source d'erreurs. Une comparaison entre les méthodes basées sur la détection de caractéristiques et celles basées sur la reconnaissance de plans ou de scènes est proposée dans (Duygulu *et al.*, 2004) où l'auteur propose d'autres méthodes de combinaison des deux approches.

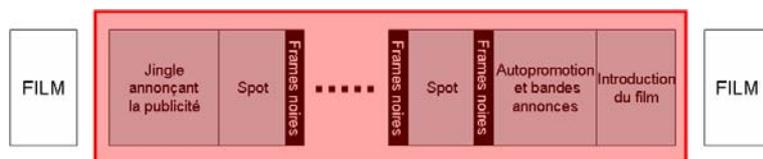


FIG. 6: Structure d'une coupure publicitaire pendant un film en Allemagne selon (Lienhart *et al.*, 1997)

Des travaux ont aussi été menés directement dans le domaine compressé. (Sadlier *et al.*, 2001) s'est intéressé à la détection des frames noires et des silences directement à partir de leurs codages en MPEG-1. Dimitrova (Dimitrova *et al.*, 2002), puis Agnihotri (Agnihotri *et al.*, 2003) proposent une détection des publicités pendant la phase d'encodage de la vidéo en se basant sur des caractéristiques du signal telles que la luminance, le format de l'image et la fréquence des plans. Agnihotri (Agnihotri *et al.*, 2003) propose un paramétrage de leur système avec des algorithmes génétiques pour atteindre une précision de 88% et un taux de rappel de 90%.

3 Classification en genres

La classification de vidéos est une autre étape de l'indexation automatique qui consiste à attribuer un ou plusieurs index (par exemple la présence d'un bateau, d'une plage, d'une route, le décollage d'un avion, une marche à pieds ou encore de la violence dans (Guironnet *et al.*, 2005)) à un document audiovisuel. La reconnaissance de genre est un cas particulier de classification qui consiste à attribuer un genre ou un sous-genre à une vidéo (Snoek et Worring, 2005).

Dans (Fischer *et al.*, 1995), Fischer s'intéresse à une classification des vidéos en quatre genres : information, sport, publicité et dessins animés. Leur système procède en trois étapes,

représentant chacune un niveau d'abstraction différent. Dans un premier temps, des propriétés syntaxiques sont extraites, dont des informations statistiques sur les couleurs, les transitions abruptes (*cuts*), des vecteurs de mouvements, la détection d'objets simples (par exemple les logos) et des statistiques audio. À partir de ces informations, des attributs stylistiques vont être construits : longueur des scènes, mouvements de caméra (panoramiques, zooms, *travelings*), les transitions entre les scènes (coupures abruptes et progressives), reconnaissance des objets (par exemple, reconnaissance des logos) et distinction entre parole et musique. Enfin, l'ensemble des informations collectées lors des phases précédentes sont comparées à des profils types pour chacun des genres. Roach (Roach *et al.*, 2001) propose une classification identique avec des résultats proches des précédents en se basant uniquement sur le mouvement des objets dans le champ de la caméra et le mouvement de la caméra. Deux résultats importants sont mis en évidence : d'une part, plus les séquences à classer sont longues, plus la classification est bonne ; d'autre part, le mouvement de caméra seul est moins discriminant que le mouvement des objets seul.

Dans (Li-Qun et Yongmin, 2003), les auteurs extraient des caractéristiques bas-niveau à partir du signal audiovisuel afin de classer les documents vidéos en cinq catégories : sport, dessin animé, information, publicité et musique. Un vecteur spatiotemporel comportant des caractéristiques audiovisuelles permet de capturer des informations inhérentes à la structure des genres. Une analyse en composante principale permet ensuite de réduire la redondance spatiotemporelle en exploitant la corrélation entre les éléments caractéristiques. Cette méthode permet de classer 86,7% des vidéos correctement. Truong (Truong *et al.*, 2000) propose une méthode similaire en extrayant un vecteur à seulement dix dimensions capturant des informations sur les transitions entre les plans, les longueurs des plans, les mouvements de caméra et les couleurs, affichant ainsi un taux de bonnes classifications de 83%.

Lu (Lu *et al.*, 2001) propose une classification de vidéos à partir de leur résumé représenté par une liste d'images clés. Ces images sont projetées sur une base qui a la particularité de ramener toute image à des conditions d'éclairage identiques ; chaque image est ensuite représentée par une signature chromatique compressée (*compressed chromaticity signature*). La méthode nécessite d'utiliser un modèle de Markov par genre télévisuel à reconnaître. La séquence de signatures constituée précédemment est donnée en entrée de chacun des modèles ; le genre est donné par le modèle de Markov attribuant la probabilité la plus grande à la séquence de signatures. Des travaux similaires sont présentés dans (Taskiran *et al.*, 2003) où le modèle de Markov caché est remplacé par un modèle hybride entre un modèle de Markov et une grammaire stochastique non-contextuelle.

Enfin, des travaux en marge de ceux que nous avons présentés jusqu'à maintenant proposent de diminuer l'intérêt porté à certaines zones de l'images. Pour cela, Rapantzikos (Rapantzikos *et al.*, 2005) se fonde sur une étude sur la perception humaine. L'attention visuelle porte sur des zones saillantes (*salient regions*) desquelles sont extraits des vecteurs d'informations spatio-temporels. Ce fonctionnement par régions d'intérêt est supposé d'une part accélérer les calculs qui ne portent plus sur l'image entière, et d'autre part augmenter la pertinence des classificateurs. Cette méthode a été appliquée à la classification de vidéos de football, de natation, de basket, de boxe et de billard. La classification des vidéos sans tenir compte des zones saillantes donne 74,4% de bonnes classifications contre 83,7% avec les zones saillantes.

Il est aussi possible de trouver des travaux présentant des classifications dichotomiques. À titre d'exemple, Glasberg (Glasberg *et al.*, 2005b,a) s'intéresse à la reconnaissance de dessins animés. Pour cela, des informations sur les couleurs, les contours et les mouvements sont intégrées par un réseau de neurones multicouche afin d'en faire une combinaison non-linéaire avec une fonction sigmoïde. Cette méthode permet de détecter correctement 80% des dessins animés qui lui sont présentés. Beaucoup de classifications dichotomiques concernent aussi la détection de

publicité (voir section 2, page 19).

4 Travaux du projet Demi-Ton

Plusieurs travaux de recherches complémentaires, dont ceux présentés dans ce manuscrit, ont été menés dans le cadre du projet Demi-Ton (présenté page 11) dont le but est de fournir une description d'un flux audiovisuel par une analyse automatique multimodale. Ces travaux vont du traitement du signal vidéo au traitement automatique du langage en passant par la reconnaissance de la parole. Le but est de décrire d'une façon multimodale le flux télévisuel afin de le structurer et l'indexer.

Huet (Huet *et al.*, 2006b,a) proposent d'utiliser conjointement des méthodes de traitement automatique de la langue française pour diminuer le taux d'erreurs de la transcription automatique. Le but de cette amélioration est de pouvoir ensuite traiter la transcription comme du texte et pouvoir en extraire des informations comme les sujets abordés pendant une émission ou son thème général.

Delakis (Delakis *et al.*, 2005, 2006) comparent l'utilisation de modèles de Markov cachés et de modèles de segments (voir section 1, page 79) pour modéliser une vidéo de match de tennis. Le but est d'utiliser non seulement des informations visuelles et sonores pour structurer la vidéo en jeux, sets et match, mais aussi d'utiliser le score afin d'éliminer les transitions impossibles d'un état à l'autre dans les modèles employés.

Enfin, Naturel (Naturel et Gros, 2005) propose une méthode de signature d'images avec laquelle il est possible d'identifier très rapidement des doublons : il est ainsi possible de retrouver dans le flux des séquences identiques à celles contenues dans une base. Partant d'un apprentissage supervisé – la base est constituée d'émissions préalablement indexées – il leur est possible de retrouver les jingles publicitaires, les publicités, les bandes-annonces et tout autre élément qui se répète dans la journée (Naturel et Gros, 2006a,b). Les guides de programmes sont utilisés ensuite pour indexer les séquences non répétées (Naturel *et al.*, 2005).

Chapitre 3

Proposition d'une approche

Sommaire

1	Flux et documents télévisuels	23
2	Motivation des travaux	24
3	Objectifs de l'approche	24
4	Inadéquation des méthodes de structuration de documents vidéo	25
5	Présentation du système SAFARI	26
5.1	Calculs des grilles de programmes possibles	27
5.2	Alignement des grilles de programmes améliorées sur le flux	29
5.3	Structuration des émissions et des tranches horaires	30

Nous présentons dans ce chapitre l'approche que nous avons retenue pour structurer les flux télévisuels reçus par l'Inathèque de France. Nous allons rappeler les objectifs de notre travail puis nous décrirons le système de structuration tel que nous l'avons conçu.

1 Flux et documents télévisuels

Le flux et le document sont des entités fortement différentes ce qui explique en partie que les méthodes qui s'appliquent à l'un ne s'appliquent pas à l'autre : le passage à l'échelle et les contraintes temporelles qui sont intrinsèquement liées à la nature fugace du flux sont des difficultés à prendre en compte.

En effet, Bachimont (Bachimont, 2003) définit un document comme « *un contenu [...] possédant une délimitation temporelle et spatiale* » ; puisqu'un document est figé, sa durée l'est aussi et on peut identifier clairement son commencement et sa fin. Il est aussi possible de se déplacer en son sein dans un sens ou dans l'autre.

Au contraire, le flux a un caractère insaisissable (le mot flux vient du latin *fluxus* qui signifie écoulement). On ne connaît pas les données qui arrivent après la donnée courante, mais on peut se souvenir des données qui ont déjà été diffusées. En français, le terme flux est ambigu car il exprime à la fois une idée quantitative (en physique, le flux est la quantité qui s'écoule à travers une aire donnée par unité de temps) et un mouvement perpétuel et insaisissable. Cependant, en anglais, on trouve deux équivalents au mot flux : le mot *flow* avec l'idée de débit et le mot *stream* pour l'idée de mouvement. Dans notre cas, nous ne nous sommes pas intéressés à la notion de

débit et de quantité mais plutôt à l'arrivée perpétuelle de nouvelles informations intrinsèquement liée à la nature du flux télévisuel.

2 Motivation des travaux

La structuration d'un flux télévisuel a pour but d'en fournir une table des matières. Il est légitime de penser que cette tâche peut être facilement accomplie par un alignement des grilles de programmes sur le flux. Une grille de programmes est propriété du programmeur et se reflète dans le conducteur d'antenne, qui est un rapport de diffusion de la chaîne. Cependant, il n'est pas possible de se procurer ces informations car aucune obligation légale n'impose aux chaînes de fournir ce genre de documents. De plus, il s'agit de la grille de programmes telle qu'elle a été conçue par la chaîne, alors que l'Inathèque et Médiamétrie sont plutôt intéressés par la grille de programmes perçue par le téléspectateur (avec les incidents techniques par exemple). De même la structuration des émissions peut être fournie par le conducteur de celles-ci, mais rien ne permet à l'INA ou Médiamétrie de les obtenir.

Les systèmes VPS (Video Program System) et PDC (Program Delivery Control) pourraient informer un système de structuration du début et de la fin de chaque émission. Ce sont deux protocoles qui permettent d'envoyer un signal pendant chaque programme, identifiant celui-ci à l'aide d'un numéro, de son heure de départ supposée et de la chaîne de diffusion. Malheureusement, seule Arte utilise ces signaux pour identifier certains de ses programmes ; ceci s'explique par le fait qu'Arte n'est pas entièrement financée par la publicité.

On pourrait également penser que les grilles prévisionnelles, proposées entre autres par la presse, sont des tables de matière du flux télévisuel. L'étude que nous proposons dans la section 6 (page 52) montre que l'imprécision et l'incomplétude de ces grilles n'en permet pas l'usage souhaité.

Aujourd'hui l'INA se procure la structuration des flux par le biais de Médiamétrie pour les chaînes hertziennes, et par le biais de ses propres documentalistes pour les autres chaînes. En effet, Médiamétrie propose à ses clients des taux d'audiences par séquences d'émissions. La société est donc obligée de procéder à une structuration manuelle des flux des télévisions hertziennes. Si la structuration est un travail nécessaire, la produire manuellement est un travail à faible valeur ajoutée, dont le résultat n'est précis qu'à la seconde près.

L'intérêt d'automatiser cette tâche de structuration des flux télévisuels est de permettre aux documentalistes, de l'INA ou de Médiamétrie, de passer plus de temps sur le travail de documentation à proprement parler.

3 Objectifs de l'approche

Le but de notre recherche est d'automatiser la structuration des flux télévisuels captés par l'Inathèque de France dans le cadre du dépôt légal de la télévision. Dans nos travaux, nous envisageons la structuration des flux télévisuels comme un processus hiérarchique. La première phase consiste en un découpage temporel des flux en éléments syntaxiques, c'est-à-dire en émissions (par exemple des fictions, des journaux télévisés), en interprogrammes (composés de publicités ou de bandes-annonces) et en tranches horaires (par exemple, *Ciné dimanche*). Nous considérons ces trois éléments comme les éléments syntaxiques élémentaires, décrits du point de vue audiovisuel dans le chapitre 6 de la deuxième partie (voir page 57).

Une fois le flux segmenté en éléments syntaxiques élémentaires, il est possible de procéder à une structuration de ces éléments eux-mêmes. Les éléments syntaxiques changent donc d'échelle

pour devenir cette fois-ci des scènes dans une émission, des publicités ou des bandes-annonces dans un interprogramme, ou encore des émissions dans une tranche horaire. La figure 7 est une représentation des divers niveaux hiérarchiques que l'on souhaite retrouver automatiquement.

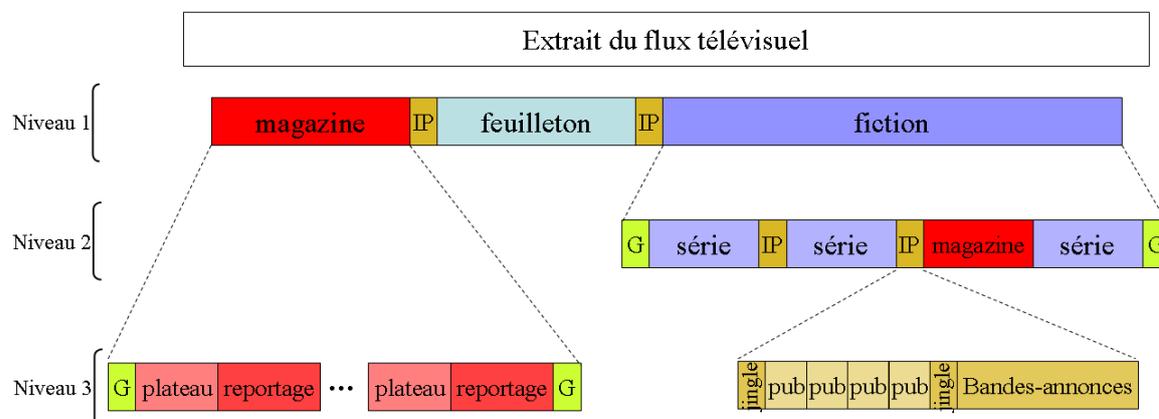


FIG. 7: Représentation de la structure hiérarchique d'un flux télévisuel (G désigne un générique et IP un interprogramme)

Nous avons choisi d'affecter à ces éléments syntaxiques une description : un type (par exemple, fiction ou reportage), un nom (le titre de l'émission) et l'espace temporel dans lequel il est diffusé. L'utilisation des guides de programmes nous permet aussi d'obtenir des informations sur les programmes, par exemple les participants (invités, présentateurs ou acteurs) ou encore le résumé.

4 Inadéquation des méthodes de structuration de documents vidéo

Nous avons présenté dans le chapitre 2 (page 13) un aperçu des méthodes de structuration appliquées aux documents vidéo (très souvent des programmes télévisés). L'ensemble des méthodes présentées analysent la vidéo image par image afin d'en trouver les plans. Puisque (Yeung *et al.*, 1996) met en évidence la présence de 300 plans dans 15 minutes de *Terminator 2 : le jugement dernier*, on peut imaginer la quantité de plans qui vont se succéder dans un flux télévisuel. Il semble assez maladroit de s'intéresser aux plans lorsqu'on désire des unités syntaxiques beaucoup plus grandes, telles que des émissions, alors même qu'aucune méthode ne permet, à notre connaissance, de rassembler ces plans en scènes puis en émissions. De plus, aucune méthode éprouvée ne permet de gérer, en terme de stockage et de fusion, l'ensemble des descriptions correspondant à chacun des plans d'un flux télévisuel.

D'un autre côté, la reconnaissance des genres des vidéos paraît également impraticable sur un flux. En effet, la lourdeur des descripteurs à extraire est prohibitive, et le nombre de genres à identifier dans ces méthodes ne convient pas aux besoins du dépôt légal de la télévision. Nous verrons plus loin que nous distinguons dans nos expérimentations plus d'une trentaine de genres alors que les méthodes de reconnaissance des genres peinent à distinguer les grands genres télévisuels comme les fictions, le journal télévisé et les dessins animés.

Les méthodes de détection des publicités basées sur les images monochromes et les silences ne permettent pas de trouver directement l'ensemble d'une coupure publicitaire en France ; la séparation obligatoire des publicités par des jingles nécessitent leur détection plutôt que celle des

publicités. En revanche, les images monochromes et les silences peuvent être utiles pour séparer certaines émissions, sans perdre de vue que leur apparition est possible au cours d'un programme. De plus, faire l'hypothèse que l'on connaît, à un moment donné, l'ensemble des jingles diffusés sur une chaîne est assez irréaliste dans la mesure où les chaînes rivalisent d'ingéniosité pour assortir leurs jingles à un événement : la fête des mères, un tournoi de sport, un film ou une saga. Cela implique que les méthodes de reconnaissance des publicités par recherche de plans similaires doivent prendre en compte la mise à jour de la base des jingles de référence.

Pour l'ensemble des raisons que nous venons d'exposer, nous proposons un système de structuration automatique des flux télévisuels qui repose sur l'exécution locale de détections appropriées pour un certain type de transition entre deux programmes.

5 Présentation du système SAFARI

Le système SAFARI (structuration automatique de flux audiovisuels pour la recherche et l'indexation) est le système que nous avons conçu (Poli et Carrive, 2005). Nous proposons une approche originale au problème de la structuration des flux qui se résume en quelques points :

- utiliser des connaissances sur les chaînes, la télévision et les émissions pour reconnaître ;
- utiliser les connaissances sur les grilles de programmes construites à partir d'observations à court et long terme ;
- minimiser les temps de calculs, en effectuant les calculs localement dans le flux ;
- utiliser des règles spécifiques pour reconnaître les éléments recherchés plutôt que des règles générales dont les résultats sont discutables.

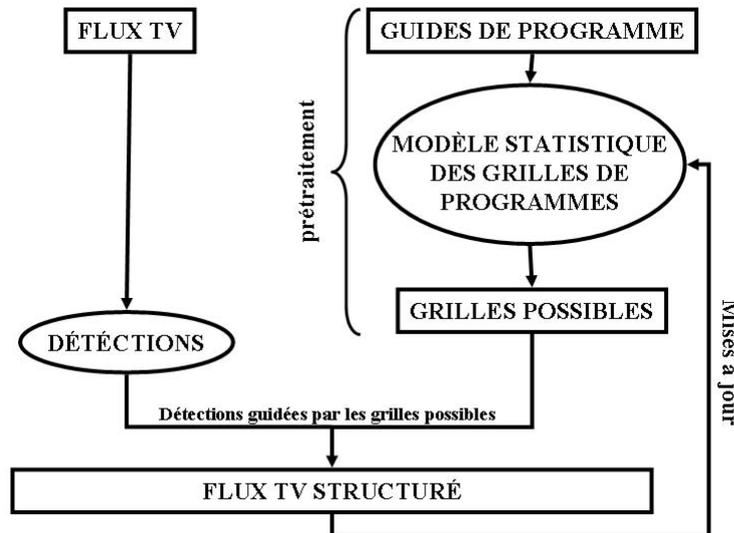


FIG. 8: Schéma général du système SAFARI

Compte tenu du contexte industriel dans lequel se situent nos travaux, nous avons pris soin de faire des hypothèses cohérentes avec les besoins de l'Inathèque et avec la réalité du paysage de l'audiovisuel français (le paysage audiovisuel français est décrit dans la partie II,

page 33). En effet, la structuration du flux télévisuel consiste avant tout à « défaire » le travail du programmeur de la chaîne, dont nous présentons quelques aspects dans le chapitre 5 de la seconde partie (page 43). Le système SAFARI fait appel à plusieurs types de connaissances :

- des connaissances sur les diffusions passées : puisque les grilles de programmes sont stabilisées en terme de genres télévisuels, nous pouvons apprendre les diffusions précédentes afin de prédire les enchaînements d'émissions pour un jour de la semaine et une chaîne donnés (voir section 5.1, page 27) ;
- des connaissances sur les diffusions présentes : ils s'agit de connaissances à court terme qui doivent pouvoir évoluer rapidement telles que les titres des émissions ;
- des connaissances sur l'habillage des chaînes : il s'agit de connaissances à moyen terme comprenant des informations sur les chartes graphiques des chaînes afin d'en détecter les jingles.

L'idée principale est de guider les détections en fonction des connaissances sur la grille de programmes d'une chaîne. La figure 8 représente une vue d'ensemble du système SAFARI. À partir des guides de programmes et du modèle statistique des diffusions passées de la chaîne, le système est capable de fournir, sous la forme d'un graphe, l'ensemble des séquences possibles pour une journée de l'année et une chaîne données. Le parcours de cet arbre se fait par confirmation de l'hypothèse faite sur la nature du programme courant dans le flux en procédant à des calculs sur le signal. Les éléments à détecter dans le flux sont donnés par les connaissances à court et moyen terme que nous avons évoquées ci-dessus. Par exemple, si l'émission suivante est censée être une coupure publicitaire, le système va essayer de détecter les jingles. Pour cela, nous avons choisi de spécifier au système des invariants de production caractéristiques de l'ensemble des jingles d'une chaîne pour une période donnée afin de les détecter : cette méthode se veut plus robuste à la présentation de nouveaux jingles que les méthodes présentées dans la section 2 du chapitre 2 (page 19). En cas d'échec, il essaiera de vérifier les autres branches possibles dans l'arbre. Si aucune branche ne donne un résultat satisfaisant lors des détections, une erreur sera annoncée à un opérateur pour résoudre le problème.

Il est donc nécessaire de trouver les bonnes règles et les bons détecteurs pour que le système fonctionne. Nous avons observé durant nos travaux que des règles spécifiques étaient détectées plus facilement mais avaient une durée de vie plus limitée que des règles générales : par exemple, pour la détection des publicités sur France 2, il est plus simple de détecter le mot *publicité* qui apparaît au même endroit sur chacun des jingles plutôt que l'ambiance sonore. Ces règles peuvent être formulées par un expert, mais nous proposerons quelques solutions pour automatiser cette construction de règles discriminantes.

Nous allons détailler les trois étapes de la structuration dans le système SAFARI : le calcul des grilles de programmes possibles, l'alignement de ces grilles sur le flux et la structuration des émissions.

5.1 Calculs des grilles de programmes possibles

La stabilité des grilles de programmes, qui sera discutée dans la section 7 (page 56), nous permet d'entraîner un modèle statistique sur les grilles de programmes des années précédentes. Il s'agit de données que l'INA sauvegarde au même titre que les émissions télévisées, ce qui nous permet de faire un tel apprentissage.

Notre idée part du principe que la structuration des flux revient à aligner les tables des

matières sur les flux. Les guides de programmes, qui peuvent se trouver dans un format papier dans des magazines spécialisés (par exemple, *Télérama*, *Télé 7 jours*) ou dans les quotidiens, ou sous une forme électronique (on parlera alors d'EPG pour *electronic program guide*, voir la section 5, page 52).

Cependant, les grilles prévisionnelles contenues dans ces guides sont incomplètes et imprécises (voir tableau 5, page 54) : les émissions courtes, généralement de moins de cinq minutes sauf pour les fictions, ne sont pas annoncées. L'originalité de notre travail vient du fait que nous nous sommes efforcés de transformer ces grilles prévisionnelles afin de remédier à leur incomplétude en modélisant les grilles de programmes des chaînes. Pour cela, il est possible de modéliser les grilles de programmes par des règles, une grammaire ou encore un modèle statistique.

Nous avons choisi d'utiliser un modèle statistique avec un formalisme très connu : les modèles de Markov cachés (Poli, 2005). Ces modèles ont déjà fait leurs preuves dans l'analyse de séquences, notamment en bio-informatique avec l'analyse de l'ADN, mais aussi en analyse de vidéos (voir le chapitre 2, page 13). Les autres possibilités restent à explorer, notamment les modèles bayésiens.

Nous considérons dans nos travaux les grilles de programmes comme étant des séries temporelles (Poli et Carrive, 2006a). Une série temporelle est une séquence de valeurs prises par une variable au cours du temps. Dans notre cas, la variable a pour valeur un genre télévisuel parmi les genres que nous considérons et les différents temps de la série représentent à la fois le jour de diffusion et l'heure de début du programme. Afin de prédire les grilles de programmes possibles, il nous faut à la fois prédire le genre de la prochaine émission mais aussi son heure de début. Nous avons remarqué que le genre de l'émission, donc sa durée (voir chapitre 5, page 43), est fortement lié à son heure de diffusion. Il nous semblait donc indispensable que l'une des variables explicatives soit l'heure de diffusion : en effet, si nous avions tenté de prédire les heures de diffusion, nous aurions perdu un excellent prédicteur. De plus, l'heure de début du prochain programme peut être déduite de l'heure de début du programme en cours auquel on ajoute sa durée. Nous avons donc choisi de prédire les durées des émissions à l'aide d'un arbre de décision. L'entraînement d'un arbre de régression est déterministe et permet de renvoyer une valeur moyenne et un écart-type : cela nous permet de déduire un encadrement de la durée d'une émission. Cet encadrement fournit une fenêtre temporelle dans laquelle le système doit chercher les changements d'émission. Cette approche permet de procéder à des calculs localement dans le flux, en sachant à l'avance ce que l'on doit chercher (un jingle de publicité ou un générique d'émission), évitant ainsi les fausses alarmes et privilégiant des détecteurs adaptés voire spécialisés.

Le résultat de cette prédiction est un arbre (voir figure 9) représentant toutes les grilles de programmes possibles pour une journée. Il suffit ensuite de fusionner ces informations avec celles contenues dans le guide de programmes : dans un premier temps, les branches ne passant par les émissions annoncées dans le guide de programmes peuvent être élaguées. Ensuite, les nœuds représentant la même émission (encadrés dans la figure 9) avec un léger décalage dans l'heure de début sont fusionnés. L'arbre devient ainsi un graphe dirigé acyclique représentant l'ensemble des séquences de programmes possibles (Poli et Carrive, 2006b). Cette prédiction revient à un prétraitement des guides de programmes.

Pour formuler simplement notre approche, il est plus simple de trouver « quand on sait ce que l'on cherche et qu'on sait où le chercher ». Nous allons voir à présent comment détecter le début et la fin des émissions en guidant les détecteurs avec les informations contenues dans le graphe des possibilités.

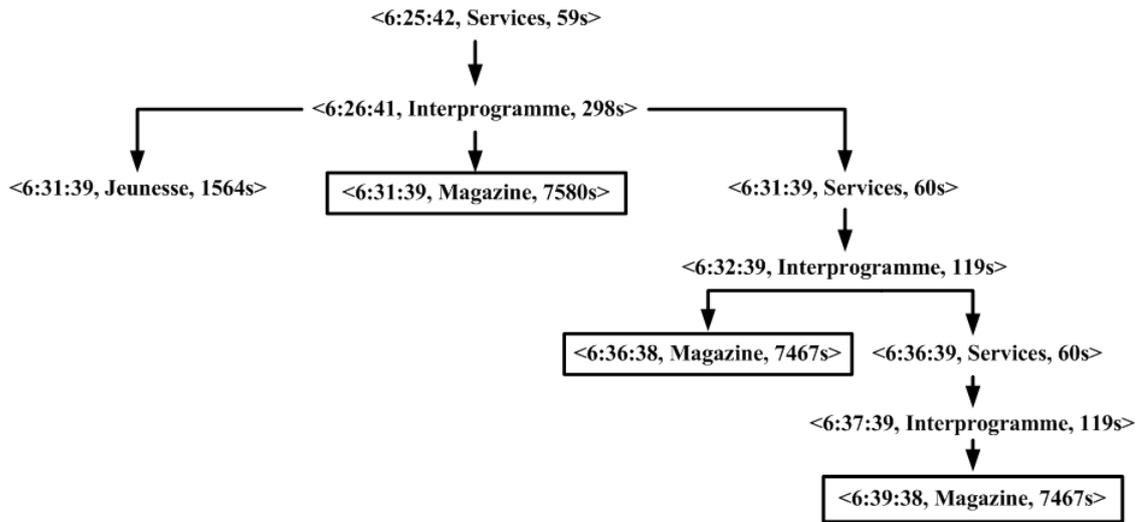


FIG. 9: Extrait d'un arbre représentant les séquences possibles de types de programmes

5.2 Alignement des grilles de programmes améliorées sur le flux

Une fois l'ensemble des possibilités d'enchaînements déduites du guide de programmes et du modèle statistique, le système doit confronter ses prédictions à la réalité du flux : c'est lors de cette confrontation que les grilles possibles sont alignées sur le flux. Chaque nœud du graphe proposé par la partie précédente représente une émission par son genre, son heure de début et un encadrement de sa durée. Il suffit donc de détecter soit la fin de l'émission en cours, soit le début de l'émission suivante, selon le cas. En effet, si l'on sait que l'émission suivante est un interprogramme (une publicité ou une bande annonce), on peut s'attendre à trouver un jingle. La détection de ce jingle se fera donc localement dans une fenêtre temporelle qui commencera à l'heure minimale de fin de l'émission courante et terminera à son heure maximale de fin.

Nous montrerons que, dans bien des cas, les silences et les images monochromes suffisent à détecter les changements de programmes. Mais le nombre de fausses alarmes de ces détecteurs ne permet pas de les utiliser seuls : en considérant les détecteurs localement, notre approche permet d'éviter bien des fausses alarmes ce qui permet de se contenter de détecteurs relativement simples.

Cependant il peut être utile de détecter les génériques de début ou de fin d'une émission si au moins une de ses modalités (son ou image) ne varie pas en fonction des diffusions : en effet, les génériques bornent de façon non ambiguë un programme. De plus, ils permettent d'identifier le programme afin de vérifier qu'il s'agit bien de ce qui a été annoncé dans le guide de programmes. Toutefois, la détection de génériques de début et de fin ne peut être systématique puisque beaucoup d'émissions n'ont pas de générique de fin ou encore ont des génériques qui changent à chaque diffusion (par exemple si le public ou les chroniqueurs présents sur le plateau sont filmés pendant le générique).

La détection des publicités est aussi une partie importante de notre système. Il est courant de penser que toutes les émissions sont séparées par des publicités, mais avec les nouvelles techniques de programmation par blocs (voir chapitre 5, page 43), ce n'est plus tout à fait exact : il peut arriver qu'une chaîne diffuse trois épisodes d'une série de façon consécutive. L'interruption par des publicités se fera alors au cours des épisodes. De plus, il est impossible de connaître avant leur diffusion les spots publicitaires ou les jingles qui annoncent les coupures publicitaires.

5.3 Structuration des émissions et des tranches horaires

Une fois que les programmes du flux ont été isolés, nous pouvons procéder à une structuration de ces émissions. Puisque la structuration du flux identifie chacun des programmes, il est possible d'utiliser des méthodes spécifiques à un genre télévisuel ou à une collection. Nous nous sommes intéressés à deux collections très différentes : le journal télévisé et une tranche horaire, le *Morning Café*.

Le journal télévisé est une collection dont les propriétés de production (générique, plateau, habillage) restent stable pendant plusieurs années. Il s'agit d'un genre qui a été énormément étudié (voir chapitre 2, page 13). Pourtant aujourd'hui, aucune méthode automatique n'est utilisée pour décrire un journal télévisé car leur structure ne se limite pas à un enchaînement de plateaux et de reportages : les journaux télévisés sont aussi composés de duplex, d'interviews, parfois même de rubriques (par exemple, *l'invité des cinq dernières minutes* du journal de 13 heures sur France 2). Selon les usages, la structure d'un journal télévisé peut changer en fonction de l'importance accordée aux plateaux ou aux reportages. Nous proposons des techniques très simples d'aide à la documentation d'un journal télévisé.

D'un autre côté, les tranches horaires (voir partie II, page 33) sont des segments temporels dans le flux dont la structuration nécessite des techniques plus évoluées que pour les journaux télévisés car les différents genres télévisuels intervenant dans une tranche horaire ne sont pas aussi différenciables qu'un plateau et un reportage. La tranche horaire à laquelle nous sommes intéressés, le *Morning Café*, est composée d'une émission en direct interrompue par d'autres émissions (flash d'information, météo, trafic routier) : sa durée doit être stable pour ne pas retarder toute la grille ce qui lui impose une certaine structure. Nous proposons ainsi d'utiliser le même modèle statistique que pour les flux afin de structurer les tranches horaires.

Deuxième partie

La programmation d'une chaîne de télévision

RÉSUMÉ

A son origine, la télévision était marquée par un monopole du service public, fortement lié à l'Etat. Elle a peu à peu évolué en un modèle commercial où la concurrence est à la fois vivace et peu courtoise, renforcée par la multiplication du nombre de chaînes qu'a entraîné la diversification des moyens de diffusion (câble, satellite, TNT, ADSL...).

Depuis la fin du monopole, la diffusion d'un programme sur les chaînes de télévision n'est plus tellement une décision de la chaîne, mais plutôt une réponse à l'attente du téléspectateur. Cette nécessité de donner un rendez-vous à son public cible a donné naissance à un objet (la grille de programmes), à une fonction (la programmation) et à un métier (programmeur).

Structurer un flux télévisuel d'une chaîne revient finalement à retrouver les idées de son programmeur. C'est pourquoi il nous semblait utile de connaître sa fonction, d'en comprendre l'origine, les enjeux et l'évolution.

Cette partie est entièrement consacrée aux diverses techniques de programmation d'une chaîne, aux différentes grilles qui existent et aux types de programmes qui entrent en jeu dans une grille. Cela nous permettra, à la fin de cette partie, de justifier nos hypothèses de travail.

Introduction

Le flux télévisuel d'une chaîne n'est pas une suite d'émissions sans logique. La façon dont les émissions se succèdent dans la journée est une stratégie affichée de la chaîne en respect de sa ligne éditoriale. Les émissions sont diffusées pour rassembler devant le poste de télévision un certain public cible, en fonction de l'emploi du temps de celui-ci. C'est le rôle du programmeur de la chaîne de satisfaire les téléspectateurs en leur présentant des programmes adaptés à leurs goûts et leurs envies. Le résultat du travail du programmeur est une grille de programmes, véritable table des matières horodatée du flux télévisuel.

Nous avons vu dans la partie précédente que la structuration d'un flux télévisuel revenait à en déduire une table des matières : la structuration automatique est donc le travail inverse du programmeur.

Cette partie, en marge des autres, peut être lue indépendamment et propose au lecteur de comprendre le monde de la télévision en s'appuyant sur l'exemple de la télévision française. Nous mettons ainsi à disposition l'expertise dont ont bénéficié nos travaux sur le métier de programmeur grâce au contexte industriel de cette thèse. Nous citons régulièrement dans cette partie plusieurs ouvrages rédigés par des professionnels de l'audiovisuel : Emmanuel Hoog, président de l'INA (Hoog, 2006), Régine Chaniac et Jean-Pierre Jézéquel (Chaniac et Jézéquel, 2005), chercheurs à l'INA, ainsi que Laurent Fonnet, ancien programmeur de TF1 (Fonnet, 2003).

Nous souhaitons que cette partie permette au lecteur de comprendre certains mécanismes de l'industrie de la télévision. Nos travaux reposant sur la compréhension de ces mécanismes, cette partie justifie l'ensemble de nos hypothèses de travail, certes contraignantes mais parfaitement adaptées au domaine.

Nous proposons de retracer l'évolution de la télévision française depuis ses débuts afin de comprendre comment le système concurrentiel que nous connaissons aujourd'hui s'est installé. Nous décrivons le cœur de métier du programmeur, c'est-à-dire l'élaboration de la grille de programmes afin de dégager quelques techniques de base de la programmation d'une chaîne. Ces techniques nous ont permis de faire certains choix dans nos travaux comme l'emploi de certains détecteurs plutôt que d'autres. Nous prendrons soin de détailler les propriétés des genres de programmes que nous retrouvons aujourd'hui dans ces grilles et nous présenterons différentes taxinomies de genres télévisuels. Cette étude des genres a été très importante dans notre modélisation des durées des programmes (voir la troisième partie, page 75) et le choix d'une taxinomie s'est imposé. Enfin, notre approche consistant à un apprentissage des grilles de programmes, nous montrerons à la fin de cette partie que les grilles de programmes se stabilisent, c'est-à-dire qu'on arrive, en théorie, à une adéquation maximale entre les programmes proposés et les attentes du public cible d'une chaîne. Pour cela, nous avons étudié les données du dépôt légal de la télévision sur plusieurs années de TF1 en étudiant notamment la constance des genres

télévisuels à certaines heures et les transitions d'un programme à un autre.

Chapitre 4

Évolution de la télévision française

Sommaire

1	La télévision de l'offre	37
2	La télévision de la demande	38
3	La télévision aujourd'hui	40
3.1	La multiplication des moyens de diffusion	40
3.2	Vendre son audience aux annonceurs	41
3.3	La télévision interactive et mobile	42

La télévision est arrivée en France assez tardivement par rapport à nos voisins européens. Son évolution est marquée par deux grandes étapes. La première, l'ère de la télévision de l'offre, ne laissait guère au téléspectateur que le choix de regarder ou non les émissions qu'elle lui proposait. La seconde, l'ère de la télévision de la demande, est celle que nous connaissons encore aujourd'hui, mais sous une forme plus évoluée qu'à ses balbutiements : le téléspectateur a la possibilité de changer de chaîne et montrer ainsi qu'il n'est pas satisfait des programmes qui lui sont présentés (Fonnet, 2003; Burtin, 2004).

Aujourd'hui, une nouvelle ère se profile : celle où le téléspectateur choisit et paie uniquement pour ce qu'il veut voir.

1 La télévision de l'offre

En France, pendant la seconde guerre mondiale, presque tous les postes de télévision étaient aux mains des Allemands. À la fin de la guerre, deux ordonnances confirment le monopole de l'État, réquisitionnant toutes les installations de postes privées. Dès 1949, l'État décide de confier la diffusion de programmes sur les ondes à un organisme public, la RTF (Radio-télévision française). Cette même année, une redevance spécifique pour les postes de télévision est créée, finançant entièrement la première chaîne. En 1974, une autre redevance sera demandée aux détenteurs de postes de télévision en couleurs. L'offre évolue alors assez lentement : une seconde chaîne rejoint la première en 1964, à l'occasion de la transformation de la RTF en l'ORTF (Office de radiodiffusion-télévision française), puis une troisième en 1973 ; parallèlement, la couverture s'étend à travers la France.

À cette époque, les chaînes possédaient un stock de programmes. Chaque programme était diffusé à des moments choisis en fonction de la présence potentielle de son public cible. L'ère de

la télévision de l'offre a fait le bonheur des producteurs : les chaînes venaient leur acheter des émissions afin de satisfaire le triptyque « informer, éduquer, divertir », sans donner un cahier des charges précis. La durée du programme importait peu dans la mesure où les chaînes n'émettaient pas de façon continue. Dans un premier temps, les émissions étaient diffusées à la mi-journée et le soir ; puis la télévision de l'après-midi est apparue. Ce n'est qu'en 1991 que TF1 commencera à diffuser en continu, suivie petit à petit par les autres chaînes.

Malgré leur rattachement à un même organisme public, une concurrence entre les trois chaînes s'installa. Quelques décisions historiques (comme par exemple l'autorisation d'émission pour de nouvelles chaînes, l'introduction de la publicité) et les progrès techniques de l'époque (comme la multiplication des canaux sur les ondes hertziennes) vont être à l'origine du développement de ce climat concurrentiel. Dès 1968, le financement des chaînes par la publicité fut autorisé en commençant par la première chaîne, puis la deuxième en 1971. L'apport des recettes publicitaires ne devaient pas dépasser 25% des ressources globales de l'organisme public : les chaînes devaient donc veiller à la diffusion et surtout à la bonne réception de leurs écrans publicitaires. L'autre tournant de l'histoire de la télévision est le démantèlement de l'ORTF en 1975, qui a rendu indépendante chacune des chaînes. Celles-ci commencèrent alors à porter une attention plus importante aux attentes du public, et le quota des ressources publicitaires fut largement dépassé. Parallèlement, les limites du réseau hertzien en terme de canaux de diffusion sont franchies : la France a la possibilité technique d'augmenter le nombre de chaînes. Il faudra attendre 1984 pour que cette possibilité devienne une volonté avec la création de Canal Plus, une chaîne au financement particulier puisque ses ressources ne proviennent que des abonnements et des publicités. Plus tard, deux chaînes privées sont introduites : la Cinq et la sixième. Enfin, le gouvernement décide de privatiser TF1 en 1987, mettant en minorité les chaînes publiques.

Durant cette ère, la France a assisté au développement de tous les modèles de télévision que l'on connaît aujourd'hui : la télévision gratuite (avec des chaînes publiques, financée par la redevance et la publicité, et des chaînes privées, financées uniquement par la publicité), et la télévision payante (financée par un forfait ou un paiement à l'acte et la publicité). Le tableau 1 permet de comparer les recettes publicitaires, les recettes publiques et les abonnements de cinq chaînes hertziennes françaises publiques (TF1 jusqu'en 1987, France 2 et France 3), privées et gratuites (TF1 à partir de 1987 et M6), et privées et payantes (Canal+), entre 1980 et 2005. Ce tableau met en évidence l'importance de ces trois catégories de recettes pour trois modèles de chaînes différents : les recettes publicitaires augmentent beaucoup plus vite pour les chaînes privées à accès gratuit que pour les autres chaînes. On note aussi que les recettes publiques pour France 2 et France 3 dépassent les recettes publicitaires d'une chaîne privée comme M6, mais restent quand même loin derrière les recettes de chaînes comme TF1 et Canal+. Le cinquième canal n'est pas représenté dans ce tableau du fait de sa diversité (la Cinq, la Cinquième, Arte et France 5).

C'est ainsi que la situation concurrentielle dans laquelle le paysage audiovisuel français se trouve encore aujourd'hui est née.

2 La télévision de la demande

Si la concurrence qui existait au temps du monopole n'était qu'une quête de prestige, la concurrence qui s'est instaurée avec l'arrivée des chaînes privées concerne plutôt les ressources de ces chaînes. L'irruption de ces chaînes privées transforme le marché de la publicité télévisée en un marché concurrentiel, ce qui n'était pas le cas au temps du monopole.

Face à cette situation nouvelle, le gouvernement a choisi de confier à une société indépen-

dante, Médiamétrie⁹, la mesure d'audience des principaux médias audiovisuels. Médiamétrie propose une mesure unique et impartiale de l'audience, contrairement aux méthodes mises en places par chacune des chaînes avant sa création. Les annonceurs ont la possibilité de connaître très rapidement les audiences des chaînes. L'introduction d'un nouvel indicateur, la part de marché (ou part d'audience), illustre parfaitement le rapport de force entre les chaînes à un moment donné.

Cette ère est totalement opposée à la précédente, car la chaîne, au travers des chiffres fournis par Médiamétrie, peut avoir une idée de l'appréciation de ses programmes par les téléspectateurs. Les progrès de Médiamétrie en terme de calculs d'audiences vont permettre aux chaînes de se rapprocher de leurs public afin d'en satisfaire les attentes. La collecte des taux d'audience sera renouvelée plus fréquemment, jusqu'à atteindre la fréquence d'une collecte par seconde. Les foyers seront divisés en plusieurs cibles (4 ans et plus, 15 ans et plus, hommes, femmes...) et l'audience sera calculée pour chacune d'elles. L'analyse des chiffres fournis par Médiamétrie est un élément de décision important pour les chaînes qui peuvent analyser la réaction du public face aux programmes qu'elles ont proposé.

		TF1 publique	TF1 privée	France 2	France 3	Canal +	M6
1980	Recettes publiques	84		96	234		
	Publicité	123		105	3		
1985	Recettes publiques	130		152	296		
	Publicité	225		207	80	8	
	Abonnements					75	
1990	Recettes publiques			213	435		
	Publicité		799	223	107	8	92
	Abonnements					789	
1995	Recettes publiques			402	521		
	Publicité		1119	348	235	67	284
	Abonnements					1077	
2000	Recettes publiques			516	623		
	Publicité		1484	329	244	101	537
	Abonnements					1410	
2005	Recettes publiques			670	781		
	Publicité		1561	381	289	91	626
	Abonnements					1377	

TAB. 1: *Comparaison des recettes (exprimées en millions d'euros) des chaînes hertziennes de 1980 à 2005 (sources : rapports financiers et annuels des chaînes)*

Grâce à la multiplication des chaînes, et donc des programmes diffusés à un même moment, le public peut enfin choisir le programme qui lui convient le mieux, indiquant à travers ses changements de chaîne, ses désirs du moment. Les relations entre les producteurs et les chaînes s'inversent : les chaînes vont maintenant exiger des productions qui sont adaptées à leur public, à leur image, et à leurs diffusions en donnant des cahiers des charges plus précis.

Les excès du modèle commercial et concurrentiel se faisant sentir, le gouvernement fixe dès 1986 des contraintes aux diffuseurs et crée un système de soutien financier pour les producteurs. La Commission nationale de la communication et des libertés, qui deviendra en 1989 le Conseil supérieur de l'audiovisuel (CSA), est chargée de veiller à l'exécution de ces consignes. Dès ses débuts, afin d'infirmer son rôle de censeur, il est décidé que le CSA ne peut intervenir qu'après la diffusion.

⁹www.mediametrie.fr

3 La télévision aujourd'hui

3.1 La multiplication des moyens de diffusion

Aujourd'hui, la télévision est une industrie où règne plus que jamais une forte concurrence. La concurrence a en effet été renforcée par la multiplication des moyens de diffusion qui favorise la multiplication des chaînes (Chaniac et Jézéquel, 2005). Cette croissance des moyens de diffusion peut parfois être à l'origine d'une mésentente entre les chaînes : par exemple, TF1, qui participe au capital de TPS, a choisi de n'être diffusée ni sur CanalSat, ni sur les bouquets ADSL autre que TPS ADSL. Cela a contribué fortement à la concurrence entre TF1 et Canal+ notamment lors de l'achat des droits de la coupe du monde de football 2006.

Dans un premier temps, le câble a permis de contourner la limitation du nombre de canaux propre au réseau hertzien. De plus, il propose une qualité d'image supérieure, des chaînes de télévision étrangère, des chaînes de télévision exclusives (payantes ou gratuites) et, dans sa seconde génération, des services interactifs. En France, les premiers travaux de câblage ont été lancés au début des années 80 mais ce n'est qu'à partir de 1990 que les réseaux sont capables de proposer une vingtaine de chaînes.

C'est à peu près au même moment que le satellite a été disponible en France. Ce n'est qu'à partir de 1996 que les bouquets proposent une diffusion exclusivement numérique. L'équipement des foyers évolue très vite par le redoublement des offres des opérateurs (premiers mois gratuits, installation de la parabole offerte). La zone de couverture étant bien plus grande que celle du câble, l'essor de la télévision par satellite n'en est que beaucoup plus rapide. Pour concurrencer le satellite, les réseaux câblés se lancent donc dans une numérisation de la diffusion.

La modernité de ces moyens de diffusion permet une multiplication fulgurante du nombre de chaînes. Entre 1996 et 1998, la France passe d'une quinzaine de chaînes françaises à plus de cent. Dans la majorité, ces chaînes sont gratuites, mais des chaînes payantes sont aussi proposées, notamment des chaînes de cinéma. On distingue parmi ces nouvelles chaînes (Chaniac et Jézéquel, 2005) :

- **les chaînes thématiques** : ce nom est injustement attribué à beaucoup de nouvelles chaînes, alors qu'elles sont en réalité bien peu à être réellement concentrées sur un thème, un sujet, un contenu. Il s'agit par exemple des chaînes consacrées à l'histoire (*Histoire*), l'information (*LCI*), la cuisine (*Gourmet TV*), les voyages (*Liberty TV*), la science fiction (*SciFi*)... Elles peuvent diffuser différents genres d'émissions (magazines, documentaires...) et toucher un public large ;
- **les chaînes dédiées à un genre** : à travers un genre télévisuel unique, elles abordent les sujets les plus variés. On peut citer *Série Club* entièrement consacrée aux séries et aux feuilletons, les chaînes documentaires comme *Odyssée* et *Planète*. La distinction entre les chaînes thématiques et les chaînes dédiées à un genre est difficile : une chaîne qui ne retransmet que des événements sportifs est une chaîne dédiée à ce type d'émissions, alors qu'une chaîne thématique sur le sport proposera en plus des retransmissions des magazines, des documentaires ;
- **les chaînes destinées à un public spécifique** : contrairement aux chaînes généralistes qui tentent de fédérer l'auditoire le plus large possible, ou qui s'adressent à un public spécifique à certains horaires, ces chaînes s'adressent à un public cible restreint. Le public peut être relativement large, par exemple les femmes (*Téva*), les jeunes filles (*Filles TV*), ou alors concerner une tranche d'âge, par exemple les enfants (*Tiji*, *Fox Kids*), ou encore

une communauté (*KTO TV*, *TFJ TV*, *Pink TV*) ;

- **les chaînes à concept** : Elles ne se distinguent ni par les genres d'émissions diffusés, ni par le public qu'elles visent, mais plutôt par le message qu'elles veulent faire passer, par une dominante, une couleur. Ainsi, la directrice de *Canal Jimmy* en 2002 déclare que les mots d'ordre de sa chaîne sont « rock, rebelle et dénicheur » ;
- **les chaînes quasi généralistes** : elles sont qualifiées de généralistes non pas pour la variété de genres d'émissions qu'elles proposent (qui sont la plupart du temps une alternance entre des fictions et du télé-achat) mais pour le large auditoire qu'elles touchent. Parmi elles, on peut citer *TMC*, *Paris Première* ou encore *TF6*.

L'avènement de la télévision numérique entraîne aussi la création de deux nouveaux moyens de diffusion. Si la télévision par ADSL, initiée à la fin de l'année 2003 par le fournisseur d'accès à Internet Free n'a pas suscité la création de nouvelles chaînes, l'introduction de la Télévision Numérique Terrestre, lancée en mars 2005, propose des canaux pour les anciennes chaînes hertziennes, pour 9 nouvelles chaînes gratuites et d'autres chaînes privées.

Le tableau 2 montre les parts de marché des chaînes hertziennes et des autres chaînes du câble et satellite entre 1989 et 2005. Dans ce tableau, le cinquième canal représente la Cinq (privée) en 1989 et 1991, puis Arte en 1993, puis Arte et la Cinq (France 5) à partir de 1995, et la catégorie Autres TV représente toutes les chaînes du câble et satellite, y compris les chaînes étrangères. Il est clairement visible que ces autres chaînes de télévision semblent gagner petit à petit l'intérêt des téléspectateurs, même s'il faut relativiser puisqu'elles sont quelques centaines à se partager 12,1% de part de marché en 2005.

	1989	1991	1993	1995	1997	1999	2001	2003	2005
TF1	41,0	42,1	41,0	35,1	35,0	35,1	32,7	31,5	32,3
France 2	23,4	21,3	24,7	22,3	23,7	22,3	21,1	20,5	19,8
France 3	10,3	11,3	14,6	16,3	17,1	16,3	17,1	16,1	14,7
Canal +	4,4	4,6	4,9	4,5	4,5	4,5	3,6	3,7	3,6
5 ^{ème} canal	13,0	10,9	0,9	3,6	3,3	3,6	3,4	4,7	4,9
M6	6,4	8,0	11,2	13,6	12,7	13,6	13,5	12,6	12,6
Autres TV	1,5	1,8	2,7	4,6	3,7	4,6	8,6	10,9	12,1

TAB. 2: *Évolution des parts de marché annuelles (en %) de 1989 à 2005*
(source : Médiamétrie)

3.2 Vendre son audience aux annonceurs

Les annonceurs n'ont à présent que l'embarras du choix pour diffuser leurs publicités. L'audience d'une émission ou d'une tranche horaire est un moyen pour les chaînes de les attirer. Aussi, les émissions qui ne réalisent pas une audience suffisante selon les critères de la chaîne sont évincées rapidement de la diffusion. Par exemple, en février 2006, TF1 diffuse *le Royaume*, une émission de télé réalité en 14 épisodes, le samedi soir à 20h50 : les parts de marché des 4 premiers épisodes étant inférieurs à 20%, l'émission est supprimée. Comme le formule Eric Macé dans (Macé, 2006), « *la télévision ne vend pas ses programmes à ses publics mais son audience à ses annonceurs* ».

Dans la course effrénée à la part de marché, les chaînes tentent de se distinguer par leur couleur et par le choix des programmes qu'elles diffusent. M6 pour rivaliser avec TF1 a introduit en France, en 2001, le concept de la télé-réalité avec *Loft Story*, réalisant des scores d'audience incroyables pour la chaîne. TF1 et Canal Plus ripostent avec de nouveaux concepts

de télé-réalité, respectivement *Nice People* et *the Simple Life*. Arte a aussi proposé des émissions de télé-réalité (*7 en route*, *la Vie de château*), mais dans un concept plus culturel : par exemple, dans *7 en route*, sept jeunes européens sillonnent l'Europe dans un bus et réalisent des reportages à chacune des étapes qui leurs sont proposées. Seul France 2, qui est la chaîne publique destinée à concurrencer TF1 malgré ses obligations de service public (les émissions religieuses du Dimanche, par exemple), refuse de rentrer dans ce jeu, introduisant le terme « télé-poubelle » (Meyer, 2006). L'obligation d'obtenir des résultats rapidement empêche les chaînes d'attendre que les programmes se rodent. En revanche, les émissions à succès sont très vite dupliquées, ce qui a tendance à user le concept de l'émission prématurément.

La concurrence se traduit également par l'envie des chaînes d'accroître leur public cible. M6, qui privilégiait jusqu'alors un public jeune (*Hit Machine*) et plutôt féminin (*Belle et zen*), a décidé depuis quelques années d'élargir son public cible au public masculin. Ainsi en 2006, M6 a diffusé pour la première fois des matchs de la coupe du monde de Football, alors que la chaîne se vantait en 1998 d'être « 0% foot ». Cette tendance a été confirmée à la rentrée 2006 par Thomas Valentin, directeur des programmes de M6, pour qui le but de la chaîne est d'« élargir [son] public à tous les moins de 50 ans » (source : Agence Page 30).

Enfin, la dernière conséquence de la concurrence est la convoitise des présentateurs. Le succès d'une émission est lié sans aucun doute à son concept, mais une part de son succès est dû aussi à son présentateur. La période estivale, durant laquelle les productions de programmes de flux sont arrêtées, est la période propice aux décisions de maintien ou pas d'une émission, de l'introduction de nouveaux programmes et du changement des présentateurs. L'été 2006 a été marqué par des transferts de présentateurs d'une chaîne à l'autre, dignes du *mercato* du Football.

C'est dans ce contexte de rivalité que coexistent les chaînes de télévision actuellement. Ce marché concurrentiel est un poids énorme sur les épaules du programmateur, dont nous explorerons plus en détail la fonction dans le chapitre 5 (page 43).

3.3 La télévision interactive et mobile

Depuis le début de la télévision, le téléspectateur a toujours été tributaire des horaires des diffusions ; dans ce cas, la télévision est un média de masse, qui touche au même moment une grande partie de la population. Avec l'apparition des magnétoscopes, cela avait un peu changé : il suffisait d'enregistrer un programme et de le visualiser plus tard. Avec l'apparition de terminaux plus évolués (set-top-box), il est aujourd'hui possible d'interrompre momentanément un programme (*time shifting*). Tout est fait pour le confort du téléspectateur : par exemple, on parle aujourd'hui d'une télévision *haute définition* avec une image et un son d'une qualité supérieure à celle d'aujourd'hui.

La convergence des technologies issues de l'informatique, de la communication et de la télévision a fait apparaître des services interactifs, comme la météo ou les guides de programmes, qui dans quelques années joueront des rôles de plus en plus importants. Ainsi, l'interactivité atteint son apogée avec la vidéo à la demande. Le téléspectateur pourra choisir les programmes qu'il veut, ou l'accès à une chaîne quand il souhaite : il en naîtra un certain individualisme quant à l'écoute et aux choix des programmes (Chaniac et Jézéquel, 2005). La télévision mobile, c'est-à-dire visible sur des téléphones mobiles, des assistants personnels ou des lecteurs multimédias portables (par exemple l'*iPod*), se développe. On parle même de l'apparition d'un format d'émission court pour la télévision mobile (INA, 2005). Par exemple, la société de production *Endemol* a profité de la diffusion de la *Star academy* pour tester le marché en proposant dans un premier temps des résumés de ces émissions dans un format très court.

Chapitre 5

La programmation d'une chaîne

Sommaire

1	Programmation et grilles de programmes	43
2	Construction d'une grille de programmes	45
3	Principes et règles de la programmation	48
3.1	Concurrence et chaîne de référence	48
3.2	Fidélisation du public	48
3.3	Changement de programmation	48
3.4	Contre-programmation et <i>blunting</i>	49
3.5	Programmation par blocs	50
3.6	Fréquence de diffusion	50
4	Impacts de différents facteurs sur les grilles de programmes . .	50
4.1	Type de chaîne	50
4.2	Législation	51
4.3	Audience et déprogrammation	51
5	Les différentes grilles de programmes	52
6	Grilles prévisionnelles et guides de programmes	52
6.1	Obtention des grilles prévisionnelles	53
6.2	Mesure de l'imprécision et l'incomplétude des grilles prévisionnelles	53
7	Stabilité des grilles de programmes	56
7.1	Explications de la stabilité	56
7.2	Éléments de mesure de la stabilité des grilles de programmes de TF1	56

Dans le chapitre 4 (page 37), nous avons pris l'exemple de la télévision française pour montrer l'évolution prévisible de cette industrie vers un système concurrentiel. Au cours de cette évolution, la rationalisation de la diffusion a donné naissance à une représentation sous forme d'une grille de programmes dont l'élaboration est devenu le cœur de métier du programmeur, ou encore éditeur de programmes (Fonnet, 2003).

1 Programmation et grilles de programmes

La programmation d'une chaîne de télévision est l'organisation de la rencontre entre une émission et un public (Fonnet, 2003; Mousseau, 1989). Elle permet à une chaîne de satisfaire

au maximum les attentes de son public cible. Il résulte de la programmation une grille de programmes.

Une grille de programmes est une représentation de la répartition, généralement sur une semaine, des émissions en fonction des jours de la semaine et des heures. On distingue deux types de programmation (et donc de grilles). La programmation horizontale consiste à diffuser tous les jours à la même heure la même émission, par opposition à la programmation verticale qui consiste à diffuser la même émission en fonction du jour de la semaine. La figure 10 présente un extrait de la grille de programmes de France 2 de janvier à juin 2006. Les couleurs de la grille permettent de distinguer très clairement la programmation horizontale (du lundi au vendredi, entre les deux journaux télévisés de 13 heures et 20 heures) et la programmation verticale (du lundi au vendredi, après le journal de 20 heures). Une grille n'est pas qu'un simple agrégat de programmes autonomes (Fonnet, 2003; Chaniac et Jézéquel, 2005; INA, 1992). Il s'agit plutôt d'une articulation d'émissions, cohérente avec la ligne éditoriale de la chaîne, où chaque programme prend sa valeur à travers les relations établies avec les autres. Le rôle du programmeur est donc de cerner les attentes, les envies d'un public cible à travers son mode de vie. Son rôle est de construire une grille de programmes type correspondant aux attentes du public visé (voir la section 2, page 45), dans le respect de la ligne éditoriale de la chaîne, puis d'acquérir les programmes audiovisuels. Cette grille de programmes a une double vocation : elle permet dans un premier temps de marquer l'identité d'une chaîne, et dans un second temps de financer l'activité de celle-ci.

Dans un système concurrentiel, les chaînes de télévision doivent se démarquer les unes des autres. La politique de programmation est le meilleur moyen de marquer son identité et fidéliser en même temps son public (Burtin, 2004). La programmation peut être vue comme un acte technique qui formalise la politique éditoriale, elle-même inscrite dans un projet d'entreprise.

L'autre but de la programmation est d'établir un lien entre la chaîne et ses annonceurs qui lui permettent de financer ses activités et la grille de programmes (voir tableau 3). Une chaîne doit non seulement les attirer mais les garder en leur offrant les audiences les plus hautes et les plus stables possible. La volonté des annonceurs est de toucher au maximum la « ménagère de moins de cinquante ans » parce qu'elle influence beaucoup les achats du foyer. Cette volonté influe évidemment sur la grille de programmes. Les propos maladroits de Patrick Le Lay, PDG de TF1, illustre tout à fait les contraintes du programmeur pour faire la grille de programmes, vis à vis du téléspectateur et des annonceurs :

« Il y a beaucoup de façons de parler de la télévision. Mais dans une perspective business, soyons réaliste: à la base, le métier de TF1, c'est d'aider Coca-Cola, par exemple, à vendre son produit [...]. Or pour qu'un message publicitaire soit perçu, il faut que le cerveau du téléspectateur soit disponible. Nos émissions ont pour vocation de le rendre disponible : c'est-à-dire de le divertir, de le détendre pour le préparer entre deux messages. Ce que nous vendons à Coca-Cola, c'est du temps de cerveau humain disponible [...]. Rien n'est plus difficile que d'obtenir cette disponibilité. C'est là que se trouve le changement permanent. Il faut chercher en permanence les programmes qui marchent, suivre les modes, surfer sur les tendances, dans un contexte où l'information s'accélère, se multiplie et se banalise »¹⁰.

Dans ce but, le programmeur dispose de plus en plus de moyens pour cerner les désirs du public (enquêtes, audiences) et les répercuter sur la grille de programmes. Le programmeur est devenu un expert du public et de ses goûts, lui permettant d'anticiper les aspirations des téléspectateurs.

¹⁰source : dépêche AFP du 9 juillet 2004

France 2 - Janvier 2006 Direction du marketing 

	lundi	mardi	mercredi	jeudi	vendredi	samedi	dimanche			
13:00	Journal de 13 heures					Journal de 13 heures	Journal de 13 heures			
13:15						Hebdo du médiateur	J'ai rdv avec vous			
13:30										
13:45										
14:00	Série Européenne : Rex					Savoir + Santé	Vivement dimanche			
14:15						Savoir+ Sciences				
14:30						Savoir + Santé				
14:45						Savoir + Santé				
15:00	Série Européenne : Un Cas pour deux					Sport	30 millions d'amis			
15:15								ou		
15:30	Série US : JAG					Film TV/Séries	Série US : Nash Bridges			
15:45								Série US : JAG		
16:00	Des chiffres et des lettres									
16:15	Tout vu tout lu									
16:30	Série US : Friends					Le grand zapping de l'humour	Stade 2			
16:45	Série US : Friends									
17:00	On a tout essayé					On a tout essayé... même le samedi	Vivement dimanche prochain			
17:15	Bande Dehouf									
17:30						Samantha Oups				
17:45										
18:00	Le journal de 20 heures									
18:15										
18:30	Film TV	Magazine (Jour après jour)	Film Cinéma	Film TV	Envoyé spécial / A vous de juger	Une soirée de polars 90' (Qual N°1, Maigret, Bd du Palais, ...)	Divertissement (Le plus grand cabaret du monde, Soirées Spéciales, ...)			
18:45								Film TV (suite du lundi)		
19:00	Mots croisés	Complément d'enquête	Mots croisés	Un œil sur la planète	Film Cinéma	Ca se discute	Documentaire	Campus	Tout le monde en parle	Faites entrer l'accusé
19:15	Journal de la nuit					Journal de la nuit	Journal de la nuit	Journal de la nuit	Journal de la nuit	Journal de la nuit
19:30	Musiques au cœur	Histoires courtes	Cine Club	Des mots de minuit	Documentaire	Série US : A la Maison Blanche	Journal de la nuit	Journal de la nuit	Journal de la nuit	Vivement Dimanche Prochain
19:45										
20:00										
20:15										
20:30										
20:45										
21:00										
21:15										
21:30										
21:45										
22:00										
22:15										
22:30										
22:45										
23:00										
23:15										
23:30										
23:45										
00:00										
00:15										
00:30										
00:45										
01:00										
01:15										
01:30										

Programation non contractuelle, sous réserve de modification de la part de l'antenne 03/01/05

LEGENDE

Magazine	Divertissement	Cinéma	Film TV	Série	Météo
Jeu	Information	Mag. d'information	Sport	Jeunesse	Documentaire

FIG. 10: Extrait de la grille de programmes de France 2 de janvier à juin 2006, fournie à ses annonceurs

2 Construction d'une grille de programmes

La construction de la grille de programmes est basée sur le mode de vie de son public cible. En fonction de celui-ci, il faut dans un premier temps découper la semaine en cases. Chacune de ces cases s'adressera à un public cible particulier en fonction de sa disponibilité et de la ligne éditoriale de la chaîne. Dans un second temps, il faut remplir ces cases avec des émissions qui devront maximiser l'audience de téléspectateurs.

Les chaînes de télévision ciblent particulièrement la « ménagère de moins de 50 ans » parce qu'elle est la personne du foyer qui s'occupe des achats. Aussi caricatural que cela puisse paraître, c'est néanmoins en fonction de son emploi du temps (comme il est perçu par le programmeur

après analyse de ses études sociologiques) que les grilles de programmes sont construites (Fonnet, 2003) :

- de 7h30 à 8h30 : la ménagère doit organiser le réveil et la toilette de ses enfants. Pendant qu'elle s'occupe de l'un d'eux, la télévision doit capter l'attention des autres. Cette case contiendra donc des émissions pour les enfants et des dessins animés (*Tfou!* sur TF1) ;
- de 8h30 à 11h30 : pour la ménagère qui ne travaille pas, il s'agit d'un moment privilégié où elle a le temps de s'occuper un peu d'elle-même. Pour participer à ce temps de loisir, la télévision lui propose du télé-achat, des séries télévisées ou des émissions de *call-tv* (*Tubissimo* sur M6) (voir le chapitre 6, page 57). Cela lui permet aussi de s'absenter le temps de quelques activités en dehors de la maison ou de pratiquer des activités ménagères en même temps ;
- de 11h30 à 13h45 : l'auditoire va peu à peu changer avec l'arrivée des enfants et du mari. Cette case comporte généralement des jeux (*les Z'amours* sur France 2) qui permettent de tenir une conversation familiale tout en suivant l'émission. Ces jeux sont généralement précédés d'émissions de service (voir chapitre 6, page 57) ou de magazines courts sur la cuisine (*Julie cuisine*, sur TF1) diffusés avant l'arrivée du mari et des enfants. À 13h, le journal télévisé est suivi devant un café, après le repas. Sa durée est plus courte que celui de 20h, car les départs du foyer vont commencer à 13h15. C'est pour cela que l'essentiel de l'information est condensé dans le premier quart d'heure qui est suivi de reportages pour découvrir des villes, des villages ou encore des métiers ;
- de 13h45 à 16h30 : la ménagère est de nouveau seule. L'après-midi est donc l'occasion de diffuser des séries et des feuilletons principalement destinés aux femmes (*Les feux de l'amour*, sur TF1) ;
- de 16h30 à 18h30 : petit à petit, les enfants rentrent de l'école, puis un peu plus tard, du collège et du lycée. Cette case correspond à un moment de loisir entre l'école et les devoirs : des séries plutôt familiales ou destinées aux enfants vont être diffusées (*Sept à la maison*, sur TF1) ;
- de 18h30 à 20h : petit à petit, toutes les personnes du foyer qui travaillent vont rentrer. Il faut pouvoir discuter de la journée tout en regardant la télévision. Des talk-shows (*On a tout essayé*, sur France 2) et des jeux (*À prendre ou à laisser*) sont donc proposés ;
- de 20h à presque 21h : il s'agit d'une case réservée à l'information, avec le journal télévisé, des bulletins météo, des magazines courts et des émissions de service ;
- de 21h à 22h30 : le type d'émission contenu dans cette case de première partie de soirée dépend plutôt du jour de la semaine (programmation verticale), mais il s'agira toujours de programmes fédérateurs. Par exemple, à partir du vendredi, les soirées sont propices au divertissement, une chaîne diffusera donc plutôt des fictions ou des émissions de variétés ;
- de 22h30 à minuit : la seconde partie de soirée est consacrée à des émissions plutôt regardées par des adultes. Il peut s'agir d'émissions qui demandent plus d'attention, comme des magazines, des talk-shows (*Ça se discute*), ou de fictions ;
- la nuit : enfin, la nuit, les programmes sont plus considérés comme une compagnie que comme un loisir. Compte tenu du faible auditoire, le programmeur ne cherche plus à

plaire au plus grand nombre. Dans cette case, les rediffusions prédominent et coexistent avec des émissions culturelles telles que des documentaires.

TF1	893,2
France 2	731,0
France 3	703,6
Canal+	1000,0
Arte	125,8
France 5	96,7
M6	258,4
Direct 8	20,0

TAB. 3: Coûts (en millions d'euros) de la grille de programmes de plusieurs chaînes en 2004 (source : rapport des chaînes)

Bien sûr, cette grille type est très caricaturale, mais si on regarde les grilles de programmes des chaînes généralistes, on peut reconnaître une démarche similaire (voir la figure 10). Les chaînes qui touchent un public moins large fonctionneront légèrement différemment : par exemple, M6, dont le public cible est jeune, va diffuser des clips musicaux à la hauteur de 30% de sa grille. France 2 et France 3, qui appartiennent au même groupe France Télévisions, vont afficher des grilles complémentaires : quand France 3 diffuse des émissions pour les enfants le matin (*Too-wam*), France 2 peut s'intéresser à un public plus âgé (*Télé matin*).

Aux États-Unis, le découpage est assez similaire et il est d'usage quand on parle de télévision d'employer les termes américains pour désigner une partie de la journée. Le tableau 4 regroupe ces termes ; on y reconnaît les deux moments stratégiques de la journée : l'*access prime time* et le *prime time*.

Early morning		6h - 9h
Day time	Morning	9h - 12h
	Afternoon	12h - 16h
Early fringe		16h - 18h
Early evening		18h - 19h
Access prime time		19h - 20h
Prime time		20h - 23h
Late fringe		23h - 23h30
Late night		23h30 - 1h
Overnight		1h - 6h

TAB. 4: Définitions américaines du découpage de la télévision selon (Mousseau, 1989)

Une grille de programmes n'est jamais figée : en fonction de l'audience, les émissions proposées peuvent être changées (voir la section 4, page 50). La grille peut évoluer en fonction du stock de programmes : si un nouveau programme vient d'être acquis, son placement dans la grille va la transformer. En fonction des fournisseurs, ces programmes peuvent être diffusés sous certaines contraintes (périodicité, horaires...). Des obligations légales et réglementaires peuvent obliger le programmateur à changer le contenu de plusieurs cases. Enfin, la grille de programmes peut évoluer en fonction de contraintes de coût (voir figure 11 et tableau 3) ; par exemple une réduction budgétaire dû à l'achat d'un programme très onéreux obligerait le programmateur à répartir la diminution du budget sur l'ensemble de la grille tout en conservant les mêmes scores d'audience. Comme le montre le tableau 3, les coûts des grilles de programmes atteignent des sommes très élevées : les chaînes ayant un large public à fidéliser sont aussi celles qui ont les coûts les plus élevés, expliqués en partie par l'achat de gros programmes fédérateurs. Notons aussi le coût

de la grille de Canal+ qui s'explique par l'achat massif de fictions américaines et de droits de retransmission de sport.

3 Principes et règles de la programmation

L'histoire et l'expérience ont mis en évidence certaines règles et certains principes de la programmation. Nous considérons ici un téléspectateur passif, c'est-à-dire qui n'a pas le choix de la programmation, contrairement à la télévision interactive. Nous présentons dans cette section les concepts qui nous ont semblé majeurs.

3.1 Concurrence et chaîne de référence

Le programmateur, quand il crée sa grille de programmes, n'est jamais isolé : il évolue dans un système concurrentiel où l'émission qu'il proposera sera diffusée en même temps que plusieurs autres. Laurent Fonnet (Fonnet, 2003), ancien directeur de la programmation de TF1, explique qu'il ne s'agit pas de « diffuser la meilleure offre de programmes mais la moins mauvaise ». Un programme particulier n'a pas une audience assurée ; son audience dépendra en effet de l'émission précédente, mais aussi des émissions qui sont proposées par les chaînes concurrentes. Toutefois, la chaîne doit s'assurer d'avoir l'exclusivité sur les meilleurs programmes afin d'empêcher les chaînes concurrentes de la menacer.

Ce phénomène de concurrence a engendré un objectif nouveau des chaînes : être la chaîne de référence. Il s'agit d'un réflexe pour un téléspectateur qui décide de regarder la télévision sans idée précise d'émission d'allumer son poste de télévision sur une chaîne qui a sa confiance, et qui est susceptible de diffuser un programme qui l'intéresse. Cet objectif est récent puisqu'avant les téléviseurs s'allumaient directement sur le premier canal.

3.2 Fidélisation du public

La fidélisation du public est nécessaire pour tout produit ou pour tout service. Pour la télévision, elle est nécessaire afin de garantir une audience minimale aux annonceurs, qui par le biais de leurs publicités vont financer les activités de la chaîne. La fidélisation repose sur deux principes. D'une part, elle suppose que le téléspectateur est attiré par une chaîne pour son offre globale et non pas parce qu'il est intéressé par un programme en particulier. Le fondement de la fidélisation est la passivité du téléspectateur, c'est-à-dire son envie de se laisser guider par la chaîne ; le téléspectateur a généralement envie de rester sur la même chaîne tout le temps de son écoute, et ne zappera qu'en cas d'insatisfaction (Jost, 2001; Fonnet, 2003; Chaniac et Jézéquel, 2005). D'autre part, la fidélisation est liée à l'image de la chaîne. Elle repose sur la promesse faite par la chaîne de satisfaire son public cible et elle est maintenue par le respect de cette ligne éditoriale. La fidélisation du public a aussi des conséquences économiques pour la chaîne, puisque le coût des programmes attractifs augmente chaque année sous les effets de la concurrence. Ainsi, une chaîne comme TF1 doit maintenir la satisfaction de son public en proposant des programmes inédits à la hauteur de ses attentes, en acceptant d'augmenter le coût de sa grille de programmes (voir figure 11).

3.3 Changement de programmation

Il faut distinguer deux types de changement de programmation. La déprogrammation de dernière minute, appelée *stunting*, n'est pas possible en France à cause des conventions signées

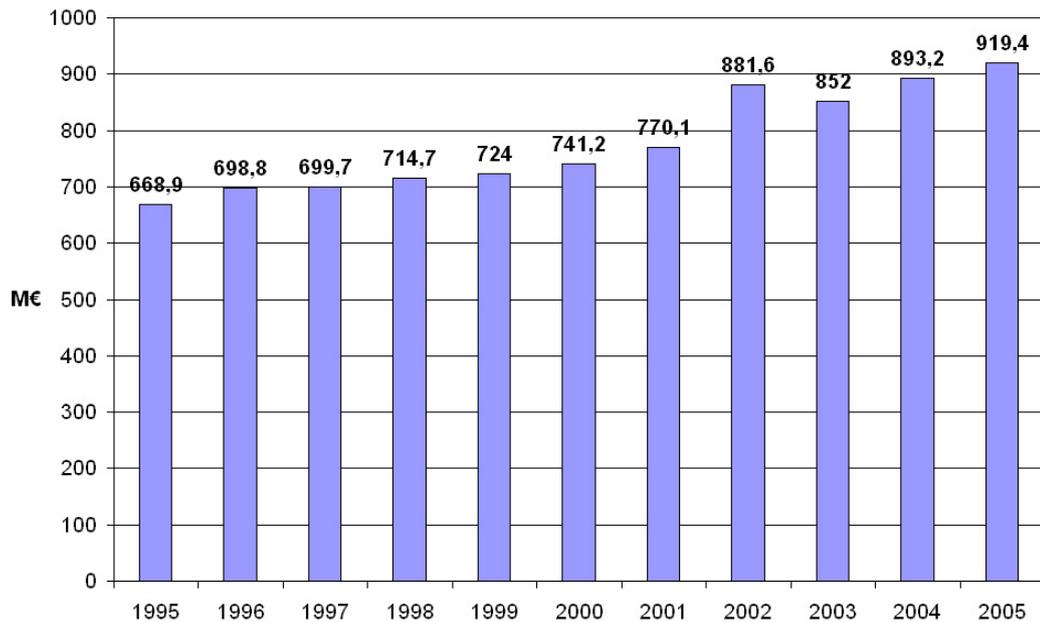


FIG. 11: Évolution du coût de la grille de programmes de TF1 entre 1995 et 2005 (source : rapports financiers de TF1)

avec le CSA ; en cas d'événement d'actualité ou sportif non prévisible, elle est toutefois tolérée. Ce changement présente un risque pour la chaîne : si le téléspectateur ne trouve pas le programme auquel il est habitué, il va souvent regarder les offres de la concurrence pour faire une comparaison avec le nouveau programme proposé. En fonction de ses choix, il faudra alors remplacer le nouveau programme par un programme plus attractif. Un changement peut être aussi nécessaire en cas de faible audience mais il est souvent périlleux : par exemple, France 2 en 2001 a arrêté la diffusion de *la Chance Aux Chansons*, perdant ainsi le public âgé sans pour autant gagner le public plus jeune (Fonnet, 2003).

3.4 Contre-programmation et *blunting*

La contre-programmation consiste à opposer à un programme *leader* un programme qui attirera un public différent (Fonnet, 2003) ou en proposant un genre totalement différent de ce qui est proposé à cet horaire. M6 s'est développée avec succès sur ce principe, en opposant un magazine (*Capital* ou *Zone interdite*) à la fiction de TF1 (*Ciné Dimanche*), ou en proposant des programmes plutôt féminin en même temps que des matchs de football.

Le contraire de la contre-programmation est appelé le *blunting*, ou contre-programmation brusque. Cette méthode consiste à diffuser un programme analogue à la principale chaîne concurrente. Cette technique est risquée car il faut être persuadé de la supériorité de son programme pour la tenter : seules les grandes chaînes peuvent se permettre de tels affrontements. TF1 propose par exemple des séries à succès le samedi soir (*LOST*) pour s'opposer au succès de la *Trilogie du samedi* proposée par M6 à la même heure.

3.5 Programmation par blocs

Nous avons vu que la grille de programmes n'était pas une succession d'émissions décorréées les unes des autres. Ce principe est d'autant plus vrai que les programmeurs font usage d'une stratégie appelée *programmation par blocs*. Un bloc est un ensemble cohérent constitué de trois parties : le *lead-in*, le *leader* et le *lead-off*. Le *lead-in* et le *lead-off* sont deux programmes dont les audiences sont fortes et assurées, et le *leader* est un programme nouveau qui doit faire ses preuves. Cette technique repose encore une fois sur la passivité du public à qui on ne donne pas l'occasion de changer de chaînes pendant la durée du bloc. Cette technique est aussi appelée *hammocking*.

3.6 Fréquence de diffusion

Deux méthodes ont été identifiées. Le *stripping* est la programmation d'une émission tous les jours, à la même heure, pour habituer le public à ce rendez-vous. En France, il s'agit souvent de séries qui ont été diffusées auparavant à un rythme hebdomadaire. Par exemple, M6 avant de diffuser une saison inédite d'une série rediffuse les saisons précédentes en *stripping*. À l'inverse, le *checkerboarding* consiste à diffuser une série différente chaque jour à la même heure. En France, il s'agit plutôt d'un *checkerboarding* hebdomadaire : par exemple TF1 diffuse une série policière différente tous les jeudis dans la tranche horaire *le policier du jeudi*.

4 Impacts de différents facteurs sur les grilles de programmes

Afin de confirmer la validité de notre approche, nous nous sommes intéressés à l'identification des facteurs de changement de la grille de programmes, afin d'évaluer la fréquence d'apparition de ces facteurs. Le type de la chaîne a des conséquences évidentes sur la grille de programmes, puisque cette dernière doit être en accord avec la ligne éditoriale de la chaîne. Les changements de réglementation, dont l'application est surveillée par le CSA, ont parfois bouleversé les grilles mais sont annoncés à l'avance. En revanche, les déprogrammations et les taux d'audience sont des éléments plus difficiles à prendre en compte du fait qu'ils interviennent respectivement au dernier moment ou après diffusion.

4.1 Type de chaîne

Nous allons présenter les éléments remarquables des grilles de programmes en fonction de la nature de la chaîne.

Chaînes généralistes gratuites

Nous distinguerons les chaînes généralistes commerciales des chaînes publiques. Les chaînes généralistes commerciales gratuites ciblent généralement la *ménagère de moins de 50 ans*. Ces chaînes considèrent la grille type que nous avons développée à la section 2 (page 45). En plus de ces grilles, les chaînes généralistes commerciales gratuites proposent régulièrement des événements exceptionnels (par exemple, le concert de Johnny Hallyday sous la tour Eiffel) pour enrichir leur grille et sortir des habitudes.

D'un autre côté, les chaînes généralistes publiques ne cherchent pas la plus large audience sur les cibles commerciales puisque la publicité n'est pas leur seule source de revenus. Elles ont en plus des obligations de service public comme par exemple la diffusion d'émissions religieuses le dimanche matin ou les décrochages régionaux. La différence avec les chaînes commerciales

ne réside pas tellement dans la structure de la grille mais plutôt par le contenu de ses cases. Les chaînes publiques peuvent en effet se permettre de diffuser, par exemple, des films moins susceptibles de faire une grande audience, mais qui « méritent » d'être vus. Comme la population âgée de plus de 50 ans représente souvent plus de 60% de l'audience totale, les émissions des chaînes publiques s'adressent principalement à ce public avec la diffusion de jeux intellectuels (*Question pour un champion* sur France 3) ou des émissions de variétés sur les vieilles chansons françaises (*la Chance aux chansons*, autrefois sur France 2).

Les chaînes payantes

L'objectif de ces chaînes n'est pas d'avoir la plus grande audience mais de satisfaire leurs clients. Cette satisfaction est le gage que ceux-ci renouvelleront leur abonnement. Généralement, ces chaînes proposent des émissions que le téléspectateur ne peut pas trouver sur les autres chaînes au même moment. L'impact n'a pas lieu sur la structure de la grille, mais plutôt sur le contenu : le nombre de rediffusions est important et la diversité des types de programmes est faible. La particularité des chaînes thématiques est d'avoir souvent une programmation horizontale.

4.2 Législation

Le secteur audiovisuel est soumis à un grand nombre de lois et de décrets. De plus, lors du renouvellement des autorisations d'émettre (pour une dizaine d'années en général), les chaînes signent un accord avec le CSA. Certaines de ces lois et certains de ces accords touchent directement la programmation : volume minimal ou interdiction de diffusion de certains types d'émission, autorisations de diffusion de certains types d'émissions à un certain horaire ou certains jours. Le programmeur doit s'assurer, après la mise en place de la grille, que celle-ci est bien conforme avec ces réglementations. À titre d'exemple, 16% du chiffre d'affaires des chaînes doivent être investis dans des œuvres audiovisuelles françaises. Les accords peuvent varier d'une chaîne à l'autre : par exemple, TF1 et M6 doivent diffuser des fictions européennes de 18h à 23h sauf le mercredi entre 14h et 23h. Il leur est aussi interdit, depuis les années 90, de diffuser des œuvres cinématographiques le mercredi à partir de 20h, et du vendredi 20h au dimanche 20h30. Nous verrons plus loin la réglementation de la publicité en France.

4.3 Audience et déprogrammation

Deux autres facteurs peuvent avoir des impacts sur les grilles de programmes : la déprogrammation et l'audience, qui peuvent être parfois liées. Nous avons dit précédemment que la diffusion d'une émission pouvait être interrompue faute d'audience suffisante (voir l'exemple donné dans la section 3.2, page 41) : la chaîne doit alors attendre 15 jours (selon un accord avec les éditeurs de magazines sur la télévision) pour arrêter sa diffusion. L'audience permet donc d'évaluer la qualité des enchaînements entre les émissions et les préférences du public par rapport aux émissions de la concurrence. Elle peut aussi avoir un effet sur le contenu d'une émission, en analysant cible par cible la réaction par rapport à chaque partie de l'émission. Le contenu d'une case peut être déplacé dans une autre case en fonction de ses résultats et de l'état de la concurrence. Nous atténuerons plus loin l'aspect imprévisible des impacts de l'audience sur une grille de programmes (voir section 7, page 56).

En revanche, la déprogrammation dite *stunting*, qui est tolérée lors d'événements d'actualité (par exemple le 11 septembre 2001 ou encore la mort du pape en 2005) ou sportifs (par exemple, un match reporté), est un casse-tête pour le programmeur lui-même. En effet, les nouvelles

émissions imposées à la grille (par exemple des bulletins d'information courts) étant souvent de durée indéterminée, le programmateur doit pouvoir combler sa grille par des émissions jusqu'à ce qu'il puisse se rattacher à une émission prévue.

Enfin, la déprogrammation d'une collection entièrement diffusée, par exemple après le dernier épisode d'une saison de série télévisée, est moins problématique. Pour ne pas bousculer les habitudes de son public, la chaîne remplace généralement ce programme par un programme de même type.

5 Les différentes grilles de programmes

Rares sont les cas où une chaîne dispose d'une grille de programmes unique. On peut distinguer les grilles de programmes suivantes :

- **la grille des vacances scolaires** (hors vacances d'été) : les matinées sont consacrées aux émissions pour les enfants (comme le mercredi hors vacances scolaires). Les autres cases de la grille sont maintenues. Parfois, les émissions quotidiennes en direct sont remplacées par des compilations de moments de ces émissions ;
- **la grille des vacances d'été** (juillet - août) : comme pour les vacances scolaires, les matinées sont réservées aux enfants. En revanche, les vacances d'été sont aussi les vacances des animateurs et des maisons de production : tous les programmes de flux¹¹ (par exemple, des magazines en direct comme *Combien ça coûte ?*) sont remplacés soit par des compilations de ces programmes, soit par des programmes de stock¹¹ (par exemple des fictions). Certains jeux télévisés se poursuivent, d'autres sont remplacés par des jeux événementiels (par exemple, *Qui veut gagner des millions ?*) ;
- **la grille TV-réalité** : ce genre de grille est apparu en 2001 en même temps que les émissions de télé-réalité qui se déclinent en plusieurs formats (*Loft story*, *Star academy*). En effet, la grille est bouleversée par les multi-diffusions des émissions quotidiennes, qui sont des résumés des dernières vingt-quatre heures de ces émissions, et par les *primes*, qui sont des émissions de format plus long diffusées en prime-time, dont le contenu est entre celui de l'émission de variétés et celui d'un magazine ;
- **la grille événementielle** : nous en avons déjà parlé, il s'agit de la grille bouleversée par un événement d'actualité ou de sport ;
- **la grille standard** : il s'agit de la grille générale utilisée pour tous les autres jours de l'année.

6 Grilles prévisionnelles et guides de programmes

Les grilles prévisionnelles, appelées aussi guides de programmes, sont des représentations partielles - plus précisément incomplètes et imprécises - des grilles de programmes. Elles ont pour vocation d'aider le téléspectateur à choisir un programme. Les programmes présentés dans ces grilles sont donc des programmes fédérateurs et dont la durée dépasse un certain seuil. En pratique, toute émission de moins de 5 minutes n'est pas présentée : il s'agit essentiellement de bulletins météo, de résultats de courses (hippiques), de résultat de loteries ou de jeux, de

¹¹voir le chapitre 6, page 57

magazines courts sponsorisés et d'émissions de service (trafic routier). Les interprogrammes, bandes-annonces et publicité, ne sont pas non plus présentés.

6.1 Obtention des grilles prévisionnelles

Il existe plusieurs manières d'obtenir ces grilles prévisionnelles, encore appelées guide de programmes :

- **la presse** : les magazines spécialisés paraissent une semaine avant la diffusion et proposent des commentaires et des résumés sur les programmes. La presse quotidienne présente généralement le guide de programmes du jour ou parfois du lendemain. Il est évident que ces guides de programmes ne peuvent bénéficier d'une mise à jour ;
- **les EPG (Electronic Program Guides)** : les guides de programmes électroniques sont diffusés sur Internet ou directement sur les écrans de télévision via le fournisseur (satellite, câble, ADSL, TNT). TV-Anytime¹² propose d'adopter un standard pour que la création d'applications pour la télévision, notamment l'accès aux guides de programmes, soit facilitée. Même s'ils ne sont pas plus précis que les précédents, ils peuvent bénéficier de mises à jour quotidiennes.

En France, la fourniture de guides de programmes est presque intégralement assurée par *les Agences TV*¹³ (regroupant *Intermonde-Prescott* et *iMédia-Press*). Ces entreprises s'occupent même de la mise en page des guides de programmes pour les magazines ou les sites Internet (par exemple celui de *Télé 7 jours*) et propose jusqu'à trois mises à jour quotidiennes.

6.2 Mesure de l'imprécision et l'incomplétude des grilles prévisionnelles

Nous avons étudié les guides de programmes afin de vérifier leur incomplétude et leur imprécision. Dans un premier temps, nous avons mesuré sur les chaînes hertziennes la différence entre le nombre d'émissions présentées dans les guides de programmes et le nombre d'émissions effectivement diffusées. Nous avons mené cette étude sur quatre années entières, de 2002 à 2005. Compte tenu du nombre de grilles prévisionnelles rentrant en jeu, nous avons dû nous contenter d'une estimation de cette incomplétude. Pour cela, nous avons construit les guides de programmes à partir des grilles de programmes passées, en retirant tous les programmes dont la durée était inférieure à cinq minutes ou dont le genre n'apparaît jamais dans un guide de programmes (par exemple, la météo). Nous avons validé cette approche en utilisant quelques semaines de guides de programmes et nous nous sommes aperçus que nous ne considérons pas quelques émissions humoristiques, à base de caméras cachées, utilisées ces dernières années pour combler les grilles de programmes (*Blagadonf* sur France 2) : celles-ci ne sont pas annoncées, mais une dizaine d'entre elles ont échappé à notre sélection parce qu'elles dureraient un peu plus de cinq minutes. Deux études ont été menées parallèlement : nous avons tantôt mesuré cette incomplétude en considérant les bandes-annonces et les publicités indépendamment les unes des autres, puis comme un ensemble du type interprogramme (voir tableau 5). Il est important de distinguer l'incomplétude en terme de nombres de programmes, qui atteint un niveau toujours supérieur à 60%, et l'incomplétude en terme de durée, qui présente un taux plus faible aux alentours de 9%. Les émissions qui ne sont pas présentées dans un guide de programmes sont donc des émissions qui ne représentent en moyenne que deux heures dans une journée entière.

¹²www.tv-anytime.org

¹³www.lesagencesTV.com

	TF1	France 2	France 3	M6
Nombre minimum d'émissions diffusées	47 / 74	38 / 58	43 / 69	40 / 91
Nombre moyen d'émissions diffusées	77 / 126	80 / 140	81 / 161	69 / 130
Nombre maximum d'émissions diffusées	100 / 159	106 / 199	104 / 215	93 / 190
Nombre minimum d'émissions non annoncées	29 / 56	27 / 55	25 / 54	24 / 75
Nombre moyen d'émissions non annoncées	51 / 100	53 / 114	50 / 130	46 / 108
Nombre maximum d'émissions non annoncées	72 / 130	78 / 169	67 / 176	67 / 167
Durée totale minimum des émissions non annoncées	1 :23 :37	1 :00 :06	0 :44 :52	0 :53 :31
Durée totale moyenne des émissions non annoncées	2 :21 :43	2 :32 :58	2 :16 :13	2 :01 :27
Durée totale maximum des émissions non annoncées	4 :16 :31	4 :12 :33	5 :15 :21	2 :53 :14
Incomplétude moyenne d'une grille prévisionnelle en terme de programmes	66,6% / 79,6%	66,9% / 81,2%	61,7% / 80,6%	67,1% / 82,6%
Incomplétude moyenne d'une grille prévisionnelle en terme de durée	9,8%	10,6%	9,4%	8,4%

TAB. 5: Estimation du nombre de programmes annoncés par rapport au nombre de programmes diffusés sur quatre chaînes hertziennes entre 2002 et 2005. Les couples de chiffres représentent les résultats si on considère les publicités et les bandes-annonces successives comme un seul interprogramme, puis les résultats si les bandes-annonces et les publicités sont considérées séparément.

Nous avons enregistré ensuite une semaine de plusieurs chaînes sur deux fournisseurs - Free (fourni par IntermondePrescott) et NOOS (fourni par iMédia Presse) - afin de vérifier l'imprécision des EPG. Pour cela, nous avons relevé certaines incohérences entre le guide et le programme diffusé (voir figure 12).

La figure 12a montre que certaines émissions sont omises dans l'EPG : l'émission *Samantha*, diffusée entre *On a tout essayé* et *Mon plus grand moment de cinéma* ne figure pas dans le guide, même lorsqu'elle est en cours de diffusion. Les figures 12b,c,d,f montrent des exemples de désynchronisation de l'EPG et du flux : le programme annoncé ne correspond pas à celui qui est diffusé à cause d'un décalage avec les horaires prévisionnels. Enfin la figure 12e est un cas plus grave où le guide n'a pas été rafraîchi. Tout cela est dû au fait que l'EPG et le télétexte sont diffusés avec le flux audiovisuel avec un taux de rafraîchissement dépendant du canal exploité sans aucun contrôle de la synchronisation.

Cette étude montre bien que les guides de programmes ne reflètent pas suffisamment le contenu d'un flux télévisuel.



(a) Exemple d'émission non annoncée par l'EPG



(b) Exemple de retard



(c) Exemple de désynchronisation de l'EPG



(d) Exemple de désynchronisation de l'EPG



(e) Exemple de non disponibilité de l'EPG



(f) Exemple de désynchronisation de l'EPG

FIG. 12: Décalages entre le guide de programmes et la diffusion

7 Stabilité des grilles de programmes

7.1 Explications de la stabilité

Afin de fédérer le plus grand nombre possible de téléspectateurs dans leur public cible, les programmes sont proposés en fonction des tranches de vie quotidiennes et l'on observe donc très tôt une certaine régularité des grilles. Même au temps de la diffusion discontinue, les programmes se sont regroupés autour du journal de 20h ; puis lorsque la diffusion a été proposée l'après-midi, ces autres émissions se sont regroupées autour du journal de 13h, et ainsi de suite jusqu'à la diffusion que nous connaissons aujourd'hui.

Lorsqu'une télévision est en régime concurrentiel, la régularité de ses grilles devient un impératif (Chaniac et Jézéquel, 2005). En effet, l'achat d'espaces publicitaires par les annonceurs intervient forcément avant la diffusion et s'appuie donc sur une anticipation de l'audience. Ainsi le diffuseur doit être capable de fournir une audience stable, lissée de toute perturbation. C'est pour cela qu'une chaîne va offrir le menu le plus constant, le plus facilement mémorisable. Laurent Fonnet annonce dès le début de son livre (Fonnet, 2003) cette régularité :

« [...] les choses changent beaucoup moins qu'on ne l'imagine. Certes, tous les films de fiction française sont différents, ainsi que deux programmes de variétés, voire deux émissions d'un même programme, mais cela est moins vrai au niveau de la case de programmation, voire des grilles de programmes d'une saison audiovisuelle. De nouveaux programmes apparaissent dans les grilles, mais globalement les grilles de programmes changent peu si l'on accepte le principe que pour un certain niveau d'analyse et de problèmes à résoudre, un divertissement reste un divertissement, une fiction française reste une fiction française, avec leurs caractéristiques communes et leurs contraintes. »

La volonté de régulariser ces grilles entraînent une sérialisation des programmes, d'où l'essor des feuilletons et des séries télévisés. Ce phénomène n'est pas propre à la télévision française mais à n'importe quelle télévision en système concurrentiel (Chaniac et Jézéquel, 2005). Toutefois, la monotonie engendrée par la régularité des grilles est parfois brisée par la diffusion d'événements exceptionnels.

7.2 Éléments de mesure de la stabilité des grilles de programmes de TF1

Si l'on prend par exemple la chaîne TF1, qui est pour les autres chaînes la cible à atteindre en terme de part de marché, on s'aperçoit que la case 19h-20h est occupée en semaine par un jeu télévisé (*l'Or à l'appel*, *le Bigdil'*, *Qui veut gagner des millions ?*, *À prendre ou à laisser*, *la Roue de la fortune* et bien d'autres se sont succédés) depuis 1997 (sauf en période de télé-réalité telle que *Star academy* ou *la Ferme célébrités*), et le dimanche par un magazine d'information depuis 1995 (*7 sur 7*, *19:00 dimanche*, *Sept à huit*). De même, le *prime time* du Dimanche soir était consacré à la tranche horaire *Ciné Dimanche* depuis 1995 ; celle-ci a pris fin à la rentrée 2006 pour être remplacée par des séries américaines (*les Experts*) pour concurrencer celles de France 2. On note aussi la présence du magazine *Combien ça coûte ?* le mercredi soir, de façon hebdomadaire, depuis 1995.

Nous avons étudié aussi les séquences de programmes remarquables, c'est-à-dire lorsque deux types d'émissions se succèdent ou non systématiquement. Ainsi, nous avons remarqué que chaque fois que les résultats des paris hippiques étaient diffusés, ils étaient systématiquement suivis d'un bulletin météorologique.

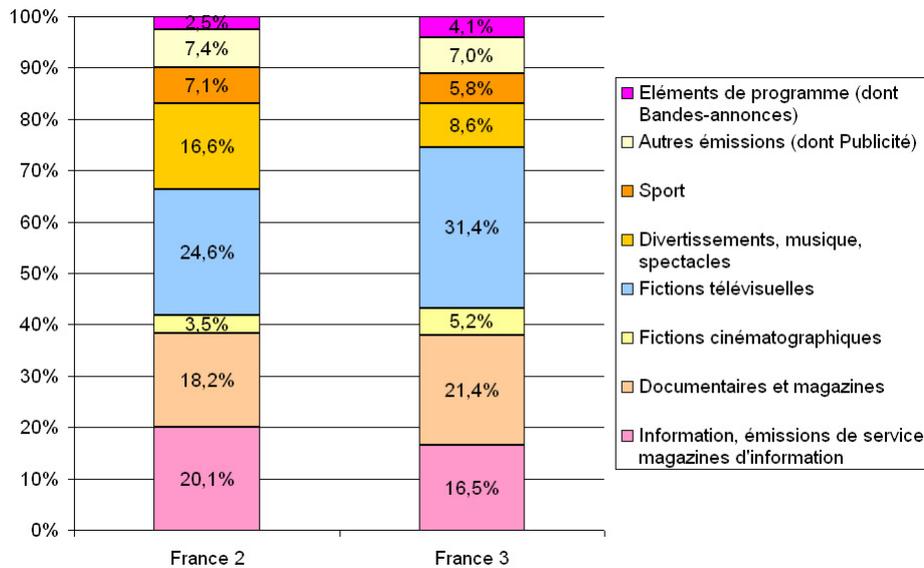
Chapitre 6

Taxinomie des programmes

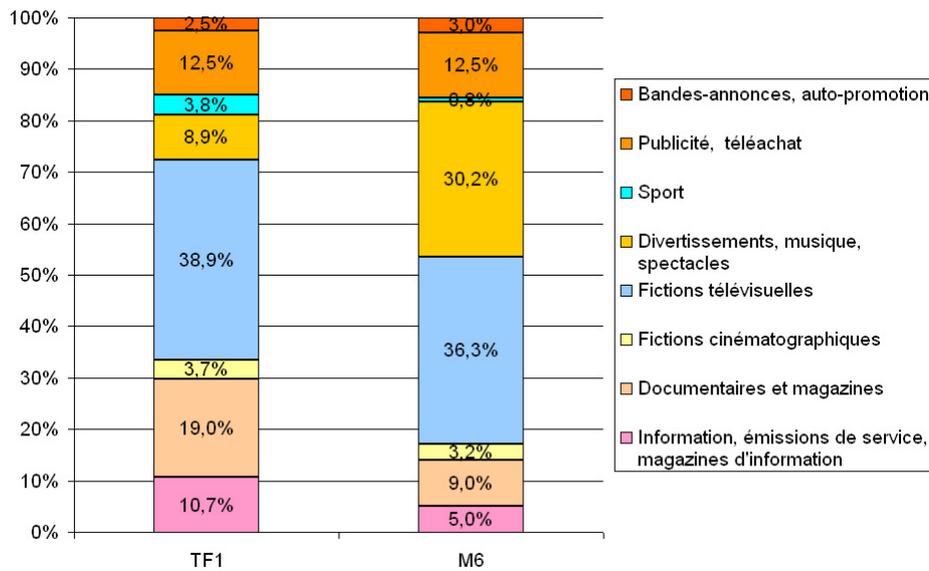
Sommaire

1	Définitions des genres télévisuels	58
1.1	Les fictions	59
1.2	L'information	60
1.3	Les divertissements	61
1.4	Les documentaires	62
1.5	Les inter-programmes et le parrainage	63
1.6	Les émissions pour la jeunesse	64
1.7	Le télé-achat	64
1.8	Le sport	64
1.9	Les tranches horaires	65
1.10	La (con)fusion des genres	65
2	Taxinomie des genres	66
2.1	Typologies de l'Institut National de l'Audiovisuel	66
2.2	Ontologies de l'audiovisuel	67
2.3	Nomenclature de Médiamétrie	69
2.4	Justification du choix	69

Dans ce chapitre, nous allons présenter dans un premier temps les différents genres de programmes afin d'en préciser quelques caractéristiques. Les définitions et les propriétés que nous décrivons dans ce chapitre sont issues de notre propre observation de la télévision et du manuel remis aux documentalistes de l'INA (INA, 2002). Nous nous intéresserons ensuite à différentes taxinomies des genres télévisuels.



(a) Chaînes publiques



(b) Chaînes privées

FIG. 13: Répartition des genres en fonction des chaînes de télévision en 2000 (source : CSA).

1 Définitions des genres télévisuels

Lorsqu'une chaîne acquiert un programme audiovisuel, elle n'achète pas le dvd ou la cassette contenant le programme mais des droits de diffusion dans une langue et sur un territoire. L'usage a distingué les programmes de stock (fictions, documentaires, spectacles), qui sont rediffusables, des programmes de flux (magazines, jeux, variétés, divertissements), qui ne sont en général pas

rediffusables dans leur intégralité. En revanche, les journaux télévisés, le sport et les émissions pour les enfants ne rentrent dans aucune de ces catégories, obéissant à des règles particulières. Les programmes de flux diffusés en direct sont la cause des retards dans les horaires. La ligne éditoriale de la chaîne et la réglementation sont des paramètres décisifs dans l'attribution du temps d'antenne à chaque genre d'émissions. La figure 13 représente de façon synthétique la proportion représentée par chacun des genres. Si les genres d'émissions sont censés représenter le triptyque « informer, cultiver, distraire » (voir page 38), ils sont avant tout un moyen pour la chaîne de vendre des espaces publicitaires : il est donc nécessaire qu'une partie de ces programmes soient coupés par des écrans publicitaires ou des auto-promotions. D'autant plus que les droits de diffusion varient énormément selon le genre, les droits les plus élevés concernant la retransmission d'événements sportifs (en particulier le football).

1.1 Les fictions

La fiction est le genre le plus répandu à la télévision. C'est aussi le genre le plus soumis à des réglementations (quelques exemples ont été cités page 51) et des quotas (par exemple, pour les œuvres cinématographiques 60% au moins doivent être européennes avec 40% de françaises, 104 long métrages au maximum diffusés en première partie de soirée).

Les œuvres cinématographiques et les téléfilms

Ces deux types de fictions peuvent être différenciés généralement par leur durée. Pour les téléfilms, le format est adapté au mode de diffusion. Aucun producteur n'ignore que son téléfilm sera placé dans une grille de programmes ; la durée est donc stabilisée autour de 90 minutes. Il peut arriver que les téléfilms fassent partie d'une *saga*, diffusée par exemple en été (*le Maître du zodiaque*, sur TF1). Le film de cinéma, lui, a une durée beaucoup plus variable puisque sa vocation première est d'être diffusé indépendamment dans une salle de cinéma. La durée moyenne est également 90 minutes, mais certains films, comme *Titanic*, durent plus de 3 heures. Ces œuvres audiovisuelles ne peuvent être coupées qu'une seule fois par une publicité, et l'auto-promotion est interdite pendant leur durée. Toutefois, dans le cas d'un film trop long, l'autorisation pour une seconde coupure peut être demandée au CSA. Les films de cinéma sont le plus souvent diffusés en prime-time sur les chaînes généralistes alors que les téléfilms sont diffusés majoritairement les après-midi et parfois en prime-time. La place de ces fictions dans les grilles de programmes des chaînes généralistes est en train de changer au profit des séries télévisées : le phénomène, qui a commencé aux États-Unis il y a quelques années, est clairement visible en France avec la disparition sur TF1 de la tranche horaire *Ciné Dimanche* destinée aux œuvres cinématographiques au profit de la diffusion de la série à succès *les Experts*.

Les séries et les feuilletons

Le feuilleton est l'un des genres les plus anciens de la télévision, directement adapté de la radio. Il s'agit d'une suite d'épisodes mettant en scène les mêmes personnages, et dont les épisodes sont intrinsèquement liés : chaque épisode présente en parallèle les histoires et les anecdotes des personnages. Le feuilleton disparaît peu à peu au profit des sagas de téléfilms et des *soap*. Ce nom est donné aux feuilletons bon marché à nombreux épisodes (*les Feux de l'amour*, *Santa Barbara*, *des Jours et des vies*) commandités à l'origine par les fabricants de lessive.

Les séries sont en revanche en plein essor. Il s'agit là aussi d'une suite d'épisodes centrés sur des personnages, mais dont chaque épisode peut être vu indépendamment. Cette indépendance entre les épisodes permet aux chaînes de diffuser les épisodes d'une série dans le désordre afin de

respecter certaines règles du CSA : souvent, deux épisodes sont inversés afin de diffuser l'épisode le plus violent des deux à un horaire plus tardif. Les épisodes sont regroupés en saisons qui lient les épisodes entre eux par une trame de fond. La série peut être policière (*Starsky et Hutch, les Experts*), dramatique (*Grey's anatomy, Summerland*), humoristique (*Desperate housewives*) ou de science-fiction (*LOST, Buffy contre les vampires*). Les saisons comportent en général une vingtaine d'épisodes de 45 minutes environ pour les séries américaines et 52 minutes pour les séries européennes. Historiquement, la différence est due au fait qu'un épisode doit être contenu dans une case de la grille de programmes qui dure une heure. La diffusion des publicités aux États-Unis étant plus abondante qu'en Europe, la durée des épisodes des séries américaines est donc réduite. Dans les séries, on trouve aussi un style particulier, la sitcom, de l'américain *situation-comedy*, qui se distingue des séries humoristiques par leur durée qui avoisine les 26 minutes (*Friends, une Nounou d'enfer*).

Les feuilletons sont diffusés à des moments de la journée où la ménagère est seule chez elle, c'est-à-dire le matin ou en début d'après-midi (voir la grille de programmes type page 46). Les séries télévisées, exceptées les sitcoms, peuvent être diffusées aussi bien en prime-time (*la Trilogie du samedi* sur M6 propose au moins trois épisodes de séries) qu'à des horaires plus calmes dans la journée. Les séries présentées tard le soir comportent en général des épisodes violents ou avec des scènes pouvant troubler le public le plus jeune. À noter qu'un nouveau format de série est apparu ; il s'agit d'un format très court (*un Gars une fille, Kaamelott*), de moins de 6 minutes, diffusé avant le journal ou à plusieurs moments de la journée.

Les courts métrages

Les courts métrages sont des films qui se distinguent par une durée de 20 à 25 minutes ; cependant, quelques rares exceptions durent jusqu'à 60 minutes. Ils sont rarement diffusés à la télévision (excepté sur Canal+). Il n'y a pas de règles ni de généralités sur leur emplacement dans une grille de programmes.

1.2 L'information

L'information à la télévision dépasse largement la simple revue de l'actualité : il peut s'agir de magazines proposant des sujets sur des faits réels ou des retransmissions d'événements (sportifs ou pas).

Les journaux télévisés

Le journal télévisé est un genre qui remonte au début de la télévision, même si son format a changé. Il s'agit principalement d'une alternance de plateaux de lancement et de reportages ou d'interviews.

Ce genre d'émission est un moyen pour les chaînes de se démarquer les unes des autres, tant au niveau de l'habillage de l'écran que du format. M6 préférera par exemple un format court, sans présentateur (*le 6 minutes*), avec une couleur dominante chaude (rouge) alors que TF1 diffusera un journal de presque 40 minutes à 20 heures, dominé par une couleur froide (bleu). La durée des journaux télévisés dépend donc de la chaîne, mais aussi de l'heure à laquelle le journal est diffusé. Certaines chaînes thématiques ne diffusent pas de journal télévisé, alors que d'autres ne diffusent que des journaux (*LCI*). Les journaux télévisés jouent souvent le rôle de pivots dans la journée.

Les magazines

Le magazine est un genre qui apparaît sous de multiples formes, sur de multiples sujets. Il est possible de distinguer les magazines de reportages des magazines de plateaux, souvent proches des talk-shows. Les formats des magazines sont très variés : historiquement, il existe un format court de 26 minutes (*C'est notre affaire*) et un format long de 52 minutes (*Sept à huit*). Mais la limitation des publicités et l'autorisation du parrainage ont fait naître un format très court (entre 1 et 4 minutes) sponsorisé par un annonceur (*Consomag, du Côté de chez vous*), faisant généralement partie d'un secteur interdit à la publicité télévisée (par exemple la grande distribution).

Les retransmissions

Les chaînes choisissent en fonction de leur ligne éditoriale des événements à retransmettre, comme des concerts ou des soirées spéciales (*la Nuit du Ramadan* sur France 2). Nous parlerons des retransmissions sportive un peu plus tard. Les durées sont imprévisibles puisqu'elles dépendent de l'événement. Ces événements spéciaux n'interviennent que très peu durant l'année et sont toujours annoncés dans les guides de programmes. De plus, les rediffusions sont assez rares.

Les émissions de service

Il s'agit généralement de formats courts (entre 2 et 6 minutes) présentant une information pratique : bulletin météo, résultats de courses ou de loteries, trafic routier.

1.3 Les divertissements

Les divertissements se déclinent en plusieurs formats selon lesquels ils seront diffusés un jour plutôt qu'un autre et à une certaine fréquence. À titre d'exemple, les émissions de variétés sont proposées le week-end, de façon hebdomadaire ou mensuelle alors que les jeux peuvent être proposés tous les jours de la semaine.

Les jeux

Il existe plusieurs sortes de jeux : les jeux de connaissance (*Question pour un champion*), les jeux d'aventure (*Fort Boyard*) ou les jeux de chance (*À prendre ou à laisser*). La durée moyenne d'un jeu en journée ou en soirée est de 26 minutes, alors que les jeux en prime-time ont une durée plus proche de 90 minutes. En journée, les jeux sont diffusés en général avant le journal de 13h et le journal de 20h.

Les émissions de variétés

Les émissions de variétés sont apparues avec la télévision, directement inspirées du music-hall. Dès les années 50, elles sont diffusées en direct et en public. Elles sont diffusées presque exclusivement en prime-time (*Champs-Élysées, Sacrée soirée*) ou en deuxième partie de soirée (*Taratata*). La durée moyenne d'une émission de variétés est de 90 minutes. Elles ont remplacé les fictions du samedi soir depuis l'interdiction des années 90.

Les émissions à base de clips

Il s'agit d'une autre forme d'émissions de variétés. Ces émissions se divisent en deux catégories : les émissions constituées seulement de clips (*M comme musique*, sur M6), et les émissions qui alternent des clips et des plateaux (*Hit Machine*, *Top of the pops*) dont le but est souvent d'établir un classement. Certaines chaînes thématiques musicales ne diffusent quasiment que ce genre d'émissions.

Les reality-shows et les émissions de télé-réalité

Les reality-shows et les émissions de télé-réalité ont pour vocation de montrer à la télévision des individus anonymes qui deviennent la vedette d'un soir. Les reality-shows vont en général exalter des personnalités (*la Nuit des héros*) ou au contraire permettre à des personnes en difficulté de surmonter leurs problèmes psychologiques ou sociaux (*Confessions intimes*), ou de connaître des situations nouvelles (*Vis ma vie*). Il peut s'agir aussi d'appels à témoins pour résoudre des enquêtes policières (*Perdu de vue*). On assiste récemment à la multiplication d'émissions de coaching où un « expert » va aider des personnes en difficulté (*le Grand frère*, *Super nanny*). Ces émissions sont diffusées en première ou deuxième partie de soirée, avec une durée de 90 minutes environ (un ou deux épisodes diffusés par soir).

D'un autre côté, la télé-réalité, apparue en France en 2001, met en avant les réactions de personnes en huis-clos ou (et) en compétition, et éliminées tour à tour. Ce genre d'émissions peut engendrer de nombreux programmes (*les Colocataires*, *Star academy*, *la Ferme des célébrités*) en fonction de la nature des candidats (vedettes ou inconnus), du lieu d'enfermement (un loft, un château, une île), du motif de la compétition (enregistrer un disque, gagner une grosse somme d'argent) et du mode d'élimination des candidats (très souvent par vote du public). Le format va donc varier en fonction de la déclinaison du genre. Généralement, par rapport au coût de l'émission, ce genre bénéficie toujours d'un prime-time hebdomadaire. Parfois, des résumés quotidiens sont diffusés en access prime-time et rediffusés plusieurs fois par jour. La durée du prime-time est d'environ 90 minutes et l'émission quotidienne dure généralement une heure.

Les talk-shows

Les talk-shows sont des émissions centrées sur un présentateur et qui reposent sur un entretien avec des personnes anonymes ou connues. Le talk-show est un programme de soirée qui est plutôt diffusé en prime-time (*Ça se discute*, *Bas les masques*) ou en seconde partie de soirée (*Tout le monde en parle*) selon son contenu et sa notoriété : *C'est mon choix*, sur France 3, était une exception. Leur durée est assez longue, au moins 90 minutes, mais très variable : nous avons remarqué en effet que *Tout le monde en parle* pouvait durer entre 1h30 et 3h00.

1.4 Les documentaires

Excepté les chaînes thématiques qui en diffusent toute la journée, ce genre ne subsiste que grâce aux quotas imposés par le CSA : le documentaire est le genre qui réalise les audiences les plus faibles. Le documentaire couvre une gamme importante de thématiques : politique, économie, faits de société, art, littérature, histoire ou faune. Il n'y a pas de structure-type pour ce genre de programmes : le contenu d'un documentaire dépend de son auteur et de la façon qu'il a d'aborder le sujet. Sur les chaînes généralistes, le documentaire est diffusé bien souvent la nuit, sauf quand il vise un public âgé (le dimanche après-midi sur France 3). La durée d'un

documentaire peut être 26, 52, ou 90 minutes. Quand ils sont diffusés à des horaires tardifs, les documentaires sont souvent tronqués afin de rattraper le retard.

1.5 Les inter-programmes et le parrainage

Les programmes peuvent être séparés par trois types d'inter-programmes : les publicités, l'auto-promotion (bandes-annonces) et les clips musicaux. En revanche, les émissions qui peuvent être interrompues (en fonction de leur genre) ne peuvent pas l'être par la diffusion d'un clip. Les clips musicaux entre les programmes sont des cas particuliers d'émissions sponsorisées par les chaînes elles-mêmes.

La particularité française est que ce domaine est très réglementé. La publicité est définie par le CSA comme l'ensemble des messages diffusés contre rémunération. L'autopromotion est une suite de cartons annonçant le programme suivant (*coming next*) ou les programmes à venir (bandes-annonces). Les chaînes du groupe France Télévisions ont le droit de promouvoir les autres chaînes du groupe. La réglementation des coupures publicitaires, en plus de règles déontologiques, comprend deux règles intéressantes qui ont orienté la suite de nos travaux :

- « *Les messages publicitaires ou les séquences de messages publicitaires doivent être aisément identifiables comme tels et nettement séparés du reste du programme, avant comme après leur diffusion par des écrans reconnaissables à leurs caractéristiques optiques et acoustiques* », nous appellerons ces écrans des jingles de publicité ;
- « *Le volume sonore des séquences publicitaires ainsi que des écrans qui les précèdent et qui les suivent ne doit pas excéder le volume sonore moyen du programme* ».

Bien sûr, d'autres règles s'appliquent aux coupures publicitaires et aux spots eux-mêmes (CSA, 2006). De plus, il existe un bureau de vérification des publicités¹⁴ (BVP) dont le rôle est de promouvoir l'autodiscipline dans la publicité en France ; toutefois, les publicitaires ne sont pas obligés légalement d'avoir recours au BVP et les régies publicitaires des chaînes peuvent accepter des spots publicitaires n'ayant pas été présentés au BVP. Les volumes de diffusion sont également limités (voir tableau 6). Certaines chaînes, comme Arte, ne profitent pas de recettes publicitaires.

	TF1	M6	France 2	France 3	France 5
Moyenne quotidienne autorisée par heure d'antenne	6 min	6 min	6 min	6 min	6 min
Durée maximale pour une heure donnée	12 min	12 min	8 min	8 min	8 min
Durée maximale de l'écran publicitaire	-	-	4 min	4 min	4 min
Durée maximale de l'écran interrompant les œuvres cinématographiques	6 min	6 min	interdite	interdite	interdite
Durée maximale de l'écran interrompant les œuvres audiovisuelles	12 min	12 min	interdite	interdite	interdite

TAB. 6: *Comparaison des volumes de publicité autorisé sur les principales chaînes hertziennes (source : rapport du CSA (CSA, 2006))*

Une autre forme de cartons peut intervenir lors de la diffusion : il s'agit du parrainage (*sponsoring*). Une chaîne a la possibilité de faire parrainer une émission ou une tranche horaire par un annonceur. Il s'agit très souvent d'annonceurs qui n'ont pas accès à la publicité télévisée

¹⁴<http://www.bvp.org>

comme ceux du secteur de la grande distribution (supermarchés, grands magasins). L'annonceur participera à l'achat des droits de l'émission pour que son logo soit diffusé avant ou après celle-ci. En revanche, les rubriques d'une émission ne peuvent pas être parrainées, excepté lorsqu'il s'agit d'émissions de service comme la météo et le trafic routier. Parfois, des magazines courts sont conçus uniquement pour être parrainés ; le parrain peut intervenir sur le choix de l'émission parrainée mais en aucun cas sur son contenu.

1.6 Les émissions pour la jeunesse

Les émissions pour la jeunesse s'adressent à des publics assez différents : les moins de 4 ans, les 4-10 ans, les 10-14 ans et les 14 ans et plus. En fonction de la cible visée, les émissions vont changer de contenu. Ces émissions sont constituées de dessins animés, de fictions pour enfants et de plateaux qui sont des séquences intercalées entre les dessins animés ou les fictions pour habiller l'ensemble de l'émission. Ce secteur est également soumis à une réglementation très contrôlée et chaque épisode de dessin animé ou de fiction est généralement contrôlé par les chaînes. Ainsi, beaucoup d'épisodes des *Castors allumés* n'ont jamais été diffusés en France, jugés trop « durs » pour les enfants. En effet, contrairement aux autres pays, la France a du mal à concevoir que les dessins animés ne s'adressent pas qu'aux enfants (*South Park* aux Etats-Unis, les mangas au Japon). La programmation de ces émissions dépend vraiment de la disponibilité des enfants : elles seront donc diffusées avant et après l'école ainsi qu'en matinée le mercredi, le samedi et le dimanche. Pendant les vacances scolaires, toutes les matinées sont destinées à ces programmes. La réglementation des publicités interdisant que les émissions pour enfants de moins de 30 minutes soient coupées par des publicités, il en résulte que les coupures publicitaires n'ont lieu qu'entre ces dessins animés ou autres fictions.

1.7 Le télé-achat

Les émissions de télé-achat sont des émissions pendant lesquelles des produits sont présentés directement au public (*Téléshopping*, *M6 boutique*). Ce secteur est soumis à une réglementation importante. Outre les règles de déontologie, la réglementation exige :

- que ces émissions ne soient pas coupées par des publicités ;
- que les chaînes ne diffusent pas plus de huit émissions de télé-achat (ou trois heures) par jour ;
- que ces émissions soient diffusées entre minuit et 11h dans la limite d'une heure par émission sur les chaînes hertziennes, à l'exception des mercredi, samedi après-midi et dimanche où elles sont interdites ;
- que ne soient vendus que des produits autorisés.

Les chaînes entièrement consacrées au télé-achat ont bien sûr une réglementation spécifique. Depuis juillet 2006, les chaînes publiques ont le droit de diffuser ce genre d'émissions.

1.8 Le sport

Le sport n'est pas un genre télévisuel en lui-même, mais plutôt un agrégat de genres : le magazine de sport et la retransmission sportive. Lors des retransmissions de sport, les grilles de

programmes sont totalement remaniées en fonction de la durée de l'événement (déprogrammation, changement d'horaires), alors que les magazines de sport s'intègrent parfaitement dans la logique de la grille de programmes standard.

Les émissions de sport sont génératrices d'audience au prix d'un budget considérable pour les chaînes. L'achat des droits de diffusion du sport est en effet la plus grosse dépense de la grille de programmes. Par exemple, la retransmission de matchs de football représente à elle seule plus de 549 heures de diffusion sur l'ensemble des chaînes hertziennes, et pour les droits du championnat de France de Football, de juillet 2005 à juin 2008, Canal+ a payé 600 millions d'euros (INA, 2006).

1.9 Les tranches horaires

Une tranche horaire est une case de la grille de programmes qui consiste en une suite de programmes référencés sous un titre unique. Les programmes audiovisuels diffusés dans ce cadre ont pour point commun un thème (un personnage, un événement, une science), un public, ou une catégorie de programmes (sport, fiction). Par exemple, *Ciné Dimanche*, le dimanche soir sur TF1, est une tranche horaire qui débute par un générique, suivi d'un sommaire, puis d'un ou deux films selon leurs durées et d'un magazine qui présente les sorties en salle.

Les tranches horaires sont des cas particuliers de collections. Une collection est un ensemble de programmes. Il existe aussi des collections d'émissions simples (*Combien ça coûte ?*), où chaque numéro est diffusable isolément mais partage une structure de production homogène, ou des collections d'émissions en plusieurs épisodes comme les séries (*Buffy contre les vampires*).

1.10 La (con)fusion des genres

Il est de plus en plus difficile de séparer les genres télévisuels tant ceux-ci se mélangent (Charaudeau, 1997). On a vu apparaître le docu-fiction, un genre à mi-chemin entre le documentaire avec son côté éducatif, et la fiction avec son côté ludique et imaginaire. Par exemple, TF1 a diffusé en décembre 2005 un docu-fiction intitulé *Dragons : et s'ils avaient vraiment existé ?*, basé sur l'étude d'espèces animales réelles et alimenté d'images de synthèse.



(a) le plateau

(b) un clip

FIG. 14: L'émission de call-TV *Tubissimo*, sur M6

De même la *call-tv* est une émission entre le jeu, le magazine et l'émission à base de clips. Ces émissions proposent de gagner une somme d'argent en répondant à des questions faciles (pour

maximiser le nombre d'appels) en appelant un numéro surtaxé ; un gagnant est ensuite tiré au sort. En attendant ce tirage au sort, des clips et des magazines sur les vedettes sont diffusés (voir figure 14).

Enfin, l'*infotainment* est une émission qui mélange plusieurs types d'actualités (mondiales, peuples) en mêlant des expériences personnelles (par exemple, le nouvel opus de *Morning café* diffusé sur M6 depuis septembre 2006). Cette évolution des genres rend plus difficile la classification des émissions.

2 Taxinomie des genres

Nous avons vu dans la sous-section 1.10 que la classification des genres télévisuels est difficile d'une part à cause de la fusion des genres, et d'autre part à cause de l'arrivée permanente de nouveaux genres. Pour (Charaudeau, 1997), la classification idéale ne tiendrait compte que de trois grands genres : l'information, le divertissement et la culture, directement déduits du triptyque de la télévision française (voir chapitre 1 page 38). En effet, dans nos travaux, nous avons cherché à différencier les genres par leurs heures de diffusion, leurs durées et les règles de montage qui marquent leur structure.

Nous nous sommes donc intéressés à plusieurs sortes de taxinomie des genres audiovisuels. Nous présentons dans cette section la méthodologie de typages des programmes appliquée à l'INA, une ontologie de l'audiovisuel et la nomenclature de Médiamétrie.

2.1 Typologies de l'Institut National de l'Audiovisuel

L'INA caractérise les programmes par une typologie à quatre listes :

- **les genres** (voir tableau 7), pour qualifier les programmes du nom des principaux genres télévisuels (documentaire, fiction, téléfilm, ...);
- **les thématiques** (voir tableau 8) dans lesquelles s'inscrit le programme (sciences, spectacles, sport, ...);
- **le public** : un ensemble de termes et de catégories qualifient le public présent lors de l'enregistrement du programme ;
- **les modes de diffusion**, pour distinguer les émissions retransmises en direct, en différé ou encore en duplex.

Adaptation	Déclaration	Micro trottoir	Retransmission
Animation	Documentaire	Mini programme	Rétrospective
Bande annonce	Entretien	Montage d'archives	Revue de presse
Best of	Extrait	Œuvres enregistrées en studio	Série
Bruitage	Feuilleton	Plateau en situation	Sketche
Causerie	Interlude	Presse filmée	Spectacle
Chronique	Jeu	Programme à base de clips	Talk show
Comédie de situation	Journal télévisé	Programme atypique	Télé achat
Conférence de presse	Journée témoin	Réalisation dans un lieu public	Téléfilm
Cours d'enseignement	Making of	Reality show	Téléréalité
Court métrage	Magazine	Récit portrait	Témoignage
Création télévisuelle	Message d'information	Reconstitution	Tranche horaire
Débat	Message publicitaire	Reportage	Vidéoclip

TAB. 7: Liste des genres possibles pour la télévision

Une collection est typée avec les quatre listes précédentes en précisant également sa périodicité (émission quotidienne, hebdomadaire, ...). Le sens de chaque terme utilisé dans cette typologie est normalisé par le thésaurus de l'INA. Ce thésaurus est une organisation hiérarchisée de plus de 10 000 noms communs. La typologie des émissions fonctionne par combinaison de ces termes. Au fil du temps, l'expérience a permis d'établir des règles pour connaître le sens des combinaisons, l'ordre des termes, et surtout les combinaisons impossibles (par exemple, retransmission et téléfilm). D'une façon générale, on peut trouver (Troncy, 2001) :

- un genre, pas de thématique : il s'agit d'une émission à contenu répétitif ou implicite (par exemple : un jeu, un journal télévisé) ou d'un programme composite (comme une tranche horaire) ;
- un genre, une thématique : il s'agit d'un type d'émission au sens ordinaire du terme (par exemple retransmission sportive, feuilleton) ou d'une émission à structure simple et à contenu unique (par exemple : magazine plateau de société) ;
- un genre, plusieurs thématiques : il s'agit d'un genre précis de fiction isolée (comme une fiction dramatique policière) ou d'une émission simple à contenu composite (par exemple, un documentaire historico-politique) ;
- deux genres, une thématique : il s'agit d'une émission appartenant à une série (par exemple, une série de documentaires), d'une genre précis (comme un documentaire prenant la forme d'un montage d'archives), d'une émission composite (par exemple, un magazine composé de reportages) ou d'une émission ayant un format de diffusion spécifique (comme un spectacle télévisé réalisé dans un lieu public) ;
- de deux à quatre genres, plusieurs thématiques : il s'agit d'une émission combinant plusieurs des critères précédents (par exemple, une série de dessins animés adaptés d'un livre).

Spectacle	Humour	Histoire	Informations pratiques
Théâtre	Science	Religion	Vie professionnelle
Fiction	Sciences naturelles	Ethnologie	Loisirs
Science fiction	Sciences exactes	Sociologie	Tourisme
Fantastique	Médecine, santé	Psychologie	Aventure, exploration
Cinéma	Sciences humaines	Géographie	Vie professionnelle
Média	Industrie	Philosophie	Loisirs
Intrigue policière	Environnement	Information	Tradition
Musique	Technologie	Politique	Urbanisme
Danse	Faune	Economie	Vie quotidienne
Variétés	Arts	Société	Erotisme
Music hall	Beaux arts	Insertion sociale	Artisanat
Cirque	Littérature	Sport	

TAB. 8: Liste des thèmes possibles pour la télévision

2.2 Ontologies de l'audiovisuel

Plusieurs terminologies ont été récemment standardisées (MPEG-7, TV Anytime) et d'autres sont toujours en cours de développement (ProgramGuideML¹⁵). Dans cette section, nous portons notre attention sur l'ontologie de l'audiovisuel¹⁶ proposée par Troncy et Isaac (Isaac et Troncy,

¹⁵ProgramGuideML se veut être le prochain standard XML pour l'échange d'informations sur les programmes télévisés ou radiophoniques (voir www.programguideml.org).

¹⁶L'ontologie de l'audiovisuel est téléchargeable à opaes.ina.fr/public/ontologies/coront/.

2004).

Une ontologie peut être vue comme une organisation hiérarchique de la connaissance sur un ensemble d'objets par leur regroupement en sous-catégories suivant leurs caractéristiques essentielles. De plus, les relations entre ces objets peuvent être spécifiées. L'ontologie qui nous intéresse propose de décrire les émissions à la fois du point de vue de sa production (par exemple qui sont les acteurs ou les auteurs) et du point de vue de sa diffusion (par exemple, combien de fois elle a été diffusée et sur quelles chaînes). L'ontologie de l'audiovisuel adapte la description de l'émission en fonction de son genre qui est caractérisé d'abord par des termes très généraux (fiction, culture, divertissement), puis par des termes de plus en plus spécialisés (sitcom, série, documentaire, ...). La figure 15 présente un extrait de la taxinomie des concepts de l'ontologie de l'audiovisuel qui permettent de caractériser des programmes télévisés. Le terme le plus spécifique dans cette taxinomie représente le genre à attribuer à l'émission.

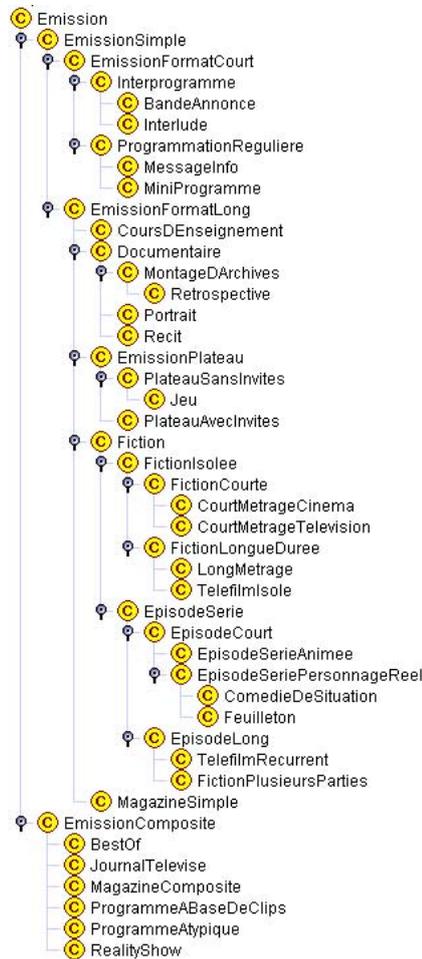


FIG. 15: Extrait de l'ontologie de l'audiovisuel

Dans le réseau d'excellence K-Space (voir page 11), une nouvelle ontologie de l'audiovisuel est en cours d'élaboration.

2.3 Nomenclature de Médiamétrie

La nomenclature de Médiamétrie (voir section 2, page 39) est fortement liée à l'usage des publicitaires puisque son but est de fournir une classification toujours plus précise aux diffuseurs et aux annonceurs. La nomenclature est composée d'un ensemble de termes regroupés en trois niveaux hiérarchiques. Un seul terme est donc utilisé pour caractériser une émission, contrairement à la typologie de l'INA, ce qui laisse présager d'une faible capacité d'adaptation. La figure 16 présente un extrait de cette nomenclature : ainsi le code ABC représente un téléfilm d'aventures. La nomenclature complète est fournie en annexe (voir annexe A, page 197). Cette nomenclature est uniquement mise à jour par Médiamétrie (la dernière mise à jour date de 2000) ce qui provoque parfois une inadéquation par rapport aux nouveaux genres (par exemple la télé-réalité est un genre toujours absent de la nomenclature).

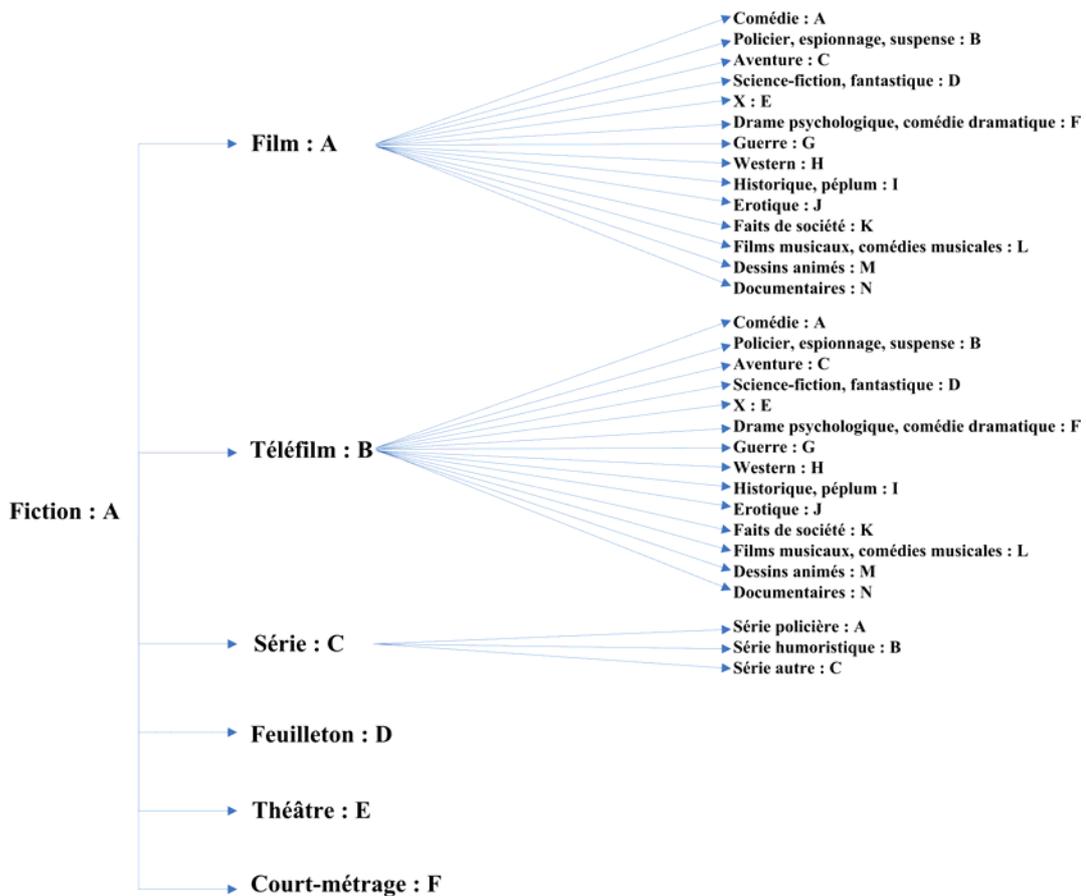


FIG. 16: Extrait de la nomenclature de Médiamétrie

2.4 Justification du choix

Nous avons dû faire un choix entre ces différentes taxinomies afin de l'utiliser dans nos travaux. Les données dont nous disposons pour nos apprentissages, contenues dans les bases de données de l'INA, possèdent un champ pour la typologie de l'institut et la nomenclature de Médiamétrie, alors que rien ne nous donne l'équivalent dans l'ontologie de l'audiovisuel. Nous présentons quelques différences entre la typologie de l'INA et la nomenclature de Médiamétrie

dans le tableau 9.

Il est possible de voir, avec l'exemple d'*un Gars, une fille*, que la typologie de l'INA peut être redondante puisque par définition une sitcom est une fiction humoristique en série. La typologie de l'INA peut prendre en compte les tranches horaires telles que *KD2A* alors qu'elle ne prend pas en compte ce genre de segment audiovisuel. Enfin, les exemples montrent les problèmes que posent la télé-réalité : Médiamétrie n'a toujours pas rajouté ce genre et considère n'importe quelle émission de télé-réalité comme un jeu interactif (puisque le téléspectateur est invité à voter pour éliminer ou conserver un candidat). Ainsi, l'émission *popstar* est considérée comme un magazine par la typologie de l'INA et comme une émission de variétés par la nomenclature de Médiamétrie, alors qu'il s'agit d'une émission de télé-réalité (mettant en jeu des personnages réels en huis-clos) relatant la formation d'un groupe de musique.

Emissions	Typologie de l'INA	Nomenclature de médiamétrie
Un gars, une fille	Sitcom, Fiction, Humour	Série humoristique
Combien ça coûte ?	Magazine, Société, Economie ou Magazine, Société	Magazine, Société
Du côté de chez vous	Miniprogramme, Informations pratiques	Magazine, Vie quotidienne
Loft story (quotidienne)	Best of	Jeu interactif
Loft Story (prime time)	TV réalité, jeu ou TV réalité, Spectacle TV	Jeu interactif
Koh Lanta	Jeu, TV réalité	Jeu interactif
Strip tease	Magazine, Reportages, Société, Vie quotidienne	Magazine, Vie quotidienne
KD2A	Tranche horaire, programme destiné à la jeunesse	Emission jeunesse
Popstar	Magazine, Montage d'archives, Société ou Magazine, Reportages, Société	Variétés

TAB. 9: Comparaison entre la typologie de l'INA et la nomenclature de Médiamétrie sur quelques programmes

L'avantage des données de Médiamétrie est leur disponibilité dès le lendemain de la diffusion et surtout l'homogénéité de la classification (une même émission aura toujours le même genre durant une saison). Cela n'est pas le cas pour la typologie de l'INA, qui intervient un peu plus tard après la diffusion, et qui est encore plus soumise à la subjectivité des documentalistes. En effet, *Combien ça coûte ?* sera tantôt affecté des thèmes société et économie, tantôt du thème société seul (voir tableau 9). De même, les émissions en *prime time* de *Loft story* seront considérées tantôt comme un jeu, tantôt comme un spectacle télévisé. Cependant, Médiamétrie manque parfois de précision puisqu'elle ne distingue pas les émissions quotidiennes des émissions hebdomadaires en *prime time* de *Loft story*. Nous avons donc choisi d'utiliser la nomenclature Médiamétrie puisque c'est cet organisme qui fournit la structure des flux des chaînes hertziennes à l'INA. Cela nous permet d'accéder directement aux données produites par Médiamétrie où chaque émission est typée de façon homogène au fil des années.

BILAN DE LA DEUXIÈME PARTIE

Cette partie sur la télévision française et la programmation d'une chaîne définit les bases de notre travail. Nous allons à partir des règles de la programmation tenter de retrouver la structure d'un flux télévisuel. De plus, les règles juridiques imposées par le CSA et les règles visuelles et sonores imposées par la concurrence seront utilisées pour reconnaître des éléments caractéristiques de certains programmes.

Le système concurrentiel dans lequel se trouvent les chaînes les oblige à stabiliser leurs grilles de programmes pour proposer des taux d'audiences lisses aux annonceurs. Ainsi, la structure de la grille de programmes reste la même pendant plusieurs années et les cases sont remplies par des émissions de même type. Toutefois, la monotonie est brisée par des événements exceptionnels (tels que les attentats du 11 septembre, la mort du Pape ou les jeux Olympiques) ou des remaniements de la grille pour s'opposer un peu plus aux programmes proposés.

Nous pouvons à présent nous intéresser à l'amélioration des guides de programmes et à la reconnaissance des émissions et des interprogrammes qui composent un flux télévisuel en tenant compte des règles qui viennent d'être citées.

Troisième partie

Approche markovienne pour la modélisation des grilles de programmes

RÉSUMÉ

Nous avons vu dans la partie précédente que les grilles de programmes proposées par les chaînes de télévision sont stables dans le temps.

Nous proposons de modéliser ces grilles de programmes en utilisant un modèle Markovien et un arbre de régression afin de capturer les habitudes de diffusion de la chaîne. Notre contribution porte sur une extension des modèles Markoviens, les modèles de Markov cachés contextuels (CHMM), permettant de faciliter le choix d'une topologie pour le modèle et permettant de diminuer le nombre de transitions possibles à partir d'un état. Les arbres de régression nous permettent d'encadrer la durée d'une émission en fonction du jour de la semaine où elle est diffusée, de son heure de diffusion et de son genre.

En partant d'une grille prévisionnelle fournie par un guide de programmes, nous proposons de calculer l'ensemble des grilles possibles pour un jour de l'année. Le résultat de cette prédiction permettra de guider les détections du système.

Introduction

Le système SAFARI (voir le chapitre 3 de la première partie, page 23) est basé sur une représentation des grilles de programmes. Nous avons choisi de les modéliser avec un modèle statistique et probabiliste. Celui-ci doit être capable de fournir pour un programme donné le genre (journal télévisé, magazine, ...) du programme suivant ainsi que sa durée. En effet, nous allons utiliser le genre de programme comme information sur l'émission et la durée va nous permettre de calculer l'heure de début de la prochaine émission.

Pour cela, le modèle est constitué de deux parties. La première partie est un modèle de Markov contextuel. Les modèles de Markov sont parfaitement adaptés à la représentation de séquences d'observations : dans notre cas, une observation sera un genre de programme. Le modèle de Markov saisit ainsi les diverses transitions d'un genre à l'autre durant la journée et sur une chaîne donnée. La seconde partie permet de prédire la durée d'une émission : nous avons utilisé un arbre de régression qui permet d'encadrer une durée pour une émission afin de réduire l'espace de recherche des transitions entre émissions à des fenêtres temporelles.

Cette modélisation va permettre de guider les détections compte tenu de l'ensemble des enchaînements d'émissions possibles pour une journée donnée. Puisque le modèle statistique est entraîné sur des grilles de programmes des années passées, il est préférable de le confronter à une représentation plus récente de la grille de programmes : le guide de programmes. Nous avons effectivement vu que le guide de programmes était une prévision incomplète et imprécise d'une journée de diffusion (voir la section 6, page 52). Nous proposons de confronter les prédictions du modèle et les guides de programmes pour :

- tenir compte d'une information plus récente, comme une déprogrammation exceptionnelle mais prévue : par exemple pour un concert, un événement sportif ;
- obtenir des informations sur l'émission : titre, résumé, ...

Nous présenterons dans un premier temps notre extension contextuelle des modèles de Markov et nous l'appliquerons à la modélisation des grilles de programmes. Nous verrons ensuite comment estimer les durées des émissions et comment évaluer la probabilité d'une certaine durée pour une émission. Enfin, nous aborderons la prédiction de grilles possibles pour une journée de diffusion à partir du modèle et des guides de programmes et nous terminerons cette partie par la présentation des résultats que nous avons obtenus.

Chapitre 7

Modélisation des grilles de programmes

Sommaire

1	Modèles de Markov	79
1.1	Chaînes de Markov	80
1.2	Modèles de Markov cachés	81
1.3	Champs de Markov	85
2	Modèles de Markov cachés contextuels	85
2.1	Définitions préliminaires	86
2.2	Modèles de Markov cachés contextuels	87
2.3	Résolution des trois problèmes de l’extension	89
3	Application à la modélisation des grilles de programmes	99
3.1	Estimation des paramètres du CHMM	101
3.2	Estimation de la probabilité d’un enchaînement d’émissions	103
4	Conclusion	104

Pour modéliser les grilles de programmes, un formalisme capable d’exprimer les transitions d’un programme à l’autre est nécessaire. Nous avons choisi de modéliser les grilles de programmes avec un modèle de Markov. Ces modèles sont parfaitement adaptés à la représentation de séquences d’observations, notamment en bioinformatique pour le traitement des séquences d’ADN (Boufounos *et al.*, 2002). De plus, les modèles markoviens ont été utilisés avec succès dans le cadre de la structuration de vidéos (voir chapitre 2, page 13).

Nous allons présenter dans une première partie différents modèles de Markov, en se limitant aux modèles permettant de modéliser des séquences d’observations à une dimension, avant d’introduire notre extension contextualisée des modèles de Markov.

1 Modèles de Markov

Les processus de décision markoviens (*Markov decision processes*) sont des modèles de décisions séquentiels. Puterman (Puterman, 1994) définit les processus markoviens comme suit : à un temps donné, un agent (*decision maker, agent, ou encore controller*) observe l’état d’un

système. Puis l'agent choisit une action en fonction de cet état ; le choix d'une action produit deux résultats : l'agent est récompensé (ou pénalisé), et le système entre dans un nouvel état en fonction d'une loi de probabilité déterminée par le choix de l'action. Ensuite, le processus recommence, mais cette fois le système est dans un autre état, avec d'autres actions à choisir.

Pour représenter un modèle séquentiel, il est donc nécessaire d'avoir :

- un ensemble de moments où les décisions doivent être prises ;
- un ensemble d'états ;
- un ensemble d'actions ;
- un ensemble de récompenses (ou de pénalités) directement liées à un état et à une action ;
- un ensemble de probabilités de transition dépendant d'un état et d'une action.

De plus, les processus de Markov sont sans mémoire, c'est-à-dire que seul l'état courant permet de connaître quel sera le prochain état. Les modèles de Markov permettent de représenter de tels processus de décision. Nous allons présenter à présent trois des modèles de Markov.

1.1 Chaînes de Markov

La chaîne de Markov est le modèle le plus simple permettant de représenter un processus markovien, et sans aucun doute le plus étudié (Norris, 1999).

Soit S un ensemble dénombrable ; chaque $s \in S$ est appelé un état et S est appelé l'espace d'états. Soit $\pi = (\pi_s : s \in S)$ une distribution sur S et X une variable aléatoire à valeurs sur S ; on pose $\pi_s = P(X = s)$. Soit $P = (p_{ij} : i, j \in S)$ une matrice stochastique. Alors on dit que $(X_n)_{n \geq 1}$ est une chaîne de Markov de distribution initiale π et de matrice de transition P si (Norris, 1999) :

- X_1 a pour distribution π ;
- pour tout $n \geq 1$, si $X_n = i$ alors X_{n+1} a pour distribution $p_{ij} : j \in S$ et est indépendant de X_1, X_2, \dots, X_{n-1} .

Autrement dit, pour tout $n \geq 1$ et $s_1, \dots, s_{n+1} \in S$:

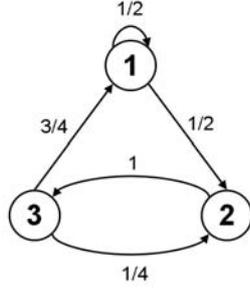
- $P(X_1 = s_1) = \pi_{s_1}$;
- $P(X_{n+1} = s_{n+1} | X_1 = s_1, \dots, X_n = s_n) = p_{s_n s_{n+1}}$.

Mathématiquement, ces deux propriétés permettent de prouver plus facilement les propriétés des chaînes de Markov. Ainsi, la probabilité de la séquence d'états $s_1, \dots, s_n \in S$ étant donnée une chaîne de Markov de matrice de transition P et de distribution initiale π s'écrit :

$$P(s_1, \dots, s_n | \pi, P) = \pi_{s_1} \times p_{s_1 s_2} \times \dots \times p_{s_{n-1} s_n}. \quad (1)$$

Ceci représente une chaîne de Markov du premier ordre. Dans une chaîne de Markov d'ordre k , les k derniers états influencent la transition de l'état courant vers l'état suivant.

Les chaînes de Markov sont graphiquement représentées par un automate à états comme le montre la figure 17. Dans cet exemple, la chaîne de Markov possède trois états et les flèches indiquent les transitions possibles d'un état à l'autre. La matrice P est la matrice de transition et le vecteur π est le vecteur représentant les distributions initiales.



$$P = \begin{bmatrix} 0,5 & 0,5 & 0 \\ 0 & 0 & 1 \\ 0,75 & 0,25 & 0 \end{bmatrix} \quad \pi = \begin{bmatrix} 0,5 \\ 0,3 \\ 0,2 \end{bmatrix}$$

FIG. 17: Représentation graphique d'une chaîne de Markov

Plus de détails sur les chaînes de Markov à temps discret ou continu peuvent être trouvés dans (Norris, 1999).

1.2 Modèles de Markov cachés

Les modèles de Markov cachés (HMM) sont des modèles où les différents états s appartenant à l'espace d'états S ne sont pas directement observables. En revanche, il est possible d'observer une variable aléatoire Y dont les valeurs sont soumises à une distribution $P(Y = y_i | s)$ dépendante de l'état dans lequel le système se trouve (Bourlard et Bengio, 2002; Rabiner, 1989). Le modèle est ainsi doublement stochastique puisque le passage d'un état à l'autre est soumis à une loi de probabilité cachée à l'observateur et l'observation d'une valeur de la variable aléatoire Y est soumise à une loi de probabilité en fonction de l'état.

Définition

Un HMM est entièrement défini par le 5-uplet $\langle S, \Sigma, \pi, A, B \rangle$ où :

- S est un espace d'états au sens défini ci-dessus (voir 1.1), de cardinal M . s_t désigne cette fois-ci l'état au temps t ;
- Σ est l'ensemble des symboles observables, valeurs de la variable aléatoire Y , de cardinal N . y_t désigne le symbole observé au temps t ;
- $\pi = (\pi_i)_{i \in \{1, \dots, M\}}$ est un vecteur stochastique représentant la distribution initiale :

$$\sum_{i=1}^M \pi_i = 1 \tag{2}$$

$$\text{et } \forall i \in \{1, \dots, M\} \pi_i = P(s_1 = i);$$

- $A = (a_{ij})_{i, j \in \{1, \dots, M\}}$ est une matrice stochastique $M \times M$:

$$\forall i \in \{1, \dots, M\} \sum_{j=1}^M a_{ij} = 1 \tag{3}$$

$$\text{et } a_{ij} = P(s_{t+1} = j | s_t = i)$$

- $B = (b_{ik})_{i \in \{1, \dots, M\}, k \in \{1, \dots, N\}}$ est une matrice stochastique $M \times N$:

$$\forall i \in \{1, \dots, M\} \sum_{k=1}^N b_{ik} = 1 \tag{4}$$

$$\text{et } b_{ik} = P(y_t = k | s_t = i).$$

Dans les HMM, les paramètres sont fixés et ne dépendent donc pas du temps t . Puisque les HMM sont toujours dans le cadre des processus markoviens, les HMM sont des modèles sans mémoire, c'est-à-dire que la transition vers l'état suivant et l'observation d'un symbole ne dépendent que de l'état courant. Ainsi la probabilité conjointe d'avoir une séquence d'états s_1, s_2, \dots, s_T et une séquence d'observations y_1, y_2, \dots, y_T , où T est la longueur de la séquence, s'écrit :

$$P(s_1, s_2, \dots, s_T, y_1, y_2, \dots, y_T) = P(s_1) \times P(y_1|s_1) \times \prod_{t=2}^T [P(s_t|s_{t-1}) \times P(y_t|s_t)]. \quad (5)$$

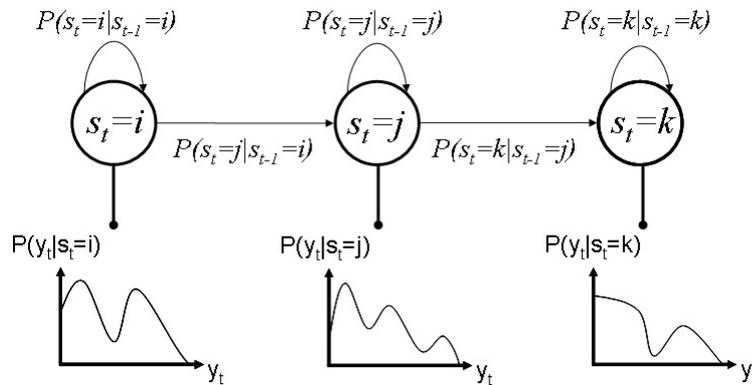


FIG. 18: Représentation graphique d'un modèle de Markov caché à trois états et à topologie gauche-droite (Bourlard et Bengio, 2002)

Les probabilités représentées par la matrice A , définie par les équations (3), sont appelées « probabilités de transition » et les probabilités représentées par la matrice B , définie par les équations (4), sont appelées « probabilités d'émission ». Généralement, une distribution paramétrique connue (comme la loi normale ou la loi de Poisson) est utilisée pour définir $P(y_t|s_t = i)$. La figure 18 illustre la représentation graphique d'un HMM à trois états et dont la topologie est de la forme gauche-droite ; cette topologie est très utilisée en reconnaissance de la parole (Bourlard et Bengio, 2002).

En pratique, l'utilisation des HMM nécessite la résolution de trois problèmes :

- estimer la probabilité d'une séquence d'observations $Y = y_1 \dots y_T$: généralement, la procédure avant-arrière (*forward-backward procedure*) est utilisée en réponse à ce problème ;
- choisir une séquence d'états telle que la probabilité conjointe de la séquence d'états s_1, \dots, s_T et de la séquence d'observations $Y = y_1 \dots y_T$ soit maximale : ce problème est résolu par l'algorithme de Viterbi (Forney, 1973) ;
- estimer les paramètres du HMM : en général, l'algorithme de Baum-Welch ou l'algorithme appelé « *segmental K-means algorithm* » est utilisé. L'algorithme de Baum-Welch (Rabiner et Juang, 1986) permet de réestimer les paramètres du HMM afin d'augmenter la probabilité des séquences d'observations contenues dans l'ensemble d'apprentissage jusqu'à ce qu'un maximum soit atteint. L'algorithme « *segmental K-means* » (Juang et L.R.Rabiner, 1990) consiste à maximiser la probabilité conjointe d'une séquence d'observations de l'ensemble d'apprentissage et de la séquence d'états qui maximise cette séquence d'observations.

L'utilisation des HMM et de leurs extensions (voir ci-dessous la sous-partie « Extensions ») dépend du critère que l'on utilise pour le décodage et l'entraînement de ces modèles. On parle de modèle de production lorsque le critère de vraisemblance est maximisé, c'est-à-dire lorsqu'on cherche à maximiser la probabilité $P(Y|M, \Lambda)$ où M désigne le modèle et Λ l'ensemble des paramètres de M . Dans ce cas, le but des HMM est de produire la séquence d'observations la plus probable. Au contraire, on parle de modèle de reconnaissance lorsqu'on s'intéresse aux probabilités *a posteriori*, soit $P(M|Y, \Lambda)$ et que la séquence d'observations est donnée en entrée du système; il a été montré (Bouclard et Morgan, 1993) que $P(M|Y, \Lambda)$ pouvait s'exprimer comme la probabilité conditionnelle de transition $P(s_t|y_t, s_{t-1})$.

Les HMM ont connu un grand succès en traitement du signal, particulièrement en reconnaissance automatique de la parole (Weber *et al.*, 2003; Bouclard et Morgan, 1993; Deller *et al.*, 1993; Jelinek, 1998) : étant donnée une séquence d'observations acoustiques, le but est d'obtenir une séquence de phonèmes ou de mots. Cependant, les hypothèses qui font des HMM des modèles efficaces et optimisés rendent parfois difficile leur généralisation (Bouclard et Bengio, 2002). Ainsi, les HMM souffrent de quelques limitations, dont entre autres :

- le choix *a priori* de la topologie du modèle et des distributions statistiques; par exemple, très souvent des mixtures de Gaussiennes sont utilisées pour représenter les probabilités d'émission;
- l'hypothèse que les séquences d'états sont des chaînes de Markov du premier ordre;
- l'hypothèse que les observations ne sont pas corrélées temporellement.

Extensions

Afin de dépasser ces limites, de nombreuses recherches ont été menées d'une part sur des extensions et d'autre part sur des algorithmes d'induction des topologies des HMM.

L'induction d'une topologie est un problème reconnu difficile par la communauté et les HMM induits sont généralement loin des HMM idéaux pour un problème donné (Dupont *et al.*, 2005). Par exemple, Thomsen (Thomsen, 2002) propose une méthode d'induction basée sur la recherche du meilleur HMM dans un espace d'HMM et emploie un algorithme évolutionniste pour trouver le HMM optimal. Dans (Stolcke et Omohundro, 1992; Stenger *et al.*, 2001), la technique employée consiste à considérer le plus grand nombre d'états possibles pour couvrir les données, puis de les fusionner entre eux. Devant la difficulté de ce problème, (Slimane *et al.*, 1996; Kwong *et al.*, 2001) proposent plutôt une optimisation de la topologie d'un HMM. Nous allons maintenant présenter quelques extensions des HMM; la figure 19 présente la hiérarchie des extensions que nous allons présenter.

Les HMM hiérarchiques (*hierarchical hidden Markov model*, HHMM) sont des HMM à structure hiérarchique permettant de représenter des séquences d'observations à différentes échelles. (ElHihi et Bengio, 1996) propose une discussion sur la nécessité de prendre en compte des niveaux hiérarchiques dans certaines modélisations, comme celle du langage naturelle, ou encore dans la reconnaissance des actions où divers niveaux d'abstraction peuvent être utilisés (Moore et Essa, 2002). Dans un HMM hiérarchique, une séquence peut contenir une sous-séquence elle-même modélisée par un HMM ou un HHMM. Néanmoins, les HHMM sont moins expressifs que les grammaires stochastiques non-contextuelles dans la mesure où la profondeur de la hiérarchie est limitée par la topologie des HHMM (Fine *et al.*, 1998).

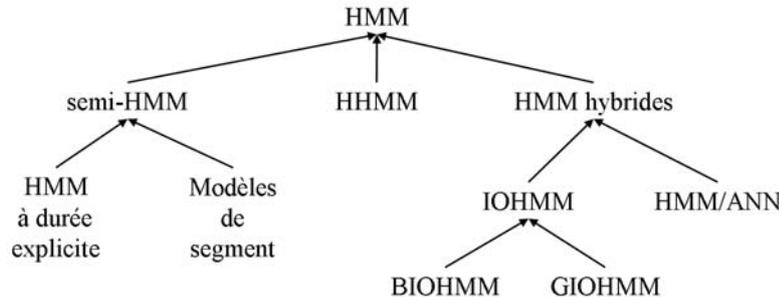


FIG. 19: Hiérarchie des extensions des modèles de Markov cachés

Les modèles semi-Markoviens cachés (*hidden semi-Markov model*, ou HSMM) sont des HMM dans lesquels chaque état peut émettre une séquence d'observations. Notons $Y(G_t)$ la sous-séquence produite par l'« état généralisé » G_t . Un état généralisé représente à la fois l'état s_t au sens classique des HMM et la longueur l_t de la séquence produite. On associe à $Y(G_t)$ la sous-séquence d'observations y_{t-l+1}, \dots, y_t . Après la production d'un segment, le prochain état est l'état G_{t_n} où $t_n = t + l_t$. La probabilité $P(Y(G_t)|s_t = k, L_t = l)$ peut suivre une distribution arbitraire. Si $P(Y(G_t)|s, l) = \prod_{i=t_l+1}^t P(y_i|s)$, le HSMM est appelé HMM à durée explicite (*explicit duration HMM*), abordés dans (Ferguson, 1980; Levinson, 1986; Mitchell *et al.*, 1995). Si $P(Y(G_t)|s, l)$ est elle-même modélisée par un HMM, on dit que le HSMM est un modèle de segment (*segment model*) introduits dans (Gales et Young, 1993; Ostendorf *et al.*, 1996). (Murphy, 2002) propose une comparaison des modèles semi-Markoviens et des HMM hiérarchiques avec d'autres extensions des HMM.

Les modèles hybrides HMM/Réseaux de neurones (*hybrid HMM/ANN* pour *hybrid Markov hidden model and artificial neural network*) permettent de profiter des avantages des HMM et des réseaux de neurones (voir la sous-section 2.1, page 108). Il a été montré (Richard et Lippmann, 1991) que si chaque unité de sortie d'un réseau de neurones était associée à un état k de l'espace d'états S du HMM, il était possible d'entraîner le réseau pour obtenir de bonnes estimations des probabilités *a posteriori* d'émissions en fonction de la séquence d'observations. En d'autres termes, notons $g_k(y_t|\Lambda)$ la valeur de la $k^{\text{ème}}$ unité de sortie du réseau de neurones lorsque le vecteur d'observations lui est présenté : $g_k(y_t|\Lambda^*) \approx P(s_t = k|y_t)$ où Λ^* est l'ensemble des paramètres du réseau de neurones entraîné.

Les *input output* HMM (Bengio et Frasconi, 1995, 1996) sont d'autres extensions hybrides dans lesquelles les probabilités d'émission et de transition sont calculées par des réseaux de neurones qui prennent en entrée une autre séquence X appelée séquence d'entrée (*input*). La séquence Y produite par le modèle est alors appelée séquence de sortie (*output*). Ainsi, les probabilités d'émission s'expriment par $P(y_t|s_t, x_t)$ et les probabilités de transition par $P(s_t|s_{t-1}, x_t)$. Si les HMM classiques sont homogènes parce qu'ils ne dépendent pas du temps, les IOHMM sont dits non-homogènes parce que la séquence d'entrée X varie en fonction du temps. Les IOHMM ont été utilisés en reconnaissance automatique de la parole et en reconnaissance d'écriture. Une extension des IOHMM, appelée IOHMM asynchrone, est introduite dans (Bengio et Bengio, 1996) et permet d'avoir une séquence d'entrée et une séquence de sortie de longueurs différentes. D'autres extensions permettent aux IOHMM de représenter des séquences spatiales ou de dimensions plus importantes : il s'agit respectivement des Bidirectional IOHMM et des Generalized IOHMM (Vullo, 2004).

Une comparaison des HMM avec les IOHMM et les HMM hybrides est proposée dans (Bour-

lard et Bengio, 2002).

1.3 Champs de Markov

Les champs de Markov fournissent une généralisation naturelle des modèles de Markov précédents. Si les chaînes de Markov ou les modèles de Markov sont représentés par un graphe dirigé, nous allons considérer à présent un graphe non dirigé (V, E) où V est l'ensemble dénombrable des sommets et E est l'ensemble des arêtes. Un champs aléatoire sur ce graphe est un ensemble de variables aléatoires $(X_i)_{i \in V}$. Si le graphe (V, E) doit être de degré borné, c'est-à-dire que chaque sommet appartient au même nombre d'arêtes, la propriété de Markov prend la forme :

$$P(X_i = s | X_j, j \neq i) = P(X_i = s | X_j, j \sim i) \quad (6)$$

où la notation $j \sim i$ signifie que les sommets i et j sont joints par une arête de E . De plus, $P(X_i = s | X_j, j \sim i)$ est strictement positif.

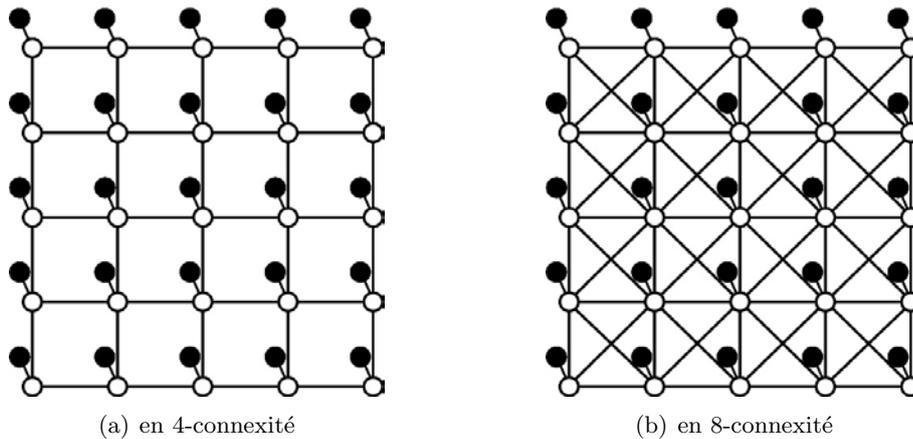


FIG. 20: Représentation de champs de Markov en 4 et 8-connexité. Les points noirs représentent les observations et les points blancs représentent les états cachés

Ainsi, dans un champs de Markov, la loi conditionnelle d'une variable X_i étant donné les autres ne dépendent que d'un voisinage borné de i : cela permet de tenir compte d'une corrélation spatiale entre les différents états. La figure 20 montre deux exemples de champs de Markov en 4-connexité et en 8-connexité.

Les champs de Markov ont de nombreuses applications dans le traitement de l'image (Li, 2001). En effet, une image est statistiquement considérée comme une réalisation d'un champs aléatoire Y indexé sur une grille. Un autre champ aléatoire caché X est associé à Y . (Gravier *et al.*, 2000; Ibáñez et Simó, 2000) proposent leur utilisation dans le cadre de la reconnaissance de la parole.

2 Modèles de Markov cachés contextuels

Le principal défaut des modèles précédents, dans notre cas, est l'hypothèse de Markov qui consiste à dire que la production d'une observation et la transition vers l'état suivant ne dépendent que de l'état courant. C'est la rigidité de cette hypothèse que les IOHMM tentent de

contourner en calculant les probabilités avec des réseaux de neurones prenant en entrée un vecteur d'observations. Cependant, les réseaux de neurones ont quelques difficultés à prendre en compte des variables symboliques (voir section 2.1, page 108) ce qui pourrait être utile pour représenter les jours de la semaine par exemple.

C'est pour palier ce défaut que nous avons étendu les HMM à un nouveau modèle, appelé modèle de Markov caché contextuel (Contextual Hidden Markov Model, ou CHMM). Il s'agit d'une extension des modèles de Markov, destinée à la production d'observations, dans laquelle les probabilités sont calculées en fonction du contexte dans lequel se trouve le modèle. En d'autres termes, si les HMM s'intéressent à $P(y_1, \dots, y_T)$ la probabilité d'observer la séquence d'observations y_1, \dots, y_T , un CHMM va s'intéresser à la grandeur $P(y_1, \dots, y_T | \theta_1)$, c'est-à-dire la probabilité de cette même séquence d'observations étant donné le contexte θ_1 dans lequel le premier symbole y_1 a été observé.

La contextualisation permet d'une part de faciliter le choix de la topologie du HMM et d'autre part d'avoir des probabilités plus proches de la réalité. Nous donnerons un exemple de ces avantages en modélisant les grilles de programmes (voir section 3, page 99). La prise en compte des contextes nécessite quelques prérequis : un contexte défini et une fonction d'évolution qui permet, lorsqu'un symbole est émis par le modèle, de mettre à jour le contexte courant (voir la sous-section suivante).

Ce contexte permet de gérer de façon transparente un nombre d'états décuplés, ce qui rend les CHMM plus facilement utilisables en pratique tout en étant aussi performants que les HMM sur le plan théorique. La prise en compte d'un contexte permet aussi d'avoir une distribution initiale des états plus vraisemblable et des probabilités de stationnement dans un état dont la forme n'est pas forcément géométrique. Les CHMM favorisent les transitions déterministes d'un état à un autre : ainsi, le nombre de séquences produites dans un contexte donné est plus faible et les séquences plus proches du phénomène observé. En contrepartie, les CHMM perdent le pouvoir de généralisation des HMM classiques. Cependant, il faut considérer que lorsque les HMM sont utilisés pour prédire des séquences (des séries temporelles ou des protéines), seule la séquence la plus probable est utilisée et cette séquence est très rarement issue de la généralisation du HMM.

Nous présentons à présent les définitions des contextes et des fonctions d'évolution avant de définir formellement un CHMM.

2.1 Définitions préliminaires

Nous introduisons dans cette section deux définitions préliminaires : le contexte et la fonction d'évolution.

Définition 1 (Contexte) *Un contexte θ est un ensemble de variables x_1, \dots, x_n à valeurs dans les domaines discrets ou continus $\{D_1, \dots, D_n\}$. Une instance θ_i de ce contexte correspond à l'affectation d'une valeur à chacune de ses variables :*

$$\forall i \in \{1, \dots, n\}, x_i = v_i \text{ avec } v_i \in D_i. \quad (7)$$

Par commodité de langage dans le reste de ce manuscrit, nous appellerons « contexte » l'instance θ_i . Il nous reste à définir les fonctions d'évolution qui permettent de passer d'un contexte au contexte suivant.

Définition 2 (Fonction d'évolution) *Soit Θ l'ensemble de toutes les instances possibles d'un*

contexte θ . On appelle fonction d'évolution une fonction F_θ définie par :

$$\begin{aligned} F_\theta &: \Theta \times D_{p_1} \times \dots \times D_{p_n} \rightarrow \Theta \\ &\theta_t, p_1, \dots, p_n \mapsto \theta_{t+1} \end{aligned} \quad (8)$$

où les p_i sont des paramètres extérieurs et les D_{p_i} leurs domaines respectifs.

La notation F_θ signifie que la fonction d'évolution s'applique au contexte θ : ainsi, F_θ doit prendre au moins en paramètre une instance du contexte θ . En effet, les fonctions d'évolution permettent de passer de l'instance du contexte au temps t à l'instance du contexte au temps $t + 1$ en prenant divers paramètres : il peut s'agir par exemple du dernier symbole émis par le modèle, ou de paramètres extérieurs.

Dans la mesure où les CHMM sont préconisés pour la prédiction de séquences et dans la mesure où la prise en compte du contexte doit permettre de diminuer le nombre de séquences « proches » de la séquence à prédire, il est préférable d'utiliser une fonction F_θ déterministe. Dans le cas contraire, si F_θ est stochastique, le nombre de séquences prédites sera augmenté ce qui est contre nature pour un CHMM. Nous nous limitons dans le reste de ce chapitre au cas des fonctions d'évolution déterministes, c'est-à-dire que nous considérons que la probabilité d'avoir le contexte θ_{i+1} étant donné le contexte θ_i $P(\theta_{i+1}|\theta_i)$ est égale à 1.

Maintenant que ces définitions ont été données, nous pouvons définir l'extension des HMM que nous avons appelée « modèle de Markov caché contextuel ».

2.2 Modèles de Markov cachés contextuels

Définition 3 (CHMM) *Un modèle de Markov caché contextuel (Contextual Hidden Markov Model, CHMM) est complètement défini par le 7-uplet $\langle S, \Sigma, \Theta, F_\theta, \pi_\theta, a_\theta, b_\theta \rangle$ où :*

- S est un espace d'états au sens défini ci-dessus (voir 1.1), de cardinal N , et s_t désigne l'état au temps t ;
- Σ est l'ensemble des symboles observables, valeurs de la variable aléatoire Y , de cardinal Z . y_t désigne le symbole observé au temps t ;
- Θ est l'ensemble des instances possibles d'un contexte θ , et on notera θ_t l'instance de θ au temps t ;
- F_θ est une fonction d'évolution pour le contexte θ : θ_{t+1} sera donc calculé à partir de F_θ et de θ_t ;
- $\pi_i(\theta_1)$ où $i \in S$ représente la distribution initiale sur les états :

$$\sum_{i=1}^N \pi_i(\theta_1) = 1 \quad (9)$$

$$\text{et } \pi_i(\theta_1) = P(s_1 = i | \theta_1) \forall i \in 1, \dots, N ;$$

- $a_{ij}(\theta_t)$ avec $i, j \in S$ est une autre distribution sur les états définie par :

$$\forall i \in 1, \dots, N, \forall \theta_t \in \Theta \sum_{j=1}^N a_{ij}(\theta_t) = 1 \quad (10)$$

$$\text{et } \forall \theta_t \in \Theta a_{ij}(\theta_t) = P(s_t = j | s_{t-1} = i, \theta_t)$$

– $b_{ik}(\theta_t)$ avec $i \in S, k \in \Sigma$ est une distribution sur les symboles définie par :

$$\begin{aligned} \forall i \in 1, \dots, N, \forall \theta_t \in \Theta \quad \sum_{k=1}^Z b_{ik}(\theta_t) &= 1 \\ \text{et } \forall \theta_t \in \Theta \quad b_{ik}(\theta_t) &= P(y_t = k | s_t = i, \theta_t). \end{aligned} \quad (11)$$

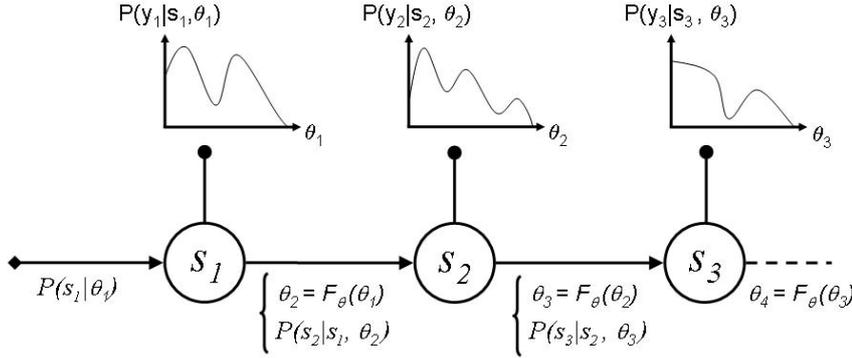


FIG. 21: Production ou classification d’une séquence d’observations avec un CHMM

Autrement dit, un CHMM est un HMM dans lequel les probabilités sont calculées en fonction d’un contexte : les matrices A et B et le vecteur π des HMM (voir page 81) sont remplacés par des fonctions représentant les diverses distributions et paramétrées par le contexte θ . Chaque fois qu’une observation y_t est générée par le modèle, le contexte θ_t est mis à jour et devient le contexte θ_{t+1} par la fonction F_θ . La figure 21 illustre la production d’une séquence d’observation Y avec un CHMM. En pratique, un CHMM revient à avoir autant de HMM qu’il y a de contextes dans Θ , partageant le même espace d’états, et à utiliser le HMM qui correspond à θ_t .

Afin de maintenir la propriété des processus de Markov, c’est-à-dire celle de représenter des processus sans mémoire, nous faisons l’hypothèse suivante :

Propriété 1 (Processus sans mémoire) Si (\mathcal{M}, Λ) est un CHMM de paramètres Λ , et si T est la longueur de l’observation $Y = y_1, \dots, y_T$ alors :

$$\forall t \in \{2, \dots, T\}, P(s_t | s_1, \dots, s_{t-1}, \theta_1, \dots, \theta_t) = P(s_t | s_{t-1}, \theta_t) \quad (12)$$

$$\forall t \in \{1, \dots, T\}, P(y_t | s_1, \dots, s_t, \theta_1, \dots, \theta_t) = P(y_t | s_t, \theta_t). \quad (13)$$

La propriété 1 signifie que les probabilités d’émission d’un symbole y_t au temps t ne dépendent que de l’état courant s_t et du contexte courant θ alors que les probabilités de transition dépendent de l’état courant s_t et du contexte suivant θ_{t+1} . La figure 22 représente le réseau bayésien qui exprime la dépendance conditionnelle des différentes variables aléatoires dans l’interprétation probabiliste d’un CHMM (figure 22.a) en comparaison à celui d’un HMM (figure 22.b) : le réseau bayésien est une interprétation graphique de la propriété précédente.

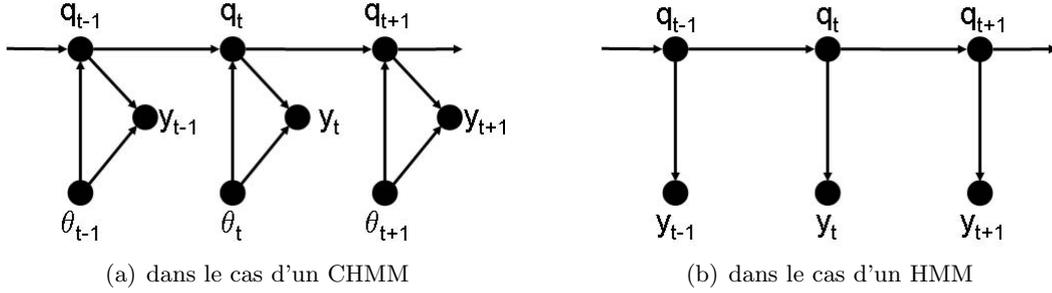


FIG. 22: Comparaison des réseaux bayésiens exprimant les dépendances conditionnelles des variables aléatoires d'un CHMM et d'un HMM

Propriété 2 (Probabilité d'une observation) Soit (\mathcal{M}, Λ) un CHMM et soit une séquence d'observations $Y = y_1, \dots, y_T$. Alors :

$$\begin{aligned}
 P(Y|\mathcal{M}, \Lambda, \theta_1) &= P(s_1|\theta_1) \times P(y_1|s_1, \theta_1) \times \prod_{i=2}^T [P(s_i|s_{i-1}, \theta_i) \times P(y_i|s_i, \theta_i)] \\
 &= \pi_1(\theta_1) \times b_{s_1 y_1}(\theta_1) \times \prod_{i=2}^T [a_{s_{i-1} s_i}(\theta_i) \times b_{s_i y_i}(\theta_i)].
 \end{aligned} \tag{14}$$

La principale différence avec un IOHMM (voir sous-section 1.2, page 84) est le fait que les probabilités ne sont pas seulement calculées en fonction de la séquence produite. Le contexte est une notion plus large, qui peut prendre des valeurs extérieures au système, comme par exemple des valeurs produites par des détecteurs ou encore des sondes. De plus, n'importe quelle méthode de régression peut être utilisée dans le calcul des probabilités : par exemple, nous verrons dans le chapitre 8 (page 105) que des arbres de régression peuvent être utilisés pour évaluer les différentes valeurs de B . Les CHMM tiennent compte d'un nombre plus élevé de relations de dépendance dans l'estimation des probabilités, comparé aux HMM qui ne tiennent compte que de la dépendance avec l'état courant.

Dans la section suivante, nous établirons des algorithmes pour cette extension des modèles de Markov.

2.3 Résolution des trois problèmes de l'extension

Toute extension des modèles de Markov doit résoudre trois problèmes – dont nous avons déjà parlé page 82 pour les HMM – afin d'être utilisable dans des applications réelles. Dans le cas des CHMM, les problèmes peuvent être formulés ainsi :

- étant donné la séquence d'observation $y = y_1, \dots, y_T$ et un CHMM (\mathcal{M}, Λ) , comment calculer $P(y|\Lambda, \theta_1)$ la probabilité d'observer la séquence y étant donné le modèle (\mathcal{M}, Λ) ?
- étant donné la séquence d'observation $y = y_1, \dots, y_T$ et un CHMM (\mathcal{M}, Λ) , comment choisir la séquence d'états $S = s_1, \dots, s_T$ optimale (c'est-à-dire celle qui explique le mieux la séquence d'observations) ?
- comment ajuster les paramètres du CHMM (\mathcal{M}, Λ) ?

Premier problème

Le premier problème est un problème d'évaluation de la probabilité d'une séquence d'observations $y = y_1, \dots, y_T$ par le CHMM (\mathcal{M}, Λ) sachant le contexte θ_1 dans lequel le premier symbole a été observé.

La méthode consistant à utiliser directement la propriété 2 des CHMMs est très coûteuse en temps de calcul. En effet, il faut additionner les probabilités de y pour chaque séquence d'états de longueur T . Si N est le nombre d'états de (\mathcal{M}, Λ) , il existe N^T séquences d'états à tester. Pour chaque séquence d'états, $2T - 1$ multiplications sont nécessaires ; enfin, $N^T - 1$ additions sont nécessaires entre les différentes séquences d'états : la complexité de cette méthode s'élèverait donc à $(2T - 1)N^T$ multiplications, $N^T - 1$ additions et $(T - 1)N^T$ évolutions de contexte (dont la complexité dépend elle-même de la fonction d'évolution F_θ).

Généralement, la procédure appelée *forward-backward* permet un calcul plus efficace de $P(y|\Lambda)$. Comme le montre la figure 23, le calcul de la variable α_T pour un HMM classique revient à un parcours de treillis dont la complexité est de l'ordre de N^2T . Les abscisses de cette figure représentent les différents états pris dans chacun des contextes et les ordonnées représentent les différentes observations de la séquence.

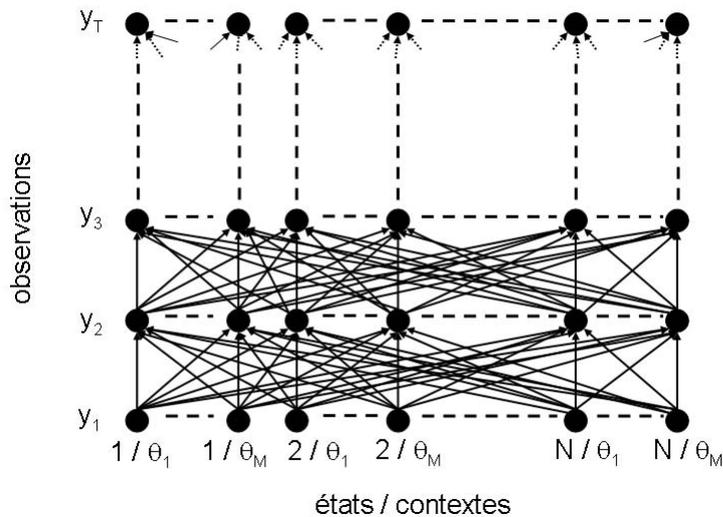


FIG. 23: Treillis représentant les calculs de la procédure *forward* pour un HMM

Nous avons donc adapté cette procédure aux CHMMs. Pour cela, posons

$$\begin{aligned} \alpha_t(i) &= P(y_1, \dots, y_t, s_t = i | \Lambda, \theta_1, \dots, \theta_t) \\ &= P(y_1, \dots, y_t, s_t = i | \Lambda, \theta_t) \end{aligned} \tag{15}$$

c'est-à-dire la probabilité de la séquence d'observations partielle y_1, \dots, y_t et du stationnement sur l'état s_i sachant que le système est dans le contexte θ_t .

Etant donné le modèle (\mathcal{M}, Λ) , la variable α peut être définie par récurrence :

$$\begin{aligned}
 \alpha_t(i) &= P(y_1, \dots, y_t, s_t = i | \theta_1, \dots, \theta_t) \\
 &= \sum_j P(y_1, \dots, y_{t-1}, s_{t-1} = j, y_t, s_t = i | \theta_1, \dots, \theta_t) \\
 &= \sum_j P(y_1, \dots, y_{t-1}, s_{t-1} = j | \theta_1, \dots, \theta_t) P(y_t, s_t = i | y_1, \dots, y_{t-1}, s_{t-1} = j, \theta_1, \dots, \theta_t) \\
 &= \sum_j P(y_1, \dots, y_{t-1}, s_{t-1} = j | \theta_1, \dots, \theta_{t-1}) P(y_t, s_t = i | y_1, \dots, y_{t-1}, s_{t-1} = j, \theta_t) \\
 &= \sum_j \alpha_{t-1}(j) P(y_t, s_t = i | s_{t-1} = j, \theta_t) \\
 &= \sum_j \alpha_{t-1}(j) P(s_t = i | s_{t-1} = j, \theta_t) P(y_t | s_{t-1} = j, s_t = i, \theta_t) \\
 &= \sum_j \alpha_{t-1}(j) a_{ji}(\theta_t) P(y_t | s_t = i, \theta_t) = \sum_j \alpha_{t-1}(j) a_{ji}(\theta_t) b_{iy_t}(\theta_t)
 \end{aligned} \tag{16}$$

en utilisant les équations (12) et (13). Pour faciliter l'écriture de la démonstration, nous avons choisi de ne pas distinguer le même état dans différents contextes. Nous allons introduire de nouvelles notations : Θ_t représente tous les contextes disponibles au temps t , θ_{t+1}^i est le contexte mis à jour par l'état i après l'émission du symbole y_t et j_{θ_t} est l'état j dans le contexte θ_t . Par exemple, $\Theta_1 = \{\theta_1\}$, et $\Theta_2 = \{\theta_2^i, 1 \leq i \leq N\}$, et ainsi de suite. À présent, l'équation (16) s'écrit :

$$\begin{aligned}
 \alpha_t(i_{\theta_t}) &= P(y_1, \dots, y_t, s_t = i | \Lambda, \theta_1, \dots, \theta_t) \\
 &= \sum_{\substack{j \in \mathcal{S}, \\ \theta_{t-1} \in \Theta_{t-1}}} \alpha_{t-1}(j_{\theta_{t-1}}) a_{ji}(F_{\theta}(\theta_{t-1})) P(y_t | s_t = i, F_{\theta}(\theta_{t-1}))
 \end{aligned} \tag{17}$$

ce qui représente la probabilité d'avoir la sous-séquence y_1, \dots, y_n et d'être arrêté sur l'état i dans le contexte $F_{\theta}(\theta_{t-1}) = \theta_t$ mais avant la mise à jour du contexte. Ainsi $P(y | \Lambda, \theta_1)$ peut s'écrire :

$$P(y | \Lambda, \theta_1) = \sum_{\substack{i \in \mathcal{S} \\ \theta_T \in \Theta_T}} \alpha_T(i_{\theta_T}) \tag{18}$$

Dans l'algorithme 1, nous avons simplifié les notations : F dénote la fonction d'évaluation du contexte et θ et θ' sont deux instances du contexte. Cet algorithme permet d'évaluer la probabilité $P(y | \Lambda, \theta_1)$ à partir de la variable α . Il est important de comprendre qu'un état est atteint dans un contexte donné : chaque état atteint dans un contexte donné doit être considéré comme un état différent. Chaque étage de notre treillis va contenir l'ensemble des états pris dans chacun des contextes produits par le niveau précédent. Si M est le nombre de contextes dans Θ , un niveau du treillis contiendra au plus $M \times N$ états. De cette observation, il est facile de déduire que la partie *itération* de l'algorithme 1 a une complexité de $(T-1)(MN)^2$ dans le pire des cas. La partie *initialisation* a une complexité de l'ordre de N et la partie *termination* de l'ordre de MN dans le pire des cas. De cela, on déduit que la complexité dans le pire des cas de l'algorithme 1 est en $O(TM^2N^2)$. Dans le pire des cas, l'utilisation des CHMM revient à l'utilisation d'un HMM classique après multiplication des états. En revanche, en pratique, le nombre moyen d'états à chaque étape est réduit : cela sera montré expérimentalement dans la section 3.4 (page 148).

Algorithme 1 Algorithme d'évaluation des probabilités

```

function  $P(y|\theta_1) \leftarrow$  EVALUATION(séquence  $y = y_1, \dots, y_T$ , contexte  $\theta_1$ )
▷ entrée : une séquence d'observation  $y = y_1, \dots, y_T$ , le contexte initial  $\theta_1$ 
▷ sortie : la probabilité de  $y$  sachant  $\theta_1$ 

  \\ Initialisation
  for  $i \leftarrow 1$  to  $N$  do
     $\alpha_1(i_{\theta_1}) \leftarrow \pi_i(\theta_1)b_{iy_1}(\theta_1)$ 
  end for

  \\ Itération
  for  $t \leftarrow 2$  to  $T$  do
    for  $j \leftarrow 1$  to  $N$  do
      for  $\theta' \in \Theta_t$  do
         $\alpha_t(i_{\theta'}) \leftarrow 0$ 
        for  $j \leftarrow 1$  to  $N$  do
          for  $\theta \in \Theta_{t-1}$  do
            if  $\theta' = F(\theta)$  then
               $\alpha_t(i_{\theta'}) \leftarrow \alpha_t(i_{\theta'}) + \alpha_{t-1}(j_{\theta})a_{j\theta i_{\theta'}}(\theta')b_{iy_t}(\theta')$ 
            end if
          end for
        end for
      end for
    end for
  end for

  \\ Fin
   $P \leftarrow 0$ 
  for  $i \leftarrow 1$  to  $N$  do
    for  $\theta \in \Theta$  do
       $P \leftarrow P + \alpha_T(i_{\theta})$ 
    end for
  end for

  return  $P$ 

end function

```

Les performances de l'algorithme d'évaluation pour un CHMM sont directement liées au nombre de contextes et d'états considérés. Cela confirme ce qui était annoncé au début de ce chapitre : les CHMMs sont de bons modèles générateurs, dans la mesure où les séquences générées sont plus proches de la réalité grâce aux contextes. En revanche, ce sont des classificateurs d'une complexité plus élevée que les HMMs classiques pour une tâche de classification. On remarque que dans le cas où l'ensemble des contextes est un singleton, on retrouve l'algorithme des HMMs classiques.

Si la variable α suffit à calculer la probabilité d'une observation, une variable β a aussi été introduite afin de calculer la probabilité *a posteriori* de cette même observation. En notant θ_{t+1}^i

le contexte mis à jour par l'état i après l'émission du symbole y_t , posons :

$$\begin{aligned}
 \beta_t(i) &= P(y_{t+1}, \dots, y_T | s_t = i, \theta_{t+1}^i) \\
 &= \sum_j P(s_{t+1} = j, y_{t+1}, \dots, y_T | s_t = i, \theta_{t+1}^i) \\
 &= \sum_j [P(s_{t+1} = j, y_{t+1} | s_t = i, \theta_{t+1}^i) \\
 &\quad \times P(y_{t+2}, \dots, y_T | s_{t+1} = j, y_{t+1}, s_t = i, \theta_{t+1}^i)] \\
 &= \sum_j [P(s_{t+1} = j | s_t = i, \theta_{t+1}^i) P(y_{t+1} | s_{t+1} = j, s_t = i, \theta_{t+1}^i) \\
 &\quad \times P(y_{t+2}, \dots, y_T | s_{t+1} = j, y_{t+1}, s_t = i, \theta_{t+1}^i)] \\
 &= \sum_j [a_{ij}(\theta_{t+1}^i) P(y_{t+1} | s_{t+1} = j, s_t = i, \theta_{t+1}^i, F(\theta_{t+1}^i)) \\
 &\quad \times P(y_{t+2}, \dots, y_T | s_{t+1} = j, y_{t+1}, s_t = i, \theta_{t+1}^i)] \\
 &= \sum_j [a_{ij}(\theta_{t+1}^i) b_{jy_{t+1}}(\theta_{t+1}^i) \beta_{t+1}(j)] \tag{19}
 \end{aligned}$$

puisque pour un j fixé, puisque F est une fonction déterministe, on a la relation

$$\begin{aligned}
 &P(y_{t+2}, \dots, y_T | s_{t+1} = j, y_{t+1}, s_t = i, \theta_{t+1}^i) \\
 &= \frac{P(y_{t+2}, \dots, y_T | s_{t+1} = j, y_{t+1}, s_t = i, \theta_{t+1}^i, F(\theta_{t+1}^i))}{P(F(\theta_{t+1}^i) | s_{t+1} = j, y_{t+1}, s_t = i, \theta_{t+1}^i)} \\
 &= P(y_{t+2}, \dots, y_T | s_{t+1} = j, y_{t+1}, s_t = i, \theta_{t+1}^i, F(\theta_{t+1}^i))
 \end{aligned}$$

Les variables α et β permettent d'exprimer plus clairement la probabilité *a posteriori* de la

transition d'un état à un autre au temps t :

$$\begin{aligned}
 & P(s_t = i, s_{t+1} = j | y, \theta_1, \dots, \theta_t, \theta_{t+1}^i) \\
 &= \frac{P(s_t = i, s_{t+1} = j, y | \theta_1, \dots, \theta_{t+1}^i)}{P(y | \theta_1, \dots, \theta_{t+1}^i)} \\
 &= \frac{P(y_1, \dots, y_t, s_t = i, s_{t+1} = j, y_{t+1}, \dots, y_T | \theta_1, \dots, \theta_{t+1}^i)}{P(y | \theta_1, \dots, \theta_{t+1}^i)} \\
 &= \frac{P(y_1, \dots, y_t, s_t = i | \theta_1, \dots, \theta_{t+1}^i) P(s_{t+1} = j, y_{t+1}, \dots, y_T | y_1, \dots, y_t, s_t = i, \theta_1, \dots, \theta_{t+1}^i)}{P(y | \theta_1, \dots, \theta_{t+1}^i)} \\
 &= \frac{P(y_1, \dots, y_t, s_t = i | \theta_1, \dots, \theta_t) P(s_{t+1} = j, y_{t+1}, \dots, y_T | y_1, \dots, y_t, s_t = i, \theta_1, \dots, \theta_{t+1}^i)}{P(y | \theta_1, \dots, \theta_{t+1}^i)} \\
 &= \frac{\alpha_t(i) P(s_{t+1} = j, y_{t+1}, \dots, y_T | s_t = i, \theta_1, \dots, \theta_{t+1}^i)}{P(y | \theta_1, \dots, \theta_{t+1}^i)} \\
 &= \frac{\alpha_t(i) P(s_{t+1} = j | s_t = i, \theta_1, \dots, \theta_{t+1}^i) P(y_{t+1}, \dots, y_T | s_{t+1} = j, s_t = i, \theta_1, \dots, \theta_{t+1}^i)}{P(y | \theta_1, \dots, \theta_{t+1}^i)} \\
 &= \frac{\alpha_t(i) a_{ij}(\theta_{t+1}^i) P(y_{t+1}, \dots, y_T | s_{t+1} = j, \theta_1, \dots, \theta_{t+1}^i)}{P(y | \theta_1, \dots, \theta_{t+1}^i)} \\
 &= \frac{\alpha_t(i) a_{ij}(\theta_{t+1}^i) P(y_{t+1} | s_{t+1} = j, \theta_1, \dots, \theta_{t+1}^i) P(\dots, y_T | y_{t+1}, s_{t+1} = j, \theta_1, \dots, \theta_{t+1}^i)}{P(y | \theta_1, \dots, \theta_{t+1}^i)} \\
 &= \frac{\alpha_t(i) a_{ij}(\theta_{t+1}^i) b_{jy_{t+1}}(\theta_{t+1}^i) P(\theta_2, \dots, y_T | y_{t+1}, s_{t+1} = j, \theta_1, \dots, \theta_{t+1}^i)}{P(y | \theta_1, \dots, \theta_{t+1}^i)} \\
 &= \frac{\alpha_t(i) a_{ij}(\theta_{t+1}^i) b_{jy_{t+1}}(\theta_{t+1}^i) \beta_{t+1}(j)}{P(y | \theta_1, \dots, \theta_{t+1}^i)} \tag{20}
 \end{aligned}$$

en utilisant la propriété de processus sans mémoire des CHMM (propriété 1, page 88).

Second problème

Le second problème à résoudre pour que les CHMM soient utilisables consiste à trouver une méthode efficace pour calculer la séquence d'états qui « explique » le mieux la séquence d'observation. Ce problème peut aussi être formulé comme la recherche d'une séquence d'états optimale. Puisqu'il existe plusieurs critères d'optimalité, ce second problème ne se résout pas d'une façon unique. Les algorithmes gloutons, c'est-à-dire qui à chaque étape choisissent la transition la plus probable, peuvent mener à des séquences optimales inattendues. L'algorithme de Viterbi (Forney, 1973), basé sur une méthode de programmation dynamique, est la solution la plus connue à ce problème. Son critère d'optimalité consiste à trouver la séquence qui maximise la probabilité conjointe de la séquence d'observations et de la séquence d'états. Dans le cas des CHMMs, cela s'écrit :

$$\begin{aligned}
 \hat{s} &= \arg \max_s P(s | y, \theta_1) \\
 &= \arg \max_s P(s | y, \theta_1) P(y | \theta_1, \dots, \theta_T) \\
 &= \arg \max_s P(s, y | \theta_1) \tag{21}
 \end{aligned}$$

La programmation dynamique permet de calculer $\hat{p}(i, t)$ qui est définie par :

$$\begin{aligned}
 \hat{p}(i, t) &= \max_{\substack{s_1, \dots, s_t \\ s_t = i}} P(y_1, \dots, y_t, s_1, \dots, s_t | \theta_1, \dots, \theta_t) \\
 &= \max_{\substack{s_1, \dots, s_t \\ s_t = i}} P(y_1, \dots, y_{t-1}, s_1, \dots, s_{t-1}, y_t, s_t | \theta_1, \dots, \theta_t) \\
 &= \max_{\substack{s_1, \dots, s_t \\ s_t = i}} [P(y_1, \dots, y_{t-1}, s_1, \dots, s_{t-1} | \theta_1, \dots, \theta_t) \\
 &\quad P(y_t, s_t | y_1, \dots, y_{t-1}, s_1, \dots, s_{t-1}, \theta_1, \dots, \theta_t)] \\
 &= \max_{\substack{s_1, \dots, s_t \\ s_t = i}} [P(y_1, \dots, y_{t-1}, s_1, \dots, s_{t-1} | \theta_1, \dots, \theta_t) \\
 &\quad P(y_t, s_t | s_{t-1}, \theta_t)] \\
 &= \max_{\substack{s_1, \dots, s_t \\ s_t = i}} [P(y_1, \dots, y_{t-1}, s_1, \dots, s_{t-1} | \theta_1, \dots, \theta_t) \\
 &\quad P(y_t, s_t | s_{t-1}, \theta_1, \dots, \theta_t)] \\
 &= \max_{\substack{s_1, \dots, s_t \\ s_t = i}} [P(y_1, \dots, y_{t-1}, s_1, \dots, s_{t-1} | \theta_1, \dots, \theta_t) \\
 &\quad P(s_t | s_{t-1}, \theta_1, \dots, \theta_t) P(y_t | s_t, s_{t-1}, \theta_1, \dots, \theta_t)] \\
 &= \max_{\substack{s_1, \dots, s_t \\ s_t = i}} [P(y_1, \dots, y_{t-1}, s_1, \dots, s_{t-1} | \theta_1, \dots, \theta_t) \\
 &\quad P(s_t | s_{t-1}, \theta_t) P(y_t | s_t, \theta_t)] \\
 &= \max_{s_1, \dots, s_{t-1}} [P(y_1, \dots, y_{t-1}, s_1, \dots, s_{t-1} | \theta_1, \dots, \theta_t)] \\
 &\quad P(s_t = i | s_{t-1}, \theta_t) P(y_t | s_t = i, \theta_t) \\
 &= \max_j \max_{\substack{s_1, \dots, s_{t-1} \\ s_{t-1} = j}} [P(y_1, \dots, y_{t-1}, s_1, \dots, s_{t-1} | \theta_1, \dots, \theta_{t-1})] \\
 &\quad P(s_t = i | s_{t-1}, \theta_t^j) P(y_t | s_t = i, \theta_t^j) \\
 &= \max_j \left[\max_{\substack{s_1, \dots, s_{t-1} \\ s_{t-1} = j}} P(y_1, \dots, y_{t-1}, s_1, \dots, s_{t-1} | \theta_1, \dots, \theta_{t-1}) \right] \\
 &\quad P(s_t = i | s_{t-1}, \theta_t^j) P(y_t | s_t = i, \theta_t^j) \\
 &= \max_j \hat{p}(j, t-1) a_{ji}(\theta_t^j) b_{iy_t}(\theta_t^j) \tag{22}
 \end{aligned}$$

Encore une fois, nous obtenons un algorithme dont la complexité est de l'ordre de $O(M^2 N^2 T)$ dans le pire des cas. En effet, dans le pire des cas, chaque Θ_t contient au plus M éléments. L'algorithme de Viterbi n'est pas très utilisé dans le cadre d'un modèle génératif mais sert surtout pour de la classification de séquences, étant donné que celle-ci doit être fournie.

Algorithme 2 Algorithme de Viterbi contextuel

```

function  $\hat{s} \leftarrow$  EVALUATION(séquence  $y = y_1, \dots, y_T$ , contexte  $\theta_1$ )
▷ entrée : une séquence d'observation  $y = y_1, \dots, y_T$ , le contexte initial  $\theta_1$ 
▷ sortie : la séquence d'états optimale  $\hat{s}$ 

  \\ Initialisation
  for  $i \leftarrow 1$  to  $N$  do
     $\hat{p}(i, 1) \leftarrow \pi_i(\theta_1) b_{iy_1}(\theta_1)$ 
  end for

  \\ Itération
  for  $t \leftarrow 2$  to  $T$  do
    for  $j \leftarrow 1$  to  $N$  do
      for  $\theta \in \Theta_t$  do
         $\hat{p}(i_\theta, t) \leftarrow \max_{j'_\theta} \hat{p}(j'_\theta, t-1) a_{ji}(\theta_t^{j'_\theta}) b_{iy_t}(\theta_t^{j'_\theta})$ 
         $prev(i_\theta, t) \leftarrow \arg \max_{j'_\theta} \hat{p}(j'_\theta, t-1) a_{ji}(\theta_t^{j'_\theta}) b_{iy_t}(\theta_t^{j'_\theta})$ 
      end for
    end for
  end for

  \\ Fin
  for  $\theta \in \Theta$  do
     $s_T \leftarrow \arg \max_{i_\theta} \hat{p}(i, T)$ 
  end for
  for  $t \leftarrow T-1$  to  $1$  do
    for  $\theta \in \Theta$  do
       $s_t \leftarrow prev(s_{t+1}, t+1)$ 
    end for
  end for

  return  $s_1, \dots, s_T$ 
end function

```

Troisième problème

Enfin, le dernier problème à résoudre est l'estimation des paramètres d'un CHMM. Nous considérons ici l'entraînement d'un CHMM lorsque son espace d'états, son espace des contextes et sa fonction d'évolution sont connus et en supposant que nous avons uniquement des séquences d'observations et leurs contextes (autrement dit, nous ne connaissons pas les séquences d'états associées aux séquences d'observations). L'algorithme que nous proposons pour l'estimation des paramètres d'un CHMM est basé sur le principe du maximum de vraisemblance. Soit \mathcal{D} l'ensemble contenant P séquences d'apprentissage :

$$\mathcal{D} = \{\theta_1(p), \dots, \theta_{T_p}(p), y_1(p), \dots, y_{T_p}(p)\} \quad (23)$$

avec $p = 1, \dots, P$ et où T_p représente la longueur de la $p^{\text{ème}}$ séquence d'apprentissage. À présent, si on note Λ l'ensemble des paramètres, la fonction de vraisemblance s'écrit :

$$L(\Lambda ; \mathcal{D}) = \prod_{p=1}^P P(y_1(p), \dots, y_{T_p}(p) | \theta_1(p), \dots, \theta_{T_p}(p) ; \Lambda) \quad (24)$$

Selon le principe du maximum de vraisemblance, les paramètres optimaux sont obtenus en maximisant l'expression (24). Ce problème peut être résolu par l'algorithme EM.

L'algorithme EM est une approche itérative du problème de l'estimation du maximum de vraisemblance (Dempster *et al.*, 1977). Chaque itération se décompose en une phase d'estimation (E) et une phase de maximisation (M) dont le but est de maximiser la fonction de log-vraisemblance définie par $l(\Lambda ; \mathcal{D}) = \log L(\Lambda ; \mathcal{D})$. L'astuce de cet algorithme est de considérer que ce problème d'optimisation pourrait être simplifié par la connaissance d'autres variables \mathcal{V} , qualifiées de « manquantes » ou « cachées ». On considère alors $l_c(\Lambda ; \mathcal{D}_c)$ la fonction complète de log-vraisemblance et \mathcal{D}_c est appelé l'ensemble d'apprentissage complet. Les variables cachées sont choisies de façon à ce que $l_c(\Lambda ; \mathcal{D}_c)$ soit facilement maximisée étant donné \mathcal{V} .

Cependant, puisque ces variables ne sont pas directement observables, l_c est une variable aléatoire et ne peut donc être maximisée directement. L'algorithme EM considère donc la fonction auxiliaire :

$$Q(\Lambda ; \hat{\Lambda}) = E_{\mathcal{V}} \left[l_c(\Lambda ; \mathcal{D}) | \mathcal{D}, \hat{\Lambda} \right] \quad (25)$$

qui représente l'espérance de la log-vraisemblance complète, étant donné l'ensemble d'apprentissage \mathcal{D} et l'ensemble des paramètres $\hat{\Lambda}$ calculés à la fin de l'itération précédente. Le calcul de Q revient en fait à compléter les variables manquantes en utilisant d'une part les données observées et d'autre part la précédente estimation des paramètres. Puisque la fonction auxiliaire Q est déterministe, il est possible de la maximiser. L'algorithme EM itère ainsi ces deux étapes :

- durant la phase d'estimation, il calcule

$$Q(\Lambda ; \Lambda^{(k)}) = E_{\mathcal{V}} \left[l_c(\Lambda ; \mathcal{D}) | \mathcal{D}, \Lambda^{(k)} \right] \quad (26)$$

- durant la phase maximisation, il met à jour les paramètres en fixant $\Lambda^{(k+1)}$ à

$$Q(\Lambda ; \Lambda^{(k+1)}) = \arg \max_{\Lambda} Q(\Lambda, \Lambda^{(k)}) \quad (27)$$

jusqu'à ce qu'un maximum local soit trouvé. En pratique, il est souvent difficile de trouver analytiquement un maximum pour $Q(\Lambda ; \Lambda^{(k)})$; dans ce cas, on cherche simplement à augmenter la valeur de Q . On parle alors d'algorithme EM généralisé. La convergence de cet algorithme est montré dans (Dempster *et al.*, 1977).

Afin d'appliquer l'algorithme EM aux CHMM, il suffit de considérer comme variables cachées les séquences d'états des différentes séquences d'apprentissage. Nous pouvons ainsi définir l'ensemble d'apprentissage complet :

$$\mathcal{D}_c = \{\theta_1(p), \dots, \theta_{T_p}(p), y_1(p), \dots, y_{T_p}(p), s_1(p), \dots, s_{T_p}(p)\} \quad (28)$$

avec $p = 1, \dots, P$.

La fonction de maximum vraisemblance complète s'écrit donc :

$$L_c(\Lambda ; \mathcal{D}_c) = \prod_{p=1}^P P(y_1(p), \dots, y_{T_p}(p), s_1(p), \dots, s_{T_p}(p) | \theta_1(p), \dots, \theta_{T_p}(p) ; \Lambda) \quad (29)$$

Avec les propriétés d'indépendance et de processus sans mémoire des CHMM, on peut écrire :

$$\begin{aligned}
 & P(y_1(p), \dots, y_{T_p}(p), s_1(p), \dots, s_{T_p}(p) | \theta_1(p), \dots, \theta_{T_p}(p) ; \Lambda) \\
 &= P(y_T(p), s_T(p) | y_1(p), \dots, y_{T_p-1}(p), s_1(p), \dots, s_{T_p-1}(p), \theta_1(p), \dots, \theta_{T_p}(p) ; \Lambda) \\
 &\quad P(y_1(p), \dots, y_{T_p-1}(p), s_1(p), \dots, s_{T_p-1}(p) | \theta_1(p), \dots, \theta_{T_p}(p) ; \Lambda) \\
 &= P(y_T(p), s_T(p) | s_{T_p-1}(p), \theta_{T_p}(p) ; \Lambda) \\
 &\quad P(y_1(p), \dots, y_{T_p-1}(p), s_1(p), \dots, s_{T_p-1}(p) | \theta_1(p), \dots, \theta_{T_p-1}(p) ; \Lambda)
 \end{aligned}$$

ce qui ramène l'expression de la fonction du maximum de vraisemblance complète à :

$$L_c(\Lambda ; \mathcal{D}_c) = \prod_{p=1}^P \prod_{t=1}^{T_p} P(y_t(p), s_t(p) | s_{t-1}(p), \theta_t(p) ; \Lambda) \quad (30)$$

Si on introduit une variable $z_{i,t}$ qui vaut 1 si $s_t = i$, 0 sinon, l'expression (30) peut se réécrire :

$$\begin{aligned}
 L_c(\Lambda ; \mathcal{D}_c) &= \prod_{p=1}^P \prod_{t=1}^{T_p} P(y_t(p), s_t(p) | s_{t-1}(p), \theta_t(p) ; \Lambda) \\
 &= \prod_{p=1}^P \prod_{t=1}^{T_p} P(y_t(p) | s_t(p), \theta_t(p) ; \Lambda) P(s_t(p) | s_{t-1}(p), \theta_t(p) ; \Lambda) \\
 &= \prod_{p=1}^P \prod_{t=1}^{T_p} P(y_t(p) | s_t(p) = i, \theta_t(p) ; \Lambda)^{z_{i,t}} P(s_t(p) = i | s_{t-1}(p) = j, \theta_t(p) ; \Lambda)^{z_{i,t} z_{j,t-1}}
 \end{aligned} \quad (31)$$

La fonction de log-vraisemblance complète s'écrit en prenant le log de l'expression (31) :

$$\begin{aligned}
 l_c(\Lambda ; \mathcal{D}_c) &= \log L_c(\Lambda ; \mathcal{D}_c) \\
 &= \sum_{p=1}^P \sum_{t=1}^{T_p} \sum_{i=1}^N z_{i,t} \log P(y_t(p) | x_t(p) = i, \theta_t(p) ; \Lambda) \\
 &\quad + \sum_{j=1}^N z_{i,t} z_{j,t-1} \log P(x_t(p) = i | x_{t-1}(p) = j, \theta_t(p) ; \Lambda)
 \end{aligned} \quad (32)$$

À partir de ce point, pour alléger les équations, nous omettrons volontairement les indices p , c'est-à-dire que par exemple $y_1(p)$ sera simplement noté y_1 . Nous pouvons donc à présent décrire la phase d'estimation qui calcule l'espérance de $l_c(\Lambda ; \mathcal{D})$ par rapport aux différentes séquences d'états, que nous noterons \mathcal{S} , étant donné l'ensemble d'apprentissage \mathcal{D} et les anciens paramètres $\hat{\Lambda}$:

$$\begin{aligned}
 Q(\Lambda ; \hat{\Lambda}) &= \sum_{p=1}^P \sum_{t=1}^{T_p} \sum_{i=1}^N E_{\mathcal{S}}[z_{i,t} | y_1, \dots, y_{T_p}, \theta_1, \dots, \theta_{T_p}, \hat{\Lambda}] \log P(y_t | s_t = i, \theta_t ; \Lambda) \\
 &\quad + \sum_{j=1}^N E_{\mathcal{S}}[z_{i,t} z_{j,t-1} | y_1, \dots, y_{T_p}, \theta_1, \dots, \theta_{T_p}, \hat{\Lambda}] \log P(s_t = i | s_{t-1} = j, \theta_t ; \Lambda) \\
 &= \sum_{p=1}^P \sum_{t=1}^{T_p} \sum_{i=1}^N \hat{g}_{i,t} \log P(y_t | s_t = i, \theta_t ; \Lambda) + \sum_{j=1}^N \hat{h}_{i,j,t} \log P(s_t = i | s_{t-1} = j, \theta_t ; \Lambda)
 \end{aligned} \quad (33)$$

avec :

$$\hat{g}_{i,t} = P(s_t = i | y_1, \dots, y_{T_p}, \theta_1, \dots, \theta_{T_p}; \hat{\Lambda}) \quad (34)$$

$$\hat{h}_{i,j,t} = p(s_t = i, s_{t-1} = j | y_1, \dots, y_{T_p}, \theta_1, \dots, \theta_{T_p}; \hat{\Lambda}) \quad (35)$$

où $\hat{\cdot}$ signifie que \hat{h} et \hat{g} sont calculées à partir des paramètres précédents $\hat{\Lambda}$.

Exprimons $\hat{g}_{i,t}$ et $\hat{h}_{i,j,t}$ en fonction de $L = P(y|\theta_1)$, α, β définis respectivement par les équations (18), (16), (19) :

$$\begin{aligned} h_{i,j,t} &= P(s_1 = i, s_{t-1} = j | y_1, \dots, y_T, \theta_1, \dots, \theta_T) \\ &= P(s_1 = i, s_{t-1} = j, y_1, \dots, y_T | \theta_1, \dots, \theta_T) / P(y | \theta_1, \dots, \theta_T) \\ &= P(y_{t+1}, \dots, y_T | s_t = i, s_{t-1} = j, y_1, \dots, y_t, \theta_1, \dots, \theta_T) \\ &\quad P(y_t | s_t = i, s_{t-1} = j, y_1, \dots, y_{t-1}, \theta_1, \dots, \theta_T) \\ &\quad P(y_1, \dots, y_{t-1}, s_{t-1} = j | \theta_1, \dots, \theta_T) \\ &\quad P(s_t = i | s_{t-1} = j, y_1, \dots, y_{t-1}, \theta_1, \dots, \theta_T) \\ &\quad / P(y | \theta_1, \dots, \theta_T) \\ &= b_{iy_t}(\theta_t) \alpha_j(t-1) \beta_i(t) a_{ij}(\theta_t) / P(y | \theta_1, \dots, \theta_T) \end{aligned} \quad (36)$$

et en utilisant l'équation (20) on obtient :

$$\begin{aligned} g_{i,t} &= P(s_t = i | y_1, \dots, y_T, \theta_1, \dots, \theta_T) \\ &= P(s_t = i, y_1, \dots, y_T | \theta_1, \dots, \theta_T) / L \\ &= P(y_{t+1}, \dots, y_T | s_t = i, y_1, \dots, y_t, \theta_1, \dots, \theta_T) P(y_1, \dots, y_t, s_t = i | \theta_1, \dots, \theta_T) / L \\ &= \alpha_i(t) \beta_i(t) / L \end{aligned} \quad (37)$$

Ces équations sont similaires à celles utilisées par l'algorithme de Baum-Welch pour les HMM et on a donc un algorithme EM de complexité en $O(N^2T)$ où T dénote la longueur obtenue en concaténant toutes les séquences d'apprentissage.

Nous ne sommes pas en mesure de décrire aussi précisément la phase de maximisation de l'algorithme EM. En effet, puisque nous souhaitons que les CHMM soient aussi généraux que possible, nous ne connaissons ni la forme ni la méthode de calcul des différentes distributions π_i, a_{ij}, b_{iy_t} . La phase de maximisation est donc très dépendante des distributions choisies.

Nous allons maintenant appliquer les CHMM à la modélisation des grilles de programmes.

3 Application à la modélisation des grilles de programmes

Le but de ces travaux est de modéliser la grille de programmes d'une chaîne (voir le chapitre 5, page 43) par un modèle markovien. Les modèles et les extensions présentées dans la section 1 (page 79) ne permettaient pas de représenter facilement toutes les propriétés d'une grille de programmes. Ainsi, la simulation des grilles était ambiguë et le choix de la topologie était arbitraire. Pour illustrer cette lacune, prenons quelques exemples.

Considérons dans un premier temps un exemple jouet dans lequel une chaîne diffuse trois journaux télévisés par jour :

- le journal de 6h, suivi par une émission pour la jeunesse ;
- le journal de 13h, suivi par un feuilleton ;

- le journal de 20h, suivi par un film.

Les modèles tels que les chaînes de Markov ou même un HMM dans lequel chaque état représenterait un type d'émission, ne permettent pas de faire la différence entre les trois journaux télévisés. Cela implique que la probabilité qu'un journal télévisé soit suivi d'un feuilleton est la même que dans le cas d'un film ou d'une émission pour la jeunesse. L'équiprobabilité pose un problème dans le cas d'un modèle de production alors qu'elle ne le serait pas pour un modèle de reconnaissance pour lequel l'observation est donnée en entrée. En effet, dans un modèle de production, il est courant de chercher à diminuer les possibilités de transition. Une solution, proposée par J.R. Norris (Norris, 1999), consiste à multiplier le nombre d'états représentant les journaux télévisés afin d'avoir un état par journal télévisé. Pour des émissions dont le nombre d'occurrences est constant chaque jour de diffusion, la multiplication d'états est assez facile à appliquer. En revanche, pour les programmes tels que les magazines sponsorisés, les publicités et les bandes annonces, le nombre d'occurrences est difficilement prévisible.

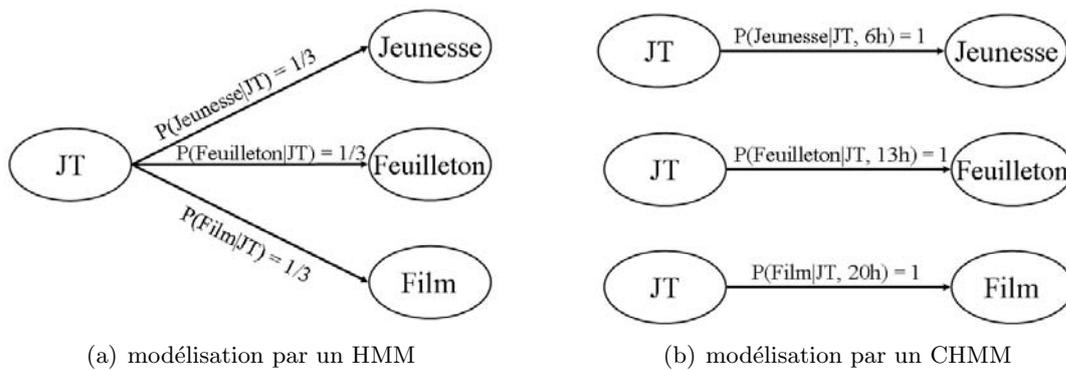


FIG. 24: Exemple jouet des mêmes transitions avec un HMM et un CHMM

Considérons à présent un exemple réel : le tiercé sur TF1 en 2004. Sur cette chaîne, le journal télévisé de 20h est toujours suivi soit du tiercé, soit de la météo ; dans le cas où ce journal est suivi du tiercé, le tiercé est alors suivi de la météo. Nous avons entraîné un modèle non-contextuel sur l'ensemble des émissions de TF1 diffusées en 2004. Nous avons obtenu, par exemple, que le journal télévisé de 20h était suivi à 72% d'une météo, et donc à 28% du tiercé. Si l'on cherche à produire la séquence la plus probable, on obtient toujours le journal de 20h suivi de la météo. En réalité, cela n'est vrai que cinq jours par semaine : en effet, le tiercé n'était diffusé que les samedi et mardi. Tenir compte de la spécificité du jour de diffusion permet à la séquence la plus probable les samedi et mardi de contenir le journal de 20h, puis le tiercé et enfin la météo, alors que la séquence la plus probable les autres jours ne contiendra pas de tiercé. Un autre problème que les modèles classiques markoviens ne permettent pas de traiter est celui des films : le même modèle entraîné précédemment nous a montré qu'un film était suivi d'un second film avec une probabilité de 68%. Ainsi, sans tenir compte du contexte de diffusion du second film, celui-ci devrait être suivi à son tour d'un troisième film avec une probabilité de 68% et ainsi de suite : la séquence la plus probable ne serait donc pas correcte.

L'utilisation des CHMM nous permet de palier plusieurs problèmes dans la modélisation des grilles de programmes. On peut ainsi définir un contexte θ représentant l'heure et le jour de la

semaine correspondant au début du programme :

$$\begin{aligned} \theta &= \{Heure, Jour\} \\ \text{et } D_{Heure} &= [0 ; 86400[\\ \text{et } D_{Jour} &= \{\text{Lundi}, \dots, \text{Dimanche}\} \end{aligned} \quad (38)$$

dans lequel les heures sont exprimées en secondes, allant de minuit (0 seconde) à 23:59:59 (86399 secondes).

Ainsi, cela nous permet de décider plus facilement de la topologie du modèle : il suffit d'attribuer à chaque genre télévisuel un état. En effet, nous avons vu que les grilles de programmes étaient stables (voir la section 7, page 56) au niveau du genre télévisuel programmé dans une case (et non pas au niveau des émissions). La contextualisation des modèles de Markov nous permettra de distinguer les journaux du matin, du midi et du soir (selon l'exemple ci-dessus) même s'ils ne sont représentés que par un seul état. De plus, par rapport à l'exemple de TF1, le fait de tenir compte du jour de la semaine nous permet de traiter le cas du tiercé et l'heure de diffusion dans le cas des films qui se succèdent.

Pour pouvoir utiliser un CHMM, il faut aussi trouver une fonction F_θ d'évolution du contexte θ : la mise à jour consiste, dans notre cas, à trouver l'heure de diffusion du programme suivant. Pour cela, il suffit d'ajouter au contexte courant la durée du programme courant. Notons \mathbb{D} l'ensemble des durées des émissions et $d \in \mathbb{D}$; alors F_θ dans notre cas est définie par :

$$F_\theta : \quad \Theta \times \mathbb{D} \rightarrow \Theta$$

$$\left\{ \begin{array}{l} \text{Heure} = h \\ \text{Jour} = j \end{array} \right\}, d \mapsto \left\{ \begin{array}{l} \text{Heure} = (d + h) \pmod{86400} \\ \text{Jour} = (j + \lfloor \frac{d+h}{86400} \rfloor) \pmod{7} \end{array} \right\}. \quad (39)$$

La fonction d'évolution prend donc en paramètre le contexte courant et la durée du programme courant pour calculer le nouveau contexte. Le contexte θ va nous permettre de modéliser correctement le phénomène « caché » qui est le phénomène de transition d'un état à l'autre, c'est-à-dire d'un genre télévisuel à l'autre. Il est possible de considérer que le phénomène observable est non pas d'avoir une émission, mais plutôt d'avoir une émission d'une certaine durée. Ainsi, l'état courant donne le genre de l'émission et l'observation générée fournit sa durée. Ainsi, avec les notations introduites ci-dessus, Σ est un ensemble continu d'entiers représentant l'ensemble \mathbb{D} des durées d'émissions, et $b_{ij}(\theta_t)$ représente la probabilité $P(\text{durée} = j | \text{type} = i, \theta_t)$, soit la probabilité qu'une émission de type i dure j secondes dans le contexte de diffusion θ_t . B est donc une fonction qui à un état, un contexte et une durée associe un réel entre 0 et 1 représentant la probabilité $b_{ij}(\theta_t)$.

Nous verrons dans le chapitre 8 (page 105) que des arbres de régressions permettent de connaître la durée minimale, la durée maximale et la durée moyenne d'une émission en fonction de son genre et de son contexte de diffusion. La distribution B est alors définie comme étant une gaussienne asymétrique paramétrée par les éléments statistiques renvoyés par l'arbre de régression.

3.1 Estimation des paramètres du CHMM

Plusieurs paramètres du modèle doivent être estimés : la distribution $a_{ij}(\theta_t)$, la distribution $b_{ik}(\theta_t)$ et la distribution $\pi_i(\theta_t)$, où i et j sont des états et k un symbole, comme définis dans la sous-section 2.2 (page 87). Les exemples d'apprentissage sont des n-uplets de la forme $\langle \text{Contexte}, \text{Genre}, \text{Durée} \rangle$, dans lequel le contexte représente le jour et l'heure de diffusion de l'exemple. Ces exemples forment une seule et même séquence : l'ensemble des programmes

diffusés sur une chaîne pendant un an. En pratique, cela n'est pas vraiment le cas dans la mesure où cette grande séquence sera parfois interrompue par un changement d'heure ou pour ne pas prendre en compte un certain genre d'émission. Puisque les états du CHMM représentent chacun un genre d'émission et puisque nous connaissons le genre télévisuel des observations, l'algorithme EM, présenté dans la section 2.3 (page 96), n'est pas utile (il n'y a plus de phénomène caché). Nous allons donc présenter un algorithme d'entraînement du modèle différent en $O(MN^2 + T)$.

Un des problèmes posés par la contextualisation des probabilités du modèle est que tout contexte plausible doit apparaître durant la phase d'apprentissage du modèle : en effet, si un contexte n'y apparaît pas, le système ne sera pas capable de produire une transition dans ce contexte là. Ce n'est pas le cas des arbres de régression grâce à leur pouvoir de généralisation (voir section 2.3, page 111). Cela nécessite donc un nombre important d'exemples.

Afin d'augmenter le nombre d'exemples considérés dans l'ensemble d'apprentissage, nous ne considérons pas la durée de l'exemple mais l'ensemble des durées fournies par un arbre de régression (voir chapitre 8, page 105). Cela nous permet de multiplier le nombre d'exemples d'entraînement en ne considérant non plus l'heure de l'émission suivante, mais l'heure du programme courant augmentée de chacune des durées possibles. En faisant varier la durée des émissions, nous introduisons ainsi un décalage temporel ; dans cette mesure, il faudrait propager ce décalage sur les émissions suivantes. En pratique, puisque nos exemples forment une seule et même séquence, il faudrait propager ce décalage sur le reste de la séquence, ce qui est trop coûteux en temps de calcul. Nous faisons donc l'hypothèse que le décalage produit, sans propagation, suffit à augmenter le nombre de contextes pouvant être pris en compte.

Dans le cas de notre modèle, l'estimation des probabilités de transition revient à l'estimation des probabilités d'une chaîne de Markov. Il s'agit donc de calculer les fréquences d'un genre télévisuel à une heure donnée :

$$\begin{aligned} \pi_i(\theta_t) &= P(G_i|\theta_t) \\ &= \frac{\text{nombre d'émissions du type } G_i \text{ débutant dans le contexte } \theta_t}{\text{nombre d'émissions débutant dans le contexte } \theta_t} \end{aligned} \tag{40}$$

$$\begin{aligned} a_{ij}(\theta_t) &= P(G_j|G_i, \theta_t) \\ &= \frac{\text{nombre d'émissions du genre } G_j \text{ débutant dans le contexte } \theta_t \text{ et précédées d'une émission du genre } G_i}{\text{nombre d'émissions précédée d'une émission du genre } G_i \text{ débutant dans le contexte } \theta_t}. \end{aligned} \tag{41}$$

L'algorithme d'estimation des paramètres nécessite donc dans un premier temps l'entraînement de l'arbre de régression, puis ensuite l'évaluation des deux probabilités ci-dessus par les formules (40) et (41). L'algorithme 3 illustre la méthode d'estimation des paramètres du CHMM utilisée pour modéliser les grilles de programmes.

Algorithme 3 Algorithme d'estimation des paramètres

```

procedure ESTIMATIONPARAMÈTRES(Arbre  $A$ , Ensemble  $T$ , CHMM  $M$ )
▷ entrées :  $A$ , l'arbre de régression entraîné et  $T$ , un ensemble d'exemples  $E_i$ 
▷ sortie : complète le CHMM  $M$  en estimant ses probabilités

  for  $i \leftarrow 2, |T|$  do
    DuréesPrécédentes  $\leftarrow A.Prédiction(E_{i-1}.Contexte, E_{i-1}.Genre)$ 
    if  $E_{i-1}.Durée \in DuréesPrécédentes$  then
      ContexteMinimum  $\leftarrow$  Mise à jour de  $E_{i-1}.Contexte$  avec  $DuréesPrécédentes.Minimum$ 
      ContexteMaximum  $\leftarrow$  Mise à jour de  $E_{i-1}.Contexte$  avec  $DuréesPrécédentes.Maximum$ 
    else
      ContexteMinimum  $\leftarrow E_{i-1}.Contexte$ 
      ContexteMaximum  $\leftarrow E_{i-1}.Contexte$ 
    end if

    for Contexte  $\leftarrow ContexteMinimum, ContexteMaximum$  do
       $A[E_{i-1}.Genre, E_i.Genre, Contexte] ++$ 
       $totalA[E_{i-1}.Genre, Contexte] ++$ 
       $PI[E_i.Genre, Contexte] ++$ 
       $totalPI[Contexte] ++$ 
    end for
  end for

  foreach Contexte  $\theta \in \Theta$  do
    foreach Genre  $G_i$  do
       $\pi_{G_i}(\theta) = PI[G_i, \theta] / totalPI[\theta]$ 
      foreach Genre  $G_j$  do
         $A_{G_i G_j}(\theta) = A[G_i, G_j, \theta] / totalA[G_i, \theta]$ 
      end foreach
    end foreach
  end foreach

end procedure

```

3.2 Estimation de la probabilité d'un enchaînement d'émissions

Déroulons à présent un exemple d'estimation de la probabilité de la grille de programmes \mathcal{G} présentée dans le tableau 10.

Jour	Heure	Genre	Durée (s)
Lundi	10:32:10	Magazine	100
Lundi	10:33:50	Jeu	1832
Lundi	11:04:22	Magazine	104
Lundi	11:06:06	Météo	87
Lundi	11:07:33	Service	74
Lundi	11:08:47	Feuilleton	1234

TAB. 10: Exemple de grille de programmes (source : Médiamétrie)

La probabilité de la grille \mathcal{G} s'exprime par :

$$\begin{aligned}
 P(\mathcal{G}) = & \\
 & P(s_1 = \text{Magazine} | \theta_1 = \{\text{Lundi}, 10 : 32 : 10\}) \times P(d = 100 | s_1 = \text{Magazine}, \theta_1) \\
 & \times P(s_2 = \text{Jeu} | s_1 = \text{Magazine}, \theta_2 = \{\text{Lundi}, 10 : 33 : 50\}) \times P(d = 1832 | s_2 = \text{Jeu}, \theta_2) \\
 & \times P(s_3 = \text{Magazine} | s_2 = \text{Jeu}, \theta_3 = \{\text{Lundi}, 11 : 04 : 22\}) \times P(d = 104 | s_3 = \text{Magazine}, \theta_3) \\
 & \times P(s_4 = \text{Meteo} | s_3 = \text{Magazine}, \theta_4 = \{\text{Lundi}, 11 : 06 : 06\}) \times P(d = 87 | s_4 = \text{Meteo}, \theta_4) \\
 & \times P(s_5 = \text{Service} | s_4 = \text{Meteo}, \theta_5 = \{\text{Lundi}, 11 : 07 : 33\}) \times P(d = 74 | s_5 = \text{Service}, \theta_5) \\
 & \times P(s_6 = \text{Feuilleton} | s_5 = \text{Service}, \theta_6 = \{\text{Lundi}, 11 : 08 : 47\}) \\
 & \times P(d = 1234 | s_6 = \text{Feuilleton}, \theta_6).
 \end{aligned} \tag{42}$$

4 Conclusion

Nous avons proposé dans ce chapitre une nouvelle extension contextuelle des HMM que nous avons appelé modèle de Markov Caché Contextuel (Contextual Hidden Markov Model ou CHMM). Cette extension a la particularité de tenir compte d'un contexte pour déterminer les probabilités de transition d'un état à l'autre et les probabilités d'émission d'un symbole. Dans notre cas, nous avons utilisé un contexte qui permet de représenter le jour et l'heure de diffusion d'une émission.

Nous avons appliqué les CHMM à la modélisation des grilles de programmes afin de représenter un enchaînement d'émissions dont les durées sont prédites par un arbre de régression. Les CHMM nous permettent d'adapter les probabilités de transition d'un état à l'autre en fonction du contexte du modèle : en conséquence, nous obtenons de nombreuses transitions déterministes, ce qui va faciliter la prédiction des grilles de programmes. En effet, afin de réduire le nombre de grilles prédites possibles pour un jour donné, nous avons rendu cette modélisation des grilles de programmes la plus déterministe possible, tout en étant conscients qu'il s'agit d'une utilisation particulière des modèles de Markov.

En parcourant le CHMM entraîné, il est possible de simuler un ensemble de grilles de programmes assez grand. Parmi ces possibilités, seule une partie d'entre elles peuvent correspondre à un jour donné. Par exemple, considérons une chaîne qui diffuse le lundi en *prime-time* tantôt un film, tantôt un match de football et tantôt un divertissement ; à un lundi particulier ne pourront correspondre que les grilles comportant soit le film uniquement, soit le match uniquement ou soit le divertissement uniquement et toutes les autres peuvent être élaguées.

Nous allons présenter dans le chapitre suivant une méthode d'estimation des durées des émissions. Ce chapitre sera suivi par la présentation d'une méthode pour élaguer les nombreuses possibilités obtenues avec le CHMM à partir du guide de programmes.

Chapitre 8

Régression des durées des émissions

Sommaire

1	Régression par des méthodes mathématiques simples	106
1.1	Régression linéaire	106
1.2	Régression non linéaire	107
1.3	Discussion sur les méthodes de régression mathématiques simples	107
2	Régression par des méthodes informatiques	108
2.1	Réseaux de neurones	108
2.2	Machines à vecteurs de support	109
2.3	Arbres de régression	111
2.4	Discussion sur les méthodes de régression informatiques	111
3	Application des arbres de régression à la prédiction des durées des émissions	112
3.1	Construction de l'arbre	112
3.2	Régression d'une durée	116
3.3	Exemples de règles et d'estimation de durées	116
4	Déduction de la probabilité des durées	118
5	Conclusion	118

Dans la modélisation des grilles de programmes, la durée des émissions joue un rôle important : en effet, elle permet non seulement de calculer l'heure de début du programme suivant mais aussi de distinguer un format court d'un format long, pour un magazine par exemple.

Nous avons vu dans la seconde partie que les durées des émissions étaient sensiblement les mêmes pour un genre télévisuel donné : les émissions sont conçues pour remplir une case de la grille de programmes et ont donc une durée à respecter (voir le chapitre 6, page 57). Cette observation n'est pas vérifiée pour certaines émissions retransmises en direct (événement sportif, magazines) ; à titre d'exemple, le magazine *Tout le monde en parle* peut durer entre une heure et demi et trois heures.

Nous considérons que la durée est une variable aléatoire continue exprimée en secondes. Nous nous intéressons dans ce chapitre à des méthodes permettant de prédire ces valeurs continues afin d'être en mesure d'estimer la durée des émissions.

La régression est une méthode mathématique qui permet, à partir de l'observation des valeurs prises par une variable aléatoire continue Y , de calculer la courbe qui reproduit au mieux les

variations de Y , appelée alors « variable expliquée ». En pratique, la courbe passe à proximité de chacune des observations de Y . L'équation de la courbe est calculée en fonction de variables appelées « régresseurs » ou « variables explicatives ». Plus formellement, si l'on considère que φ est une fonction de régression de Y en X , alors φ est égale à l'espérance conditionnelle de Y sachant X , notée $\varphi(x) = E(Y|X = x)$, où X est un régresseur de Y (Saporta, 1990).

Il existe de nombreuses méthodes de régression. Sans soucis d'exhaustivité, nous pouvons citer les méthodes paramétriques, les méthodes semi-paramétriques et enfin les méthodes non paramétriques. Une méthode de régression est dite paramétrique si ses paramètres ont un rapport avec la distribution des données. À titre d'exemple, la loi normale est un modèle paramétrique puisque ses deux paramètres sont l'écart-type et la moyenne. En revanche, un histogramme est un modèle non paramétrique car ses paramètres, l'origine et la largeur des cases, ne sont pas interprétables en termes de propriétés globales de la distribution.

La section suivante permettra au lecteur non initié de se familiariser avec la notion de régression en comprenant comment, lorsque la nature de la distribution est connue, il est possible de proposer une approximation d'une variable aléatoire. Nous verrons aussi que la distribution des durées d'émission par rapport à divers critères ne semble pas appartenir une famille connue.

Nous présenterons ensuite des méthodes de régression non paramétriques très utilisées en informatique : les réseaux de neurones, les machines à vecteurs de support et nous finirons par les arbres de régression.

1 Régression par des méthodes mathématiques simples

Ces méthodes, paramétriques (régression linéaire) ou non (régression polynomiale), consistent généralement à calculer les valeurs d'une variable aléatoire Y lorsque la distribution des valeurs de celle-ci se rapproche d'une fonction mathématique linéaire ou non-linéaire.

1.1 Régression linéaire

Parmi les méthodes de régression linéaire, on distingue les méthodes linéaires simples, c'est-à-dire qui n'expliquent la variable Y que par une seule variable explicative X , et les méthodes linéaires multiples, qui calculent la courbe en fonction de plusieurs variables X_1, \dots, X_n (Savy, 2006). Dans les deux cas, la courbe calculée par la méthode est une fonction affine des variables explicatives. Nous préférons décrire une méthode de régression linéaire multiple puisqu'il s'agit de la généralisation à n variables explicatives de la méthode de régression linéaire simple.

Soit Y une variable aléatoire continue et X_1, \dots, X_n les variables aléatoires explicatives de Y . Notons Y_k avec $k \in 1, \dots, K$ la $k^{\text{ème}}$ observation de Y . La régression linéaire multiple consiste à déterminer les coefficients a_0, \dots, a_n à partir des K observations de Y :

$$Y_k = a_0 + a_1 X_{1k} + \dots + a_n X_{nk} \quad (1)$$

Les paramètres a_i sont généralement estimés en minimisant la somme des moindres carrés S :

$$S = \sum_{k=1}^K (y_k - a_0 - \dots - a_n x_{nk})^2 \quad (2)$$

Les variables x_i doivent être indépendantes. Si ce n'est pas le cas, il s'agit alors d'une régression non linéaire (voir section 1.2).

1.2 Régression non linéaire

Dans le cas où les variables explicatives ne sont pas toutes indépendantes – par exemple si $x_2 = x_1^2, x_3 = x_1^3$, etc, . . . –, on parle de régression non linéaire (Huet *et al.*, 1996). Très souvent, il s’agit d’une régression par une courbe polynomiale, de la forme

$$Y = a_0 + a_1X + \dots + a_nX^n. \quad (3)$$

L’estimation des paramètres a_i se fait de la même façon que pour les modèles linéaires, en minimisant la somme des moindres carrés. Lorsqu’il s’agit d’une courbe d’allure exponentielle, il est possible de ramener le problème à celui de l’estimation des paramètres d’une droite (pente et biais). Par exemple, lorsque la variable aléatoire Y est définie par $Y = ba^X$, il est facile de se ramener à une fonction affine en écrivant :

$$\log Y = \log b + X \log a. \quad (4)$$

1.3 Discussion sur les méthodes de régression mathématiques simples

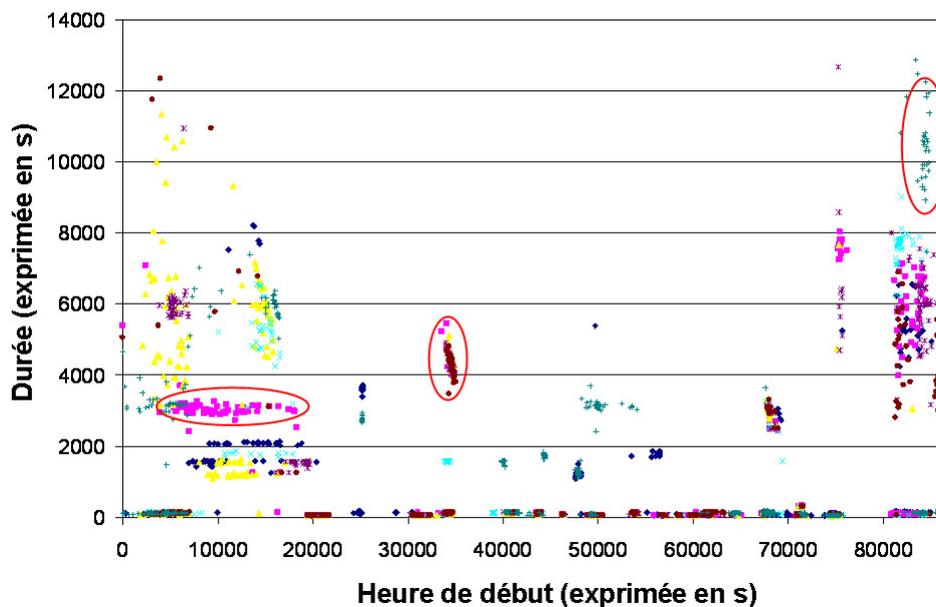


FIG. 25: Durées (en secondes) des magazines en fonction de leur heure de diffusion, sur France 2 en 2004

Dans un premier temps, ces méthodes de régression classiques, linéaires ou non-linéaires, nécessitent une étude ou une visualisation des données assez précises pour connaître la variation de la variable aléatoire Y . En fonction de sa tendance, on peut choisir un modèle linéaire ou non-linéaire avec par exemple une fonction polynomiale, cubique ou exponentielle. La seconde étape consiste à estimer les paramètres du modèle choisi.

Dans notre cas, lorsque nous projetons les durées d’un genre télévisuel diffusé sur une chaîne donnée et pendant toute une année, nous nous apercevons qu’il est difficile de trouver une allure générale en fonction des heures de diffusion. Les méthodes mathématiques classiques de régression ne s’appliquent donc pas à la prédiction de durées. La figure 25 montre que l’entropie

des durées des magazines de France 2 en 2004, par rapport à leur heure de diffusion, est très élevée. Les jours de la semaine sont représentés par des différentes couleurs et formes de points. Deux cas de figure apparaissent sur cette illustration : les groupes positionnés horizontalement sont des magazines dont l'heure de début est peu stable, c'est-à-dire souvent retardée ou avancée, mais des durées stables, par opposition aux groupes verticaux qui ont une heure de départ stable mais des durées très variables. Par exemple, le groupe de gauche, repéré par un ovale rouge, correspond aux magazines *C'est au programme* et celui de droite à *Tout le monde en parle*. Une structure horizontale étalée sur toute la journée et proche de l'axe des abscisses est nettement représentée : il s'agit des magazines courts sponsorisés (*CD'aujourd'hui, un Jour, un arbre*). Une autre structure horizontale, également entourée en rouge, représente la collection de magazines « Thé ou café » diffusés le même jour de la semaine. Sur cette figure, il apparaît nettement que les groupements horizontaux se produisent surtout la nuit (l'origine de l'axe des abscisses représente minuit) : cela signifie que les heures de diffusion des émissions sont un peu plus aléatoires que dans la journée.

Nous allons présenter maintenant des méthodes issues de méthodes informatiques, comme les réseaux de neurones et les machines à vecteur support.

2 Régression par des méthodes informatiques

Nous allons aborder dans cette section trois techniques de régression issues de recherches informatiques. Les réseaux de neurones et les machines à vecteurs de support sont des techniques pouvant être utilisées soit dans le cadre de la classification soit dans le cadre de la régression et leurs applications sont nombreuses. La régression par des arbres est une technique moins courante issue des arbres de décision en classification.

2.1 Réseaux de neurones

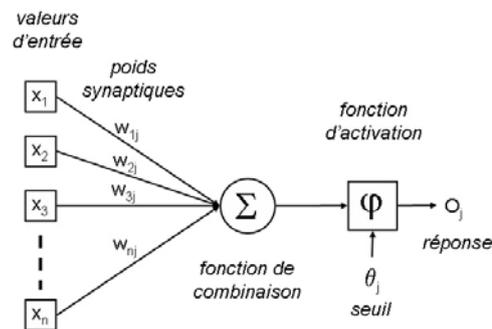


FIG. 26: Structure d'un neurone formel

Un réseau de neurones (*artificial neural network*) est un formalisme de calcul inspiré du fonctionnement des neurones dans le cerveau. (McCulloch et Pitts, 1943) propose un modèle simplifié du modèle biologique, appelé « neurone formel », pouvant réaliser des fonctions logiques et arithmétiques. La figure 26 présente le fonctionnement d'un neurone formel : il s'agit d'une combinaison des différentes valeurs d'entrée x_i , pondérées par des poids w_{ij} appelés « poids synaptiques », et représentant le poids entre le neurone i et le neurone j . La combinaison est opérée par une fonction spécifique et la sortie du neurone formel est conditionnée par une fonction d'activation φ . Ainsi, la valeur de sortie y_j du neurone j en fonction des valeurs d'entrée x_i , avec

pour fonction de combinaison une somme pondérée, s'écrit :

$$O_j = \varphi\left(\sum_{i=1}^n w_{ij}x_i + \theta_j\right). \quad (5)$$

La première application, le Perceptron, est présentée dans (Rosenblatt, 1958) ; il s'agit du premier système capable d'apprendre à partir d'exemples comprenant quelques erreurs. Ce type de réseau possède une couche d'entrée, appelée « rétine » et une couche de sortie. Les topologies de réseaux qui sont apparus après le Perceptron peuvent se classer en deux catégories : les réseaux à couches et les réseaux récurrents. Ils ont pour but de palier l'impossibilité d'utiliser les réseaux de neurones monocouches pour un problème non-linéaire (Minsky et S.Papert, 1969).

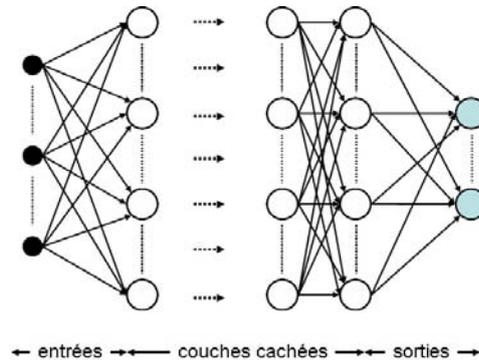


FIG. 27: Structure d'un réseau de neurones multicouche

La figure 27 montre l'architecture d'un réseau de neurones multicouches. Ces réseaux sont composés de trois couches au moins : une couche d'entrée (en noir sur la figure 27), une couche de sortie (en gris sur la figure 27), et entre ces deux couches, au moins une couche dite « cachée » (en blanc sur la figure 27). Un neurone d'une couche inférieure ne peut être relié qu'à des neurones d'une des couches suivantes. L'information se propage donc couche par couche sans jamais remonter dans le réseau.

La figure 28 montre cette fois-ci un exemple de réseau de neurones récurrent : la couche de sortie peut être directement considérée comme une entrée du réseau et influencer les prochaines sorties (contexte). L'information se propage cette fois-ci dans les deux sens, ce qui permet d'obtenir des comportements plus complexes qu'avec un réseau multi-couche.

Il existe différents algorithmes pour entraîner un réseau de neurones à une tâche de régression ou de classification. Le but de ces algorithmes est d'ajuster les poids synaptiques des réseaux. On peut citer, à titre d'exemple, la règle de rétropropagation (Rumelhart *et al.*, 1986) ou la règle de corrélation en cascade (Fallman et Lebière, 1990). Les réseaux de neurones, leurs applications et leurs algorithmes d'entraînement sont plus profondément décrits dans (Bishop, 1999; Personnaz et Rivals, 2003).

2.2 Machines à vecteurs de support

Les machines à vecteurs de support (*Support Vector Machines*, SVM) ont été développées par Vapnik à partir de ses anciens travaux en apprentissage statistique (Vapnik et Lerner, 1963), qui avaient pour but de généraliser les propriétés d'un ensemble de données d'apprentissage afin de caractériser de nouvelles données. Le développement de la version actuelle des SVM s'est fait dans un contexte industriel dans un but de reconnaissance automatique des caractères (*Optical*

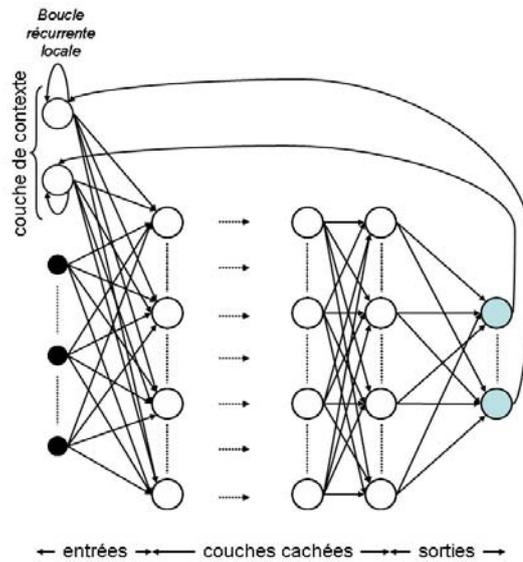


FIG. 28: Exemple de réseau de neurones récurrent

Character Recognition, OCR), c'est-à-dire pour une tâche de classification. Les SVM ont ensuite été adaptées à la régression et la prédiction de séries temporelles (Vapnik *et al.*, 1997).

La ϵ -régression par SVM consiste à trouver une fonction f qui passe à au plus une distance ϵ de tous les points Y_i de l'ensemble d'apprentissage. Ainsi, un exemple Y_i va changer l'équation de f seulement s'il se situe à une distance de f supérieure à ϵ . Afin de donner un exemple, supposons que f soit une fonction linéaire :

$$f(x) = \langle w, x \rangle + b \quad (6)$$

avec $w \in \mathbb{X}$, l'espace des exemples, et $b \in \mathbb{R}$ et $\langle \cdot, \cdot \rangle$ dénote le produit scalaire. Alors l'algorithme de ϵ -régression par une SVM peut se formuler comme suit (Vapnik, 1995) :

$$\begin{aligned} &\text{minimiser} && \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i + \xi_i^* \\ &\text{sous contrainte que} && \begin{cases} y_i - \langle w, x_i \rangle - b \leq \epsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0. \end{cases} \end{aligned} \quad (7)$$

La constante $C > 0$ permet de changer la tolérance de l'algorithme par rapport aux points qui sont situés à une distance de f supérieure à ϵ . Ainsi l'algorithme tient compte d'une fonction $|\xi|_\epsilon$ de coût insensible à ϵ près définie par :

$$|\xi|_\epsilon = \begin{cases} 0 & \text{si } |\xi| \leq \epsilon \\ |\xi| - \epsilon & \text{sinon.} \end{cases} \quad (8)$$

La distance entre f et le point le plus proche dans l'ensemble d'apprentissage est appelée marge. Les variables C et ξ dans l'équation (7) permettent de passer d'une marge dure (*hard margin*) à une marge molle (*soft margin*), selon que l'on accepte que certains points se situent en dehors de cette marge ou pas (Vincent, 2003).

Comme pour les réseaux de neurones, les fonctions non-linéaires sont atteintes en utilisant des noyaux qui permettent d'augmenter la dimension des vecteurs d'entrée.

Pour de plus amples informations sur la régression avec des SVM, le lecteur est encouragé à lire (Smola et Schölkopf, 2004).

2.3 Arbres de régression

Les arbres de régression sont une adaptation au cas des classes continues d'une technique de classification appelée « arbre de décision » (Breiman *et al.*, 1984). L'idée est d'utiliser un arbre dont les nœuds sont des tests logiques booléens et dont les branches représentent les diverses valeurs de ces tests (vrai ou faux). La construction de l'arbre a pour but de diviser l'espace de représentation des données en hyperrectangles et d'associer à chacune de ces régions de l'espace une moyenne et un écart-type de la valeur à prédire.

Dans les extensions les plus simples des arbres de régression, les feuilles contiennent une valeur moyenne et un écart-type de la variable aléatoire à approcher ; mais ces valeurs peuvent être remplacées par des fonctions d'approximation locales (Quinlan, 1993b; Karalic, 1992; L.Torgo, 1997), ce qui fait des arbres de régression une technique de régression très performante (Quinlan, 1993a).

Contrairement aux réseaux de neurones, la phase d'apprentissage propose non seulement le calcul des valeurs contenues dans les feuilles mais aussi la topologie de l'arbre. Nous aborderons la phase d'apprentissage plus en détail dans la section 3 (page 112).

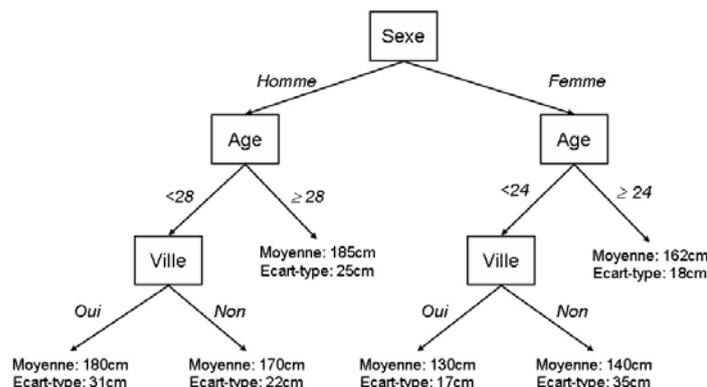


FIG. 29: Exemple d'arbre de régression

La figure 29 présente un exemple d'arbre sur des données factices. La variable à approcher est la taille des individus en fonction de leur sexe, leur âge, et leur lieu d'habitation. Dans cet exemple, l'âge est une variable continue alors que le sexe et le lieu d'habitation sont des variables symboliques. Ceci met en évidence le problème de la discrétisation des variables continues dans la construction de l'arbre (voir section 3, page 112).

2.4 Discussion sur les méthodes de régression informatiques

Le choix d'une méthode de régression dépend de l'application souhaitée (Quinlan, 1994). Dans notre cas, la difficulté de ce choix réside dans le fait que, comme le montre la figure 25, les durées sont trop dispersées dans l'espace, ce qui revient à un ensemble d'apprentissage très bruité. En effet, le phénomène observé est la durée des émissions, bruité, entre autres, par la durée de la coupure publicitaire pendant l'émission. La plupart des méthodes de régression permettent, pour une valeur cible, d'obtenir une valeur approchée, tandis que l'arbre de régression fournit

plusieurs éléments statistiques (minimum, maximum, moyenne, écart-type) permettant de tenir compte d'émissions dont la durée peut énormément varier (comme *Tout le monde en parle*).

Les réseaux de neurones représentent un formalisme largement éprouvé et qui est assez facile à mettre en œuvre. L'inconvénient est qu'ils sont en général peu adaptés à des valeurs symboliques, notamment pour les vecteurs d'entrée dans notre cas : il faut en effet découper le segment $[0; 1]$, intervalle à privilégier pour les entrées des réseaux de neurones, en autant de segments qu'il y a de valeurs symboliques. Il est donc fastidieux de tenir compte de la trentaine de genres télévisuels. De plus, non seulement l'apprentissage des réseaux de neurones nécessite d'être exécuté par un expert, mais aussi il n'est pas possible de connaître l'architecture idéale du réseau pour un problème donné (Bishop, 1999), ce qui ne facilite pas non plus l'automatisation de l'apprentissage.

Les machines à vecteurs de support nécessitent également une normalisation des valeurs d'entrée pour garantir de bons résultats. Cette normalisation ramène ces valeurs dans l'intervalle $[-1; 1]$ ou $[0; 1]$, ce qui n'est pas pratique pour les valeurs symboliques. De plus, l'entropie des durées est tellement importante que les résultats que nous avons obtenus n'étaient pas significatifs.

Finalement, les arbres de régression se sont montrés de bons candidats à la régression des durées des émissions : ils permettent de prendre en compte des grandeurs symboliques ou continues. De plus, leur consultation permet d'obtenir une valeur moyenne et un écart-type. L'arbre déduit de l'ensemble d'apprentissage est humainement exploitable, c'est-à-dire que l'on peut vérifier que la construction s'est bien passée et corriger facilement l'arbre sans recommencer la phase d'apprentissage. Nous verrons que l'algorithme d'induction de l'arbre est déterministe, contrairement à l'entraînement des réseaux de neurones et des SVM.

3 Application des arbres de régression à la prédiction des durées des émissions

Nous avons choisi de prédire les durées des émissions en utilisant des arbres de régression. Pour cela, nous allons utiliser les régresseurs suivant :

- l'heure de début de l'émission ;
- le genre d'émission ;
- le jour de la semaine où l'émission est diffusée.

Parmi ces régresseurs, l'heure est une variable continue alors que le genre d'émission et le jour de diffusion sont des variables discrètes (symboliques).

3.1 Construction de l'arbre

La construction d'un arbre de régression part de la racine de l'arbre et produit les autres nœuds de façon récursive. À chaque nœud correspond un sous-ensemble de l'ensemble d'apprentissage et un test qui partitionnera ce sous-ensemble en plusieurs sous-ensembles. Nous avons choisi d'attribuer à chaque nœud un ensemble de caractéristiques statistiques (minimum, maximum, moyenne et écart-type) sur l'ensemble qui lui est rattaché afin de pouvoir traiter les cas qui n'apparaissent pas dans l'ensemble d'apprentissage (voir sous-section 3.2). Ainsi, au départ de l'induction de l'arbre, la racine contient les statistiques de tout l'ensemble d'apprentissage.

L'algorithme 4 présente une idée générale de l'algorithme que nous décrivons plus en détail dans cette section : il s'agit de l'adaptation de l'algorithme proposé par (Breiman *et al.*, 1984).

Algorithme 4 Algorithme d'induction d'un arbre de régression

```

procedure DÉVELOPPERNOEUD(Noeud  $N$ , Ensemble  $T$ , Entier  $\sigma$ )
▷ entrées :  $N$ , le noeud à développer,  $T$ , un ensemble d'exemples
           et  $\sigma$ , le paramètre d'arrêt de la construction de l'arbre
▷ sortie : complète le noeud  $N$  qui est soit une feuille, soit un noeud-test

   $N.moyenne \leftarrow moyenne(T.durée)$ 
   $N.minimum \leftarrow minimum(T.durée)$ 
   $N.maximum \leftarrow maximum(T.durée)$ 
  if  $N.maximum - N.minimum \leq \sigma$  then
     $N$  est une feuille
  else
    foreach  $test_i$  in  $Tests$  do
      calculer  $\Delta_i$ , la réduction de l'erreur attendue
    end foreach
     $TestMax \leftarrow \operatorname{argmax}_{test_i} \Delta_i$ 
    Partitionner  $T$  en  $T_i$  en fonction de  $TestMax$ 
    foreach  $T_i$  do DévelopperNoeud(new Noeud  $N_i, T_i, \sigma$ )
    end foreach
  end if
end procedure

function INDUCTIONARBRE(Ensemble  $A$ , Ensemble  $T$ ,  $\sigma$ ,  $\delta Erreur$ )
▷ entrées :  $A$  l'ensemble d'apprentissage,  $T$  l'ensemble de test
▷ paramètres :  $\sigma$  le critère d'arrêt et  $\delta Erreur$  le seuil d'augmentation de l'erreur
           pour l'élagage
▷ sortie : un arbre de régression

   $Arbre \leftarrow DévelopperNoeud(\mathbf{new}$  Noeud  $N, A, \sigma)$ 
  return ElaguerArbre( $Arbre$ ,  $\delta Erreur$ )
end function

```

Choix d'un test

Les arbres de régression se construisent à partir d'un ensemble de tests à effectuer sur l'ensemble d'apprentissage (voir la section 2.3, page 111). Nous avons choisi de prédire les durées d'une émission à partir d'informations disponibles à partir d'un contexte du modèle. Nous avons donc pensé aux tests suivants :

- à quelle heure l'émission a-t-elle été diffusée ?
- quel jour a-t-elle été diffusée ?
- le jour de diffusion est-il un jour férié ?
- le jour de diffusion tombe-t-il pendant les vacances scolaires ?
- le jour de diffusion est-il en été ?
- quel était le type de l'émission précédente ?
- quel est le type de l'émission ?

Afin de savoir si les tests étaient significatifs, nous avons calculé le coefficient de corrélation de Pearson de ces variables explicatives et de la durée des émissions des chaînes hertziennes en 2004. Plus ce coefficient est proche de 1, plus le test est significatif. Le tableau 11 résume les coefficients obtenus. L'heure et le type de l'émission ont les coefficients de Pearson les plus élevés et expliquent bien la durée. Le type de l'émission précédente, lorsqu'on le connaît, s'avère être aussi un bon prédicteur. En revanche les autres tests sont inutiles compte tenu de leurs coefficients.

Nom du test	Coefficient de Pearson avec les durées
type	0,470
heure	0,412
type précédent	0,351
été	0,017
vacances	0,013
férié	0,011
jour	0,007
week-end	0,004

TAB. 11: Coefficients de corrélation de Pearson des tests et des durées

Lors de l'induction de l'arbre, le choix du test est effectué selon des critères qui peuvent être très différents selon l'application. Par exemple, dans (Breiman *et al.*, 1984), le test choisi est celui qui réduit le plus l'écart-type des données du sous-ensemble. (Quinlan, 1992) utilise le critère de réduction de l'erreur attendue (*expected error reduction*). Si T dénote l'ensemble des exemples d'apprentissages avant l'application d'un test et que T_i dénote le $i^{\text{ème}}$ sous-ensemble de T correspondant à la $i^{\text{ème}}$ valeur du test, alors la réduction de l'erreur attendue est mesurée par (Quinlan, 1992) :

$$\Delta_{\text{erreur}} = sd(T) - \sum_i \frac{|T_i|}{|T|} \times sd(T_i) \quad (9)$$

où $sd(T)$ dénote l'écart-type des durées de l'ensemble T et $|T|$ dénote le nombre éléments dans T . Le quotient $\frac{|T_i|}{|T|}$ pénalise les sous-ensembles T_i qui ne contiennent qu'un élément et pénalise plus globalement un test qui va avoir tendance à isoler les éléments de T .

Après avoir calculé Δ_{erreur} pour chaque test potentiel, l'algorithme choisit celui qui maximise la réduction de l'erreur attendue. Dans notre but, c'est-à-dire encadrer les durées d'une émission, l'écart-type des durées n'est pas assez représentatif de l'ensemble T . En effet, le domaine $[\bar{T} - sd(T) ; \bar{T} + sd(T)]$, où \bar{T} est la moyenne des durées de l'ensemble T , ne contient pas, par définition, les durées minimales ou maximales. Pourtant, ces valeurs sont nécessaires pour pouvoir prédire un éventail de durées suffisamment proche des durées réelles. Nous avons donc voulu réduire l'écart entre la durée minimale et la durée maximale de l'ensemble afin d'encadrer plus précisément les durées lors de la prédiction :

$$\Delta_{\text{erreur}} = |\max(T) - \min(T)| - \sum_i \frac{|T_i|}{|T|} \times |\max(T_i) - \min(T_i)| \quad (10)$$

où $\max(T)$ et $\min(T)$ désignent respectivement la durée maximale et la durée minimale de l'ensemble T . Les durées étant toujours positives, $|\max(T) - \min(T)| = \max(T) - \min(T)$, mais nous laissons les valeurs absolues pour la lisibilité.

Une fois que le test a été choisi, les données sont partitionnées en fonction des valeurs du test et autant de branches que de partitions sont créées. Plutôt que de se limiter à un arbre binaire,

nous avons choisi de créer autant de branches que de valeurs pour une variable explicative symbolique. En revanche, pour une variable explicative continue, le test est toujours binaire : une branche représente l'ensemble des exemples dont la valeur est inférieure à une valeur ν , et l'autre branche les exemples dont la valeur est supérieure ou égale à ν . La difficulté est de trouver la valeur ν qui sépare les valeurs continues en deux sous-ensembles et qui maximise le critère (10).

De nombreux articles s'intéressent à la discrétisation des variables continues dans le cadre de l'apprentissage symbolique (Nguyen, 2003) : cela consiste à partitionner l'espace de la variable continue en intervalles. Plusieurs méthodes ont été introduites selon que l'on souhaite obtenir deux (Breiman *et al.*, 1984; Quinlan, 1986; Elomaa et Rousu, 2002) ou plusieurs partitions (Janssens *et al.*, 2006; Boulle, 2004; Grzymala-Busse, 2004) : la plupart sont basées sur des critères statistiques (réduction de l'entropie ou de l'écart-type), mais on trouve aussi des méthodes basées sur des algorithmes de clustering. Nous avons choisi d'adapter aux arbres de régression la méthode proposée dans (Elomaa et Rousu, 2002), pour sa facilité de mise en œuvre et sa rapidité d'exécution. Cette méthode consiste à séparer l'ensemble T avec chacun des seuils potentiels ν_i ; le seuil retenu sera celui qui maximise le critère (10). Les ν_i potentiels sont généralement les valeurs intermédiaires entre deux valeurs de T consécutives.

L'utilisation de tests n-aires plutôt que binaires ne permet cependant pas le regroupement des valeurs des tests. Par exemple, les tests binaires peuvent permettre de regrouper les jours de la semaine en séparant les jours ouvrés et le week-end. Afin de remédier à ce problème, nous avons ajouté une variable *week-end* booléenne dont la valeur est calculée à partir de la variable représentant le jour de la semaine. Cependant, le tableau 11 montre que ce nouveau test a un coefficient de Pearson très faible et, en pratique, ce test n'est jamais utilisé.

Malgré ce désavantage, les tests n-aires produisent un arbre qui peut être lu facilement, voire même modifié manuellement – par exemple, pour l'ajout d'un nouveau genre télévisuel.

Critères d'arrêt de la construction de l'arbre

Un critère d'arrêt est nécessaire pour éviter que chaque feuille ne contienne qu'un seul élément. En général, un critère naturel est utilisé : si l'expression (10) vaut 0 ou est en dessous d'un certain seuil, aucun test ne peut être choisi, et par conséquent la construction de l'arbre est arrêtée.

On utilise aussi un critère d'arrêt additionnel pour éviter d'attendre que le critère ci-dessus se réalise. Il existe différents critères d'arrêt additionnels de l'induction de l'arbre : certains critères limitent la profondeur de l'arbre, d'autres s'intéressent aux propriétés statistiques de l'ensemble correspondant à la feuille. Dans notre cas, nous avons choisi d'arrêter la construction de l'arbre dès que la valeur $|\max(T) - \min(T)|$ devient inférieure à un seuil σ . Cela nous évite en effet d'avoir des intervalles de durées insuffisamment larges qui seraient sensibles à toute variation inattendue de la durée d'une émission.

Élagage de l'arbre

Comme beaucoup d'algorithmes d'apprentissage, l'algorithme d'induction des arbres de régression conduit parfois à un sur-apprentissage. Ce phénomène est caractérisé par un taux d'erreur extrêmement faible sur l'ensemble d'apprentissage et une inadaptation à de nouvelles données. Plusieurs algorithmes sont proposés pour élaguer l'arbre après sa construction, le but étant toujours de trouver un sous-arbre de l'arbre induit dont le taux d'erreur reste acceptable. Les premiers algorithmes (Breiman *et al.*, 1984; Quinlan, 1986) nécessitent un ensemble de test,

contenant des exemples qui n'ont pas été présentés lors de l'induction de l'arbre, et un seuil qui indique à l'algorithme de combien il peut augmenter le taux d'erreur. Les méthodes plus récentes ne diffèrent que par l'évaluation de l'erreur en cas d'élagage d'une branche : plusieurs méthodes sont comparées dans (Torgo, 1998).

3.2 Régression d'une durée

La régression d'une durée se fait en parcourant l'arbre depuis la racine et en effectuant les tests situés aux nœuds de l'arbre, puis en empruntant la branche qui correspond à la valeur du test. La valeur moyenne et les extrema locaux sont renvoyés lorsqu'une feuille est atteinte ou lorsque le dernier test que peut satisfaire le vecteur d'attributs est atteint. Ce dernier cas peut survenir lorsqu'un genre d'émission n'a jamais été vu auparavant ou lorsqu'une émission n'a jamais été diffusée à cette heure.

3.3 Exemples de règles et d'estimation de durées

```

SI genre = météo
    moyenne = 134 s, minimum = 69 s, maximum = 285 s

SI genre = feuilleton
    SI heure < 17:43:51
        moyenne = 2624 s, minimum = 2549 s, maximum = 2691 s
    SINON
        SI heure < 20:58:24
            moyenne = 5978 s, minimum = 5803 s, maximum = 6084 s
        SINON
            SI heure < 20:59:25
                moyenne = 6237 s, minimum = 5862 s, maximum = 6618 s
            SINON
                moyenne = 6112 s, minimum = 6074 s, maximum = 6150 s
    
```

FIG. 30: Extrait de l'arbre de régression de TF1 entraîné sur les émissions de 2004

```

SI genre = météo
    moyenne = 159 s, minimum = 75 s, maximum = 311 s

SI genre = émission de service
    SI heure < 11:55:30
        moyenne = 60 s, minimum = 41 s, maximum = 116 s
    SINON
        SI heure < 14:30:14
            moyenne = 206 s, minimum = 59 s, maximum = 311 s
        SINON
            moyenne = 83 s, minimum = 73 s, maximum = 174 s
    
```

FIG. 31: Extrait de l'arbre de régression de France 2 entraîné sur les émissions de 2004

Les figures 30 et 31 présentent des extraits d'arbres de régression respectivement entraînés sur les émissions de TF1 et de France 2 en 2004. Les deux arbres montrent que les durées des

3. Application des arbres de régression à la prédiction des durées des émissions

Date et heure	Titre	Genre	Durée réelle	Durée prédite		
				moyenne	minimum	maximum
06 jan. 14:49:26	Entre haine et amour	Film	5452 s	5365 s	5242 s	5492 s
19 jan. 20:58:24	Combien ça coûte	Magazine	8502 s	7164 s	5827 s	8573 s
01 avr. 19:50:02	À vrai dire	Service	118 s	124 s	100 s	176 s
26 avr. 20:46:10	Météo	Météo	126 s	134 s	69 s	285 s
20 mai 11:12:08	La ferme célébrités	Jeu	2796 s	2709 s	2639 s	2806 s
25 mai 20:39:20	Ligue des champions	Sport retransmis	11351 s	7816 s	7535 s	8007 s
27 juin 20:58:08	Dolmen	Feuilleton	5889 s	5978 s	5803 s	6084 s
06 nov. 12:58:46	13 Heures	Journal télévisé	1367 s	1336 s	1223 s	1473 s
12 sep. 08:30:16	Céline Dion	Clip	91 s	94 s	15 s	206 s
25 sep. 20:57:54	Ciné Dimanche	Fiction	13533 s	14526 s	13522 s	15520 s

TAB. 12: Exemples de durées prédites pour quelques émissions de TF1 en 2005

météos sont facilement encadrables, c'est-à-dire qu'un seul test est nécessaire pour atteindre la distance souhaitée entre la durée minimale et la durée maximale.

Sur la figure 30, nous montrons que peu de tests sont nécessaires pour l'estimation des durées des feuilletons. Nous voyons que le dernier test (« si heure < 20:59:25 ») produit par l'induction de l'arbre n'est pas réellement nécessaire puisque les deux groupes créés à l'issue du test ont des propriétés relativement proches. Ce test est cohérent avec la politique de diffusion de TF1 : il existe en effet les feuilletons du matin et de l'après-midi (*En toute amitié, les Feux de l'amour*) dont la durée est d'une quarantaine de minutes et le feuilleton diffusé en première partie de soirée (la « saga de l'été », *Zodiaque*) dont la durée est proche d'une heure et quarante minutes.

Sur la figure 31, nous avons choisi de développer la branche permettant d'estimer la durée des émissions de service. France 2 en compte deux : *Inforoute*, diffusée plusieurs fois par jour, et *Consumag*, diffusée une fois par jour peu avant 14h. Les règles induites sont simples et ne dépendent que de l'heure de diffusion. La première règle « si heure < 11:55:30 » et la troisième « si heure \geq 14:30:14 » permettent d'écarter tous les *Inforoute*. En revanche, les émissions de service diffusées entre 11:55:30 et 14:30:14 sont soit des *Inforoute*, soit des *Consumag*. L'arbre de régression ne permet pas de distinguer ces deux émissions, ce qui explique que la distance entre la durée minimale et la durée maximale soit plus importante.

Le tableau 12 présente quelques exemples de prédiction des durées des émissions de TF1 en 2005 avec l'arbre de régression entraîné sur les données de 2004. Nous avons sélectionné les exemples afin de montrer quelques cas où les arbres de régression étaient performants et quelques cas où, au contraire, la capacité des arbres de régression à généraliser n'est pas suffisante. En effet, la durée de la tranche horaire *Ciné Dimanche* est encadrée par une fenêtre temporelle dont l'envergure dépasse une demi-heure. À cet exemple, nous ajoutons une mauvaise estimation de la durée d'un match de football pour montrer que les durées des retransmissions sportives sont difficilement estimables. À l'inverse, nous avons montré précédemment que l'arbre de régression estimait bien la durée des feuilletons (voir figure 31) : cela est confirmé par l'estimation de la durée de *Dolmen*, saga de l'été 2005, obtenue par l'apprentissage des durées du feuilleton *Zodiaque*, saga de l'été 2004.

Nous allons décrire dans la section suivante comment, à partir des informations de régression fournies par l'arbre, il est possible d'estimer la probabilité d'une durée pour une émission donnée.

4 Dédution de la probabilité des durées

Dans la modélisation probabiliste que nous utilisons (voir chapitre 7), il est nécessaire de connaître la probabilité d'une durée d'émission. Or les arbres de régression nous donnent des informations statistiques sur les durées lors de leur prédiction : la durée minimale, la durée maximale et la moyenne. Nous avons donc choisi d'approcher la distribution empirique des durées par une distribution connue, paramétrée à partir de ces données statistiques. La figure 32 présente la distribution des durées sur trois feuilles d'un arbre de décision entraîné sur l'ensemble des émissions de France 2 diffusées en 2004. Nous représentons ces distributions par des histogrammes dont chaque tranche représente dix valeurs. L'envergure de ces histogrammes montre aussi l'envergure de l'encadrement des durées : par exemple, nous serons en mesure de prédire la durée d'un jeu télévisé matinal à moins de cent secondes près pour une durée minimale d'environ trente minutes. La nature asymétrique de la distribution est clairement visible sur ces figures. Compte tenu des distributions empiriques, nous avons choisi une distribution en forme de cloche qui attribue à la durée moyenne la probabilité la plus forte et des probabilités de plus en plus faibles au fur et à mesure que la durée se rapproche des extrema.

Nous avons ainsi choisi d'interpréter ces distributions comme des gaussiennes asymétriques. (Bennett, 2003) présente une nette amélioration de ses classifieurs en utilisant des gaussiennes asymétriques plutôt que la loi normale. Soit μ la moyenne, M le maximum et m le minimum des durées sur une feuille de l'arbre de décision. On pose :

$$\begin{aligned}\sigma_l &= \mu - m \\ \sigma_r &= M - \mu \\ \sigma &= M - m\end{aligned}\tag{11}$$

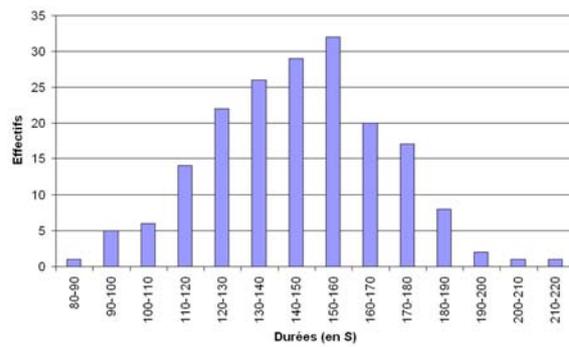
alors la probabilité d'une durée d est donné par la fonction p suivante :

$$p(d|\mu, \sigma_l, \sigma_r) = \begin{cases} \frac{2}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(d-\mu)^2}{2\sigma_l^2}\right] & \text{si } d \leq \mu \\ \frac{2}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(d-\mu)^2}{2\sigma_r^2}\right] & \text{sinon.} \end{cases}\tag{12}$$

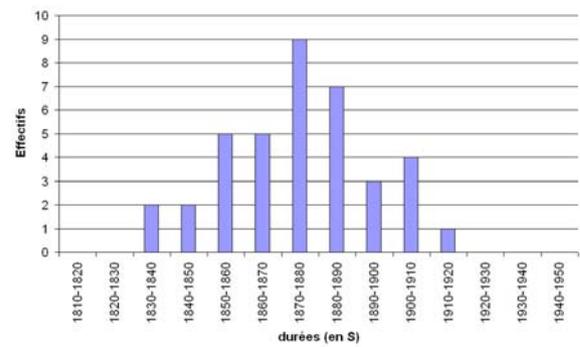
5 Conclusion

Nous avons adapté dans ce chapitre une méthode de régression basée sur le principe des arbres de décision. Ces arbres de régression permettent, à partir d'informations symboliques telles que le jour de la semaine, le genre de l'émission, et à partir d'informations continues, telle que l'heure de diffusion, d'encadrer la durée de l'émission. Sur chaque nœud de l'arbre, une durée est représentée par des extrema et une moyenne, ce qui permet de déduire la probabilité d'une durée donnée en approximant la distribution empirique par une gaussienne asymétrique.

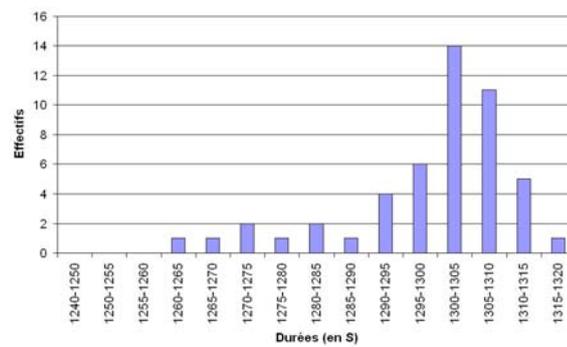
Dans le chapitre suivant, nous allons modéliser les grilles de programmes en utilisant la régression des durées.



(a) La météo du soir en semaine



(b) Un jeu télévisé diffusé le jeudi matin vers 5h30



(c) Une série diffusée en semaine vers 11h

FIG. 32: Distribution des durées sur une feuille de l'arbre de décision (source : France 2, 2004)

Chapitre 9

Prédiction de grilles de programmes

Sommaire

1	Fusion des prédictions et du guide des programmes	122
1.1	Méthode par développement complet du graphe	122
1.2	Méthode par développement itératif	124
2	Réduction du graphe des possibilités	125
3	Zones d’incertitude	126
4	Étiquetage des émissions	127
5	Conclusion	127

Dans les deux chapitres précédents, nous avons modélisé les grilles de programmes d’une chaîne de télévision à l’aide d’un modèle markovien. Ce modèle, entraîné sur des grilles de programmes de journées passées, permet de reproduire n’importe quelle combinaison de programmes ayant eu lieu par le passé.

Le tableau 13 (page 122) présente un extrait de la grille prévisionnelle d’un *guide* de programmes (source : *Télérama*) comparé à l’extrait correspondant de la grille de programmes réellement diffusée ce jour-là. Si l’incomplétude et l’imprécision des guides de programmes (voir section 6, page 52) ne nous permet pas de les aligner directement sur le flux, ces derniers présentent cependant quelques avantages : non seulement ils apportent des informations sur les émissions qui vont être diffusées (titres, descriptions, . . .), mais en plus leur grille prévisionnelle a été décidée une semaine avant la diffusion. En effet, cela permet de prendre en compte une actualité à moyen terme, comme par exemple la diffusion d’un match de football après une victoire la semaine précédente. De plus, la grille prévisionnelle peut nous permettre de diminuer le nombre de grilles prédites par le modèle, puisque seule les grilles prédites contenant la grille prévisionnelle peuvent être de bonnes candidates pour l’alignement sur le flux (Poli et Carrive, 2007).

Nous présentons dans ce chapitre une méthode de fusion des informations issues du guide de programmes et du modèle statistique.

Heure	Genre	Titre	Heure	Genre	Titre
5:55:00	Jeu	Les Z'amours	5:55:13	Jeu	Les Z'amours
			6:24:54	Interprogrammes	
			6:25:41	Service	Point route
6:30:00	Magazine	Télématin	6:26:36	Interprogrammes	
			6:29:53	Magazine	Télématin
			8:30:50	Interprogrammes	
			8:34:09	Service	Point Route
			8:35:11	Magazine	Un livre,...
8:35:00	Feuilleton	Des jours et...	8:36:23	Interprogrammes	
	⋮		8:37:14	Feuilleton	Des jours ...
				⋮	
			19:54:18	Météo	
			19:57:55	Interprogrammes	
20:00:00	Information	20H le journal	19:58:57	Information	20H le journal
			20:37:57	Interprogrammes	
			20:40:04	Magazine	Mission prix
			20:41:07	Interprogrammes	

TAB. 13: Extraits de la grille prévisionnelle (à gauche) et de la grille de programmes (à droite) du 21 Septembre 2005 sur France 2

1 Fusion des prédictions et du guide des programmes

Soit (M, Λ) un CHMM, comme défini dans le chapitre précédent et entraîné sur les grilles de programmes des années précédentes. L'ensemble des séquences d'observations produites à partir du CHMM peut être considéré comme un graphe dirigé acyclique dont l'origine est l'émission prédite par le premier état et où les poids des arêtes sont les probabilités de transition d'une émission à l'autre (c'est-à-dire d'un état du CHMM à l'autre). Une émission est déduite d'un état en utilisant le genre télévisuel qu'il représente, le contexte dans lequel il a été atteint, et les durées prédites par l'arbre de régression étant donnés le contexte et l'état courants.

Chaque chemin de ce graphe orienté acyclique est une grille de programmes possible. Pour sélectionner les grilles candidates, il suffit de sélectionner dans le graphe les chemins qui passent par les émissions présentées par la grille prévisionnelle. Deux méthodes peuvent être employées.

1.1 Méthode par développement complet du graphe

La première méthode consiste à développer le graphe jusqu'à ce que l'heure du dernier programme de la grille prévisionnelle soit atteinte. Cela revient à un parcours en profondeur dans lequel chaque fois qu'un nouveau nœud est atteint, de nouveaux nœuds adjacents sont créés en fonction des probabilités de transition du CHMM (voir algorithme 5). Si l'on note N le nombre d'états du modèle et p la profondeur de l'arbre que l'on construit, la complexité de l'expansion de l'arbre est, dans le pire des cas, en $O(N^p)$. On considère ensuite chaque chemin allant de l'origine jusqu'à ce dernier programme comme une série temporelle, et les grilles prédites candidates sont celles qui englobent la grille prévisionnelle.

Une méthode de comparaison de séries temporelles peut être employée dans le but d'identifier si la grille prévisionnelle est incluse dans le graphe des grilles de programmes possibles. La méthode la plus connue est appelée « *dynamic time warping* » (Myers et Rabiner, 1981) et permet de trouver le meilleur alignement entre deux séquences, autrement dit la transformation temporelle la plus faible d'une séquence source en une séquence cible; l'inconvénient de cette méthode est que chaque élément de la séquence cible doit avoir un correspondant dans la séquence

Algorithme 5 Algorithme de développement complet du graphe

procédure DÉVELOPPEMENTCOMPLETGRAPHE(Modèle M , Guide G , Nœud N)
 ▷ entrées : M un CHMM entraîné et $G = g_1, \dots, g_n$ un guide de programmes
 ▷ sortie : N le nœud développé

```

Ajouter  $g_1$  à  $N$ 
File  $\leftarrow g_1$ 

repeat
    NoeudCourant  $\leftarrow$  Défile(File)
    NoeudCourant.Fils  $\leftarrow \emptyset$ 
    ContexteCourant  $\leftarrow$  NoeudCourant.Contexte
    for  $s \in M$ .Etats do
        if  $P(s|NoeudCourant.Genre, NoeudCourant.Contexte) > 0$  then
            NoeudCourant.Fils  $\leftarrow$  NoeudCourant.Fils  $\cup s$ 
            File  $\leftarrow$  File +  $s$ 
        end if
    end for
until ContexteCourant >  $g_n$ .Contexte ou File =  $\emptyset$ 
end procedure
    
```

source.

Nous avons finalement utilisé la méthode appelée « *minimal variance matching* » (MVM) et introduite dans (Latecki *et al.*, 2005) en réponse à cet inconvénient. Considérons deux séquences $a = (a_1, \dots, a_m)$ et $b = (b_1, \dots, b_n)$ de longueurs différentes n et m telles que $m < n$: le but de la méthode MVM est de trouver une sous-séquence b' de b , de longueur m telle que a puisse correspondre au mieux à b' . Pour cela, il faut définir une fonction f de correspondance telle que :

$$\left\{ \begin{array}{l} f : \{1, \dots, m\} \rightarrow \{1, \dots, n\} \\ f(i) < f(i+1) \quad \forall i \in \{1, \dots, m\} \\ a_i \text{ corresponde à } b_{f(i)} \quad \forall i \in \{1, \dots, m\}. \end{array} \right. \quad (1)$$

L'ensemble d'indices $\{f(1), \dots, f(m)\}$ définit la sous-séquence b' de b . Une fois que la correspondance entre les deux séquences a été trouvée, le calcul de la distance est effectué par n'importe quelle distance $d_f(a, b)$. Finalement, le but de la méthode MVM est de trouver une fonction de correspondance f telle que $d_f(a, b)$ soit minimale. Pour résoudre ce problème, il faut dans un premier temps calculer la matrice $m \times n$ des différences $r_{ij} = (b_j - a_i)$. La meilleure correspondance entre les deux séquences peut être trouvée en résolvant le problème du chemin minimal sur la matrice des différences. Pour cela, (r_{ij}) doit être considérée comme un graphe dirigé acyclique dans lequel r_{ij} est directement lié à r_{kl} si et seulement si $k - i = 1$ et $j < l$. Ces deux conditions permettent de s'assurer respectivement que l'on se déplace toujours vers la colonne suivante et que l'on puisse sauter des colonnes sans pour autant pouvoir aller en arrière. Le coût de l'arête entre r_{ij} et r_{kl} est défini par la fonction $poids(r_{ij}, r_{kl}) = (r_{kl})^2$.

Exemple 1 Soit les deux séquences $a = (1, 2, 8, 6, 8)$ et $b = (1, 2, 9, 3, 3, 5, 9)$. Si l'on considère la distance euclidienne, la matrice r s'écrit :

$$r = \begin{bmatrix} \boxed{0} & 1 & 8 & 2 & 2 & 4 & 8 \\ -1 & \boxed{0} & 7 & 1 & 1 & 3 & 7 \\ -7 & -6 & \boxed{1} & -5 & -5 & -3 & -1 \\ -5 & -4 & 3 & -3 & -3 & \boxed{-1} & 3 \\ -7 & -6 & 1 & -5 & -5 & -3 & \boxed{1} \end{bmatrix}.$$

Le plus court chemin est encadré dans la matrice ci-dessus ; en suivant les éléments encadrés, la fonction de correspondance optimale \hat{f} est donnée par :

$$\hat{f}(1) = 1, \hat{f}(2) = 2, \hat{f}(3) = 3, \hat{f}(4) = 6, \hat{f}(5) = 7.$$

La distance entre a et b est donc égale à $\sqrt{3}$.

Pour appliquer la méthode MVM à la comparaison de grilles de programmes prédites et de grilles de programmes prévisionnelles, nous proposons d'utiliser la mesure de similarité suivante à la place de la distance euclidienne :

$$d(e, e') = \begin{cases} \infty & \text{si les genres de } e \text{ et } e' \text{ sont différents} \\ |\theta_e - \theta_{e'}| + |Durée(e) - Durée(e')| & \text{sinon} \end{cases} \quad (2)$$

où e et e' dénotent deux émissions. Ainsi, la mesure de similarité est infinie si les émissions sont d'un genre différent et est égale à la différence de leurs heures de départ à laquelle on ajoute la différence de leurs durées.

Avec la méthode MVM introduite ci-dessus, il est possible de sélectionner tous les chemins du graphe dirigé acyclique produit par le CHMM en vérifiant si la grille prédite, représentée par un chemin dans ce graphe, contient bien la grille prévisionnelle.

L'inconvénient de cette méthode de prédiction des grilles de programmes est naturellement la production de grilles inutiles qui ne seront éliminées qu'après comparaison avec le guide de programmes. Compte tenu de la complexité en $O(N^p)$, même si le nombre de transitions possibles est diminué grâce au contexte du CHMM, l'expansion de l'arbre entier est trop coûteuse en temps de calculs (voir la section 3.3, page 138).

1.2 Méthode par développement itératif

La seconde méthode évite le développement total du graphe. Pour cela, nous proposons de considérer deux émissions consécutives de la grille prévisionnelle et de développer l'ensemble des chemins qui conduisent de la première à la seconde. Lorsqu'une branche est développée à cet effet et que l'heure de début du nœud courant dépasse l'heure de l'émission cible d'un délai Δ , on abandonne le développement de cette branche : Δ correspond au retard maximum toléré entre une émission et sa prévision dans un guide de programmes. Pour nos expérimentations, nous avons fixé Δ à trente minutes, ce délai étant le retard maximum observé sur les données de l'INA concernant TF1 lors d'une journée « normale » (sans événement extraordinaire). L'algorithme consiste donc à développer les branches du graphe tant que :

- le nœud courant n'est pas similaire à l'émission cible, au sens de la mesure de similarité (2) ;

- la différence entre l’heure de départ de l’émission représentée par le nœud courant et l’émission cible de la grille prévisionnelle est inférieure à Δ .

Ainsi, il est possible de se rendre compte assez rapidement que la branche en cours de construction ne correspond pas à une grille prédite candidate. Toutefois, le nombre de branches à explorer reste encore trop élevé. Par exemple, tous les chemins du graphe partant d’une émission de la grille prévisionnelle à l’émission suivante mènent chacun à des nœuds différents. Pourtant ceux-ci représentent bien la même émission de la grille prévisionnelle avec des heures de départ légèrement différentes. Nous proposons dans la section suivante deux heuristiques qui ont pour but de contrôler la croissance du graphe dirigé acyclique.

2 Réduction du graphe des possibilités

Afin de réduire la taille du graphe des grilles prédites, en terme d’arcs et de nœuds, et de ne conserver que des « bonnes » candidates pour une journée donnée, nous avons utilisé deux heuristiques.

La première tire profit des connaissances que nous avons sur les grilles prévisionnelles (voir la section 6, page 52). En effet, si certains genres télévisuels ne sont pas annoncés dans un guide de programmes, d’autres, au contraire, le sont toujours : c’est le cas, entre autres, des fictions, des magazines longs et des jeux télévisés. Ainsi, si un chemin entre deux émissions consécutives annoncées dans le guide de programmes passe par un nœud représentant une émission qui devrait être annoncée, alors ce chemin peut être supprimé. En effet, ce cas de figure ne peut se produire dans un guide de programmes.

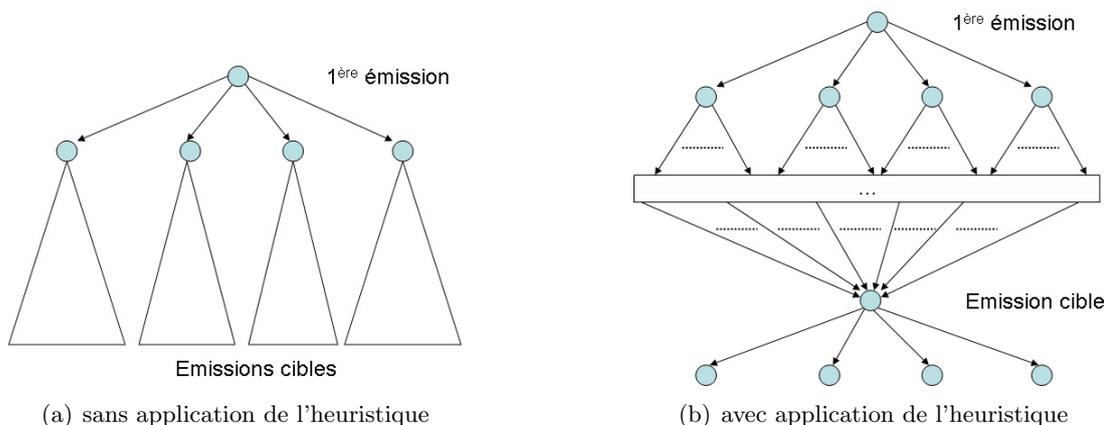


FIG. 33: Schéma représentatif d’un arbre des grilles possibles avant et après l’application de la deuxième heuristique. La fusion des différents nœuds cibles diminue le temps de construction de l’arbre.

La seconde heuristique concerne le problème évoqué dans la sous-section 1.2 : le graphe produit par la simulation du CHMM croît très rapidement. En effet, plusieurs nœuds correspondent au programme cible p_i du guide. Ainsi, lorsque l’on tente de prédire les chemins possibles entre le programme p_i et le programme p_{i+1} , plusieurs nœuds de départ sont à envisager. Cela a pour conséquence d’augmenter la taille de l’arbre inutilement. Afin de réduire la complexité du graphe, le but de la seconde heuristique est de diminuer le nombre de nœuds correspondants aux différentes émissions cibles en les fusionnant en un seul (voir figure 33). Le nouveau nœud

résultant de la fusion des N nœuds n_i a les propriétés suivantes :

- le contexte de départ de l'émission est égale à une combinaison des contextes des nœuds n_i (cela équivaut à une moyenne) ;
- l'ensemble des durées est égale à l'union des ensembles des durées des nœuds n_i ;
- la moyenne des durées est égale à la moyenne des moyennes des durées des nœuds n_i .

Nous avons présenté les deux heuristiques qui permettent respectivement d'arrêter le développement des branches inutiles et de limiter la largeur de l'arbre en créant des « goulots d'étranglement », évitant ainsi les branches redondantes. Nous montrerons l'application des heuristiques sur un cas réel dans la section 3.3 du chapitre 10 (page 138).

Nous allons voir à présent comment le système traite le cas où une émission imprévue est annoncée dans le guide de programmes.

3 Zones d'incertitude

Nous abordons dans cette section un cas marginal, différent de la déprogrammation : l'annonce d'une émission qui n'apparaît pas dans le modèle statistique. C'est le cas d'émissions événementielles dans un contexte d'actualités particulier mais suffisamment importantes pour être annoncées dans les guides de programmes. Par exemple, les campagnes électorales de mai 2005 sur France 2 étaient diffusées à des heures précises et régulières, donc annoncées dans les grilles prévisionnelles, mais n'avaient jamais été diffusées auparavant. Aucun chemin de notre arbre des grilles possibles ne peut donc passer par un nœud représentant un tel genre d'émission, dont il est même impossible de prédire la durée. De plus, il est impossible de savoir si ces programmes sont insérés dans la grille ou s'ils remplacent un programme existant.

	Heure	Genre
E_1	5:55:00	Jeu
E_2	6:30:00	Magazine
E_3	8:35:00	Campagne électorale
E_4	8:45:00	Feuilleton
E_5	9:05:00	Feuilleton

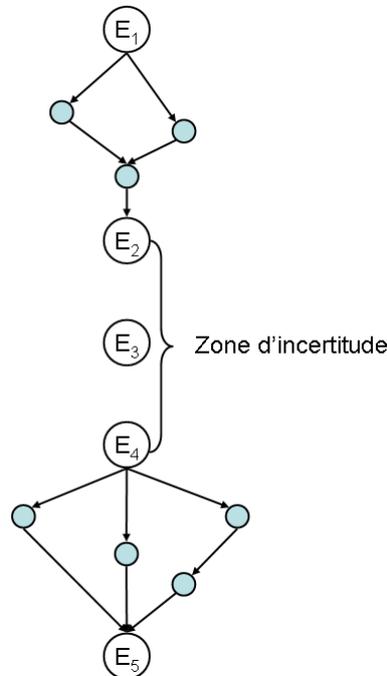


FIG. 34: Illustration d'une zone d'incertitude

Néanmoins, ce n'est que localement autour de ce genre de programmes que la prédiction est incertaine. On définit donc une zone d'incertitude entre le dernier programme de la grille prévisionnelle correctement prédit et le programme suivant l'émission imprédictible. La figure 34 montre un extrait de guide de programmes $\{E_1, \dots, E_5\}$ annonçant une émission de « campagne électorale » exceptionnelle. Le modèle a réussi à développer les chemins entre E_1 et E_2 ; en revanche, il n'est pas en mesure d'en trouver un entre E_2 et E_3 à cause du caractère exceptionnel de celui-ci. E_2 est donc la dernière émission du guide à avoir été atteinte et E_4 est la première à partir de laquelle il sera possible de poursuivre le développement de l'arbre : une zone d'incertitude est donc créée entre E_2 et E_4 exclues et peut se noter $]E_2 ; E_4[$.

La prise en compte des zones d'incertitude dans notre système permet de faire appel à un opérateur pour structurer ces portions de flux tout en continuant à structurer le reste de la journée automatiquement.

4 Étiquetage des émissions

C'est aussi lors de cette phase de rapprochement entre le guide des programmes et les nœuds du graphe des grilles possibles, qu'il est possible de récupérer les données issues du guide de programmes telles que les artistes de l'émission et le résumé : ces informations peuvent être affectées au nœud du graphe des grilles possibles correspondant à l'émission pointée dans le guide des programmes.

Pour donner un titre aux émissions qui ne sont jamais présentées dans les guides de programmes, le système nécessite une base de connaissance à moyen-terme, qui peut être mise à jour souvent et facilement, contenant tous les titres des émissions n'apparaissant pas dans les guides. Cette base est amenée à changer au moins une fois par saison audiovisuelle, au moment où les nouveaux programmes sont proposées par les chaînes.

5 Conclusion

L'intérêt de ce chapitre est de montrer le rapprochement entre des connaissances issues des diffusions passées, représentées par un modèle statistique, et des connaissances complémentaires issues des guides de programmes (environ une semaine avant la diffusion). Le croisement des deux sources d'information peut être source de conflit dans le cas où le guide de programme présente des émissions qui n'ont jamais été diffusées. Cependant, dans la majeure partie des cas, le guide de programmes est une confirmation des informations statistiques et permet non seulement de lever les ambiguïtés en réduisant le nombre de grilles de programmes prédites mais aussi d'ajouter des descriptions à certains programmes.

Chapitre 10

Évaluation du modèle et des prédictions

Sommaire

1	Environnements d'apprentissage	130
1.1	Environnement d'apprentissage optimal	130
1.2	Environnement utilisé	131
2	Évaluation des arbres de régression	131
2.1	Taux d'erreur global	131
2.2	Taux d'erreur par genres	132
2.3	Précision temporelle des prédictions	133
2.4	Impact de σ sur les arbres de régression	133
2.5	Nombre d'exemples d'apprentissage	135
3	Évaluation de la prédiction des grilles	135
3.1	Capacité d'expression du CHMM seul	135
3.2	Fusion des guides de programmes et des prédictions du CHMM	137
3.3	Exemple de construction d'un graphe	138
3.4	Effets de la contextualisation sur le nombre moyen de transitions possibles	148
3.5	Impact du modèle sur la recherche des transitions	148
4	Conclusion	149

Nous nous intéressons dans ce chapitre à l'évaluation des différentes parties du modèle. Nous présentons dans un premier temps les environnements d'apprentissage que nous avons utilisés. Puis nous décrirons les arbres de régressions en termes de nombre de nœuds et de feuilles, d'envergure des fenêtres et nous les évaluerons sur plusieurs chaînes de télévision françaises. Dans cette partie, nous appelons envergure d'une fenêtre la distance entre la durée minimale et la durée maximale contenue dans un nœud de l'arbre de régression; en effet, c'est dans cette zone que le système cherchera une rupture entre deux programmes, d'où la désignation de fenêtre temporelle ou simplement de fenêtre. Nous nous intéressons par la suite au CHMM et à sa capacité d'exprimer des grilles de programmes. Nous testerons ensuite la prédiction des grilles de programmes avec la fusion des prédictions du CHMM et des guides de programmes.

1 Environnements d'apprentissage

Chacun des apprentissages nécessite l'utilisation de la taxinomie des genres audiovisuels (voir la section 2, page 66). Nous avons vu que la nomenclature de Médiamétrie, utilisée dans le cadre de nos travaux, utilise des genres parfois trop spécifiques qu'il est difficile de distinguer automatiquement, comme par exemple les films et les téléfilms.

Nous avons donc utilisé des « environnements d'apprentissage ». Un environnement d'apprentissage est un moyen de représenter les connaissances sur la taxinomie des genres audiovisuels. Il est constitué de trois ensembles : l'ensemble des classes, l'ensemble des genres à ignorer et l'ensemble des règles d'alias. Les classes sont les genres qui seront utilisés dans les algorithmes de prédiction ou comme états du CHMM. Les règles d'alias permettent de pallier la surspécialisation de la nomenclature de Médiamétrie et permettent de faire en sorte que téléfilm et film soient considérés comme une seule et même classe.

Nous avons cherché dans un premier temps à trouver l'environnement optimal pour une chaîne, c'est-à-dire celui qui maximise les prédictions du modèle. Cependant, devant la complexité de la tâche, nous nous sommes résolus à créer des environnements manuellement.

1.1 Environnement d'apprentissage optimal

La nomenclature de Médiamétrie contient 170 désignations de genres télévisuels. Un environnement d'apprentissage peut être vu mathématiquement comme une k -partition de ces genres, c'est-à-dire une partition de la nomenclature en k sous-ensembles disjoints, où k est le nombre de classes considérées. Le nombre de k -partitions d'un ensemble à n éléments est donné par le nombre de Stirling de seconde espèce :

$$\left\{ \begin{matrix} n \\ k \end{matrix} \right\} = \frac{1}{k!} \sum_{j=1}^k (-1)^j \binom{k}{j} (k-j)^n. \quad (1)$$

Le nombre de Bell, qui donne le nombre de partitions d'un ensemble à n éléments peut se calculer par :

$$B_n = \sum_{k=1}^n \left\{ \begin{matrix} n \\ k \end{matrix} \right\}. \quad (2)$$

L'espace des environnements possibles est donc très grand puisqu'il comporte $B_{170} \approx 1,05 \times 10^{209}$ éléments. La recherche de l'environnement optimal est donc un problème de combinatoire dont il faut définir le but. Dans notre cas, un environnement d'apprentissage est optimal s'il permet de séparer les émissions de telle sorte que les résultats des arbres de régression et des CHMM soient maximaux. Pour tester un environnement, nous construisons un arbre de décision et nous entraînons un CHMM, puis nous testons le tout sur un ensemble de validation. Devant de tels problèmes, des méthodes stochastiques sont souvent employées. Il s'agit de sélectionner aléatoirement un individu de l'espace et d'exécuter une fonction de test dessus.

Dans notre cas, nous proposons de choisir un k aléatoirement, tel que $1 < k < 170$, de regrouper aléatoirement les genres en k partitions, de chercher un environnement qui permet de créer un arbre de régression et d'entraîner un CHMM plus performants que ceux que nous avons obtenus avec un environnement créé manuellement. À ce jour, un seul environnement aussi performant que celui défini manuellement a été trouvé. Cependant, il est équivalent à celui que nous avons créé manuellement excepté quelques genres jamais diffusés qui ont été déplacés de l'ensemble des genres ignorés à l'ensemble des classes.

1.2 Environnement utilisé

Afin de pouvoir définir manuellement des environnements d'apprentissage, nous avons implémenté un éditeur d'environnement basé sur la nomenclature de Médiamétrie.

L'environnement que nous avons utilisé est présenté dans l'annexe B (page 201). Il comporte trente-six classes et ignore deux éléments de la nomenclature de Médiamétrie.

2 Évaluation des arbres de régression

Dans cette section, nous évaluons les arbres de régression. L'arbre de régression optimal dans notre cas est un arbre de régression dont chaque feuille encadre bien les durées correspondantes, même pour de nouveaux exemples, et dont l'intervalle d'encadrement est le plus étroit possible. En effet, le but de notre méthode étant d'exécuter les détections (détection de logo, images monochromes et silences) localement dans le flux, des intervalles d'encadrement trop grands ne sont pas souhaitables. En revanche, si les intervalles sont trop étroits, les durées seront mal prédites.

2.1 Taux d'erreur global

Pour mesurer le taux d'erreur des arbres de régression, nous les avons entraînés sur les grilles de programmes des chaînes hertziennes de l'année 2004 avec le paramètre σ fixé à 300 secondes (voir section 3.1, page 115), puis nous les avons testés sur l'année 2005. Le tableau 14 présente les résultats obtenus sur les différentes chaînes hertziennes : nous considérons qu'une prédiction est bonne si la durée réelle est bien incluse dans l'intervalle prédit par l'arbre de régression. Les résultats obtenus sont très convaincants dans la mesure où ils sont supérieurs à 90% alors même que la base des grilles de programmes de 2005 n'a pas été filtrée ; on y retrouve, à titre d'exemple, des événements exceptionnels comme Roland Garros, pendant lequel la durée des matchs n'est pas stable, le Téléthon sur France 2 et d'autres émissions de ce genre.

	Nombre d'émissions testées	Taux de bonnes prédictions	Envergure moyenne
TF1	27 729	94,1 %	338 s
France 2	29 366	96,8 %	409 s
France 3	29 485	96,6 %	465 s
Arte	6 182	91,1 %	411 s
France 5	11 239	98,6 %	397 s
M6	26 729	92,1 %	384 s

TAB. 14: *Taux d'erreur et envergure des arbres de régression des chaînes hertziennes*

Le faible taux obtenu sur Arte est dû à la nature de la chaîne qui diffuse beaucoup de documentaires dans leur intégralité, contrairement aux chaînes (par exemple, TF1) qui diffusent partiellement des documentaires la nuit, donc de manière plus formatée du point de vue des durées. En effet, nous avons vu que les documentaires faisaient partie des genres pour lesquels il était délicat de prédire une durée (voir la section 2, page 66). À l'opposé, le taux de France 5 est le plus élevé car la chaîne diffuse peu d'émissions en direct et ne prend pas de retard dans la mesure où elle ne diffuse que peu d'interprogrammes. L'envergure moyenne, exprimée en secondes, est la taille moyenne de la fenêtre dans laquelle le système aura à rechercher le début

de l'émission suivante. On voit que la moyenne des envergures de ces fenêtres est assez faible, puisque les valeurs tournent autour de cinq à huit minutes.

2.2 Taux d'erreur par genres

Nous nous sommes également intéressés à déterminer expérimentalement les genres d'émission dont la durée était le plus facilement encadrable – c'est-à-dire dont la durée est la plus stable (voir la section 2, page 66). Pour cela nous avons utilisé les prédictions sur les six chaînes hertziennes gratuites. Le tableau 15 présente un classement des genres les plus prédictibles.

Genre télévisuel	Taux de bonnes prédictions	Envergure moyenne
Météo	100,0 %	285 s
Émissions de service	100,0 %	283 s
Loteries	99,8 %	258 s
Interprogrammes, clip	99,1 %	254 s
Journaux télévisés	98,9 %	335 s
Émissions culturelles	98,7 %	965 s
Interprogrammes	97,9 %	275 s
Séries	97,6 %	354 s
Jeux	95,6 %	395 s
Émissions de télé-achat	95,6 %	351 s
Magazines	95,1 %	657 s
Feuilletons	92,8 %	679 s
Émissions jeunesse	92,6 %	757 s
Émissions d'humour	92,5 %	447 s
Magazines politiques, économiques	92,4 %	368 s
Émissions à base de clips	92,4 %	918 s
Décrochages régionaux	91,6 %	1137 s
Magazines de sport	89,7 %	402 s
Films	88,6 %	565 s
Variétés	83,2 %	801 s
Musique classique	79,7 %	881 s
Courts-métrages	76,2 %	502 s
Fictions	74,6 %	1017 s
Documentaires	62,2 %	544 s
Cirques	51,7 %	341 s
Talk-shows	48,1 %	329 s
Pièces de théâtre	41,7 %	2736 s
Concerts modernes	35,0 %	1164 s
Évènements sportifs	26,3 %	989 s

TAB. 15: Classement des genres télévisuels en fonction de leur prédictibilité

Les envergures moyennes sont exprimées en secondes. Les quatre premiers genres de la liste sont des émissions pour lesquelles il est facile de prédire une durée et de l'encadrer précisément à moins de cinq minutes près. Il émerge également de ce tableau que les émissions enregistrées ont des durées suffisamment stables pour être encadrées assez étroitement tandis que les émissions en direct (retransmission sportive, talk-shows) ou indirectement destinées à la télévision (cirque, concerts, pièces de théâtre) nécessitent un encadrement plus large pour des résultats moins bons. Enfin, les chiffres confirment le problème des documentaires, que nous abordions plus haut, et qui s'explique par le fait que le même documentaire peut être diffusé plusieurs fois tout en ayant des durées différentes (voir section 1.4, page 62). De plus, nous avons vu que le documentaire est un genre qui n'est pas formaté.

2.3 Précision temporelle des prédictions

La précision temporelle d'une prédiction d'un arbre de régression est la distance entre la valeur moyenne renvoyée par la feuille et la valeur réelle si la durée a été correctement encadrée. Nous avons calculé la précision temporelle des prédictions sur l'ensemble des chaînes hertziennes (voir tableau 16).

Précision temporelle	Proportion des durées prédites avec cette précision
inférieure à 10 minutes	98,472 %
entre 10 et 20 minutes	0,770 %
entre 20 et 30 minutes	0,270 %
entre 30 et 40 minutes	0,187 %
entre 40 et 50 minutes	0,141 %
entre 50 et 60 minutes	0,055 %
entre 60 et 70 minutes	0,038 %
entre 70 et 80 minutes	0,027 %
entre 80 et 90 minutes	0,014 %
entre 90 et 100 minutes	0,006 %
entre 100 et 110 minutes	0,006 %
entre 110 et 120 minutes	0,005 %
entre 120 et 130 minutes	0,006 %
au-delà de 130 minutes	0,003 %

Précision temporelle	Proportion des durées prédites avec cette précision
inférieure à 1 minute	72,557 %
entre 1 et 2 minutes	18,427 %
entre 2 et 3 minutes	5,698 %
entre 3 et 4 minutes	0,777 %
entre 4 et 5 minutes	0,252 %
entre 5 et 6 minutes	0,187 %
entre 6 et 7 minutes	0,150 %
entre 7 et 8 minutes	0,176 %
entre 8 et 9 minutes	0,142 %
entre 9 et 10 minutes	0,106 %
au-delà de 10 minutes	1,528 %

TAB. 16: *Précision temporelle des arbres de régression sur les chaînes hertziennes*

Les résultats sont très encourageants car la quasi-totalité (98,5 %) des durées moyennes prédites sont à moins de dix minutes des durées réelles ce qui montre que la régression est assez efficace pour notre application. Le tableau 16 présente aussi les mêmes résultats à une autre échelle : il en ressort que le modèle est très précis puisque 72,6 % des durées moyennes sont à moins d'une minute des valeurs réelles.

2.4 Impact de σ sur les arbres de régression

σ représente l'envergure d'une fenêtre à partir de laquelle le prolongement d'une branche est arrêté. Ce paramètre d'arrêt de la construction d'une branche peut être déterminé expérimentalement ; dans notre cas, σ est une durée en secondes.

Ainsi, si σ est trop petit, l'arbre ne généralisera pas suffisamment les durées et échouera souvent sur de nouveaux exemples. En revanche, si σ est trop grand, le taux de prédiction

pourra atteindre 100 % mais les fenêtres temporelles seront très grandes. Il ne faut pas perdre de vue que notre but est d'avoir des fenêtres assez petites pour minimiser les temps de calcul mais suffisamment grandes pour être générales. Le paramètre σ permet ainsi de gérer la priorité entre la généralisation et le temps de calcul.

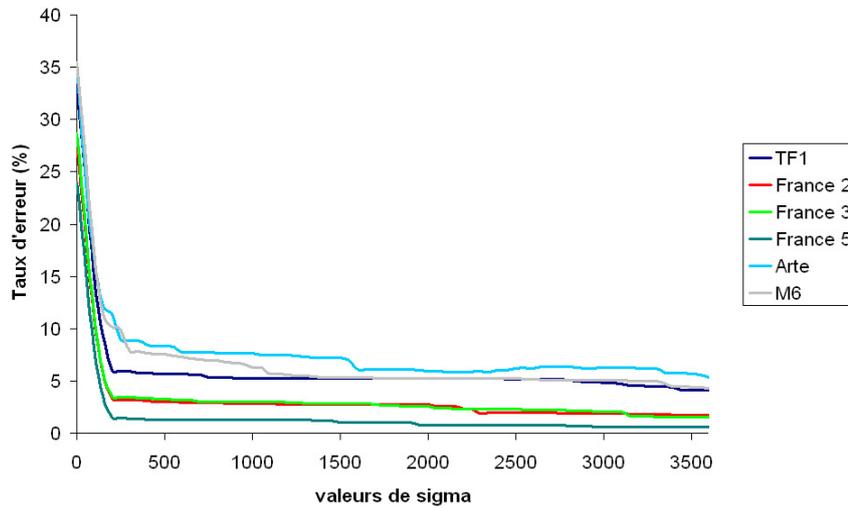


FIG. 35: Taux d'erreur en fonction du paramètre σ pour les chaînes hertziennes

La figure 35 montre que le taux d'erreur en fonction du paramètre σ décroît rapidement. On voit aussi que quelle que soit la chaîne, le paramètre θ « optimal » peut être sensiblement le même. Nous avons choisi de prendre un θ proche de 300 secondes, c'est-à-dire après la partie de la courbe à forte pente.

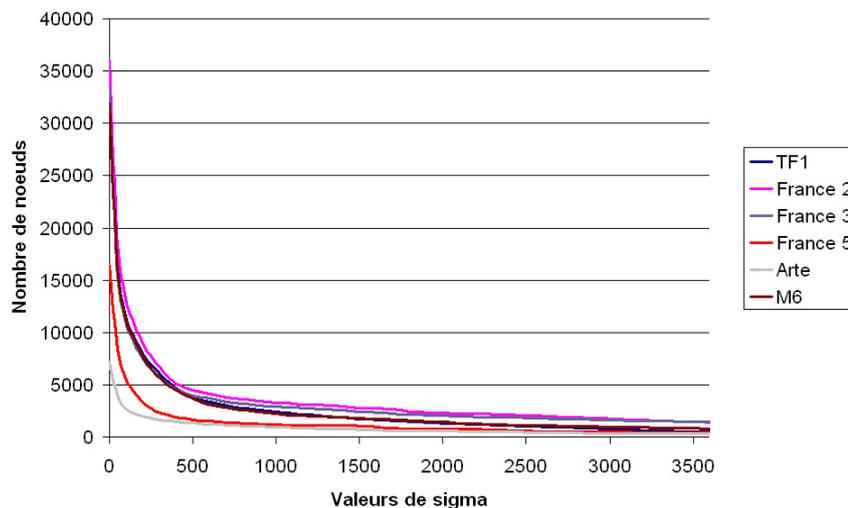


FIG. 36: Nombre de nœuds dans l'arbre de régression pour les chaînes hertziennes en fonction du paramètre σ

La figure 36 montre le nombre de nœuds dans les arbres de régression pour chacune des chaînes hertziennes en fonction de différentes valeurs du paramètre σ . On voit que le nombre de

nœuds suit une courbe asymptotique qui décroît quand σ augmente.

2.5 Nombre d'exemples d'apprentissage

Nous avons construit différents arbres de régression pour une même chaîne (France 2) en prenant plusieurs années d'émissions comme ensemble d'apprentissage : de 2003 à 2004, de 2002 à 2004, de 2001 à 2004 et enfin de 2000 à 2004. Les résultats ont révélé qu'une année d'émissions était largement suffisante et qu'en augmentant le nombre d'années, le taux d'erreur augmentait lentement. Cela est dû à l'ajout de bruit dans l'ensemble d'apprentissage. Prendre en compte des émissions trop vieilles ne rendraient pas forcément mieux compte des durées des émissions d'aujourd'hui puisqu'on peut vérifier que les épisodes des séries, par exemple, durent plus longtemps en moyenne aujourd'hui qu'il y a dix ans. De plus, certains formats n'existaient pas encore : par exemple, ces formats d'émissions particulièrement courts comme les séries humoristiques (« un Gars une fille », « Caméra café », « Avez-vous déjà vu ? »).

3 Évaluation de la prédiction des grilles

La prédiction des grilles de programmes est plus difficile à évaluer. Nous avons commencé par tester expérimentalement l'expressivité du CHMM, c'est-à-dire sa capacité à représenter des grilles de programmes de l'année 2005 sans l'aide du guide de programmes. Nous évoquerons ensuite l'apport du guide de programmes lorsque la journée n'est pas prédictible avec le CHMM seul. Enfin nous donnerons quelques exemples de graphes obtenus qui seront l'occasion de discuter des effets des heuristiques.

3.1 Capacité d'expression du CHMM seul

Compte tenu que l'algorithme d'apprentissage du CHMM repose sur la capacité de l'arbre de régression à prédire les durées, le taux d'erreur de l'arbre va avoir un impact sur les performances du CHMM.

Chaîne	Nombre de jours exprimables	Nombre de jours avec une diffusion exceptionnelle	Nombre de jours avec au moins une erreur
TF1	345 (94,6 %)	14 (3,8 %)	6 (1,6 %)
France 2	277 (75,9 %)	59 (16,2 %)	29 (7,9 %)
France 3	255 (69,9 %)	36 (9,9 %)	74 (20,2 %)
Arte	341 (93,4 %)	14 (3,8 %)	10 (2,8 %)
France 5	338 (92,6 %)	23 (6,3 %)	4 (1,1 %)
M6	335 (91,8 %)	20 (5,5 %)	10 (2,7 %)

TAB. 17: Nombre de journées entières de 2005 exprimables à l'aide du CHMM entraîné sur 2004

Le tableau 17 résume le nombre de jours qui ont été entièrement exprimés à l'aide du CHMM. Nous avons considéré chaque jour de 2005 séparément (c'est-à-dire 365 jours) et nous avons calculé la probabilité de ces séquences d'émissions à l'aide du CHMM : si la probabilité est nulle, la journée n'est pas exprimable. L'évaluation du pourcentage de réussite est assez pessimiste dans la mesure où nous ne considérons pas ici les journées partiellement exprimables : une seule émission imprédictible entraîne l'invalidation de la journée entière. Les nombres de journées exprimables par le CHMM sont donc les journées où d'une manière ou d'une autre, on peut retrouver la structure du flux sans nécessiter l'intervention d'un opérateur. Nous verrons dans la quatrième

partie combien de ces journées peuvent être structurées entièrement automatiquement. Nous avons choisi de considérer séparément les erreurs dues à la diffusion d'une émission concernant un événement d'actualité.

Le tableau 17 montre aussi que le nombre de journées qui ne peuvent pas être exprimées par le CHMM est dépendant de la chaîne de télévision. En effet, il est visible que France 2 et France 3 sont plus difficilement prédictibles qu'une chaîne comme TF1, aussi surprenant que cela puisse paraître. Néanmoins, l'explication est assez simple : de par leur mission de service public, France 2 et France 3 ont tendance à diffuser des événements spéciaux plus fréquemment que les autres chaînes.

L'année 2005 a été marquée par plusieurs événements qui ont bousculé les grilles de programmes : la mort du Pape Jean-Paul II et du Prince Rainier III en avril, le référendum pour la constitution européenne en mai, la libération de la journaliste Florence Aubenas en juin, les attentats (à Londres et à Charm-El-Cheikh en juillet), la candidature de Paris à l'organisation des Jeux Olympiques de 2012 et les problèmes dans les banlieues en novembre. Ces actualités sont les causes des erreurs présentées dans la deuxième colonne du tableau 17, dans la mesure où ils ont perturbé les grilles de programmes des chaînes par la déprogrammation des émissions habituelles pour la diffusion de flashes d'information ou d'émissions spéciales ou encore par l'ajout de messages politiques sur les chaînes publiques pour le référendum.

Toutefois, ces événements spéciaux ne sont pas les seules sources d'erreurs. Les erreurs strictement dues à une erreur de modèle sont indiquées dans la troisième colonne du tableau 17. Une fois encore, les résultats sont bruts et nécessitent des explications. Dans les données de Médiamétrie, des incidents techniques sont représentés sous la forme d'une émission du type « AUTRE » (voir annexe A, page 197). Ces erreurs sont plus ou moins fréquentes selon les chaînes : elles concernent 2 journées de TF1, 18 journées de France 2, 74 journées de France 3, 10 journées d'Arte, 4 journées de France 5 et 8 journées de M6. On voit donc que nos résultats saut considérablement faussés par ces incidents techniques. Ensuite, nous avons cherché la cause des autres erreurs qui sont réellement des erreurs de prédiction :

- sur TF1, les 4 erreurs restantes sont dues à des transitions entre un concert classique et un documentaire nocturnes qui n'avaient jamais été vues en 2004 ;
- sur France 2, 9 erreurs sont dues à la retransmission de Roland Garros dont la durée des matchs est difficilement prédictible, ce qui change les transitions vers les émissions suivantes ; toutefois, 5 des journées de Roland Garros ont été correctement prédites, sans doute parce qu'elles étaient similaires à des journées de l'année précédente ; les deux erreurs restantes sont dues au Téléthon ;
- les erreurs restantes de M6 sont dues à la diffusion de 2 programmes courts quotidiens multidiffusés de 36 secondes environ intitulés *Bonne fête*.

Chaîne	5h30 - 13h	13h - 14h	14h - 18h	18h - 20h
TF1	358 (98,1 %)	359 (98,3 %)	365 (100 %)	363 (99,4 %)
France 2	333 (91,2 %)	361 (98,9 %)	357 (97,8)	341 (93,4 %)
France 3	339 (92,9 %)	359 (98,3 %)	347 (95,0 %)	334 (91,5 %)
Arte	-	-	-	356 (97,5 %)
France 5	360 (98,6 %)	359 (98,3 %)	361 (98,9 %)	358 (98,1 %)
M6	353 (96,7 %)	361 (98,9 %)	358 (98,1 %)	357 (97,8 %)

Chaîne	20h - 21h	21h - 1h	nuît
TF1	358 (98,1 %)	363 (99,4 %)	357 (97,8 %)
France 2	355 (97,3 %)	356 (97,5 %)	353 (96,7 %)
France 3	356 (97,5 %)	340 (93,1 %)	354 (97,0 %)
Arte	365 (100 %)	356 (97,5 %)	359 (98,3 %)
France 5	-	-	358 (98,1 %)
M6	361 (98,9 %)	362 (99,2 %)	363 (99,4 %)

TAB. 18: Nombre de tranches horaires entières de 2005 exprimables à l'aide du CHMM entraîné sur 2004

Le tableau 18 montre cette fois-ci le nombre de tranches horaires correctement prédites en considérant les tranches horaires suivantes :

- de 5h30 à 13h, pendant laquelle la grille de programmes est horizontale (voir la section 1, page 43) ;
- de 13h à 14h, qui représente une tranche horaire dédiée à l'information ;
- de 14h à 18h, pendant laquelle la grille de programmes est toujours horizontale ;
- de 18h à 20h qui peut être vue comme l'*access prime-time* ;
- de 20h à 21h, qui représente une nouvelle tranche horaire dédiée à l'information ;
- de 21h à 1h, pendant laquelle la grille est cette fois-ci verticale ;
- enfin, la nuit, où la grille est un enchaînement de programmes permettant de rattraper le retard accumulé dans la journée.

Contrairement à ce que l'on pouvait penser, notre modèle n'est pas sensible à la tranche horaire nocturne. Ce sont en effet les tranches horaires de grande écoute, comme le matin et l'*access prime-time* qui vont être le plus perturbées. La tranche horaire entre 20 heures et 21 heures est la plus stable. Quant aux incidents techniques, leurs occurrences sont totalement aléatoires : aucune tranche horaire n'est plus perturbée qu'une autre.

3.2 Fusion des guides de programmes et des prédictions du CHMM

Si certaines grilles ne peuvent pas être directement prédites par le CHMM, elles peuvent néanmoins l'être partiellement : nous avons préconisé dans le chapitre précédent le signalement des zones d'incertitude autour du programme annoncé dans le guide mais imprédictible par le modèle.

Le nombre de journées exprimables en tenant compte des guides de programmes change peu. Il est important de comprendre que le but de l'utilisation du guide de programmes n'est pas d'augmenter le nombre de journées exprimables mais de diminuer le nombre de chemins dans le

graphe des grilles possibles.

Lorsqu'un programme est annoncé dans le guide de programmes sans qu'il puisse être prédit par le CHMM, une zone d'incertitude est créée, partant du nœud du graphe représentant le dernier programme annoncé et correctement prédit jusqu'au prochain programme annoncé et correctement prédit. Dans cette zone, nous ne sommes pas en mesure de savoir si le programme ajouté est entouré ou non d'interprogrammes, ou encore si ce programme en remplace un autre. Si le programme est simplement inséré dans la grille, sans interprogramme, alors la prédiction sera bonne. Nous avons vu quelques cas isolés sur France 2 où un programme de quelques minutes – pour inciter les citoyens français à voter au référendum – a été inséré sans interprogramme.

3.3 Exemple de construction d'un graphe

L'emploi d'heuristiques (voir la section 2, page 125) s'est avéré incontournable lorsque nous avons tenté de prédire les grilles de programmes possibles d'une journée de TF1. Plus de vingt-quatre heures de calculs sur un PC standard ont été nécessaires au développement du graphe jusqu'au quatrième programme du guide. Une semaine après, le cinquième programme n'avait toujours pas été atteint. Les deux heuristiques ont pour but d'élaguer des branches du graphe des grilles possibles afin d'éviter sa croissance exponentielle. En effet, puisque nous considérons qu'en moyenne il y a trois transitions possibles par état (voir section 3.4), l'arbre ainsi construit aura 3^p nœuds à la profondeur p .

Nous allons présenter un exemple de fonctionnement de l'algorithme de prédiction de la grille de programmes pour une journée de télévision. Nous nous intéressons seulement à une partie du 10 mai 2005 et à un extrait du guide de programmes présenté dans le tableau 19.

Heure	Titre	Genre
6:45	Journal télévisé du matin	Journal télévisé
6:50	TFOU!	Émission jeunesse
8:35	Télésopping	Télé-achat
9:25	Julia Corsi, commissaire	Série

TAB. 19: Extrait du guide de programmes du 10 mai 2005 sur TF1

La figure 37 représente un extrait du graphe acyclique dirigé représentant l'ensemble des grilles de programmes possibles pour la journée du 10 mai 2005 sur TF1, à partir du journal du matin (vers 6h45) et jusqu'à la première série (vers 9h25). Le chemin représenté en gras est le chemin qui correspond à la succession d'émissions réelle. On peut voir sur cette figure que la branche la plus à gauche est absurde puisque constituée d'un chemin qui passe par une météo, un interprogramme, une nouvelle météo et un nouvel interprogramme. De telles branches sont inévitables si on se limite à un modèle de Markov d'ordre 1 ; elles sont de plus très coûteuses en temps de calcul car il peut s'agir d'une succession de programmes courts, et la séquence ainsi formée peut être très longue avant de dépasser le retard de 30 minutes autorisé par le seuil Δ .

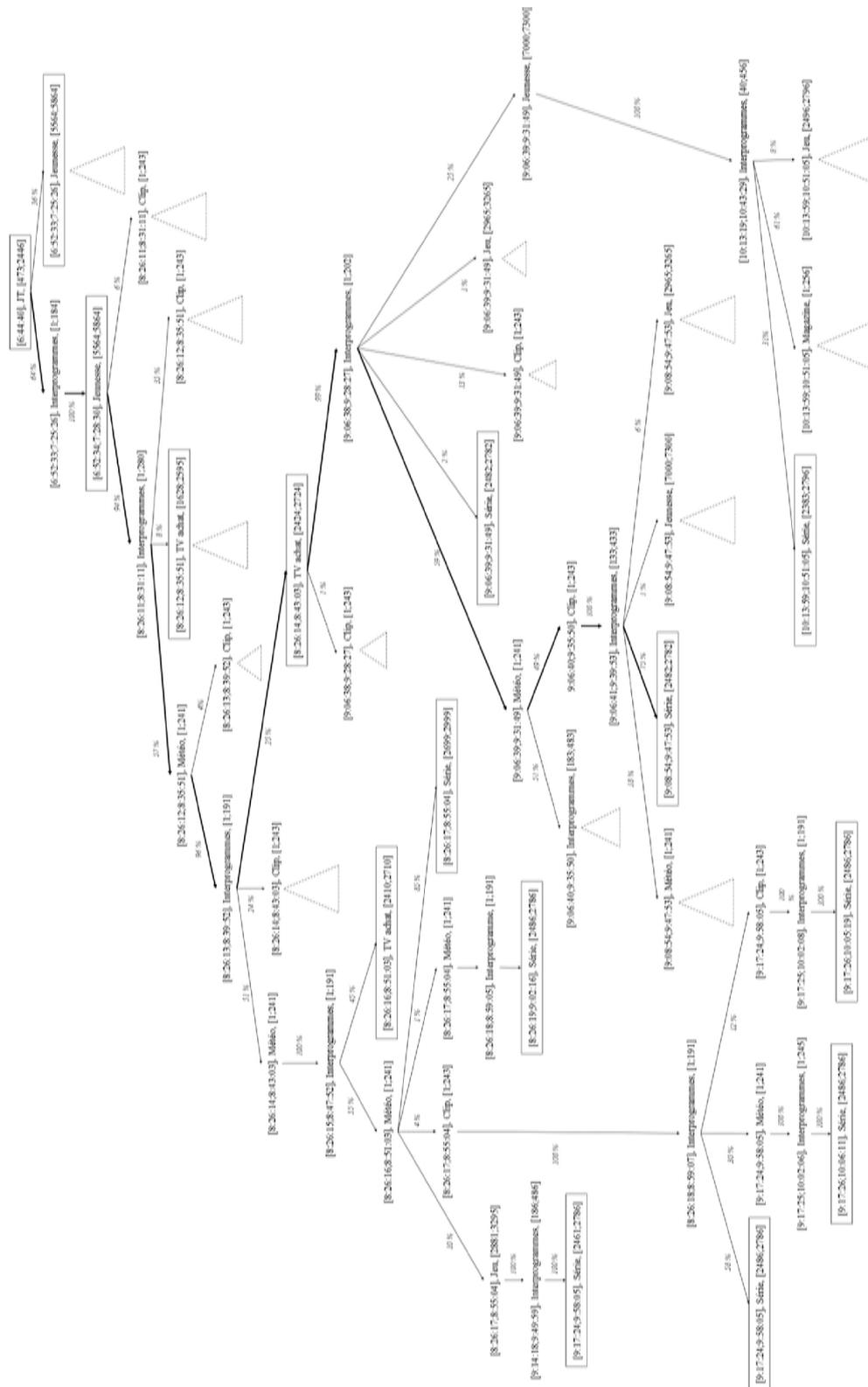


FIG. 37: Extrait du graphe acyclique dirigé représentant les grilles de programmes possibles du 10 Mai 2005 sur TF1

La figure 37 permet de se rendre compte que le graphe des grilles possibles possède un grand nombre de nœuds pour représenter les grilles de programmes passant par les quatre premiers programmes du guide. De plus, cette figure montre qu'il y a une certaine redondance dans l'information portée par ces nœuds puisqu'une même émission peut être représentée par plusieurs nœuds.

Nous allons maintenant développer le graphe des grilles possibles avec la méthode itérative (voir section 1.2, page 124). Dans notre exemple (voir figure 38), le journal du matin commence à 6:44:40 (six heures, quarante-quatre minutes et quarante secondes). L'arbre de régression propose une durée comprise entre 473 secondes et 2446 secondes. Ceci est un bon exemple de sur-généralisation des arbres de régression : le journal du matin dure généralement 8 minutes environ mais il a dû arriver en 2004 qu'il dure environ 41 minutes pour couvrir un événement spécial. Le graphe de la figure 38 comporte deux chemins possibles pour arriver à une émission pour les enfants.

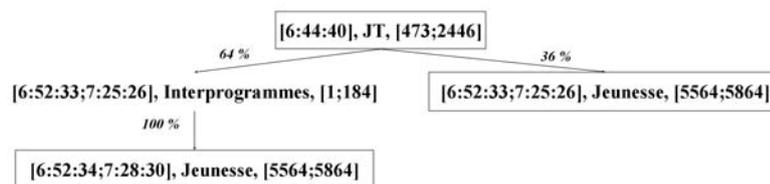


FIG. 38: Première étape : construction du graphe entre les deux premières émissions du guide de programmes

Il est possible d'appliquer sur le graphe de la figure 38 la deuxième heuristique (voir section 2, page 125) qui consiste à fusionner les nœuds du graphe représentant la même émission. Les deux extrémités de ce graphe correspondent en effet à la même émission cible fournie par le guide de programmes. Le résultat de l'application de cette heuristique est montré dans la figure 39.

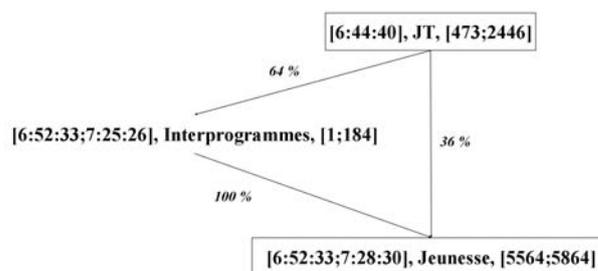


FIG. 39: Deuxième étape : application de la deuxième heuristique sur le graphe

L'algorithme calcule à présent tous les chemins allant du nœud résultant de l'application de la deuxième heuristique jusqu'au troisième programme du guide : l'émission de télé-achat. Les possibilités sont beaucoup plus nombreuses (voir figure 40). Cependant, certaines branches ne conduisent pas à une émission de télé-achat passé le délai Δ fixé à 30 minutes : elles peuvent donc être élaguées (voir figure 41). En effet, certains chemins passent par un nœud représentant la série télévisée de 9h25 sans passer par l'émission de télé-achat. La première heuristique que nous employons consiste justement à supprimer ces branches, considérant que le télé-achat doit être rencontré avant la série télévisée. La figure 42 montre le graphe qui résulte de l'application de la première heuristique.

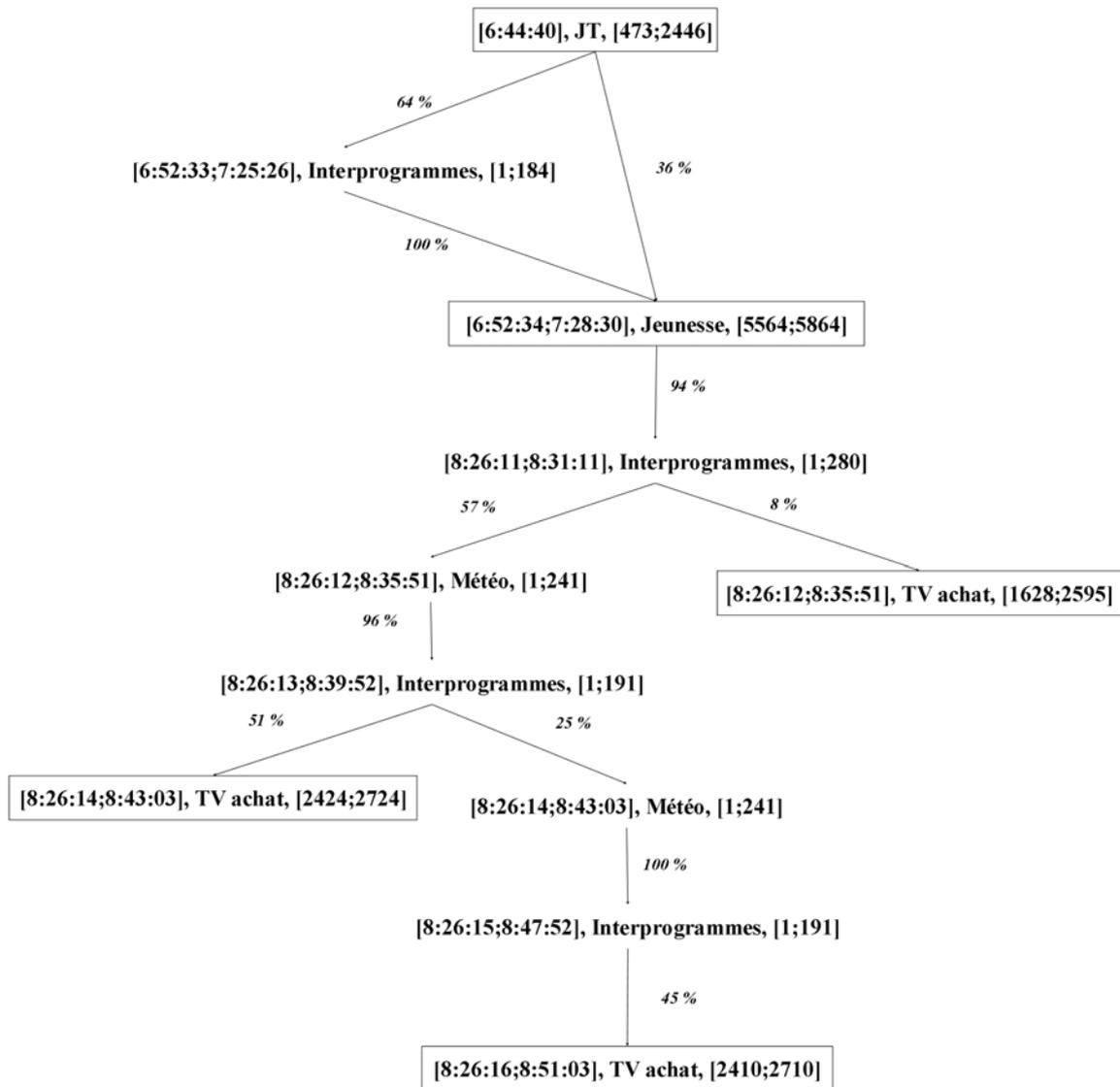


FIG. 42: Cinquième étape : application de la première heuristique sur le graphe

Ensuite, il est à nouveau possible d'appliquer la seconde heuristique afin de n'obtenir qu'un nœud correspondant à la troisième émission du guide de programmes (figure 43).

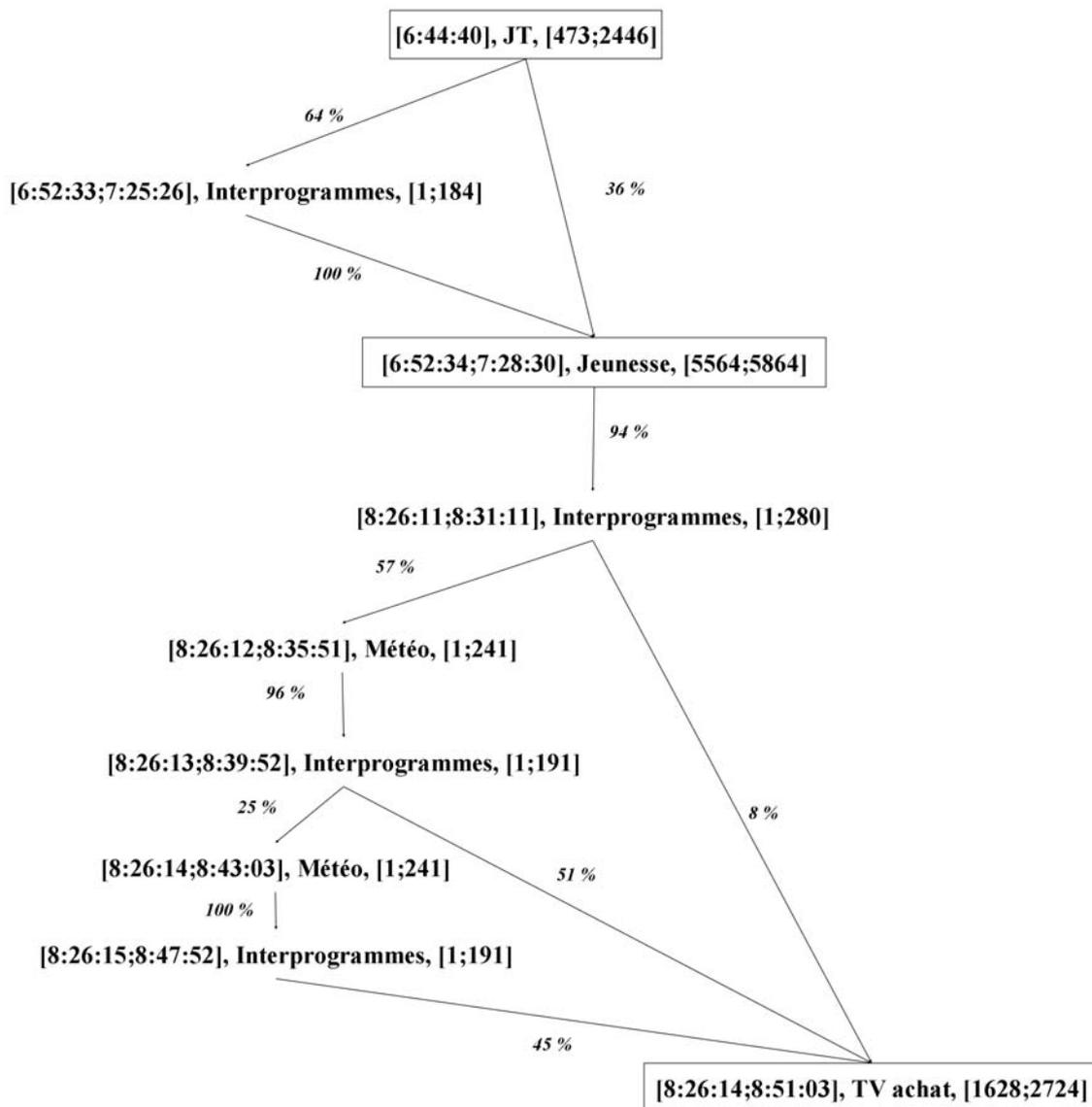


FIG. 43: Sixième étape : application de la deuxième heuristique sur le graphe

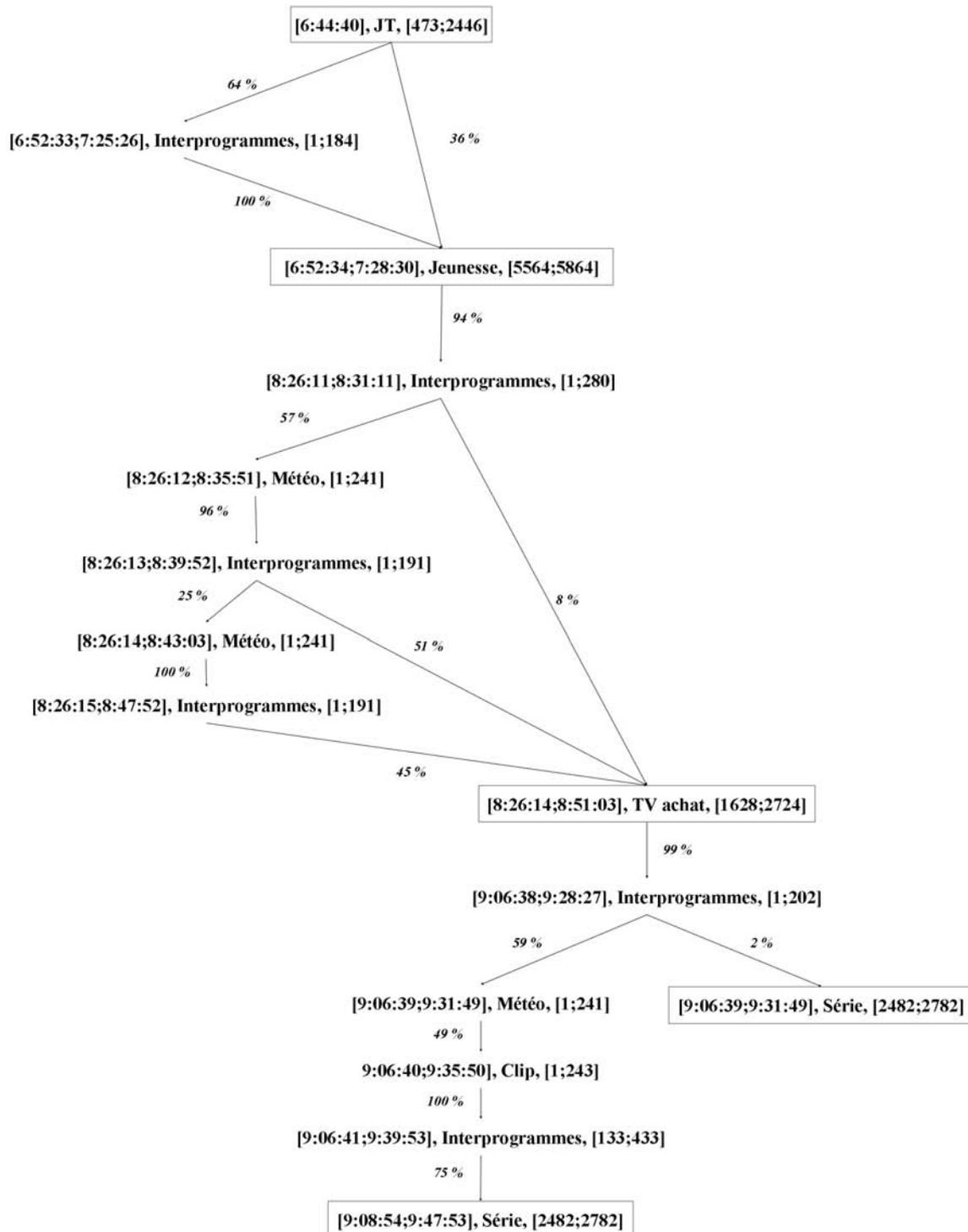


FIG. 45: Huitième étape : application de la première heuristique et élimination des chemins impossibles

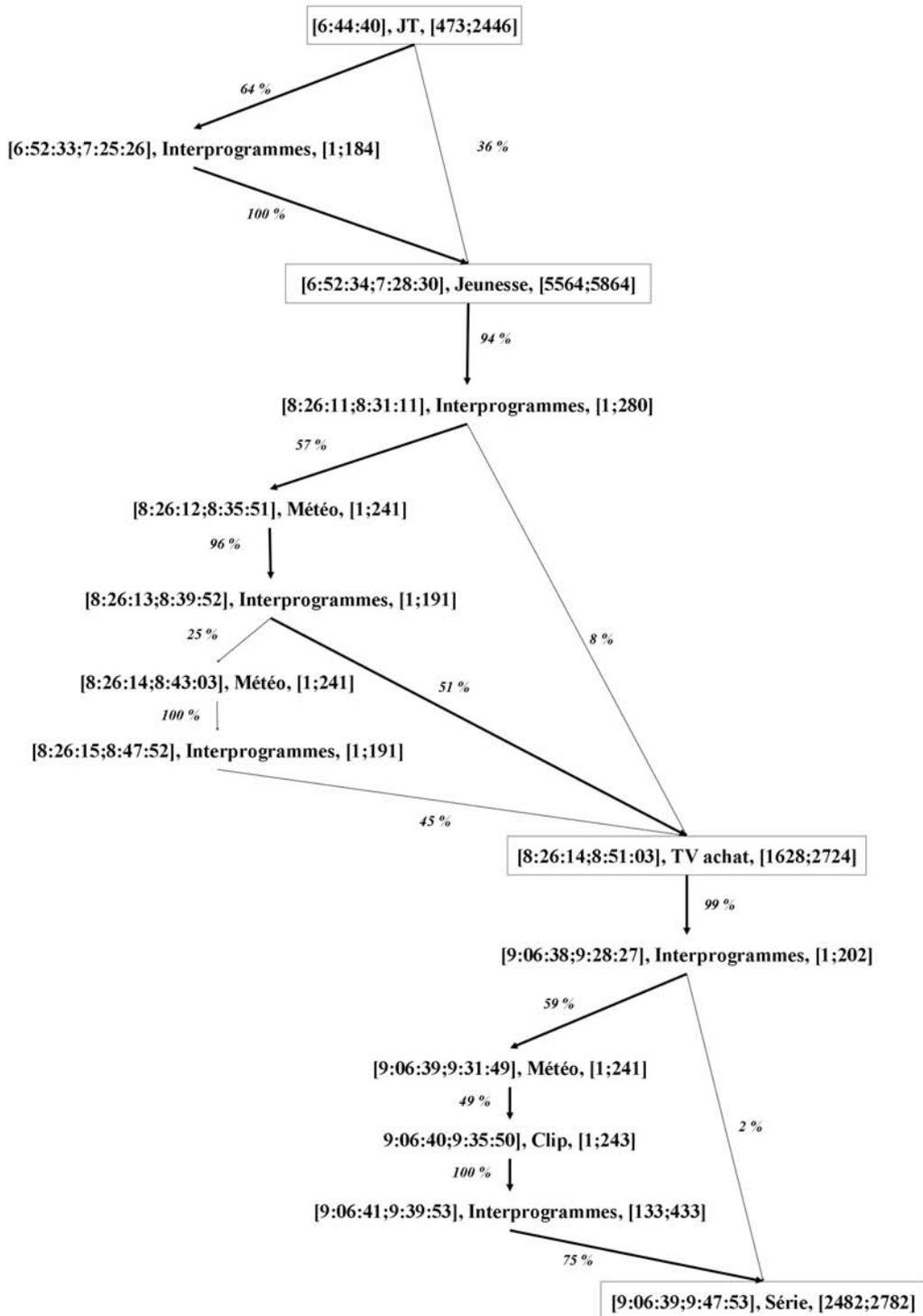


FIG. 46: Dernière étape : application de la deuxième heuristique sur le graphe

L'algorithme poursuit la croissance du graphe afin d'atteindre la série en quatrième position du guide de programmes (figure 44). Les possibilités sont nombreuses (plus de 10 branches différentes), mais après élagage des branches impossibles et application de la première heuristique, nous obtenons le graphe simplifié de la figure 45.

Enfin, la figure 46 présente le graphe obtenu à la fin de l'algorithme, après application de la seconde heuristique. Le chemin en gras est le chemin correspondant à l'enchaînement réel des émissions pour la journée du 10 mai 2005 sur TF1.

Notons aussi qu'après élimination de certaines branches, les probabilités ne sont pas mises à jour. Ainsi, les figures 41 à 46 peuvent surprendre dans la mesure où la somme des probabilités des branches issues d'un nœud n'est pas égale à 1. La réestimation des durées n'est pas nécessaire en pratique dans la mesure où les branches sont parcourues par ordre de probabilité décroissante, et la réestimation de ces probabilités n'en changera pas l'ordre.

L'emploi des heuristiques et de la méthode itérative de construction du graphe nous permet donc de réduire le nombre de nœuds nécessaires pour représenter les différentes séquences d'émissions possibles pour une journée donnée.

3.4 Effets de la contextualisation sur le nombre moyen de transitions possibles

La contextualisation a pour effet de diminuer le nombre de transitions possibles d'un état du CHMM à l'autre dans un contexte donné. Nous avons pris chaque état dans chaque contexte afin de compter le nombre possible de transitions. Nous avons ainsi mesuré sur un CHMM – entraîné sur les données de TF1, France 2, France 3 et M6 en 2004 – que le nombre moyen de transitions possibles dans un CHMM est égal à trois, pouvant aller de une à six transitions. À titre de comparaison, nous avons entraîné une chaîne de Markov d'ordre un sur les mêmes données et nous avons compté le nombre moyen de transitions possibles par état : il s'élève cette fois à dix-neuf, pouvant varier d'une transition à trente-quatre (pour l'état représentant la publicité).

Ainsi, la contextualisation nous permet de diminuer le nombre de grilles possibles lors de la construction du graphe.

3.5 Impact du modèle sur la recherche des transitions

Le graphe dirigé acyclique représentant les grilles possibles pour une journée d'une chaîne donnée nous permet de naviguer dans le flux afin de trouver les ruptures entre deux émissions en calculant une fenêtre temporelle qui se déplace au sein du flux. Cette fenêtre temporelle est formée par la durée minimale et la durée maximale prédites.

Que la détection des ruptures soit manuelle ou automatique, notre approche permet de réduire le nombre de secondes qu'il est nécessaire d'observer pour trouver la transition entre deux programmes. Nous avons simulé le déroulement de notre système en tentant de prédire les grilles de programmes de chacune des chaînes et pour chacun des jours dont la grille est prédictible par le modèle : pour cela, nous avons parcouru le graphe en commençant par les branches de plus fortes probabilités et additionné les longueurs des encadrements temporels associés à chaque nœud parcouru. Ainsi, nous avons été en mesure de calculer le nombre de secondes qu'il est nécessaire d'observer pour structurer le flux :

- si on ne considère que la partie parcourue de ces fenêtres, c'est-à-dire de la durée minimale à la durée réelle, seuls 29% du flux sont observés ;

- si on part de la durée moyenne et qu'on progresse parallèlement vers les durées extrémales, 28% du flux suffisent à la recherche des transitions entre les émissions.

La seconde méthode n'est pas forcément praticable car il faut être sûr de pouvoir lire le flux de la durée moyenne vers la durée minimale, donc à l'envers. Quelle que soit la méthode, la complexité moyenne du parcours d'une fenêtre temporelle est de l'ordre de $O(\frac{L}{2})$ où L est la longueur de la fenêtre.

Finalement, la seconde expérimentation montre bien que le modèle nous permet d'effectuer des calculs localement, à la hauteur de 29% du flux.

4 Conclusion

Dans ce chapitre, nous avons présenté dans un premier temps les différentes expérimentations que nous avons menées sur les arbres de régression appliqués à la régression des durées des émissions. Il en ressort des taux d'erreur assez faibles. Cependant cette méthode de régression nécessite environ une année d'émissions afin de pouvoir prédire correctement de nouveaux cas.

Le CHMM, tout en éliminant des séquences d'émissions impossibles, permet de prédire une grande partie des grilles de programmes des chaînes. La prise en compte des guides de programmes permet de créer des zones d'incertitude autour d'un programme exceptionnel et d'atténuer ainsi les problèmes posés par ce genre de programmes.

Nous verrons qu'une des perspectives de cette thèse est de maintenir à jour les probabilités du CHMM et les arbres de régression afin que le système puisse « capturer » au mieux les nouvelles tendances des chaînes de télévision.

BILAN DE LA TROISIÈME PARTIE

Dans cette partie, nous avons proposé une nouvelle extension des modèles de Markov appelée modèle de Markov caché contextuel. Nous avons appliqué ce modèle à la modélisation des grilles de programmes d'une chaîne de télévision.

Dans cette modélisation, la probabilité d'une durée d'émission est estimée par un arbre de régression. Les arbres de régression sont en effet adaptés à notre problème grâce à leur prise en compte de données symboliques et continues.

En fusionnant les observations produites par le modèle et les guides de programmes, il est possible de calculer l'ensemble des grilles de programmes possibles pour une journée sous la forme d'un graphe acyclique orienté.

Le graphe ainsi obtenu permet au système de procéder à des détections localement dans le flux afin de se déplacer dans le graphe jusqu'à ce que tout le flux soit structuré.

À ce stade de notre recherche, nous avons donc un système semi-automatique de structuration qui permet à un opérateur de naviguer dans le flux en se rendant dans les portions du flux susceptibles de contenir une rupture entre deux programmes.

Nous nous proposons dans la partie suivante d'automatiser cette recherche de ruptures.

Quatrième partie

**Expérimentation du système de
structuration**

RÉSUMÉ

Dans la partie précédente, nous avons proposé une modélisation des grilles de programmes d'une chaîne afin d'être en mesure de calculer l'ensemble des séquences d'émissions possibles sous la forme d'un graphe dirigé acyclique. Les guides de programmes sont utilisés conjointement avec le modèle dans le but de supprimer les branches impossibles dans le graphe, et également pour pré-annoter les émissions en utilisant les titres et les résumés qu'ils proposent.

Dans cette partie, nous proposons de détecter automatiquement les transitions entre deux émissions ; pour cela nous distinguons les interprogrammes (publicité, bandes-annonces et autopromotion) des autres programmes. Nous nous proposons aussi de vérifier que certains programmes annoncés dans le guide ont bien été diffusés. Nous nous intéressons ensuite à la structuration d'une tranche horaire matinale en utilisant la méthode proposée pour les flux et à l'extraction des plateaux d'un journal télévisé par une méthode de regroupement d'images clés.

Introduction

Dans la partie précédente, nous avons décrit un moyen pour enrichir statistiquement les informations apportées par les grilles prévisionnelles. Nous avons employé un modèle de Markov contextualisé qui permet de décrire à la fois les enchaînements possibles entre les différents genres d'émissions (par ses probabilités de transition d'un état à un autre) et les durées des émissions (par ses probabilités d'émission d'un symbole). La probabilité des durées est estimée à partir de la moyenne et des valeurs extrêmes fournies par un arbre de régression.

Cette modélisation permet de tirer parti de connaissances complémentaires puisque le modèle statistique est entraîné sur les structurations des années précédentes et que la prédiction du modèle de Markov contextuel est guidé par les guides de programmes, qui fournissent en plus des informations sur les émissions. Les avantages de cette modélisation sont doubles. La régression de la durée des émissions permet de définir des fenêtres temporelles dans lesquelles les détecteurs seront appliqués, ce qui permet d'une part d'éviter certaines fausses alarmes et d'autre part de réduire le temps de calcul en observant qu'une partie du flux (voir la section 3.5, page 148). De plus, la connaissance des transitions permet d'appliquer des règles précises et adaptées; autrement dit, cela permet d'employer uniquement les détecteurs adéquats à la recherche d'une frontière entre deux émissions d'un certain genre. Par exemple, quelle que soit l'émission qui précède une coupure publicitaire, il suffit de détecter le jingle publicitaire. La figure 47 illustre la structuration d'une matinée entre un journal télévisé et une émission pour enfants.

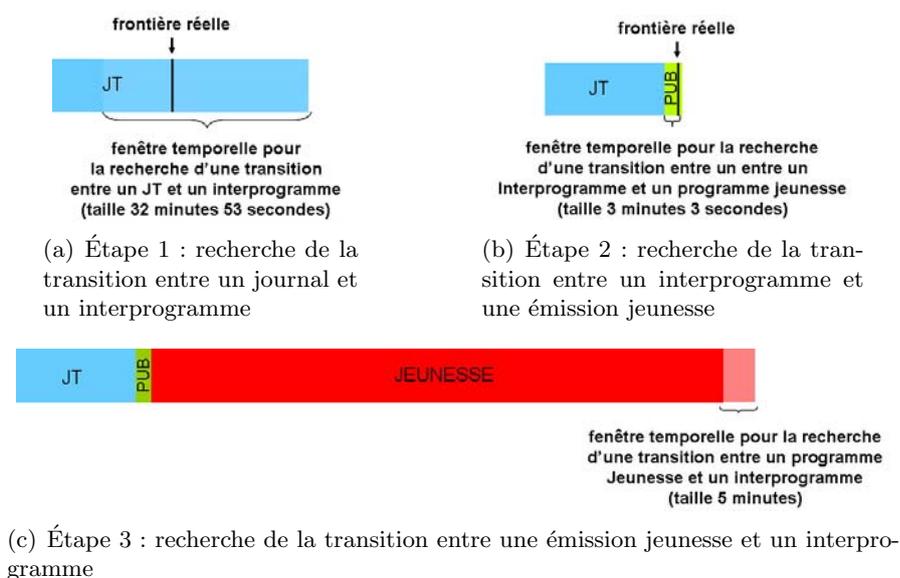


FIG. 47: Structuration du flux télévisuel par suivi d'un chemin dans le graphe prédit

Dans le premier chapitre, nous nous intéressons à la recherche de la transition entre deux émissions en l'effectuant uniquement dans les fenêtres temporelles prédites. Pour cela, nous allons considérer des règles simples, comme par exemple :

- lorsqu'une émission est suivie d'un interprogramme, chercher soit la marque d'une bande-annonce, soit la marque d'une coupure publicitaire (jingle) ;
- lorsqu'une émission débute par un générique, chercher ce générique ;
- ...
- sinon chercher une « rupture » dans le flux.

De plus, la structuration d'un flux télévisuel, qu'elle soit manuelle ou automatique, est un processus hiérarchique : le flux est décomposé en unités de plus en plus fines. Dans les parties précédentes, nous avons divisé le flux en trois types de composantes :

- les interprogrammes ;
- les émissions indépendantes ;
- les tranches horaires.

En découpant le flux en fonction de ces trois catégories de segments, nous obtenons bien une partition de celui-ci, c'est-à-dire qu'il n'y a pas de recouvrement entre deux segments consécutifs et que tous les segments sont adjacents.

Cependant, ces éléments peuvent eux aussi être structurés. Les émissions telles que les retransmissions de matchs peuvent être décomposés en jeux (pour le tennis), en mi-temps (pour le football), en actions. De nombreuses méthodes existent pour cela et ont été décrites dans le chapitre 2 (page 13). Les émissions de plateaux peuvent être découpées en plateaux, performances et reportages. De même, les publicités, les jingles et les bandes-annonces peuvent être séparées au sein d'un interprogramme. Une fois les bornes de l'interprogramme trouvées, sa structuration n'est donc plus un problème.

En revanche, la structuration des tranches horaires n'a jamais été étudiée. Elles sont pourtant très présentes dans un flux télévisuel. Une tranche horaire est délimitée par un générique de début et un générique de fin. Elle est ensuite composée de plusieurs émissions, généralement unies par un même genre ou une même thématique. L'habillage d'une tranche horaire est très souvent caractéristique ; par exemple, un logo peut apparaître et les jingles publicitaires sont modifiés. Les émissions d'une tranche horaire peuvent aussi être articulés par un plateau ou un présentateur. À titre d'exemples, voici une liste de quelques tranches horaires :

- *Ciné Dimanche* sur TF1 est une tranche horaire dédiée au cinéma, composée d'un générique, d'un sommaire, et d'un ou deux films séparés par le *Journal des sorties* ;
- la tranche horaire religieuse de France 2 le dimanche matin est elle aussi annoncée par un générique et chaque documentaire, magazine ou messe est annoncé par un plateau ;
- *la Trilogie du samedi soir* sur M6 propose trois (ou quatre) épisodes de séries télévisées séparées par des cartons ;
- *le Morning Café*.

Il est donc important de considérer les tranches horaires au premier niveau hiérarchique de la structuration du flux télévisuel. En effet, si l'on ne considère que les émissions des tranches horaires, que faut-il faire des génériques, des sommaires et des articulations? La difficulté de leur structuration est un peu similaire à celle du flux : par rapport à une émission, la tranche horaire est constituée de plusieurs parties hétérogènes.

Nous proposons d'étudier dans le chapitre 2 la structuration d'une tranche horaire matinale, le *Morning Café*, et l'extraction des plateaux dans un journal télévisé.

Chapitre 11

Alignement des grilles prédites sur le flux

Sommaire

1	Détection des transitions	159
1.1	Détection des transitions entre deux programmes	159
1.2	Détection des jingles publicitaires	164
1.3	Choix de la transition parmi les transitions candidates	165
1.4	Discussion	167
1.5	Conclusion	168
2	Identification d'un programme	168
2.1	Description de la méthode	169
2.2	Résultats expérimentaux	171
3	Conclusion	172

1 Détection des transitions

Dans cette section, nous nous proposons de détecter la transition entre deux émissions. Dans un premier temps, nous étudions quelques détecteurs simples : il s'agit d'un détecteur de silence, d'un détecteur d'images monochromes et enfin d'un détecteur de « cuts » (voir section 1.1). Ensuite, nous proposerons une méthode simple pour détecter le début et la fin des interprogrammes. Nous désignons, dans le reste de cette partie, la transition entre émission par le terme « rupture » et nous allons vérifier si de tels détecteurs permettent de reconnaître ces ruptures.

Nous présentons dans un premier temps les méthodes de détection d'images monochromes, de « cuts », puis de silences dans le flux télévisuel. Nous étudierons ensuite leur présence dans le flux issu d'une semaine de France 2. Le but de cet étude est de savoir si ces ruptures suffisent à séparer des émissions.

1.1 Détection des transitions entre deux programmes

Les ruptures d'un flux télévisuel peuvent être détectées par l'application conjointe d'un détecteur d'images monochromes ou d'un détecteur de « cuts » et d'un détecteur de silence. Nous

Jours	Nombre de programmes correctement localisés	Nombre de fausses alarmes
Samedi	100% (sur 70)	28356
Dimanche	100% (sur 67)	24177
Lundi	100% (sur 90)	25552
Mardi	100% (sur 81)	25926
Mercredi	100% (sur 77)	26322
Jeudi	100% (sur 93)	24970
Vendredi	100% (sur 95)	25543

TAB. 20: *Évaluation de la détection des « cuts » pour la localisation de programmes*

allons décrire ces trois détecteurs puis la fusion de leurs résultats. Étant donné que nous paramétrons ces algorithmes de façon à privilégier les fausses alarmes par rapport aux oublis, il faut s’attendre à ce que la précision de ces détecteurs soit très faible et que le rappel soit plutôt élevé.

Détection de « cuts »

Le « cut » est la transition entre deux plans la plus simple à détecter (voir section 1.1, page 14). Nous avons choisi de calculer l’histogramme basé sur la luminosité de chacune des images et de comparer les histogrammes deux à deux. Si la différence entre deux histogrammes est supérieure à un certain seuil, un « cut » est détecté. Nous avons fixé le seuil sur une vidéo de quelques heures de France 2, qui n’appartient pas à notre corpus de test, de façon à ce que tous les débuts de plans soient détectés, sans tenir compte du nombre de fausses alarmes.

Nous avons remarqué, lors de nos expérimentations, que le détecteur que nous utilisons trouve un « cut » même lorsque la transition entre deux plans est progressive (voir figure 5, page 15) ; en revanche, le début du plan ne sera pas précisément localisé. En général, le nouveau plan est détecté à la fin de la transition progressive.

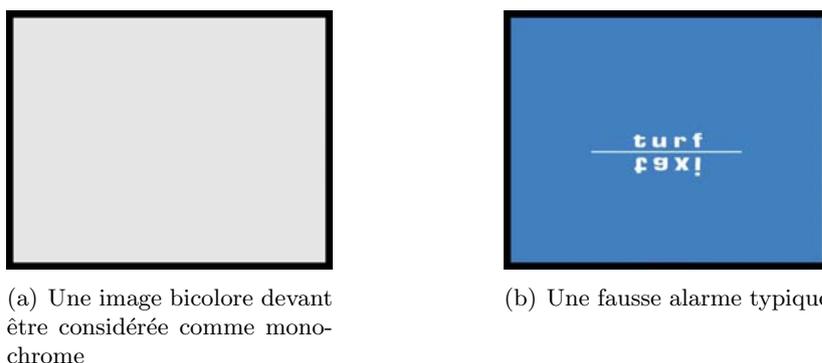
Nous n’avons pas été en mesure d’évaluer le détecteur de « cuts » car nous n’avions pas de vérité terrain. En revanche, le tableau 20 montre les résultats de l’évaluation du détecteur de « cuts » en tant que détecteur de transitions entre émissions ; ces chiffres correspondent à une précision de 0,3%, due au nombre élevé de fausses alarmes, et à un rappel de 100%. Ces résultats sont facilement explicables : un nouveau programme commence par un nouveau plan. Si ce type de détecteur permet de retrouver la totalité des débuts de programmes, il est cependant impossible de l’utiliser seul compte tenu du nombre de fausses alarmes trop élevé.

Détection d’images monochromes

Par définition, une image monochrome est une image dont tous les pixels sont à la même intensité (ou même couleur). Dans un flux télévisuel, ce ne sera que très rarement le cas : l’image est souvent encadrée par un fond noir afin d’être mise au format d’un écran de télévision. Par exemple, la figure 48a montre une image qui doit être considérée comme monochrome alors qu’elle est encadrée de noir (et du coup bicolore). Pire encore, des artefacts de compression compliquent leur détection.

Pour pouvoir détecter de telles images, (Naturel et Gros, 2005) propose d’utiliser comme critère l’entropie de l’histogramme de la luminosité des pixels. On rappelle que l’entropie est utilisée en statistiques pour mesurer la dispersion des valeurs et qu’elle se calcule par :

$$E = - \sum_k P_k \times \log P_k. \quad (1)$$

FIG. 48: *Détection des images monochromes*

Dans notre cas, P_k dénote la fréquence de la $k^{\text{ème}}$ composante de l'histogramme.

Un seuil est ensuite nécessaire pour décider si une image est monochrome ou pas. Pour le déterminer, une image « monochrome » du flux télévisuel est choisie et le seuil est fixé à la valeur de l'entropie de l'histogramme de la luminosité de ses pixels. Le détecteur que nous avons développé est une implémentation de cette méthode dont la précision est de 87% et le rappel de 92%. Le tableau 21 montre les résultats obtenus avec la détection des images monochromes pour la détection des débuts de programmes, correspondant à une précision de 12,2% et un rappel de 83,9%.

Jours	Nombre de programmes correctement localisés	Nombre de fausses alarmes
Samedi	78,6% (sur 70)	563
Dimanche	71,6% (sur 67)	472
Lundi	72,2% (sur 90)	448
Mardi	95,1% (sur 81)	399
Mercredi	76,6% (sur 77)	591
Jeudi	93,5% (sur 93)	552
Vendredi	95,8% (sur 95)	434

TAB. 21: *Évaluation de la détection des images monochromes pour la localisation de programmes*

Les résultats obtenus méritent d'être commentés. Ces chiffres montrent en effet que la détection des images monochromes produit beaucoup moins de fausses alarmes que le détecteur de « cuts ». La plupart des fausses alarmes sont dues soit à des cartons publicitaires (voir figure 48b), soit à des génériques sur fonds monochromes, soit enfin à la « pause » observée entre deux publicités. Cependant, un tel détecteur ne permet pas de retrouver l'ensemble des émissions. Enfin, il est important de préciser que ce bon taux de reconnaissance est probablement dû au fait que la chaîne articule ses émissions avec un effet de transition qui consiste en une page blanche qui se tourne (voir la figure 49). Les erreurs sont localisées principalement la nuit lorsque France 2 enchaîne brutalement ses émissions ou alors lorsque la page tournante est si rapide qu'on n'observe pas l'image représentant la page entière.

Détection des silences

Il est parfois possible d’observer des silences entre deux programmes. Ainsi, Naturel (Naturel et Gros, 2005) propose la détection des silences comme repère, étendant ainsi les méthodes de détections des publicités aux émissions. La détection des silences peut se faire en calculant la log-énergie des échantillons sonores du flux. L’énergie d’un échantillon est simplement égal à son carré et il suffit donc de calculer l’énergie moyenne d’une fenêtre sonore de quelques millisecondes (dans notre cas 20ms, soit deux fois moins que le temps d’observation d’une image fixe à l’écran).

Il suffit de fixer un seuil en-dessous duquel la faible activité sonore peut être considérée comme un silence. Ce seuil est fixé expérimentalement de façon à n’oublier aucun silence au détriment du nombre de fausses alarmes. Nous ne sommes pas en mesure d’évaluer le détecteur de silences car nous n’avons pas de corpus et de vérité terrain. Cependant, le tableau 22 présente l’évaluation de la détection des silences dans le cadre de la localisation des programmes.

Jours	Nombre de programmes correctement localisés	Nombre de fausses alarmes
Samedi	95,7% (sur 70)	27231
Dimanche	83,6% (sur 67)	24204
Lundi	88,9% (sur 90)	27329
Mardi	100% (sur 81)	27217
Mercredi	83,1% (sur 77)	25684
Jeudi	96,8% (sur 93)	28303
Vendredi	100% (sur 95)	26946

TAB. 22: Évaluation de la détection des silences pour la localisation de programmes

Le détecteur de silences est assez sensible et nous obtenons ainsi de nombreuses fausses alarmes : la précision de ce détecteur pour la localisation des programmes est de 0,28% et le rappel est de 93,0%. Nous n’avons pas cherché à intégrer les résultats temporellement, par exemple en filtrant les silences très courts. Malgré sa sensibilité, le détecteur de silences ne permet pas de localiser correctement le début de tous les programmes, pour des raisons identiques au détecteur d’images monochromes. La nuit, deux programmes peuvent s’enchaîner brutalement et le générique de fin du premier est ainsi directement suivi par le générique de début du second. De plus, l’effet de transition de France 2 est accompagné d’un son qui accompagne la page blanche qui se tourne.

Fusion des détections

Il est possible de diminuer le nombre de fausses alarmes des détecteurs précédents en fusionnant ces descripteurs (Naturel et Gros, 2005). Il est possible d’envisager deux cas :

- la détection simultanée des images monochromes et du silence ;



FIG. 49: Effet visuel de transition utilisé par France 2

- la détection d'un nouveau plan précédé par une période de silence.

Les résultats peuvent se déduire des tableaux 20, 21, 22 mais nous les avons regroupés dans les tableaux 23 et 24.

Jours	Nombre de programmes correctement localisés	Nombre de fausses alarmes
Samedi	78,6% (sur 70)	281
Dimanche	71,6% (sur 67)	298
Lundi	72,2% (sur 90)	373
Mardi	95,1% (sur 81)	347
Mercredi	76,6% (sur 77)	292
Jeudi	93,5% (sur 93)	314
Vendredi	95,8% (sur 95)	367

TAB. 23: *Évaluation de la détection simultanée des silences et des images monochromes pour la localisation de programmes*

Jours	Nombre de programmes correctement localisés	Nombre de fausses alarmes
Samedi	95,7% (sur 70)	3605
Dimanche	83,6% (sur 67)	2228
Lundi	88,9% (sur 90)	2873
Mardi	100% (sur 81)	3240
Mercredi	83,1% (sur 77)	3165
Jeudi	96,8% (sur 93)	2993
Vendredi	100% (sur 95)	3072

TAB. 24: *Évaluation de la détection de « cuts » précédés de silence pour la localisation de programmes*

Pour simplifier la fusion, nous avons choisi d'exécuter la détection des silences en utilisant des fenêtres de 20 ms, soit la moitié du temps de présentation d'une image fixe dans nos vidéos. Les résultats montrent que les deux fusions donnent de bons résultats tout en réduisant le nombre de fausses alarmes.

La méthode qui consiste à détecter les images monochromes et les silences localise parfaitement les transitions entre deux publicités : les fausses alarmes obtenues sont majoritairement des ruptures entre deux publicités. Cette méthode affiche donc une précision de 17,5% et un rappel de 84,1%. Cette détection pourrait être très utile mais il semble difficile de savoir si la première image monochrome détectée est celle qui précède le jingle de la publicité ou celle qui sépare deux publicités. Ces détections sont porteuses d'information mais sont source d'ambiguïté. De plus, certains génériques de films peuvent être une source de fausses alarmes dans la mesure où ils consistent en une apparition et une disparition de textes sur fond monochrome, accompagnés d'un silence ou encore d'une musique très faible ; ainsi, le film *Possession*, diffusé lors de notre semaine de test, nous a montré le danger de telles détections.

Enfin la méthode détectant les « cuts » précédés d'un silence génère plus de fausses alarmes que celle qui utilise les images monochromes, mais localise un peu plus de débuts de programmes. Sa précision est de 2,4% et son rappel est de 93,0%. Les détections sont donc moins porteuses d'information que les précédentes.

Jour	Nombre de fausses alarmes	Jour	Nombre de fausses alarmes
Samedi	276	Samedi	766
Dimanche	289	Dimanche	487
Lundi	327	Lundi	543
Mardi	315	Mardi	559
Mercredi	267	Mercredi	612
Jeudi	312	Jeudi	737
Vendredi	293	Vendredi	596

(a) Détection d'images monochromes et de silences

(b) Détection de « cuts » et de silences

TAB. 25: Nombre de fausses alarmes pour les détections locales

Détections locales

Les arbres de régression (voir troisième partie, page 75) permettent d'encadrer les durées par un intervalle de valeurs possibles : si les détections que nous venons de présenter sont exécutées localement, dans ces fenêtres, les bénéfiques sont multiples. D'une part, nous avons montré que seul un tiers du flux nécessite d'être examiné (voir section 3.5, page 148). D'autre part, cela permet d'éviter les fausses alarmes qui ont lieu en dehors des fenêtres temporelles. En revanche, puisque la grille de programmes de la semaine test est exprimable par le modèle, cela ne change pas le taux de bonnes détections. Le tableau 25 indique le nombre de fausses alarmes pour les deux méthodes multimodales présentées dans la section précédente.

Le tableau de gauche montre que le fait d'appliquer les détecteurs localement diminue le nombre de fausses alarmes de la détection conjointe des images monochromes et des silences d'en moyenne 8%. En revanche, le tableau de droite montre une diminution de 80% en moyenne du nombre de fausses alarmes du détecteur de plans précédés d'un silence. La modélisation des grilles de programmes permet donc d'utiliser un détecteur tel que le détecteur de plans précédés d'un silence qui produit de nombreuses fausses alarmes.

Ainsi, dans une fenêtre temporelle, plusieurs transitions candidates peuvent être trouvées ; ce sera très souvent le cas lors d'une coupure publicitaire puisque les publicités sont séparées par des silences et des frames monochromes. Afin de diminuer le nombre de transitions candidates, nous proposons de détecter les jingles publicitaires.

1.2 Détection des jingles publicitaires

Nous avons présenté dans la section 2 (page 19) différentes méthodes de détection des coupures publicitaires. Compte tenu des spécificités françaises, à savoir qu'un écran publicitaire doit être séparé du reste par des jingles publicitaires, les méthodes par reconnaissance sont les plus fiables en France. Nous rappelons qu'elles consistent à garder dans une base l'ensemble des signatures des jingles des chaînes et de les comparer à celles des images candidates. Le principal inconvénient de ces méthodes est la mise à jour de la base. En effet, l'ensemble des jingles d'une chaîne évolue dans le temps même si l'habillage général de la chaîne reste le même pendant plusieurs années. À titre d'exemple, l'habillage courant de France 2 est le même depuis 2002, et en fonction des événements (fêtes de fin d'années, films particuliers, événements sportifs, ...), de nouveaux jingles sont produits. Ainsi, la première fois qu'un jingle est rencontré, ces méthodes sont défaillantes et une mise à jour de la base est nécessaire.

Il est cependant possible de minimiser le nombre de mises à jour en utilisant des invariants propres à l'habillage de la chaîne. L'avantage est que l'invariant a la même durée de vie que l'habillage, dont la durée moyenne est de dix années. Prenons le cas, par exemple, des chaînes du

groupe France Télévisions, ou TF1, dont l'invariant est une combinaison du logo de la chaîne et de texte occupant toujours la même place dans l'image. La détection est ainsi simplifiée puisqu'il s'agit d'une détection d'identité : il n'y a ni mise à l'échelle, ni transformation, ni déplacement à prendre en compte. Les figures 50a à 50f sont des images extraites de différents jingles de différentes chaînes présentant de tels invariants (encadrés en rouge). Nous proposons ainsi de détecter ces invariants afin de localiser les jingles publicitaires. La méthode utilisée pour cela est décrite dans la section 2.1 (page 169). Cependant, toutes les chaînes n'ont pas d'invariants aussi simples à détecter ; par exemple, les figures 50g et 50h montrent différents jingles de la chaîne M6.

Le tableau 28 (page 171) montre les résultats obtenus pour la détection de l'invariant des jingles publicitaires de France 2 (le mot « publicité » et le logo de France 2). De même sur France 2, la détection des bandes annonces et de l'autopromotion peut être détectée de la même façon, l'invariant étant cette fois le logo de France 2 seul.

La détection des jingles permet une détection plus précise et moins ambiguë d'une telle partie du flux. En effet, dans une fenêtre temporelle correspondant à une coupure publicitaire, le nombre de jingles détectés sera limité. La section suivante établit la méthode de choix de la transition entre deux émissions parmi les transitions candidates détectées.

1.3 Choix de la transition parmi les transitions candidates

Les fenêtres temporelles dans lesquelles les détections sont effectuées permettent d'éviter les fausses alarmes qui ont lieu en dehors de celles-ci. En revanche, il se peut que des fausses alarmes aient lieu dans ces fenêtres. En fonction du nombre de fausses alarmes produites par les détecteurs, il est nécessaire de choisir parmi les détections candidates celle qui correspondra à la transition entre deux programmes. Dans notre expérimentation, nous utilisons des détecteurs très sensibles, et le choix d'une transition unique est donc nécessaire.

L'avantage du modèle que nous utilisons est qu'il permet de prédire le genre de transition que nous devons rechercher. Cela nous permet d'utiliser les règles indiquées dans le tableau 26, dans lequel le symbole * peut désigner n'importe quel genre d'émission (autrement dit n'importe quel état du modèle). Ce tableau montre que certaines transitions sont fortement marquées par l'habillage. Par exemple, la météo et les émissions de services étant sponsorisées, une image monochrome peut être détectée avant et après l'annonce du sponsor. Dans ces cas là, nous n'avons pas observé de fausses alarmes dans la fenêtre temporelle.

Nature de la transition		Détection
	* \mapsto interprogramme	logo de France 2
interprogramme	\mapsto *	disparition du logo de France 2
	* \mapsto publicité seule	invariant publicitaire
publicité seule	\mapsto	disparition de l'invariant publicitaire
	* \mapsto météo	image monochrome et silence
météo	\mapsto *	image monochrome et silence
	* \mapsto service	image monochrome et silence
service	\mapsto *	image monochrome et silence
	* \mapsto *	image monochrome / « cuts » et silence

TAB. 26: Règles utilisées pour la détection de France 2

Dans les autres cas, qui sont finalement rares, le choix d'une transition parmi celles détectées doit être envisagée. En utilisant le modèle, il est possible d'attribuer à chaque transition une probabilité. En effet, une transition marquera la fin de l'émission courante et déterminera sa



(a) Jingle France 2



(b) Jingle France 2 temporaire



(c) Jingle France 3



(d) Jingle France 3 temporaire



(e) Jingle TF1



(f) Jingle TF1 temporaire



(g) Jingle M6



(h) Jingle M6 temporaire

FIG. 50: Exemples de jingles publicitaires

durée. Or le modèle permet d'estimer la probabilité de la durée d'une émission dans un contexte donné. Cependant, cette méthode privilégiera systématiquement la transition qui donnera à l'émission courante sa durée moyenne dans ce contexte. En pratique, les fausses alarmes sont assez éloignées des ruptures réelles : le modèle attribuera ainsi une faible probabilité aux fausses alarmes.

Journée	Nombre de segments	Correction	Précision temporelle
Samedi	70	97,1 % (68 / 70)	7 s
Dimanche	67	94,0 % (63 / 67)	26 s
Lundi	90	97,8 % (88 / 90)	17 s
Mardi	81	98,8 % (80 / 81)	0 s
Mercredi	77	93,5 % (72 / 77)	46 s
Jeudi	93	97,8 % (91 / 93)	17 s
Vendredi	95	96,8 % (92 / 95)	0 s

TAB. 27: Correction et précision temporelle des structures produites automatiquement

Les résultats obtenus pour la structuration de la semaine de test sont montrés dans le tableau 27. Afin de valider une structuration produite automatiquement, nous nous intéressons à deux grandeurs. D'une part, la correction de la structure est la mesure du nombre de programmes et d'interprogrammes qui ont été correctement retrouvés. D'autre part, la précision temporelle mesure son décalage par rapport à la réalité (de la vérité terrain) ; il s'agit donc de la somme des retards de chacun de ses segments. Étant donné que la vérité terrain que nous utilisons est précise à la seconde près (données Médiamétrie) alors que la structuration que le système produit est à l'image près, nous ne donnons dans le tableau 27 qu'un ordre de grandeur de cette précision temporelle puisqu'on ne peut pas la calculer à l'image près. De plus, cette précision temporelle ne tient pas compte des légers retards de détection dûs aux effets de transition. Le détail de la structuration d'une journée est montré dans l'annexe C (voir page 203).

1.4 Discussion

Avec les détections de logos et les détections de ruptures, nous sommes en mesure de retrouver la quasi totalité de l'ensemble des émissions du flux de France 2 diffusées en dehors des tranches horaires nocturnes. Toutefois, la précision temporelle mériterait d'être améliorée.

Les sponsors, par exemple ceux des bandes-annonces, ne posent pas de problème s'il y a un carton de France 2 avant. En effet, nous avons vu que les bandes-annonces étaient détectées par l'apparition du logo de France 2 à un emplacement précis. Si l'émission présentée dans la bande annonce est sponsorisée (par exemple, un tournoi de rugby), le sponsor étant diffusé avant la bande-annonce et sans carton de France 2, nous allons considérer que celle-ci commence après le carton du sponsor.

Un autre problème que nous n'avons pas pris en compte est la possibilité pour les chaînes de France Télévisions de diffuser des bandes-annonces pour les autres chaînes du groupe. Dans ce cas, le logo de France 2 n'apparaît plus (il est remplacé par le logo des autres chaînes). Ainsi, si un interprogramme termine par une bande-annonce de ce genre, il sera tronqué au début de cette bande-annonce, augmentant ainsi la précision temporelle. En revanche, si ce type de bande-annonce apparaît au milieu de l'interprogramme, cela ne pose pas de problème grâce aux fenêtres temporelles prédites (on n'exécute les détecteurs que vers la fin de l'interprogramme). Ce type de bandes-annonces inter-chaînes est un cas particulier de France Télévisions. Une détection des logos des chaînes de France Télévisions permettrait de remédier à ce problème.

Le détail de la structure produite quasi-automatiquement par le système pour le mardi 25 novembre 2005 est montré dans l'annexe C (page 203). La précision temporelle est exceptionnellement nulle car le seul segment à ne pas avoir été détecté est un interprogramme composé uniquement d'une bande-annonce pour France 5. En général, ce type d'interprogramme aug-

mente la mesure de la précision temporelle.

Nous n'avons pas non plus considéré les concours liés à une émission où une question est posée sur le contenu du programme : le logo de France apparaît et le concours sera ainsi considéré comme une bande-annonce. C'est le cas le dimanche après la diffusion de *30 millions d'amis* ; le carton du concours dure 25 secondes.

Lors de nos expérimentations, nous avons remarqué que certaines émissions n'avaient pas été détectées automatiquement parce qu'elles n'étaient pas précédées d'un silence (ou le silence n'a pas été détecté). Par exemple, pour détecter la transition entre un journal et la météo nous avons spécifié (voir tableau 26) qu'il suffisait de détecter des images monochromes et du silence : cela n'a pas été le cas pour les météos du samedi à 20h41 et du dimanche à 0h28.

De plus, nous avons vu dans la partie précédente qu'il existait plusieurs grilles prédites possibles pour une même journée : ainsi, une émission peut être suivie par différents genres d'émissions (voir figure 46, page 147). Nous avons choisi de parcourir les branches les plus probables en premier (approche gloutonne). Dans notre semaine de test, le parcours glouton du graphe permet toujours de prédire la bonne grille. Toutefois, le cas contraire peut se produire. Cependant, puisque les émissions que nous prédisons sont des émissions courtes, il est peu probable de trouver des transitions telles que des images monochromes au cours d'une émission. Dans le pire des cas, si le système accumule des erreurs de prédiction, étant donné que le modèle est contextuel, il se trouvera rapidement dans l'incapacité de trouver une transition. Cet état peut conduire à l'appel d'un opérateur pour une vérification manuelle.

1.5 Conclusion

Nous avons montré que la modélisation des grilles de programmes permettait d'obtenir de bons résultats de structuration automatique sur France 2 avec des détections très simples (silences, images monochromes, logos) et effectuées localement. En effet, seules les émissions nocturnes, qui s'enchaînent de façon brutale, n'ont pas été détectées.

Pour obtenir de meilleurs résultats, il est possible d'employer des détecteurs plus spécifiques pour caractériser un type de transition entre deux programmes. À titre d'exemple, sur France 2, au lieu d'utiliser un détecteur d'images monochromes, on pourrait se limiter à un détecteur d'images blanches (l'effet visuel étant une page blanche), ce qui éviterait d'autres fausses alarmes (telles que celles causées par le générique du film *Possession*). Le modèle des grilles de programmes permettant de choisir quel détecteur sera employé, les performances ne seront pas affectées puisqu'il évite de lancer la totalité des détecteurs sur l'ensemble du flux.

Puisque nous nous basons sur un guide de programmes et un modèle statistique pour structurer, il sera parfois nécessaire de vérifier l'étiquetage des émissions. Le chapitre suivant décrit une des techniques possible pour vérifier la structuration : la reconnaissance de logos.

2 Identification d'un programme

Une émission est facilement reconnaissable par son habillage : ses génériques, ses jingles, et l'ensemble des éléments de son aspect visuel et auditif.

Ce chapitre relate nos expérimentations sur la détection des logos dans un flux TV. En effet, les émissions ont souvent un logo caractéristique qui remplace ou vient en supplément de celui de la chaîne. Détecter et reconnaître le logo d'une émission permet de l'identifier de façon unique et de vérifier la correspondance entre une émission du guide de programmes et l'émission courante dans le flux.

Cette expérimentation a été réalisée en partenariat avec Stephan Duffner et Christophe Garcia du département recherche de France Télécom. Nous allons dans un premier temps décrire la méthode proposée dans (Duffner et Garcia, 2006) puis décrire les résultats obtenus.

2.1 Description de la méthode

Lorsque l'on parle de logo dans un flux télévisuel, il faut distinguer les trois types suivants :

- les logos fixes opaques ;
- les logos fixes translucides ;
- les logos animés (transparents ou opaques).

De nombreuses méthodes proposent la détection et la reconnaissance de logos fixes dans une vidéo, le cas le plus simple étant les logos fixes opaques (Albiol *et al.*, 2004b; Meisinger *et al.*, 2005; Seeber *et al.*, 2007). À notre connaissance, il n'existe pas de méthode spécifique pour les logos animés. Généralement, les méthodes de détection des logos opaques fixes sont basées sur la constance d'une image à l'autre des valeurs des pixels (en supposant que la vidéo n'est pas constituée d'une seule image fixe). Pour les logos translucides, cette propriété n'est pas envisageable. La méthode que nous avons utilisé prend en compte les logos fixes en général (Duffner et Garcia, 2006) et est basée sur les réseaux à convolution (*convolutionnal neural network*) (LeCun *et al.*, 1998; Garcia et Delakis, 2004).

Les réseaux à convolution sont des cas particuliers de réseaux de neurones (voir section 2.1, page 108) qui ont la capacité de produire des extracteurs de caractéristiques du signal à partir d'exemples positifs et négatifs de logos. Une fois entraîné, le réseau se comporte comme un enchaînement de convolutions et de sous-échantillonnage de l'image d'origine. Ces réseaux fonctionnent avec une fenêtre glissante qui parcourt l'image originale et qui correspond à un champ de vision restreint du signal global.

Architecture du réseau à convolution

L'architecture d'un réseau à convolution se compose de trois types de couche (voir figure 51). Les couches de convolution (notées C_i sur la figure) permettent d'appliquer des convolutions de différents noyaux. La couche C_1 est directement reliée à la rétine, c'est-à-dire à la partie de l'image que l'on souhaite classer (dans notre cas, en logo de France 2 ou non). Contrairement aux perceptrons, les neurones des couches de convolution ne sont reliés qu'à un voisinage restreint de neurones de la couche précédente afin de se rapprocher du fonctionnement de la rétine et du cerveau humains (Poisson *et al.*, 2002). Afin d'augmenter la capacité du réseau à généraliser, son architecture est contrainte par un partage de poids : cela revient à calculer plusieurs fois la même fonction mais avec des entrées différentes. Cette technique réduit aussi le nombre de paramètres à calculer. Chaque couche possède plusieurs groupes de neurones, chacun correspondant à un détecteur de caractéristiques du signal.

Une fois qu'un élément caractéristique a été détecté, sa localisation exacte est moins importante. C'est pour cela que, généralement, une couche de convolution est suivie par une couche de sous-échantillonnage (notées S_i). Cette phase permet d'être plus robuste à des transformations telles que la translation, la différence d'échelle ou encore la déformation du logo.

Tous les paramètres de l'architecture, c'est-à-dire le nombre de couches, le nombre de groupes par couche et leur connexité, la taille de la fenêtre glissante sont choisis de façon empirique. Pour

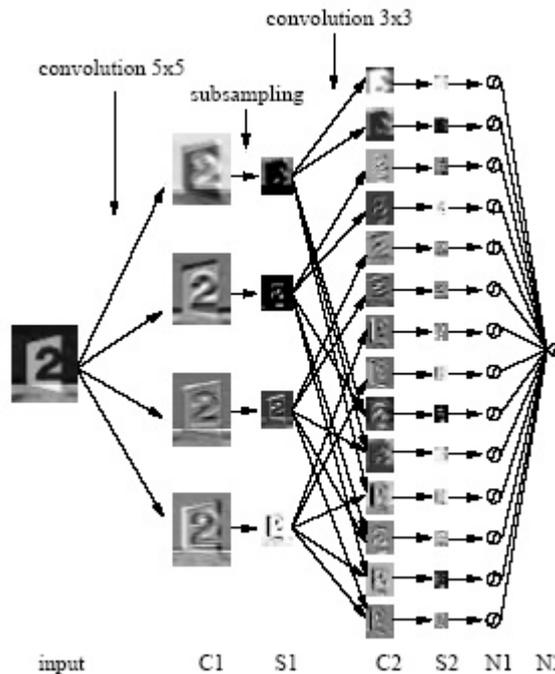


FIG. 51: Architecture du réseau à convolution (Duffner et Garcia, 2006)

la détection des logos, (Duffner et Garcia, 2006) propose que les couches C_1 et C_2 soient des convolutions de masques 5×5 et 3×3 respectivement, dont les poids sont des paramètres du modèle. La couche C_1 exécute quatre convolutions sur l'image en entrée. La couche C_2 possède quatorze groupes de neurones : chaque résultat de la couche C_1 se voit appliquer deux convolutions différentes. Les six autres groupes de C_2 correspondent à la fusion de l'ensemble des couples formés par les groupes de C_1 . Les couches N_1 et N_2 sont de simples fonctions sigmoïdes qui permettent la classification après extraction des caractéristiques et réduction de la dimension des informations.

Dans le cas général, la fenêtre glissante doit parcourir l'ensemble de l'image d'entrée. Dans notre cas, puisque nous connaissons l'emplacement du logo et puisque nous souhaitons reconnaître un logo à un emplacement particulier, la fenêtre glissante ne parcourt qu'une zone limitée de l'image.

À titre d'exemple, pour le logo de France 2 seul, la taille de la rétine du réseau est de 38×46 neurones. Dans ce cas, le nombre de paramètres à optimiser lors de la phase d'entraînement du réseau à convolution s'élève à 1147 paramètres.

Dans la section suivante, nous décrivons la méthode d'apprentissage d'un tel réseau de convolution.

Entraînement du réseau à convolution

Pour que le réseau à convolution soit opérationnel, il nécessite d'être entraîné sur un ensemble d'exemples positifs et négatifs. La sortie du réseau, c'est à dire la sortie du neurone de la couche N_2 (voir figure 51) est fixée à +1 si l'image d'entrée contient le logo désiré et -1 dans le cas contraire.

Afin d'estimer les paramètres du réseau de convolution, l'algorithme présenté dans (LeCun

et al., 1998) est utilisé. Il s'agit d'un algorithme qui dérive de la rétropropagation du gradient adaptée aux réseaux à convolution. Nous fournissons un ensemble d'exemples positifs et négatifs issus d'une même journée de diffusion de France 2. La difficulté réside dans la constitution de l'ensemble d'exemples négatifs dans la mesure où celui-ci doit être aussi représentatif que possible.

La solution à ce problème consiste en une approche dite *bootstrapping* (Sung et Poggio, 1998) : le système est entraîné de façon itérative à partir des fausses alarmes qu'il génère. Dans un premier temps, un ensemble de validation est créé à partir d'un certain nombre d'images de l'ensemble d'apprentissage. Cet ensemble de validation permettra de tester la généralité du réseau, c'est-à-dire son comportement face à de nouvelles images. Si l'ensemble de validation est constant pendant tout le reste de la phase d'entraînement, l'ensemble d'apprentissage va évoluer au fil des itérations. Toutefois, à chaque itération de l'algorithme d'apprentissage, le nombre d'exemples positifs et négatifs présentés au réseau à convolution seront les mêmes.

Pour constituer le nouvel ensemble d'exemples négatifs, l'algorithme utilise les fausses alarmes générées par la phase précédente. Un seuil *ThrFa* sur la réponse du réseau à convolution permet de considérer dans un premier temps les fausses alarmes dont la réponse est fortement proche de +1. À chaque itération, le seuil *ThrFa* est diminué de façon à forcer le système à redéfinir son hyperplan de séparation en fonction de ces fausses alarmes et d'autres exemples négatifs sont choisis aléatoirement dans l'ensemble d'apprentissage de façon à compléter l'ensemble des exemples négatifs. L'itération s'arrête lorsque une convergence est observée.

2.2 Résultats expérimentaux

	Taux de détections	Taux de fausses alarmes
Logo de France 2	93,5 %	0 %
Logo de France 2 et « publicité »	94,1 %	0,4 %
KD2A	93,2 %	0,9 %
Télématin	99,9 %	0 %

TAB. 28: Récapitulatif des résultats obtenus sur les quatre logos

L'ensemble d'apprentissage de départ (avant le prélèvement de l'ensemble de validation) est constitué de plus de 46000 images annotées à la main et représentant une image toutes les deux secondes dans une journée du flux de France 2 contenant chacune des émissions visées. Nous avons testé les réseaux à convolution sur quatre logos :

- le logo de France 2 seul ;
- le logo de France 2 accompagné du mot « publicité » ;
- le logo de KD2A, une émission jeunesse de France 2 ;
- et enfin le logo de Télématin, la tranche horaire matinale de France 2.

Les résultats¹⁷ présentent tous un taux de fausses alarmes très faible et un taux de bonnes détections très élevé, montrant ainsi la fiabilité de l'approche. Le tableau 28 résume l'ensemble des résultats obtenus. Les logos qui n'ont pas été détectés sont souvent peu visibles à l'oeil nu : ils sont en général affichés pendant l'effet de transition de France 2, donc énormément blanchis. La figure 52 présente des exemples de fausses alarmes.

¹⁷Ces expérimentations ont été menées par Stephan Duffner et Christophe Garcia de France Télécom R&D



(a) logo de *Télématin* non détecté



(b) fausse alarme du jingle publicitaire



(c) fausse alarme du logo de *KD2A*

FIG. 52: Exemples de logos mal détectés par les réseaux à convolution

3 Conclusion

Afin que le système puisse retrouver le début et la fin des programmes du flux télévisuel, il est nécessaire de caractériser les transitions d'un genre à l'autre par des règles faisant intervenir des détecteurs.

Chaque fois qu'un nœud du graphe des grilles de programmes possibles a plusieurs successeurs il suffit de rechercher dans les fenêtres temporelles fournies les éléments caractéristiques intervenant dans les règles. Ces règles reposent sur l'habillage de la chaîne qui reste le même pendant plusieurs années, ce qui évite une mise à jour trop fréquente des règles.

Dans nos expérimentations sur France 2, avec quelques simples règles de caractérisation, nous sommes en mesure de retrouver plus de 94% des émissions. Ces résultats peuvent encore être améliorés en considérant davantage de règles et davantage de détecteurs.

Chapitre 12

Structuration d'émissions et de tranches horaires

Sommaire

1	Structuration d'une tranche horaire	173
1.1	Prédiction du conducteur du <i>Morning Café</i>	174
1.2	Détections pour l'alignement du conducteur sur le flux	175
1.3	Résultats expérimentaux	179
2	Isolation des plateaux dans un journal télévisé	182
2.1	Description de la méthode	182
2.2	Résultats expérimentaux	183
3	Conclusion	184

1 Structuration d'une tranche horaire : exemple du *Morning Café*

Cette section décrit une étude menée en collaboration avec Jérémy Philippeau et Julien Pinquier¹⁸ dans le cadre de la structuration d'une tranche horaire (Poli *et al.*, 2007). Le *Morning Café* est la tranche horaire matinale de M6, diffusée du lundi au vendredi de 7 heures à 9 heures. Elle est composée de plusieurs segments hétérogènes :

- des génériques ;
- des flashes d'information ;
- des bulletins météo ;
- des informations sur le trafic routier ;
- une émission principale composée de plateaux, de reportages, de clips musicaux et de séries (*Kaamelott*), diffusée en deux parties (de 7 heures à 8 heures, puis de 8 heures à 9 heures) ;

¹⁸IRIT, équipe Samova

- une série animée humoristique (*Avez-vous déjà vu ?*);
- et des coupures publicitaires.

Morning Café est retransmis en direct : compte tenu du fait qu'il s'agit d'une tranche horaire matinale, tout retard dans le conducteur de l'émission aurait pour conséquence un retard dans toute la grille de programmes de la journée. Le conducteur de cette tranche horaire est donc assez formaté. La modélisation des grilles de programmes, introduite dans la troisième partie (page 75), peut être appliquée au conducteur du *Morning Café*. En revanche, la dynamique de la tranche horaire et de l'enchaînement de ses composantes ne permet pas l'utilisation de détecteurs aussi simples que ceux utilisés dans le premier chapitre (page 159).

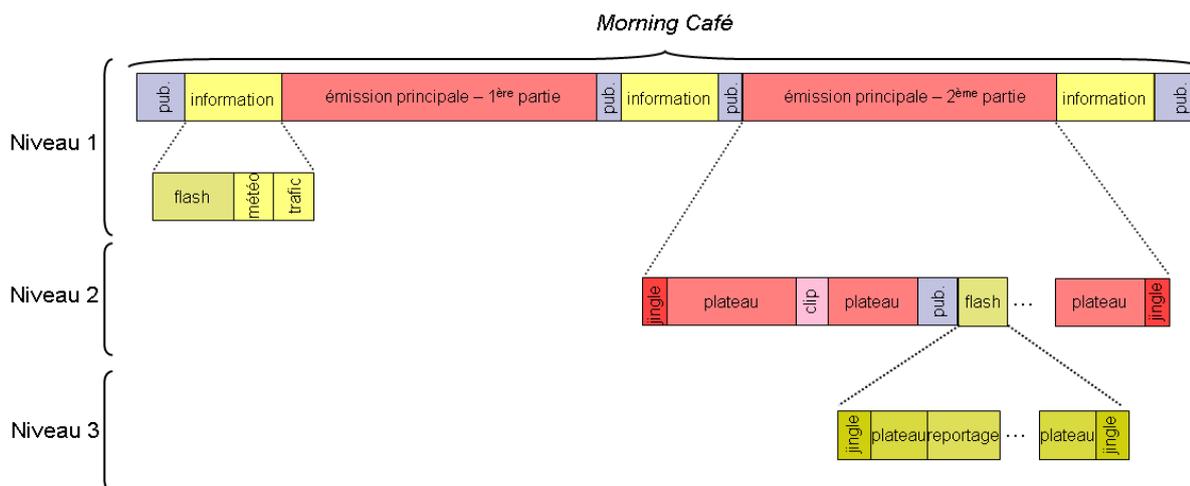


FIG. 53: Structure hiérarchique du *Morning Café*

Nous envisageons la structuration du *Morning Café* de façon hiérarchique (voir figure 53). Lors de la première étape, la tranche horaire est décomposée en journaux télévisés, météo, émissions de trafic routier et émissions principales. Au second niveau, les parties de l'émission principale sont découpées en génériques, plateaux, clips musicaux et séries. Finalement, les journaux télévisés sont découpés en plateaux et en reportages.

Nous allons dans un premier temps décrire la prédiction du conducteur du *Morning Café*. Nous présenterons par la suite les détecteurs utilisés et nous terminerons ce chapitre par les résultats expérimentaux.

1.1 Prédiction du conducteur du *Morning Café*

Pour prédire le conducteur du *Morning Café*, nous allons utiliser les mêmes méthodes présentées précédemment pour la prédiction des grilles de programmes. Les états du modèle de Markov contextuel (CHMM) représente cette fois les différents types de segments qui composent la tranche horaire.

Cependant, les ambiguïtés ne peuvent être résolues qu'à partir de détections faites sur le flux. En effet, nous n'avons pas de guides de programmes concernant cette tranche horaire. De plus, la variation du conducteur est bien moins grande que celle d'une grille de programmes.

Pour pouvoir entraîner le modèle, nous avons utilisé cinq *Morning Café* structurés manuellement : cela s'est avéré suffisant pour nos expérimentations. Pour traiter notre corpus, nous avons

gardé le même contexte que pour les grilles de programmes : l'heure et le jour de diffusion. En effet, le jour de la semaine est important car le conducteur change le mercredi, jour de sortie des films cinématographiques en salle. Pour le reste, le comportement du modèle est essentiellement le même que pour les grilles de programmes. La figure 54 montre un extrait d'un graphe correspondant aux conducteurs prédits. Pour ne pas surcharger la figure, nous avons omis les heures de diffusion.

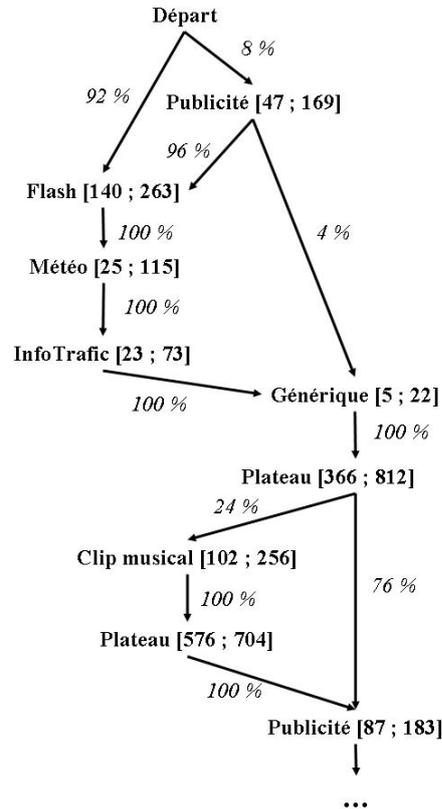


FIG. 54: Extrait du graphe représentant le début du *Morning Café*

1.2 Détections pour l'alignement du conducteur sur le flux

Puisque les segments qui composent un *Morning Café* sont hétérogènes et puisque nous n'avons pas de guide de programmes pour diminuer le nombre de chemins possibles dans le graphe, les détecteurs que nous devons employer doivent être plus spécifiques. La nature d'une tranche horaire facilite le travail : dans le cadre du *Morning Café*, les jingles publicitaires sont tous identiques et sont constitués par un extrait du générique. De même, les génériques du flash d'information, de la météo et du trafic routier sont identiques à chaque fois. Il est ainsi possible de reconnaître chacun de ces jingles pour caractériser les transitions possibles entre ces parties de la tranche horaire. Nous allons présenter dans cette section les différents détecteurs utilisés.

Détection et reconnaissance de jingles sonores

Le processus de reconnaissance des jingles sonores peut se décomposer en trois modules fréquents dans les problèmes de reconnaissance des formes (Piquier et André-Obrecht, 2006) :

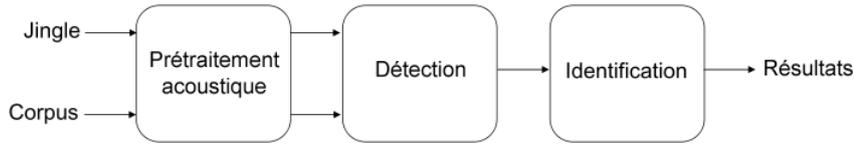


FIG. 55: Système de reconnaissance des jingles

dans un premier temps le document audio passe par un module de prétraitement, puis par un module de détection et enfin par un module de reconnaissance (voir figure 55).



FIG. 56: Analyse spectrale

Le prétraitement acoustique consiste en une analyse spectrale du signal afin de le représenter sous forme de vecteurs ; autrement dit, à partir d'un signal, il s'agit d'obtenir les informations pertinentes pour la tâche souhaitée (Pinquier, 2004). Dans un premier temps, les aigus du signal audio sont accentués, puis le signal est découpé en fenêtres de Hamming de 32 ms avec un recouvrement de 16 ms. Les coefficients spectraux sont alors créés à la suite du calcul des énergies dans 28 filtres, après le module de la FFT (Transformée de Fourier rapide) et une pondération triangulaire (filtrage). Le vecteur représentatif d'une trame est ainsi constitué de ses 28 coefficients spectraux et de son énergie.

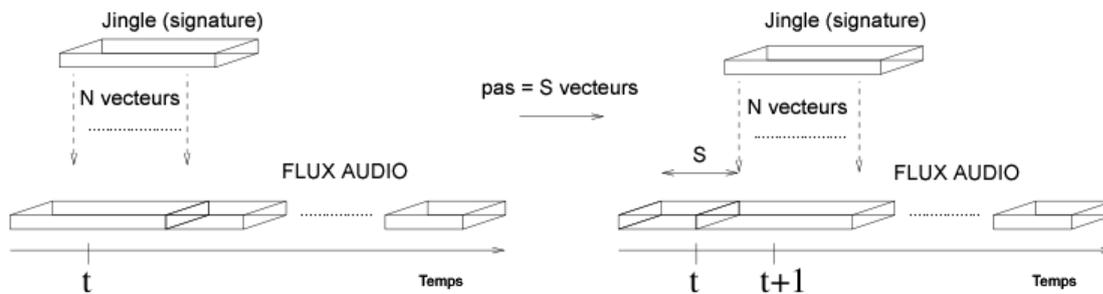


FIG. 57: Comparaison du jingle de référence et du flux audio

Chaque jingle est représenté par sa signature qui consiste en une séquence de N vecteurs spectraux, où N est le nombre de trames analysées. La détection consiste ainsi à trouver cette séquence dans le flux audio ; pour cela, les vecteurs d'information sont extraits du flux et comparés à ceux du jingle de référence par une distance euclidienne. À chaque comparaison, la séquence de vecteurs adjacents du flux est décalée de S vecteurs (voir figure 57).

Les séquences se présentant comme de bonnes candidates sont obtenues en sélectionnant les distances euclidiennes minimales (voir figure 58). Pour affiner cette sélection, nous calculons la moyenne M de ces distances. Si la distance courante est inférieure à $\frac{M}{2}$, alors cette distance correspond à une bonne séquence candidate. La figure 58 montre un exemple de distances euclidiennes obtenues par comparaison d'un jingle de référence avec trois minutes d'un flux audio. Dans un premier temps, cinq minima sont sélectionnés. Les deux premiers correspondent bien

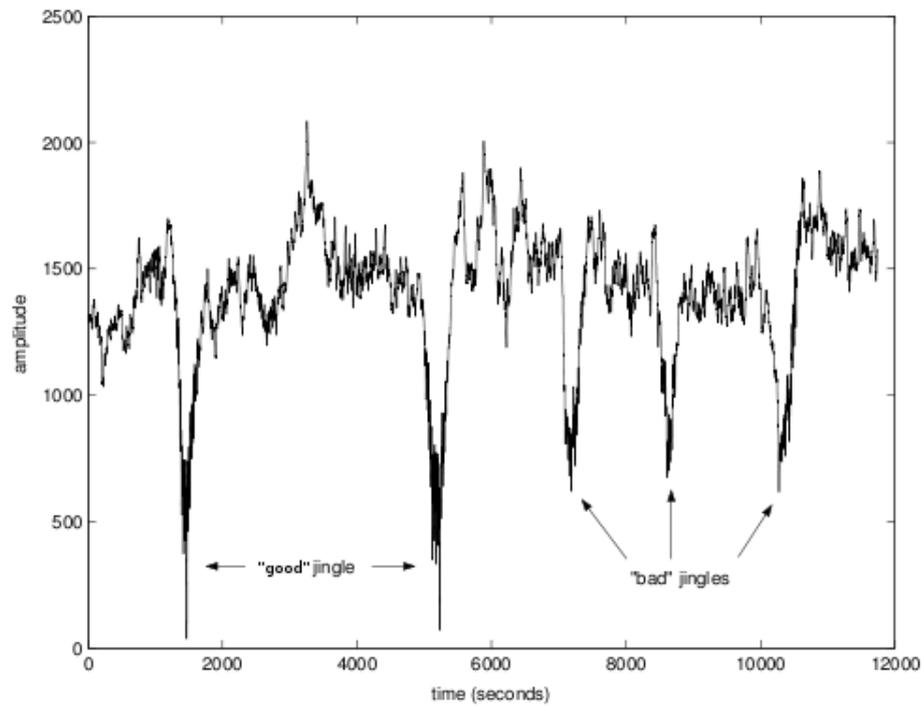


FIG. 58: Exemple de distances euclidiennes issues de la comparaison d'un jingle avec trois minutes de flux (Pinquier, 2004)

à un jingle de référence. En revanche, les trois derniers minima montrent bien la présence d'un jingle mais qui n'est pas le jingle recherché.

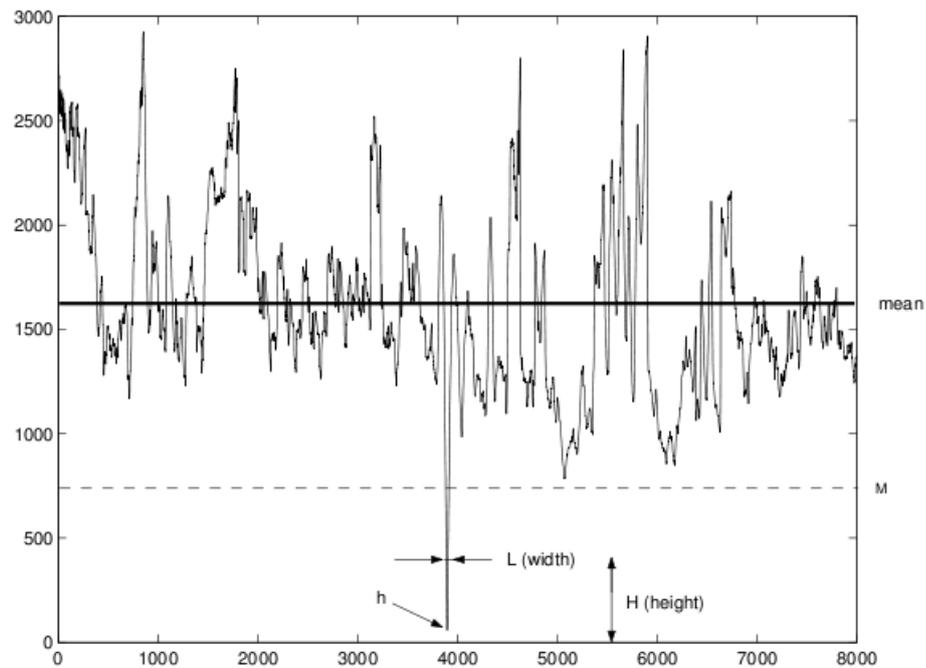


FIG. 59: Distance entre le jingle de référence et le flux audio (Pinquier, 2004)

Afin de reconnaître le jingle de référence parmi les séquences candidates, (Pinquier et André-Obrecht, 2004) propose de caractériser la largeur des pics (figure 58). Ainsi, pour chacun des pics correspondant à un minimum local, il faut calculer :

- la valeur courante h du minimum local ;
- la largeur L du pic à la hauteur H , où H est la hauteur du pic où l'on calcule sa largeur (voir figure 59).

Classification clip musical ou non clip

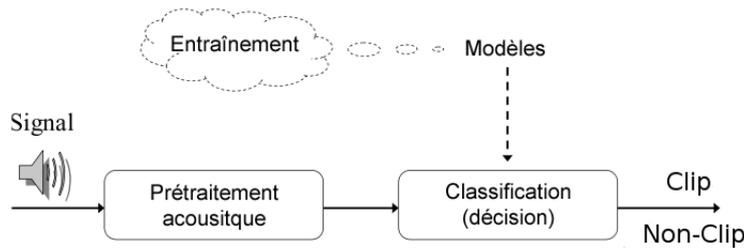


FIG. 60: Module de classification en clip musical ou non clip

Nous nous sommes intéressés à un autre détecteur qui nous permettra d'extraire un clip musical du plateau de l'émission principale. Nous rentrons ainsi dans le domaine de la détection de la musique. La figure 60 présente un aperçu de la méthode.

On retrouve trois modules similaires à ceux employés dans le détecteur d'applaudissements de (Pinquier et André-Obrecht, 2005) : un module de prétraitement (voir section 1.2), un module d'apprentissage et un module de classification.

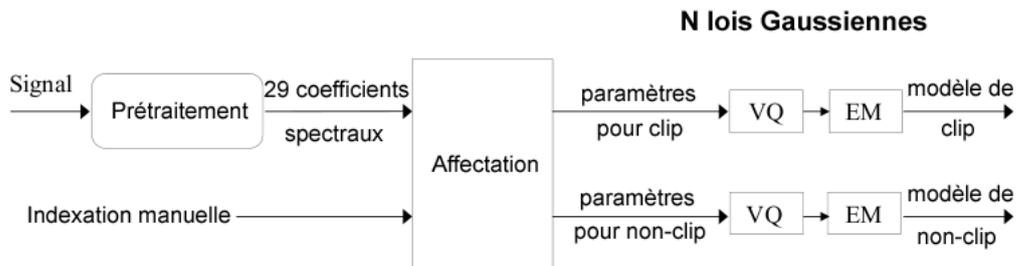


FIG. 61: Apprentissage du système de classification

Afin de représenter les éléments du signal audio comme des clips musicaux ou des non-clips, deux modèles de mélanges de lois gaussiennes sont nécessaires (GMM). Les GMM sont des outils probabilistes qui nécessitent un apprentissage supervisé soigné. L'apprentissage des paramètres des GMM est généralement réalisé par un algorithme EM qui se déroule en deux étapes. La première est une initialisation du modèle par Quantification Vectorielle (algorithme VQ) fondée sur l'algorithme de Lloyd (Lloyd, 1982). La seconde phase est une optimisation des paramètres du mélange de gaussiennes. Toutefois, le nombre de gaussiennes dans les GMM est à déterminer de façon empirique (voir section 1.3). Cette fois-ci, les trames sont représentées par un vecteur constitué à partir d'une analyse cepstrale.

La classification des segments du flux audio en clip ou non-clip se fait en plusieurs étapes à partir des deux modèles. Au début, on procède à une classification de trame de 10 ms en calculant la log-vraisemblance entre chaque vecteur d'information et chacun des modèles. Les vecteurs sont classés en fonction du modèle qui leur attribue le plus haut score. Compte tenu de la taille des trames, deux étapes d'intégration temporelle suivent la première phase. Dans un premier temps, les segments adjacents classés de la même façon sont fusionnés. Ensuite, une phase de lissage est nécessaire pour supprimer les segments considérés comme des clips musicaux de moins de 10 secondes.

Reconnaissance du plateau de l'émission principale

Le problème du détecteur employé pour la détection des clips musicaux est qu'il est basé uniquement sur le flux audio de la tranche horaire. Ainsi, si le clip démarre en arrière-plan alors que les chroniqueurs du *Morning Café* sont toujours filmés, ce plan sera classé comme faisant partie du clip. Pour éviter ce genre d'erreurs, nous avons introduit un détecteur de plateaux. En effet, le plateau de l'émission principale du *Morning Café* se distingue visuellement du reste par ses couleurs vives (voir figure 62).

Dans un premier temps, nous utilisons le détecteur de « cuts » introduit à la section 1.1 (page 160). Nous représentons ensuite l'image-clé de chaque plan par un vecteur à trois composantes : la luminosité moyenne et la teinte des deux couleurs dominantes (dans le domaine de couleurs HSL). À partir d'un échantillon d'images issues de plateaux du *Morning Café*, nous avons constitué un vecteur de référence α . Afin de déterminer si un plan appartient à un plateau ou non, nous vérifions la distance euclidienne entre le vecteur représentatif de ce plan et le vecteur α . Si cette distance est inférieure à un seuil δ estimé expérimentalement, le plan est considéré comme un plateau.



FIG. 62: Différentes vues du plateau de *Morning Café*

1.3 Résultats expérimentaux

Notre corpus de *Morning Café* est constitué de quatorze vidéos d'environ deux heures chacune. Nous allons dans un premier temps définir les détections à exécuter pour détecter certaines transitions d'un type de segment à un autre. Nous décrivons ensuite les expérimentations menées dans le but de déterminer les paramètres du module de classification en clip ou non-clip. Nous poursuivrons par l'évaluation de chacun des détecteurs et nous terminerons en évaluant la structuration des *Morning Café*.

Définition des règles pour caractériser une transition

Comme nous l'avons fait pour le flux de France 2, il faut associer des détecteurs aux différents types de transitions. L'ensemble de ces règles est répertorié dans le tableau 29.

Transitions		Détections
Plateau	→ Flash	jingle du flash
Flash	→ Météo	jingle de la météo
Météo	→ Trafic routier	jingle du trafic routier
Trafic	→ Plateau	jingle publicité
Publicité	→ Plateau	jingle publicité
Plateau	→ Publicité	jingle publicité
Plateau	→ Clip	NON plateau ET clip
Clip	→ Plateau	plateau
Flash	→ Plateau	jingle publicité
Plateau	→ Flash	jingle publicité
Publicité	→ <i>Avez-vous déjà vu ?</i>	générique d' <i>Avez-vous déjà vu ?</i>
<i>Avez-vous déjà vu ?</i>	→ Publicité	jingle publicité
Plateau	→ <i>Kaamelott</i>	générique <i>Kaamelott</i>
<i>Kaamelott</i>	→ Plateau	plateau

TAB. 29: Règles utilisées pour détecter les transitions

Ce que nous appelons *jingle publicité* dans le tableau 29 est en fait un générique court extrait du générique principal. Ainsi pour détecter le générique principal, on détecte deux fois le générique court.

Estimation des paramètres du module de classification en clip ou non-clip

Nombre de gaussiennes	Morning 1	Morning 2	Nombre de gaussiennes	Morning 1	Morning 2
8	78.75 %	80.60 %	8	80.56 %	81.53 %
16	77.85 %	83.19 %	16	79.51 %	81.67 %
32	79.98 %	83.57 %	32	81.47 %	81.12 %
64	84.84 %	84.89 %	64	82.04 %	81.48 %
128	83.58 %	84.41 %	128	83.77 %	84.46 %
256	83.00 %	83.23 %	256	79.25 %	79.89 %

a) Avec un vecteur à 12 coordonnées

b) Avec un vecteur à 26 coordonnées

TAB. 30: Taux de recouvrement de segments détectés automatiquement sur les segments annotés manuellement en fonction du nombre de gaussiennes et du nombre de composantes utilisées

Nous avons testé différents nombres de gaussiennes et différentes dimensions pour les vecteurs d'information sur les deux premières tranches horaires. En effet, les vecteurs sont constitués de 12 coordonnées (12 coefficients cepstraux) ou de 26 coordonnées (12 coefficients cepstraux, leurs dérivées, l'énergie et sa dérivée). Les meilleurs résultats ont été obtenus avec le modèle à 64 gaussiennes et les vecteurs à 12 coordonnées.

Évaluation du système de structuration

L'évaluation du système a été menée sur un ensemble de 9 *Morning Café*. Dans un premier temps, nous avons évalué les performances du détecteur de jingles qui a été entraîné sur :

- le générique de la tranche horaire ;
- les jingles de publicité ;
- le jingle de flashes d'information ;
- le jingle des bulletins météo ;
- le jingle du trafic routier ;
- le générique de la série *Avez-vous déjà vu ?* ;
- le générique de la série *Kaamelott*.

Les résultats sont présentés dans le tableau 31. Les précisions et les rappels sont très élevés pour chacun des jingles à détecter. Certains génériques de début ont néanmoins été oubliés et nous avons noté une fausse alarme pour le générique de *Kaamelott* et quelques une pour le générique principal.

Jingles	Précision	Rappel
Générique principal court	91.7 %	84.6 %
Flash d'information	100.0 %	100.0 %
Météo	100.0 %	100.0 %
Trafic routier	100.0 %	100.0 %
<i>Avez-vous déjà vu ?</i>	100.0 %	100.0 %
<i>Kaamelott</i>	100.0 %	99.9 %

TAB. 31: Évaluation de la reconnaissance des jingles

Le CHMM entraîné permet de représenter chaque jour du corpus de validation. Chaque nœud du graphe a au plus deux successeurs, ce qui signifie qu'au plus deux détections seront nécessaires pour déterminer la fin du segment courant.

Nos vérités terrain sont précises à l'image près dans la mesure où elles ont été faites manuellement. Ainsi, afin de valider notre approche, nous sommes en mesure de considérer à la fois la correction et la précision temporelle de la structuration automatique. La correction d'une structure produite automatiquement est une mesure du nombre de segments qui ont été correctement retrouvés. La précision temporelle d'une structure se calcule en sommant le retard accumulé par chacun des segments. Le tableau 32 indique la correction et la précision temporelle des structures obtenues sur les tranches horaires du corpus de validation.

Video	Nombre de segments		Correction	Précision temporelle
Morning	6	44	100.0 %	4 s
Morning	7	43	95.4 %	3 s
Morning	8	52	100.0 %	1 s
Morning	9	47	97.9 %	5 s
Morning	10	47	93.6 %	7 s
Morning	11	47	95.7 %	7 s
Morning	12	46	100.0 %	5 s
Morning	13	41	100.0 %	9 s
Morning	14	42	97.6 %	3 s

TAB. 32: Correction et précision temporelle des structures produites automatiquement

Les résultats obtenus sont tout à fait satisfaisants. Cependant, ils sont fortement liés à la qualité des détecteurs. La précision temporelle ne dépasse pas 10 secondes : elle est généralement due aux clips musicaux dont la musique démarre avant ou finit après le lancement du clip alors que le plan n'est pas considéré comme un plateau. Les problèmes de correction sont plutôt dûs à des clips qui n'ont pas été retrouvés, cette fois-ci à cause du module de classification en clip ou non-clip. Une publicité n'a pas pu être détectée à cause d'une erreur de détection du jingle.

Une fois encore, les avantages du modèle utilisé sont mis en avant. En effet, non seulement les détections ne se font que sur 32,7% de la vidéo en moyenne, mais de plus toutes les fausses alarmes sont évitées car elles ont lieu en dehors des fenêtres temporelles prédites.

Dans ce chapitre, nous n'avons pas traité tous les niveaux de la hiérarchie de la structure d'un *Morning Café*. En effet, nous n'avons pas traité l'extraction des plateaux dans les flashes d'information de la tranche horaire. Nous proposons dans la section suivante de remédier à cette omission en traitant le cas plus général des plateaux de journaux télévisés.

2 Isolation des plateaux dans un journal télévisé

L'INA bénéficie d'un accord avec France Télévisions qui lui donne les droits d'exploitation des journaux télévisés produits par le groupe. Ces journaux bénéficient d'un traitement documentaire approfondi dont la première étape consiste à les découper. Dans le cadre d'un groupe de travail avec les secteurs opérationnels, il est apparu que la simple isolation des plateaux dans un journal télévisé serait d'une grande aide.

La méthode par prédiction de conducteurs, comme présentée dans la section précédente, ne peut s'appliquer à un journal télévisé. En effet, les différents segments sont très courts et le nombre de reportages et de plateaux par journal n'est pas régulier.

Toutefois, les journaux télévisés ont longtemps attiré l'attention des chercheurs (voir le chapitre 2, page 13). En effet, des structurations très fines (en plateaux, reportages, duplex, interviews et autres) peuvent être fournies par des méthodes plus complexes telles que celles proposées dans (Demarty, 2000). Cependant l'isolation de plateaux ne nécessite pas une telle complexité. Nous proposons une méthode simple et rapide, qui ne nécessite aucun apprentissage, basée sur la comparaison des images et issue d'une étude préalable (Poli, 2003).

2.1 Description de la méthode

La figure 63 montre des images extraites de différents plateaux de journaux télévisés. Dans (Poli, 2003), nous avons essayé plusieurs techniques de comparaison d'images pour isoler les plateaux. Nous allons décrire la méthode qui a donné les meilleurs résultats et qui revient à une classification de plans en deux classes : plateaux et non-plateaux.

Dans un premier temps, le journal est découpé en plans avec la méthode proposée page 160. Une image clé est prélevée dans chacun des plans (dans notre cas, l'image centrale). Nous représentons chacune de ces images par un histogramme à 64 plages de valeurs. Pour cela, seuls les 2 bits de poids fort de chacune des composantes RVB (espace de couleur Rouge-Vert-Bleu) d'un pixel sont utilisés.

L'algorithme des k -moyennes est ensuite exécuté sur l'ensemble des vecteurs obtenus en fixant k à 2. Cela revient à regrouper les images clés en deux groupes en minimisant les distances intra-groupes et maximisant les distances inter-groupes. Les images des plateaux sont tellement proches les unes des autres qu'elles sont systématiquement regroupées entre elles. Pour étiqueter les deux groupes obtenus, il suffit de considérer le plus petit groupe comme les plateaux : en effet, (Poli, 2003) évalue à 30% la proportion maximum des plateaux dans un journal télévisé.



FIG. 63: Plateaux de différents journaux télévisés

Les résultats sont ensuite traités de la façon suivante : les plans adjacents dont les images clés portent la même étiquette sont fusionnés. Ensuite, une étape de lissage peut être utilisée pour supprimer les résultats incohérents. Nous avons fixé deux seuils de durée minimale pour un plateau et pour un non-plateau ; si la durée du segment obtenu à la dernière étape est inférieure au seuil, alors celui change d'étiquette. Dans la cadre de nos expérimentations, ces seuils ont été fixés à 2,5 secondes.

2.2 Résultats expérimentaux

Nous avons testé cette méthode sur 7 journaux télévisés issus de différentes collections : le 20 heures de TF1, le 20 heures de France 2 et le flash de 7 heures du *Morning Café* (voir le chapitre précédent). L'algorithme est de faible complexité : les vecteurs d'informations sur les images clés ne contiennent que 64 valeurs, le nombre de plans dans un journal est assez limité (aux alentours de 560 plans en moyenne), ce qui permet d'exécuter un algorithme de clustering dans de très bonnes conditions. L'algorithme des k -moyennes peut ainsi s'exécuter en 5 itérations moyenne.

Les résultats obtenus sont tout à fait satisfaisants : les corrections sont toutes égales à 100 %, les fausses alarmes varient entre 2 et 4, et les précisions temporelles sont inférieures à 2 secondes. Un exemple détaillé est fourni dans l'annexe E (page 207). Nous calculons la correction et la précision temporelle comme indiqué au chapitre précédent. Comme la vérité terrain est à l'image près, nous sommes en mesure de calculer la précision temporelle à l'image près. En revanche, puisque les détections ne sont plus guidées par un modèle, nous avons dû introduire une nouvelle grandeur : le nombre de fausses alarmes. Le nombre de fausses alarmes est le nombre de plateaux

Journal	Correction	Fausses alarmes	Précision temporelle
2 février 2006	100%	2	1,04 s
3 février 2006	100%	4	0,36 s
4 février 2006	100%	2	0,96 s
5 février 2006	100%	0	0,16 s
6 février 2006	100%	4	0,48 s
7 février 2006	100%	4	1,36 s
8 février 2006	100%	0	0,28 s

a) le 20 heures de TF1

Journal	Correction	Fausses alarmes	Précision temporelle
6 mars 2006	100 %	0	0,48 s
7 mars 2006	100 %	4	0,96 s
8 mars 2006	100 %	0	0,36 s
9 mars 2006	100 %	2	0,96 s
10 mars 2006	100 %	0	0,28 s
11 mars 2006	100 %	4	0,88 s
12 mars 2006	100 %	0	0,32 s

b) le 20 heures de France 2

Vidéo	Correction	Fausses alarmes	Précision temporelle
Morning 6	100%	0	0 s
Morning 7	100%	0	0 s
Morning 8	100%	0	0 s
Morning 9	100%	0	0 s
Morning 10	100%	2	0 s
Morning 11	100%	0	0 s
Morning 12	100%	0	0 s

c) le flash du *Morning Café*

TAB. 33: Correction, fausses alarmes et précision temporelle des structures obtenues

et de non-plateaux trouvés dans la vidéo. Ce nombre est forcément pair dans la mesure où si un plateau a été détecté en trop, il y a forcément un non-plateau qui lui est associé. La précision temporelle s'explique par les erreurs du détecteur de « cuts » vis à vis des transitions progressives. Le flash du *Morning Café* n'a pas de transition progressive et sa structure est donc très précise.

Nous avons donc un algorithme très simple qui permet d'extraire tous les plateaux mais qui génère quelques fausses alarmes. Du point de vue opérationnel, des plateaux non détectés rendraient l'algorithme inexploitable alors que les fausses alarmes peuvent être corrigées plus facilement. Toutefois, cet algorithme ne fonctionne pas sur le *12.50* de M6 (voir la figure 63.c) puisque ce journal télévisé à la particulier de diffuser des images de reportage derrière le présentateur.

3 Conclusion

Nous nous sommes intéressés dans ce chapitre à la structuration des tranches horaires et des journaux télévisés. Nous avons choisi d'appliquer la méthode de structuration des flux télévisuels à la structuration du *Morning Café*, une tranche horaire matinale sur M6. Les résultats obtenus sont convaincants. Toutefois, ces résultats doivent être considérés avec précaution : la taille du

corpus n'est pas assez grande pour pouvoir généraliser ces résultats.

En effet, ce corpus est composé des quatorze dernières émissions de la saison 2005–2006. Cela a plusieurs implications sur la capacité de prédiction du modèle : d'une part, la structure peut être particulièrement stable puisque les chroniqueurs ont bien pris l'habitude du minutage de l'émission, et d'autre part aucun évènement ne vient modifier le contenu de l'émission (par exemple *la Nouvelle star* faisait l'objet d'un traitement particulier dans le *Morning Café*). La taille du corpus de test peut aussi expliquer les très bonnes précisions obtenues par les différents détecteurs de jingles.

Nous nous sommes ensuite intéressés aux journaux télévisés pour lesquels la méthode précédente n'est pas applicable puisque la structure d'un journal n'est pas stable. Cependant, les journaux télévisés ont des structures bâties sur l'alternance entre les plateaux et les reportages. Nous avons donc proposé une méthode d'extraction des plateaux qui repose sur un regroupement des images clés des plans en deux groupes. Les expérimentations menées sur trois collections de journaux télévisés montrent que la méthode produit parfois un peu plus de plateaux qu'il n'en existe dans l'émission.

BILAN DE LA QUATRIÈME PARTIE

Nous avons montré dans cette partie que les transitions entre deux émissions pouvaient être caractérisées par des règles basées sur le résultat de détections (silences, images monochromes, plans et logos). La structuration du flux télévisuel consiste ainsi à parcourir le graphe acyclique dirigé et à exécuter les détections liées à la règle correspondant à la transition courante. Le système permet ainsi de piloter des algorithmes de détection à partir des grilles de programmes prédites.

Ce pilotage des algorithmes permet d'employer des détecteurs spécifiques dont le nombre de fausses alarmes et le coût d'exécution sont moins pénalisants grâce à leur exécution locale dans le flux. Nous avons ensuite appliqué cette méthode de structuration des flux à la structuration du *Morning Café*, tranche horaire matinale de M6, dont la particularité est d'être composée de segments très variés. Cela permet d'envisager la structuration du flux comme un processus hiérarchique.

Conclusion

Conclusion

Dans le cadre de l'indexation automatique, et plus précisément de la structuration automatique de contenus audiovisuels, nous proposons une méthode de structuration des flux télévisuels. Il s'agit d'une tâche que des instituts tels que Médiamétrie, pour les chaînes hertziennes, et l'Institut National de l'Audiovisuel, pour les autres chaînes entrant dans le cadre du dépôt légal, exécutent de façon manuelle.

La structuration des flux est un domaine de recherche abordé pour la première fois, à notre connaissance, dans le projet Demi-Ton (voir page 11). C'est pour cela qu'une des étapes préalables à ces travaux a été de poser le problème de la structuration des flux télévisuels. Il s'agit, pour l'INA, d'un catalogage des émissions isolées, des interprogrammes et des tranches horaires diffusés sur une chaîne de télévision.

Nos travaux sont partis du travail du programmeur, c'est-à-dire celui qui, au sein de la chaîne, construit une grille de programmes en fonction du rendez-vous qu'il souhaite créer avec son public cible. Cette étude de la télévision actuelle (seconde partie, page 33) nous a mené à l'observation qui est à la base de nos travaux : sur une période de plusieurs années, les grilles de programmes sont suffisamment stables pour pouvoir les prédire automatiquement. Les notices de l'Inathèque de France nous ont en effet permis d'étudier cette stabilité sur une quinzaine d'années.

Nous avons ainsi proposé un système de structuration des flux télévisuels, en fonction des contraintes imposées par les secteurs opérationnels de l'INA, qui se compose de plusieurs modules. Notre approche consiste à ramener le problème de la structuration des flux télévisuels à un alignement temporel d'une grille de programmes sur ceux-ci.

Dans un premier temps, un module de prédiction permet de prédire l'ensemble des grilles de programmes du flux traité (troisième partie, page 75). Pour cela, nous avons étendu les modèles de Markov cachés afin que l'évaluation des probabilités dépendent du contexte de diffusion d'une émission. Cette extension, appelée modèle de Markov caché contextuel (CHMM), est un cadre formel général mieux adaptés à la production de séquences de par la prise en compte d'un contexte. Nous avons doté ces CHMM d'algorithmes équivalents à ceux des HMM et notamment d'un canevas d'algorithme EM (Espérance-Maximisation) pour leur apprentissage. Dans le cas de la modélisation des grilles de programmes, les CHMM permettent de représenter un genre télévisuel par un état pouvant être atteint dans divers contextes, plutôt que de le représenter par autant d'états que de contextes. La prédiction des grilles de programmes consiste ensuite en un parcours du CHMM et à une mise à jour du contexte de diffusion chaque fois qu'un symbole – une émission – est produit. Le graphe dirigé acyclique (DAG) ainsi construit peut être réduit en utilisant une grille prévisionnelle (par exemple l'Electronic Program Guide) pour forcer le passage par certains nœuds. Afin d'évaluer la probabilité des durées des émissions, nous avons utilisé un arbre de régression capable de prédire les durées minimales, maximales et moyennes des émissions en fonction de leur genre et de leur contexte de diffusion : ces trois paramètres permettent de représenter les probabilités des durées par une loi gaussienne asymétrique. Une

grille de programmes prédite est alors un parcours du DAG d'un bout à l'autre. Une telle modélisation des grilles de programmes, entraînée sur les grilles de programmes des chaînes hertziennes en 2004, permet de prédire 94 % des journées de 2005.

Dans un second temps, le système utilise les grilles de programmes prédites pour retrouver dans le flux les débuts ou les fins des émissions, des interprogrammes et des tranches horaires. Les transitions d'un nœud à l'autre du DAG – par exemple d'un nœud représentant un journal télévisé à un nœud représentant une météo – permettent de mettre en évidence des éléments caractéristiques de ce genre de transition et de n'employer que les détecteurs adéquats (détection d'images monochromes ou de silences, reconnaissance des logos). De plus, l'arbre de régression permet de définir des fenêtres temporelles à partir des valeurs minimales et maximales des durées. Ce n'est qu'à l'intérieur de ces fenêtres temporelles que le système va exécuter les détecteurs ; cela permet non seulement de n'exécuter des détections que sur le tiers du flux en moyenne, mais aussi de diminuer de 80% en moyenne le nombre de fausses alarmes d'un détecteur très sensible (quatrième partie, page 159). Nous avons testé des règles et des détecteurs très simples sur une semaine de France 2 et nous avons retrouvé jusqu'à 98% des segments qui composaient une journée. Les détections multimodales comprenaient un détecteur de silence, un détecteur d'images monochromes et un détecteur de « cuts ». Un réseau à convolution a été employé pour détecter les coupures publicitaires et les bandes annonces en détectant le logo de France 2.

Une fois les segments indépendants isolés dans le flux, nous nous sommes intéressés à leur propre structuration afin d'atteindre le niveau de structure utilisé par l'INA (quatrième partie, page 153). Nous nous sommes intéressés aux journaux télévisés qui sont des émissions dont la structure est fortement contrainte. Nous proposons aussi d'utiliser une approche similaire à celle utilisée pour les grilles de programmes afin de modéliser les conducteurs des tranches horaires telles que *Morning Café*. Il s'agit d'une tranche horaire matinale composée de plusieurs parties hétérogènes. Des détecteurs plus spécifiques sont utilisés et guidés par le modèle afin d'obtenir plus de 93% des segments avec une précision temporelle d'une seconde pour l'ensemble de la tranche horaire.

Les perspectives à court terme sont de reprendre les expérimentations menées sur France 2. En effet, il est possible de rendre le système plus performant en définissant de nouvelles règles pour caractériser les diverses transitions et en leur associant des détecteurs plus spécifiques : à titre d'exemple, un détecteur de logos des chaînes de France Télévisions permettrait de détecter l'auto-promotion inter-chaîne et un détecteur de génériques défilants afin de détecter la fin des émissions nocturnes. Pour expérimenter le système sur d'autres chaînes, il faudra définir d'autres règles ; en effet, notre étude préliminaire sur TF1 et M6 montre que la détection des silences et des images monochromes ne permet de séparer que quelques émissions. Sur M6, par exemple, les émissions sont séparées par des fondus enchaînés très courts – de trois images – détectables facilement. Sur TF1, les cartons de parrainage peuvent être détectés par la présence du logo de la chaîne. La question de la production de ces règles se pose : un expert de l'audiovisuel est nécessaire afin de remarquer tous les éléments d'articulation de l'habillage des chaînes et un expert en informatique doit savoir quels sont les éléments détectables et quels détecteurs il est nécessaire d'appliquer au flux.

Le système que nous proposons pour la structuration des flux du dépôt légal de la télévision peut être adapté à moindre coût au dépôt légal de la radio. En effet, une radio fonctionne plus ou moins comme une chaîne de télévision : une grille de programmes est établie par le programmeur de la radio et un guide de programmes est disponible. De plus, les grilles de programmes des radios sont horizontales, c'est-à-dire qu'on retrouve la même émission tous les jours à la même heure, excepté le week-end où cela peut changer. La définition des règles de transition consistent cette fois encore à la détection des jingles et des génériques des différentes

émissions proposées par la radio.

De plus, le système peut être intégré au schéma de fonctionnement actuel de l'Inathèque. Aujourd'hui, l'Inathèque capte les flux télévisuels, les décompresse, les recomprime, les stocke puis les décrit. La phase de décompression et recompression permet à l'Inathèque de maîtriser l'algorithme d'encodage du flux. Notre système a été créé pour se placer entre la phase de décompression et la phase de réencodage. Une mémoire tampon permettrait d'accumuler les données décompressées du flux. Le système ne serait actif que lorsque suffisamment de données seraient accumulées pour atteindre la fenêtre temporelle correspondant à la prochaine transition à trouver.

Les modèles que nous avons introduits, les CHMM, sont des modèles de Markov adaptés à la production de séquences. De nouvelles applications des CHMM sont envisageables : par exemple, la prédiction de séries temporelles avec un CHMM devrait permettre de tenir compte plus facilement de leur saisonnalité. Ce genre de propriétés doit être montré par une étude spécifique et une comparaison avec les modèles de Markov classiques.

Enfin, nous n'avons pas eu l'occasion d'étudier comment les structurations produites par le système et vérifiées par un opérateur peuvent mettre à jour les modèles statistiques. À plus long terme, la mise à jour du CHMM et de l'arbre de régression seraient une bonne voie à explorer : cela rendrait le système robuste aux changements de programmation, comme celle de la rentrée 2006 (par exemple, la suppression de *Ciné Dimanche* sur TF1 une semaine sur deux). De plus, afin d'améliorer les fenêtres temporelles obtenues par les arbres de régression, ces derniers devraient pouvoir tenir compte de la collection à laquelle l'émission appartient. Par exemple, si l'émission dont on essaie de prédire la durée est un épisode de la série *Friends*, on peut calculer plus précisément les bornes des fenêtres temporelles.

L'ensemble de ces améliorations devraient permettre au système SAFARI d'être testé en situation réelle, c'est-à-dire en installant le système directement à la réception du flux satellite.

Annexes

Annexe A

Nomenclature de Médiamétrie

Le premier niveau de la hiérarchie (dont les codes ont une seule lettre) est très rarement utilisé seul du fait de sa généralité. La catégorie J est utilisée pour quelques émissions exceptionnelles comme les décrochages pour l'opération « pièces jaunes » sur TF1 ou les problèmes techniques. La catégorie Z est une catégorie spécialement ajoutée pour l'INA, sans laquelle les clips seraient aussi du type J.

Code	Désignation
A	FICTION
AA	FILM
AAA	Comédie
AAB	Policier/Espionnage/Suspense
AAC	Aventure
AAD	Science-fiction/Fantastique
AAE	X
AAF	Drame psychologique/Comédie dramatique
AAG	Guerre
AAH	Western
AAI	Historique/Peplum
AAJ	Erotique
AAK	Faits de société
AAL	Films musicaux/Comédie musicale
AAM	Dessins animés
AAN	Documentaires
AB	TELEFILM
ABA	Comédie
ABB	Policier/Espionnage/Suspense
ABC	Aventure
ABD	Science-fiction/Fantastique
ABE	X
ABF	Drame psychologique/Comédie dramatique
ABG	Guerre
ABH	Western
ABI	Historique/Peplum
ABJ	Erotique
ABK	Faits de société
ABL	Films musicaux/Comédie musicale

Code	Désignation
ABM	Dessins animés
ABN	Documentaires
AC	SERIE
ACA	Série policière
ACB	Série humoristique
ACC	Série autre
AD	FEUILLETON
AE	THEATRE
AF	COURT-METRAGE
B	MUSIQUE CLASSIQUE
BA	OPERA/BALLET
BAA	Opéra
BAB	Ballet
BB	CONCERT CLASSIQUE
C	VARIETE DIVERTISSEMENT
CA	EMISSION A BASE DE CLIPS
CAA	Succession de clips
CAB	Clips avec animateur
CAC	Clips + jeu
CB	JEU
CBA	Jeux interactifs
CBB	Jeux non interactifs
CC	HUMOUR
CD	TALK-SHOW
CE	VARIETES
CF	CONCERT MODERNE
CFA	Variétés/rock/...
CFB	Jazz
CG	CIRQUE
D	INFORMATION
DA	JOURNAL NATIONAL
DAA	Avec présentateur
DAB	Sans présentateur
DB	JOURNAL REGIONAL
DBA	Avec présentateur
DBB	Sans présentateur
DC	FLASH
DCA	Avec présentateur
DCB	Sans présentateur
DD	MAGAZINE(économique/politique/sociologique)
DDA	Plateau
DDB	Reportages
DDC	Magazine (SP)
DE	EMISSION POLITIQUE (Campagne électorale...)
DF	EVENEMENT EXCEPTIONNEL (Voyage du Pape...)
DG	METEO
DH	BOURSE
E	CULTURE CONNAISSANCE
EA	DOCUMENTAIRE
EAA	Science et technique
EAB	Médecine
EAC	Nature/animaux

Code	Désignation
EAD	Art/peinture/musique
EAE	Littérature
EAF	Loisir/tourisme/géographie/sport
EAG	Histoire
EAH	Cinéma
EAI	Théâtre
EAJ	Société
EAK	Mode/beauté
EAL	Vie quotidienne/pratique
EAM	Architecture/urbanisme
EAN	Autres
EAO	Spectacles
EAP	Documentaire-feuilleton
EB	MAGAZINE
EBA	Science et technique
EBB	Médecine
EBC	Nature/animaux
EBD	Art/peinture/musique
EBE	Littérature
EBF	Loisir/tourisme/géographie/sport
EBG	Histoire
EBH	Cinéma
EBI	Théâtre
EBJ	Société
EBK	Mode/beauté
EBL	Vie quotidienne/pratique
EBM	Architecture/urbanisme
EBN	Autres(sans précision)
EBO	plateau (invité)
EBP	Spectacle
EC	EMISSION RELIGIEUSE
ED	EMISSION EDUCATIVE
EE	AUTRE
EF	FORMATION PROFESSIONNELLE
F	SPORT
FA	RETRANSMISSION D'ÉVÉNEMENTS SPORTIFS
FAA	Football
FAB	Rugby
FAC	Tennis
FAD	Neige - glace
FAE	Cyclisme
FAF	Golf
FAG	Sports mécaniques
FAH	Athlétisme
FAI	Sports équestres
FAJ	Sports de combat
FAK	Corrida
FAL	Avant/après coulisses
FAM	Autres sports
FB	MAGAZINE SPORTIF (avec extraits)
FBA	Football
FBB	Rugby

Code	Désignation
FBC	Tennis
FBD	Neige - glace
FBE	Cyclisme
FBF	Golf
FBG	Sports mécaniques
FBH	Athlétisme
FBI	Sports équestres
FBJ	Sports de combat
FBK	Corrida
FBL	Autres sports
FBM	Multisports/non spécialisé
FC	MAGAZINE MIXTE
FD	RESULTATS/PRONOSTICS (TIERCE/LOTO SPORTIF)
G	JEUNESSE
GA	EMISSION A BASE DE DESSINS ANIMES
GB	EMISSION JEUNESSE
GC	EDUCATIF
GD	JEUX (JEUNESSE)
GDA	Jeux interactifs
GDB	Jeux non interactifs
GE	DIVERS (PLATEAUX)
H	PUBLICITE
HA	ECRAN PUBLICITAIRE
HB	MAGAZINE PUBLICITAIRE
HC	SPONSORING-PARRAINAGE
I	DIVERS
IA	AUTO PROMOTION DE LA CHAINE
IB	PRESENTATION PROGRAMME
IC	LOCATION D'ANTENNE
ID	HABILLAGE D'ANTENNE
IE	DECROCHAGE REGIONAUX
IF	EMISSION DE SERVICE
IG	LOTERIES
IH	TELE-ACHAT
II	COMMUNIQUES (EXTERIEURS A LA CHAINE)
J	AUTRE
Z	SPECIAL INA
ZA	INTERPROGRAMME PUBLICITE
ZB	INTERPROGRAMME CLIP

Annexe B

Environnement d'apprentissage utilisé

Dans le cadre de nos expérimentations, nous avons utilisé l'environnement d'apprentissage décrit ci-dessous.

Les classes correspondent aux genres télévisuels que nous utilisons à la fois dans le modèle de Markov et dans l'arbre de régression. La classe PRESENTATION PROGRAMME correspond en pratique aux interprogrammes composés de publicités et bandes-annonces tandis que la classe PUBLICITE est utilisée pour représenter les coupures publicitaires isolées. Cette distinction est utile pour la détection des jingles publicitaires marquant le début et la fin des coupures publicitaires alors que le système tentera de détecter le logo de France 2 pour les interprogrammes. Les classes du premier niveau de la hiérarchie de Médiamétrie sont utilisées pour les tranches horaires ; par exemple, *Ciné Dimanche* est classé en FICTION.

FICTION	METEO
FILM	BOURSE
SERIE	CULTURE CONNAISSANCE
FEUILLETON	DOCUMENTAIRE
THEATRE	MAGAZINE
COURT-METRAGE	SPORT
MUSIQUE CLASSIQUE	RETRANSMISSION D'EVENEMENTS SPORTIFS
VARIETE DIVERTISSEMENT	MAGAZINE SPORTIF (avec extraits)
EMISSION A BASE DE CLIPS	RESULTATS/PRONOSTICS (TIERCE/LOTO SPORTIF)
JEU	JEUNESSE
HUMOUR	DIVERS
TALK-SHOW	PRESENTATION PROGRAMME
VARIETES	DECROCHAGE REGIONAUX
CONCERT MODERNE	EMISSION DE SERVICE
CIRQUE	LOTERIES
INFORMATION	TELE-ACHAT
JOURNAL NATIONAL	PUBLICITE
MAGAZINE(éco./politique/socio.)	CLIP

Certains genres ont dû être ignorés pendant les phases d'apprentissage. En effet ces genres

EMISSION POLITIQUE (débat/interview/campagne électorale...)
EVENEMENT EXCEPTIONNEL (Voyage du Pape/soirée électorale)
AUTRES.

ne sont utilisés que ponctuellement et ne font pas partie des habitudes de diffusion de la chaîne.

Tous les autres genres, qui ne sont ni utilisés dans la liste des classes ni dans celle des genres à ignorer, doivent pointer vers une classe ou un genre à ignorer.

Annexe C

Résultat de la structuration du mardi 25 octobre sur France 2

La ligne en gras correspond à un élément non détecté qui a nécessité une intervention de l'opérateur.

Guide	Heure de début		Guide	Etiquette	
	Médiamétrie	Détectée		Base de Connaissance	Etat
05:55:00	05:49:47	05:49:47:05	Les z'amours	Un livre	Magazine
	05:50:54	05:50:53:22			Publicité
	05:53:58	05:53:57:22			Jeu
	06:24:53	06:24:53:03			Interprogramme
06:30:00	06:25:45	06:25:45:10	Télé matin	Point route	Service
	06:26:31	06:26:31:05			Publicité
	06:29:47	06:29:47:07			Magazine d'info
	08:31:23	08:31:23:02			Publicité
08:35:00	08:34:31	08:34:31:06	Des jours et des vies	Point route	Service
	08:35:21	08:35:21:01		Un livre	Magazine
	08:36:36	08:36:36:04			Feuilleton
	08:59:14	08:59:13:12			Interprogramme
09:00:00	09:04:09	09:04:08:18	Amour, gloire et ...		Feuilleton
	09:25:40	09:25:39:18			Publicité
	09:28:01	09:28:00:20		CD'Aujourd'hui	Magazine
	09:30:09	09:30:08:24			Interprogramme
09:30:00	09:30:43	09:30:43:16	KD2A		Jeunesse
10:50:00	10:48:27	10:48:27:00			Interprogramme
	10:51:54	10:51:54:12	Flash		Journal National
	10:54:57	10:54:57:05			Interprogramme
11:00:00	10:58:12	10:58:12:05	Motus		Jeu
	11:29:01	11:29:01:11			Interprogramme
11:30:00	11:34:38	11:34:38:13	Les z'amours		Jeu
	12:05:51	12:05:51:23			Interprogramme
	12:06:43	12:06:43:17		CD'Aujourd'hui	Magazine
	12:08:51	12:08:51:23			Publicité
12:15:00	12:12:48	12:12:48:21	La cible		Jeu
	12:48:37	12:48:37:09			Interprogramme
12:50:00	12:51:58	12:51:58:11	Millionnaire		Jeu

Annexe C. Résultat de la structuration du mardi 25 octobre sur France 2

Guide	Heure de début		Guide	Etiquette Base de Connaissance	Etat
	Médiamétrie	DéTECTÉE			
13:00:00	12:53:29	12:53:29:23	Journal de 13h		Publicité
	12:56:20	12:56:20:23			Météo
	12:58:52	12:58:51:21			Interprogramme
	12:59:16	12:59:15:16			Journal National
	13:43:19	13:43:19:13			Interprogramme
	13:47:27	13:47:27:21			Météo
	13:49:27	13:49:27:10			Interprogramme
13:50:00	13:50:39	13:50:39:11	Expression directe		Magazine
13:55:00	13:53:08	13:53:08:22	Rex		Série
	14:38:07	14:38:07:22			Interprogramme
14:45:00	14:46:45	14:46:45:13	Le renard		Série
	15:44:13	15:44:13:07			Interprogramme
15:50:00	15:50:48	15:50:48:00	Washington police		Série
	16:32:57	16:32:57:00			Interprogramme
	16:33:57	18:08:12:12		Un livre	Magazine
	16:35:12	16:35:12:13			Interprogramme
16:40:00	16:40:37	16:40:37:16	Des chiffres et des lettres		Jeu
	17:07:29	17:07:29:23			Interprogramme
17:15:00	17:13:13	17:13:14:05	Tout vu tout lu		Jeu
	17:49:21	17:49:21:10			Interprogramme
17:55:00	17:54:02	17:54:02:09	Friends		Série
	18:15:29	18:15:29:20			Interprogramme
18:20:00	18:19:19	18:19:19:23	Friends		Série
	18:40:18	18:40:18:13			Interprogramme
	18:41:58	18:41:58:08		CD'Aujourd'hui	Magazine
	18:44:06	18:44:06:12			Publicité
18:50:00	18:48:00	18:48:00:12	On a tout essayé		Magazine
	19:40:05	19:40:05:00			Interprogramme
	19:42:15	19:42:15:23		Mon plus grand...	Magazine
	19:43:51	19:43:51:17			Publicité
19:50:00	19:48:33	19:48:33:16	Un gars, une fille		Série
	19:55:30	19:55:30:00			Interprogramme
	19:55:42	19:55:42:15			Météo
	19:58:05	19:58:05:21			Interprogramme
20:00:00	19:58:59	19:58:59:09	20h le journal		Journal National
	20:35:48	20:35:38:19			Interprogramme
	20:38:59	20:38:59:06			Météo
	20:41:28	20:41:28:02			Interprogramme
20:45:00	20:44:38	20:44:38:06	Plus jamais comme ça		Humour
	20:46:01	20:46:01:11			Publicité
20:50:00	20:50:50	20:50:50:18	Comme au cinéma		Fiction
	00:43:38	00:43:38:00			Interprogramme
	00:50:05	00:50:04:15			Journal National
	01:13:41	01:13:41:23			Météo
01:20:00	01:16:10	01:16:10:03			Publicité
	01:20:28	01:20:28:03		CD'Aujourd'hui	Magazine
	01:22:37	01:22:36:09	Comme au cinéma		Fiction
	03:15:25	03:15:25:00			Publicité
	03:17:20	03:17:20:00	Chanter la vie		Variété
04:05:00	04:07:22	04:07:21:24	30 millions d'amis		Magazine
	04:38:21	04:38:14:18			Journal National
05:00:00	05:01:57	05:01:57:01			Météo
	05:04:27	05:04:27:17	L'enquêteur		Série

Annexe D

Résultat de la structuration d'un *Morning Café*

Pour faciliter la fusion des différents détecteurs utilisés, nous avons arrondi les résultats des détecteurs audio en considérant des fenêtres de 40 ms (temps d'exposition d'une image lorsque le débit est de 25 images par seconde). Nous pouvons ainsi noter les différentes heures sous la forme hh:mm:ss:ii, indiquant ainsi le nombre d'heures (hh), de minutes (mm), de secondes (ss) et d'images (ii).

Les vidéos du *Morning Café* commencent à 6h50 et finissent à 9h15. Les heures indiquées sont les heures relatives au début du fichier.

Vérité terrain		Structuration automatique		Précision temporelle cumulée
Début	Etiquette	Début	Etiquette	
00:00:00:00	Autre	00:00:00:00	Autre	00:00:00:01
00:14:11:02	Fondu enchaîné			
00:14:11:08	Flash Info	00:14:11:07	Flash Info	
00:17:16:01	Météo	00:17:16:01	Météo	
00:18:37:18	La route en direct	00:18:37:18	La route en direct	
00:19:37:12	Interprogrammes	00:19:37:12	Interprogrammes	
00:21:51:16	Plateau	00:21:51:16	Plateau	00:00:00:06
00:24:07:07	Fondu enchaîné			
00:24:07:18	Clip musical	00:24:07:23	Clip musical	
00:27:15:23	Fondu enchaîné			
00:27:16:09	Plateau	00:27:16:00	Plateau	
00:34:16:11	Kaamelott	00:34:16:11	Kaamelott	
00:38:05:03	Plateau	00:38:05:03	Plateau	00:00:00:15
00:42:28:18	Interprogrammes	00:42:28:18	Interprogrammes	
00:44:39:17	Autre	00:44:39:17	Autre	
00:45:06:20	Plateau	00:45:06:20	Plateau	
00:45:28:23	Flash Info	00:45:28:23	Flash Info	
00:48:25:00	Plateau	00:48:25:00	Plateau	
00:56:56:05	Fondu enchaîné			00:00:00:15
00:56:56:09	Clip musical	00:56:56:09	Clip musical	
00:58:58:12	Fondu enchaîné			
00:58:58:20	Plateau	00:58:58:20	Plateau	
01:03:39:06	Fondu enchaîné			
01:03:39:15	Clip musical	01:03:39:15	Clip musical	
01:05:35:09	Fondu enchaîné			

Annexe D. Résultat de la structuration d'un Morning Café

Vérité terrain		Structuration automatique		Précision temporelle cumulée
Début	Etiquette	Début	Etiquette	
01:05:35:21	Plateau	01:06:35:10	Plateau	00:00:01:01
01:12:24:17	Fondu enchaîné			
01:12:25:10	Clip musical	01:12:25:10	Clip musical	
01:14:15:05	Fondu enchaîné			
01:14:15:09	Plateau	01:14:15:09	Plateau	
01:14:52:15	Flash Info	01:14:52:15	Flash Info	
01:17:58:01	Météo	01:17:58:01	Météo	
01:19:19:18	La route en direct	01:19:19:18	La route en direct	
01:20:20:01	PlateauPlateau	01:20:20:01	PlateauPlateau	
01:21:04:06	Interprogrammes	01:21:04:06	Interprogrammes	
01:23:20:02	Fondu enchaîné			
01:23:20:06	Plateau	01:23:20:06	Plateau	
01:25:32:17	Fondu enchaîné			
01:25:33:00	Clip musical	01:25:33:00	Clip musical	
01:27:27:01	Fondu enchaîné			
01:27:27:11	Plateau	01:27:11:19	Plateau	00:00:01:19
01:34:15:19	Kaamelott	01:34:15:19	Kaamelott	
01:38:03:21	Plateau	01:38:03:21	Plateau	
01:41:05:17	Interprogrammes	01:41:05:17	Interprogrammes	
01:43:18:22	Fondu enchaîné			
01:43:19:01	Plateau	01:43:19:01	Plateau	
01:43:43:06	Flash Info	01:43:43:06	Flash Info	
01:46:40:21	Plateau	01:46:40:21	Plateau	
01:53:14:18	Fondu enchaîné			
01:53:15:16	Clip musical	01:53:15:16	Clip musical	
01:55:52:05	Fondu enchaîné			
01:55:52:10	Plateau	01:55:52:10	Plateau	
02:02:05:11	Fondu enchaîné			
02:02:05:18	Clip musical	02:02:05:18	Clip musical	
02:04:18:12	Fondu enchaîné			
02:04:18:19	Plateau	02:04:18:19	Plateau	
02:15:24:06	Flash Info	02:15:24:06	Flash Info	
02:18:13:03	Générique	02:18:13:03	Générique	
02:18:36:08	Interprogrammes	02:18:36:08	Interprogrammes	
02:23:37:18	Météo	02:23:37:18	Météo	00:00:01:19

Annexe E

Résultat de la structuration d'un journal de 20 heures sur France 2

Le journal télévisé est contenu dans un fichier séparé. Les heures indiquées sont donc relatives au début de ce fichier. Le type de segment « AUTRE » désigne les génériques et les éléments de la vidéo ne faisant pas partie du journal télévisé. Les résultats et la vérité terrain étant à l'image près, la notation hh:mm:ss:ii désignera le nombre d'heures (hh), le nombre de minutes (mm), le nombre de secondes(ss) et le nombre d'images (ii). De plus, la vérité terrain tient compte des transitions progressives qui sont des frontières imprécises entre un segment du type plateau et un segment de type non-plateau. La dernière colonne indique la précision temporelle cumulée de la structure produite automatiquement.

Vérité terrain		Structuration automatique		Précision temporelle cumulée
Début	Etiquette	Début	Etiquette	
00:00:00:00	Autre	00:00:00:00	Autre	00:00:00:00
00:00:04:13	Plateau	00:00:04:13	Plateau	
00:00:08:22	Sommaire	00:00:08:22	Non-plateau	
00:01:31:01	Fondu enchaîné			00:00:00:02
00:01:31:10	Plateau	00:01:31:08	Plateau	
00:01:53:22	Reportage	00:01:53:22	Reportage	
00:03:28:14	Plateau	00:03:28:14	Plateau	
00:03:40:03	Reportage	00:03:40:03	Reportage	
00:04:07:17	Plateau	00:04:07:17	Plateau	
00:04:18:19	Reportage	00:04:18:19	Reportage	
00:06:03:20	Plateau	00:06:03:20	Plateau	
00:06:06:13	Reportage	00:06:06:13	Reportage	
00:06:25:05	Fondu enchaîné			
00:06:25:14	Plateau	00:06:25:11	Plateau	
00:06:47:03	Reportage	00:06:47:03	Reportage	
00:08:22:23	Plateau	00:08:22:23	Plateau	
00:08:30:00	Reportage	00:08:30:00	Reportage	
00:10:10:21	Plateau	00:10:10:21	Plateau	
00:10:21:15	Reportage	00:10:21:15	Reportage	
00:10:48:03	Plateau	00:10:48:03	Plateau	
00:11:03:22	Reportage	00:11:03:22	Reportage	
00:12:24:03	Plateau	00:12:24:03	Plateau	
00:12:27:11	Reportage	00:12:27:11	Reportage	
00:12:57:08	Plateau	00:12:57:08	Plateau	

Annexe E. Résultat de la structuration d'un journal de 20H sur France 2

Vérité terrain		Structuration automatique		Précision temporelle cumulée
Début	Etiquette	Début	Etiquette	
00:13:01:12	Reportage	00:13:01:12	Reportage	
00:13:19:01	Fondu enchaîné			
00:13:19:07	Plateau	00:13:19:07	Plateau	
00:13:41:00	Reportage	00:13:41:00	Reportage	
00:17:20:14	Plateau	00:17:20:14	Plateau	
00:17:23:03	Reportage	00:17:23:03	Reportage	
00:17:51:03	Fondu enchaîné			
00:17:51:09	Plateau	00:17:51:09	Plateau	
00:17:55:18	Reportage	00:17:55:18	Reportage	
00:18:06:16	Plateau	00:18:06:16	Plateau	
00:18:14:21	Reportage	00:18:14:21	Reportage	
00:20:00:19	Plateau	00:20:00:19	Plateau	
00:20:07:13	Reportage	00:20:07:13	Reportage	
00:20:14:03	Plateau	00:20:14:03	Plateau	
00:20:22:04	Reportage	00:20:22:04	Reportage	
00:21:41:07	Plateau	00:21:41:07	Plateau	
00:21:43:22	Reportage	00:21:43:22	Reportage	
00:21:57:11	Fondu enchaîné			
00:21:57:17	Plateau	00:21:57:16	Plateau	00:00:00:06
00:22:07:12	Reportage	00:22:07:12	Reportage	
00:23:32:04	Plateau	00:23:32:04	Plateau	
00:23:47:23	Reportage	00:23:47:23	Reportage	
00:24:07:13	Fondu enchaîné			
00:24:07:19	Plateau	00:24:07:18	Plateau	00:00:00:07
00:24:16:09	Reportage	00:24:16:09	Reportage	
00:25:38:08	Plateau	00:25:38:08	Plateau	
00:26:01:03	Reportage	00:26:01:03	Reportage	
00:27:37:06	Plateau	00:27:37:06	Plateau	
00:27:55:11	Reportage	00:27:55:11	Reportage	
00:29:54:19	Plateau	00:29:54:19	Plateau	
00:30:11:17	Fondu enchaîné			
00:30:11:23	Reportage	00:30:11:18	Reportage	00:00:00:12
00:31:43:05	Plateau	00:31:43:05	Plateau	
00:31:50:01	Reportage	00:31:50:01	Reportage	
00:33:20:18	Plateau	00:33:20:18	Plateau	
00:33:41:24	Reportage	00:33:41:24	Reportage	
00:35:37:11	Plateau	00:35:37:11	Plateau	
00:35:40:08	Reportage	00:35:40:08	Reportage	
00:36:17:05	Fondu enchaîné			
00:36:17:11	Autre	00:36:17:11	Autre	00:00:00:12

Glossaire

Grille de programmes :

les grilles de programmes sont des représentations sur une semaine de la répartition des programmes en fonction des jours et des heures. La grille de programmes *a priori* est produite par le programmeur de la chaîne et la grille de programmes *a posteriori* est un horodatage précis de chaque émission après diffusion (aussi appelé conducteur d'antenne).

Grille de programmes prévisionnelle :

il s'agit d'une vue partielle de la grille de programmes *a priori*. Seuls les programmes fédérateurs sont indiqués à des horaires imprécis.

Guide de programmes :

il s'agit d'une documentation sur la grille de programmes contenant une grille prévisionnelle, des résumés et des informations sur les programmes principaux. Les guides de programmes se trouvent soit dans la presse spécialisée (*Télé 7 jours*, *Télérama*), soit par voie électronique ; on parle alors d'EPG (Electronic Program Guides).

Bibliographie

- L. AGNIHOTRI, N. DIMITROVA, T. MCGEE, S. J. MIN, S. SCHAEFFER, et J. NESVADBA. Evolvable visual commercial detector. *Computer Vision and Pattern Recognition*, 25:79–84, 2003.
- P. AIGRAIN, P. JOLY, et V. LONGUEVILLE. Medium knowledge-based macrosegmentation of video into sequences. *Intelligent Multimedia Information Retrieval*, pages 159–173, août 1995.
- A. A. ALATAN, A. N. AKANSU, et W. WOLF. Incorporating audio cues into dialog and action scene detection. *ACM Multimedia Tools and Applications*, 14:137–151, 2001.
- A. ALBIOL, M.J. FULLA, A. ALBIOL, et L. TORRES. Commercials detection using hmms. In *Proceedings of the International Workshop on Image Analysis for Multimedia Interactive Services*, avril 2004.
- A. ALBIOL, M.J. FULLA, A. ALBIOL, et L. TORRES. Detection of tv commercials. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, avril 2004.
- H. AOKI, S. SHIMOTSUJI, et O. HORI. A shot classification method of selecting effective key frames for video browsing. In *Proceedings of ACM Multimedia 1996*, novembre 1996.
- J. ASSFALG, M. BERTINI, A. DEL BIMBO, W. NUNZIATI, et P. PALA. Soccer highlights detection and recognition using hmm's. In *Proceedings of IEEE International Conference on Multimedia and Expo 2002*, pages 825–828, août 2002.
- N. BABAGUCHI, Y. KAWAI, et T. KITAHASHI. Event based indexing of broadcasted sports video by intermodal collaboration. *IEEE Transactions on Multimedia*, pages 68–75, 2002.
- B. BACHIMONT. Indexation audiovisuelle : une problématique en pleine évolution. *L'objet*, 6(2):171–190, juin 2000.
- B. BACHIMONT. *L'indexation multimédia*, Chapitre 1, pages 15–49. Hermes Science Publications, 2003.
- S. BENGIO et Y. BENGIO. An em algorithm for asynchronous input/output hidden markov models. In *Proceedings of the International Conference on Neural Information Processing*, pages 328–334, septembre 1996.
- Y. BENGIO et P. FRASCONI. An input/output hmm architecture. In *Advances in Neural Information Processing Systems 7*, pages 427–434. MIT Press, novembre 1995.
- Y. BENGIO et P. FRASCONI. Input-output hmm's for sequence processing. *IEEE Transactions on Neural Networks*, 7:1231–1249, 1996.

- P. N. BENNETT. Using asymmetric distributions to improve text classifier probability estimates. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, juillet 2003.
- C.M. BISHOP. *Neural networks and machine learning*. Springer-Verlag Telos, 1999.
- J.S. BORECZKY et L.A. ROWE. Comparison of video shots boundary detection techniques. In *Storage and retrieval for image and video databases*, pages 170–179. SPIE, 1996.
- P. BOUFONOS, S. EL-DIFRAWY, et D. EHRLICH. Hidden markov models for dna sequencing. In *Workshop on Genomic Signal Processing and Statistics 2002*, octobre 2002.
- M. BOULLE. Khiops: a statistical discretization method of continuous attributes. *Machine Learning*, 55(1):53–69, 2004.
- H. BOURLARD et S. BENGIO. *Hidden Markov Models and other finite state automata for sequence processing*. MIT Press, deuxième édition édition, 2002.
- H. BOURLARD et N. MORGAN. *Connectionist speech recognition : a hybrid approach*. Kluwer Academic Press, 1993.
- L. BREIMAN, J.H. FRIEDMAN, R.A. OLSHEN, et C.J. STONE. Classification and regression trees (cart). Rapport Technique, Wadsworth International Group, 1984.
- P. BROWNE, A. SMEATON, N. MURPHY, N. O’CONNOR, S. MARLOW, et C. BERRUT. Evaluating and combining digital video shot boundary detection algorithms. In *Irish machine vision and image processing conference*, septembre 2000.
- R. BRUNELLI, O. MICH, et C.M. MODENA. A survey on the automatic indexing of video data. *Journal of Visual and Image Representation*, 10:78–112, 1999.
- J. BURPIN. La programmation télévisuelle : une arme stratégique dans la guerre des chaînes. *COMMPosite*, page <http://commposite.org/2004/articles/burpin.html>, 2004.
- R. CABASSON et A. DIVAKARAN. Automatic extraction of soccer video highlights using a combination of motion and audio features. In *Symposium on Electronic Imaging: Science and Technology: Storage and Retrieval for Media Databases*, pages 272–276, janvier 2002.
- M. CAILLET, J. CARRIVE, V. BRUNIE, et C. ROISIN. Pace : an experimental web based audio-visual application using FDL. In *Web Document Analysis 2005*, août 2005.
- J. CARRIVE, F. PACHET, et R. RONFARD. Clavis - a temporal reasoning system for classification of audiovisual sequences. In *Proceedings of Content-Based Multimedia Information Access (RIAO) Conference*, Paris, France, 2000.
- P. CHANG, M. HAN, et Y. GONG. Extract highlights from baseball game video with hidden markov models. In *Proceedings of IEEE International Conference on Image Processing*, pages 609–612, septembre 2002.
- R. CHANIAC et J.P. JÉZÉQUEL. *La télévision*. Repères. La découverte, 2005.
- P. CHARAUDEAU. Les conditions d’une typologie des genres télévisuels d’information. *Réseaux*, 81, 1997.

-
- CSA. *Publicité, parrainage et télé-achat à la télévision et à la radio*. Les brochures du CSA, www.csa.fr, juillet 2006.
- M. DELAKIS, G. GRAVIER, et P. GROS. Multimodal segmental-based modeling of tennis video broadcasts. In *Proceedings of IEEE international Conference on Multimedia and Expo*, pages 546–549, juillet 2005.
- M. DELAKIS, G. GRAVIER, et P. GROS. Score oriented viterbi search in sport video structuring using hmm and segment models. In *Proceedings of the International Workshop on Multimedia Signal Processing*, octobre 2006.
- J. DELLER, J. PROAKIS, et J. HANSEN. *Discrete-time processing of speech signals*. MacMillan, 1993.
- C.H. DEMARTY. *Segmentation et structuration d'un document vidéo pour la caractérisation et l'indexation de son contenu sémantique : application aux journaux télévisés*. Thèse de doctorat, Ecole des mines de Paris, 2000.
- A.P. DEMPSTER, N.M. LAIRD, et D.B. RUBIN. Maximum-likelihood from incomplete data via the em algorithm. *Journal of Royal Statistical Society*, 39:1–38, 1977.
- N. DIMITROVA, S.J. NIN, J. NESVADBA, T. MCGEE, L. AGNIHOTRI, et G. MEKENKAMP. Real-time commercial detection using mpeg features. In *Proceedings of the 9th International Conference On Information Processing and Management of Uncertainty in knowledge-based systems*, pages 481–486, juillet 2002.
- A. DIVAKARAN, K. MIYARAH, K.A. PEKER, R. RADHAKRISHNAN, et Z. XION. Video mining using combinations of unsupervised and supervised learning techniques. In *Proceedings of SPIE Conference on Storage and Retrieval for Multimedia Databases*, pages 235–243, janvier 2004.
- S. DUFFNER et C. GARCIA. A neural scheme for robust detection of transparent logos in tv programs. In *Proceedings of the 16th International Conference on Artificial Neural Networks*, pages 14–23, septembre 2006.
- P. DUPONT, F. DENIS, et Y. ESPOSITO. Links between probabilistic automata and hidden markov models : probability distributions, learning models and induction algorithms. *Pattern recognition*, 38(9):1349–1371, 2005.
- P. DUYGULU, M. CHEN, et A. HAUPTMANN. Comparison and combination of two novel commercial detection methods. In *Proceedings of IEEE International Conference on Multimedia and Expo*, pages 1267– 1270, juin 2004.
- S. ELHIHI et Y. BENGIO. Hierarchical recurrent neural networks for long-term dependencies. In *Advances in Neural Information Processing Systems 8*, pages 493–499. MIT Press, juin 1996.
- T. ELOMAA et J. ROUSU. Fast minimum training error discretization. In *Proceedings of the nineteenth International Conference on Machine Learning*, pages 131–138, juillet 2002.
- S.E. FALHMAN et C. LEBIÈRE. The cascade-correlation learning architecture. In *Neural Information Processing Systems*, volume 2, pages 524–532, 1990.

- B. FAUVET, P. BOUTHEMY, P. GROS, et F. SPINDLER. A geometrical key-frame selection method exploiting dominant motion estimation in video. In *Int. Conf. on Image and Video Retrieval, CIVR 2004*, volume 3115 de *Lecture Notes in Computer Science*, pages 419–427, Dublin, Eire, juillet 2004.
- J.D. FERGUSON. Variable duration models for speech. In *Proceedings of the Symposium on Application of HMMs to text and speech*, pages 143–179, octobre 1980.
- S. FINE, Y. SINGER, et N. TISHBY. The hierarchical hidden markov model : analysis and applications. *Machine learning*, 32(1):41–62, 1998.
- S. FISCHER, R. LIENHART, et W. EFFELSBERG. Automatic recognition of film genres. In *Proceedings of the third ACM international conference on Multimedia*, pages 295–304, novembre 1995.
- L. FONNET. *La programmation d'une chaîne de télévision*. Dixit, 2003.
- G.D. FORNEY. The viterbi algorithm. *Proceedings of IEEE*, 61(3):263–278, 1973.
- M.J.F. GALES et S.J. YOUNG. The theory of segmental hidden markov models. Rapport Technique, Université de Cambridge, 1993.
- C. GARCIA et M. DELAKIS. Convolutional face finder: a neural architecture for fast and robust face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11):1408–1423, 2004.
- J.M. GAUCH et A. SHIVADAS. Finding and identifying unknown commercials using repeated video sequence detection. *Computer Vision and Image Understanding*, pages 80–88, juin 2006.
- J.M. GAUCH et A. SHIVIDAS. Identification of new commercials using repeated video sequence detection. In *Proceedings of IEEE International Conference on Image Processing*, septembre 2005.
- R. GLASBERG, A. SAMOUR, K. ELAZOUZI, et T. SIKORA. Cartoon-recognition using video and audio-descriptors. In *Proceedings of European Signal Processing Conference*, septembre 2005.
- R. GLASBERG, A. SAMOUR, K. ELAZOUZI, et T. SIKORA. Cartoon-recognition using visual-descriptors and a multilayer-perceptron. In *Proceedings of Workshop on Image Analysis for Multimedia Interactive Services*, avril 2005.
- G. GRAVIER, M. SIGELLE, et G. CHOLLET. A markov random field model for automatic speech recognition. In *Proceedings of the 15th international conference on pattern recognition*, pages 254–257, septembre 2000.
- P. GROS. *L'indexation multimédia – description et recherche automatiques*, Chapitre Introduction, pages 15–19. Hermes, 2007.
- J.W. GRZYMALA-BUSSE. Three strategies to rule induction from data with numerical attributes. *Lecture Notes in Computer Science*, 3135:54–62, 2004.
- M. GUIRONNET, D. PELLERIN, et M. ROMBAUT. Video classification based on low-level feature fusion model. In *Proceedings of European Signal Processing Conference*, septembre 2005.

-
- B. GUNSEL, A. FERMAN, et A. TEKALP. Temporal video segmentation using unsupervised clustering and semantic object tracking. *Journal of Electronic Imaging*, 7(3):592–604, 1998.
- A.G. HAUPMANN et M.J. WITBROCK. Story segmentation and detection of commercials in broadcast news video. In *Proceedings of the Advances in Digital Libraries Conference*, pages 168–180, mai 1998.
- E. HOOG. *L'INA. Que sais-je?* Presses universitaires de France, 2006.
- X.S. HUA, L. LU, et H.J. ZHANG. Robust learning-based tv commercial detection. In *Proceedings of IEEE International Conference on Multimedia and Expo*, juillet 2005.
- J. HUANG, Z. LIU, et A. ROSENBERG. Automated semantic structure reconstruction and representation generation for broadcast news. In *Proceedings of SPIE Storage and Retrieval for Image and Video Databases VII*, pages 50–62, janvier 1999.
- J. HUANG, Z. LIU, et Y. WANG. Integration of audio and visual information for content-based video segmentation. In *Proceedings of IEEE International Conference on Image Processing*, octobre 1998.
- S. HUET, A. BOUVIER, M.A. GRUET, et M.A. JOLIVET. *Statistical tools for nonlinear regression (a practical guide with S-plus examples)*. Series in Statistics. Springer, 1996.
- S. HUET, G. GRAVIER, et P. SÉBILLOT. Are morphosyntactic taggers suitable to improve automatic transcription? In *Proceedings of Text, Speech, and Dialogue*, septembre 2006.
- S. HUET, P. SÉBILLOT, et G. GRAVIER. Introduction de connaissances linguistiques en reconnaissance de la parole : un état de l'art. Rapport Technique 1804, IRISA, mai 2006.
- M.V. IBÁNEZ et A. SIMÓ. Parameter estimation in markov random field image modeling with imperfect observations : a comparative study. *Patter recognition letters*, 24(14):2377–2389, 2000.
- INA. *Les dossiers de l'audiovisuel*, volume 41. INA, 1992.
- INA. Connaissance des fonds : histoire des programmes par genre. Rapport Technique, Institut National de l'Audiovisuel, 2002.
- INA. *Les nouveaux dossiers de l'audiovisuel*, volume 4. INA, mai 2005.
- INA. *Les nouveaux dossiers de l'audiovisuel*, volume 11. INA, juillet 2006.
- A. ISAAC et R. TRONCY. Designing an audio-visual description core ontology. In *Workshop on Core Ontologies in Ontology Engineering, 14th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2004)*, 2004.
- D. JANSSENS, T. BRIJS, K. VANHOOF, et G. WETS. Evaluating the performance of cost-based discretization versus entropy- and error-based discretization. *Computers and Operations Research*, 33(11):3107–3123, 2006.
- F. JELINEK. *Statistical methods for speech recognition*. MIT Press, 1998.
- F. JOST. *La télévision du quotidien*. De Boeck Université, 2001.

- B.H. JUANG et L.R. RABINER. The segmental k-means algorithm for estimating the parameters of hidden markov models. In *IEEE Transaction on Acoustics, Speech and Signal Processing*, pages 1639–1641, septembre 1990.
- A. KARALIC. Employing linear regression in regression tree leaves. In *Proceedings of the european congress on artificial intelligence 1992*, pages 440–441, août 1992.
- E. KIJAK, L. OISEL, et P. GROS. Audiovisual integration for tennis broadcast structuring. In *Proceedings of the IEEE Third International Workshop on Content-Based Multimedia Indexing 2003*, 2003.
- E. KIJAK, L. OISEL, et P. GROS. Hierarchical structure analysis of sport videos using hmms. In *IEEE International Conference on Image Processing, ICIP'03*, volume 2, pages 1025–1028. IEEE Press, 2003.
- S. KWONG, C. CHAU, K. MAN, et K. TANG. Optimisation of hmm topology and its model parameters by genetic algorithms. *Pattern recognition*, 34:509–522, 2001.
- L.J. LATECKI, V. MEGALOOIKONOMOU, Q.W., R. LAKAEMPER, C. A. RATANAMAHATANA, et E. KEOGH. Partial elastic matching of time series. In *Proceedings of the IEEE International Conference on Data Mining*, pages 701–704. IEEE Computer Society, novembre 2005.
- M. LAZARESCU, S. VENKATESH, et G WEST. On the automatic indexing of cricket using camera motion parameters. In *Proceedings of IEEE International Congerence on Multimedia and Expo 2002*, pages 809–812, août 2002.
- Y. LECUN, L. BOTTOU, Y. BENGIO, et P. HAFFNER. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- S.E. LEVINSON. Continuously variable duration hidden markov models for automatic speech recognition. *Computer speech and language*, 1:29–45, 1986.
- LI et M.I. SEZAN. Event detection and summarization in american football broadcast video. In *Symposium on Electronic Imaging: Science and Technology: Storage and Retrieval for Media Databases*, pages 202–213, janvier 2002.
- S.Z. LI. *Random Markov Fields modeling in image analysis*. Springer-Verlag, 2001.
- Y. LI et C.C. JAY KUO. *Video content analysis using multimodal information for movie content extraction, indexing and representation*. Kluwer Academic Publishers, 2003.
- X. LI-QUN et L. YONGMIN. Video classification using spatial-temporal features and pca. In *Proceedings of IEEE International Conference on Multimedia and Expo*, pages 485–488, juillet 2003.
- L. LIANG, H. LU, X. XUE, et Y.P. TAN. Program segmentation for tv videos. In *IEEE International Symposium on Circuits and Systems*, pages 1549–1552, mai 2005.
- R.W. LIENHART. Comparison of automatic shots boundary detection algorithms. In *Proceedings of the SPIE Storage and Retrieval for Still Image and Video Databases*, pages 290–301, décembre 1998.

-
- R. LIENHART, C. KUHMÜNCH, et W. EFFELSBERG. On the detection and recognition of television commercials. In *Proceedings of the 1997 International Conference on Multimedia Computing and Systems*, pages 509–516, juin 1997.
- R. LIENHART, S. PFEIFFER, et W. EFFELSBERG. Scene determination based on video and audio features. In *Proceedings of IEEE Conference on Multimedia Computing and Systems*, pages 685–690, juin 1999.
- S.P. LLOYD. Least squares quantization in pcm. *IEEE Transactions Information Theory (Special Issue on Quantization)*, 28(2):129–137, 1982.
- L.TORGO. Functional models for regression tree leaves. In *Proceedings of the fourteenth International conference on machine learning*, pages 385–393, juillet 1997.
- C. LU, M.S. DREW, et J. AU. Classification of summarized videos using hidden markov models on compressed chromaticity signatures. In *Proceedings of ACM Multimedia*, pages 479–482, septembre 2001.
- E. MACÉ. *La société et son double - une journée ordinaire de télévision*. Médiacultures. Armand Colin, 2006.
- M. MAYBURY, M. MERLINO, et J. RAYSON. Segmentation, content extraction and visualization of broadcast news video using multistream analysis. In *Proceedings of ACM multimedia 1996*, 1996.
- W. MCCULLOCH et W. PITTS. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943.
- K. MEISINGER, T. TROEGER, M. ZELLER, et A. KAUP. Automatic tv logo removal using statistical based logo detection and frequency selective inpainting. In *Proceedings of the 13th of the European Signal Processing Conference*, septembre 2005.
- A. MERLINO, D. MOREY, et M. MAYBURY. Broadcast news navigation using story segmentation. In *Proceedings of ACM Multimedia 1997*, 1997.
- M. MEYER. *Le livre noir de la télévision*. Grasset, 2006.
- M.L. MINSKY et S.PAPERT. *Perceptrons : an introduction to computational geometry*. MIT Press, 1969.
- C.D. MITCHELL, M.P. HARPER, et L.H. JAMIESON. On the complexity of explicit duration hmms. *IEEE transactions on speech and audio processing*, 3(3):213–217, 1995.
- M. MIZUTANI, M. EBADOLLAHI, et S.F. CHANG. Commercial detection in heterogeneous video streams using fused multi-modal and temporal features. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 157–160, mars 2005.
- S. MONCRIEFF, C. DORAI, et S. VENKATESH. Detecting indexical signs in film audio for scene interpretation. In *Proceedings of 2001 IEEE International Conference on Multimedia and Expo*, pages 251–254. IEEE Computer Society, août 2001.
- D. MOORE et I. ESSA. Recognizing multitasked activities using stochastic context-free grammar. In *Proceedings of the 18th national conference on Artificial Intelligence*, pages 770–776, juillet 2002.

- J. MOUSSEAU. La programmation d'une chaîne de télévision. *Communication et langage*, 80:74–89, 1989.
- K. MURPHY. *Dynamic Bayesian Networks: representation, inference and learning*. Thèse de doctorat, Université de Californie - Berkeley, 2002.
- C.S. MYERS et L.R. RABINER. A comparative study of several dynamic time-warping algorithms for connected word recognition. *The Bell system technical journal*, 60(7):1389–1409, 1981.
- J. NAM, M. ALGHONIEMY, et A. H. TEWFIK. Audio-visual content-based violent scene characterization. In *Proceedings of International Conference on Image Processing*, pages 351–357, octobre 1998.
- J. NAM, J. CETIN, et A.H. TEWFIK. Speaker identification and video analysis for hierarchical video shot classification. In *Proceedings of International Conference on Image Processing 1997*, octobre 1997.
- X. NATUREL, G. GRAVIER, et P. GROS. Étiquetage automatique de programmes de télévision. In *Compression et représentation des signaux audiovisuels*, pages 17–22, novembre 2005.
- X. NATUREL et P. GROS. A fast shot matching strategy for detecting duplicate sequences in a television stream. In *Proceedings of the 2nd ACM SIGMOD International Workshop on Computer Vision meets DataBases*, juin 2005.
- X. NATUREL et P. GROS. Detecting repeats for video structuring. Rapport Technique 1790, IRISA, mars 2006.
- X. NATUREL et P. GROS. Fast structuring of large television streams using program guides. In *Proceedings of the 4th International Workshop on Adaptive Multimedia Retrieval*, juillet 2006.
- S. NEPAL, U. SRINIVASAN, et G. REYNOLDS. Automatic detection of goal segments in basketball videos. In *Proceedings of ACM Multimedia*, pages 261–269, septembre 2001.
- H.S. NGUYEN. *On exploring soft discretization of continuous attributes*, Chapitre 13, pages 333–350. Cognitive technologies. Springer-Verlag, 2003.
- J.R. NORRIS. *Markov chains*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, 1999.
- M. OSTENDORF, V. DIGALAKIS, et O. KIMBALL. From hmms to segment models: a unified view of stochastic modeling for speech recognition. *IEEE Transactions on speech and audio processing*, 4(5):360–378, 1996.
- L. PERSONNAZ et I. RIVALS. *Réseaux de neurones formels pour la modélisation, la classification et la commande*. CNRS Editions, 2003.
- C. PETERSOHN. Fraunhofer heinrich hertz institute at trecvid 2004: shot boundary detection system. In *Proceedings of TRECVID 2004*, 2004.
- M. PETKOVIC, V. MIHAJLOVIC, M. JONKER, et S. DJORDJEVIC-KAJAN. Multi-modal extraction of highlights from tv formula 1 programs. In *Proceedings of IEEE International Congerence on Multimedia and Expo 2002*, pages 817–820, août 2002.

-
- J. PINQUIER. *Indexation sonore : recherche de composantes primaires pour une structuration audiovisuelle*. Thèse de doctorat, Université Toulouse III, 2004.
- J. PINQUIER et R. ANDRÉ-OBRECHT. Jingle detection and identification in audio documents. In *ICASSP'2004*, pages 329–332, mai 2004.
- J. PINQUIER et R. ANDRÉ-OBRECHT. Audio classification tools for multimedia indexing. In *Fourth International Workshop on Content-Based Multimedia Indexing*, juin 2005.
- J. PINQUIER et R. ANDRÉ-OBRECHT. Audio indexing : primary components retrieval - robust classification in audio documents. *Multimedia Tools and Applications*, 30(3):313–330, septembre 2006.
- A. POISSON, C. VIARD-GAUDIN, P.M. LALLICAN, A. BELAID, et H. Ben GHÉZALA. Combinaison de réseaux de neurones à convolution pour la reconnaissance de caractères manuscrits en-ligne. In *Colloque international francophone sur l'écrit et le document 2002*, pages 315–324, octobre 2002.
- J.P. POLI. Segmentation plateaux/reportages de journaux télévisés. Rapport de DEA, Université Pierre et Marie Curie, 2003.
- J.P. POLI. Predicting program guides for video structuring. In *Proceedings of the 17th IEEE International Conference on Tools with Artificial Intelligence*, pages 407–411, novembre 2005.
- J.P. POLI et J. CARRIVE. Proposition d'une architecture pour un système de structuration automatique de flux audiovisuels. In *Actes des journées CORESA 2005*, novembre 2005.
- J.P. POLI et J. CARRIVE. Prédiction de séries temporelles : application à la structuration des flux audiovisuels. In *Actes des journées CORESA 2006*, novembre 2006.
- J.P. POLI et J. CARRIVE. Tv stream structuring with program guides. In *Proceedings of the 8th IEEE Symposium on Multimedia*, pages 329–334, décembre 2006.
- J.P. POLI et J. CARRIVE. Modeling tv schedules for tv stream structuring. In *Proceedings of ACM Multimedia Modeling*, pages 680–689, janvier 2007.
- J.P. POLI, J. PHILIPPEAU, J. PINQUIER, et J. CARRIVE. Fast hierarchical structuring of morning drives. In *Proceedings of IEEE Fifth International Workshop on Content-Based Multimedia Indexing 2007(à paraître)*, juin 2007.
- K.M. PUA, J. M. GAUCH, S.E. GAUCH, et J. Z. MIADOWICZ. Real time repeated video sequence identification. *Computer Vision and Image Understanding*, 93(3):310–327, 2004.
- M.L. PUTERMAN. *Markov decision processes*. Wiley series in probability and mathematical statistics. John Wiley & sons, 1994.
- J.R. QUINLAN. Induction of decision trees. *Machine Learning*, 1:81–106, 1986.
- J.R. QUINLAN. Learning with continuous classes. In *Proceedings of Artificial Intelligence Conference*, pages 343–348, décembre 1992.
- J.R. QUINLAN. A case study in machine learning. In *Proceedings of the sixteenth australian computer science conference*, pages 731–737, february 1993.

- J.R. QUINLAN. Combining instance-based and model-based learning. In *Proceedings of the Tenth International Conference on Machine Learning*, pages 236–243, avril 1993.
- J.R. QUINLAN. *Comparing connectionist and symbolic learning methods*, Chapitre 15, pages 445–456. MIT Press, 1994.
- L.R. RABINER. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of IEEE*, 77(2):257–286, 1989.
- L.R. RABINER et B.H. JUANG. An introduction to hidden markov models. *IEEE ASSP Magazine*, 3(1):4–16, 1986.
- K. RAPANTZIKOS, Y. AVRITHIS, et S. KOLLIAS. On the use of spatiotemporal visual attention for video classification. In *Proceedings of International Workshop on Very Low Bitrate Video Coding*, septembre 2005.
- M. RICHARD et R. LIPPMANN. Neural network classifiers estimate bayesian a posteriori probabilities. *Neural computation*, 3:461–483, 1991.
- M. ROACH, J. MASON, et M. PAWLEWSKI. Video genre classification using dynamics. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, mai 2001.
- F. ROSENBLATT. The perceptron : probabilistic model for information storage and organization in the brain. *Psychological Review*, 65:386–408, 1958.
- Y. RUI, T.S. HUANG, et S. MEHROTRA. Constructing table of content for video. *ACM Journal of Multimedia Systems*, 7(5):359–368, 1998.
- Y. RUI, T.S. HUANG, et S. MEHROTRA. Exploring video structure beyond the shots. In *Proceedings of IEEE International Conference on Multimedia Computing and Systems*, pages 237–240, 1998.
- D.E. RUMELHART, G.E. HINTON, et R.J. WILLIAMS. Learning representations by back-propagation errors. *Nature*, 323:533–536, 1986.
- D.A. SADLIER, S. MARLOW, N.E. O’CONNOR, et N. MURPHY. Automatic tv advertisement detection from mpeg bitstream. In *Proceedings of the 1st International Workshop on Pattern Recognition in Information Systems*, pages 14–25, juillet 2001.
- D. SADLIER et N. O’CONNOR. Event detection in field sports video using audio-visual features and a support vector machine. In *IEEE transactions on circuits systems and video technology*, pages 1225–1233, juillet 2005.
- J.M. SANCHEZ, X. BINEFA, J. VITRIA, et P. RADEVA. Local color analysis for scene break detection applied to tv commercials recognition. *Lecture Notes In Computer Science*, 1614:237–244, 1999.
- G. SAPORTA. *Probabilités, analyse des données et statistique*. Technip, 1990.
- D. D. SAUR, Y-P. TAN, S.R. KULKARNI, et P.J. RAMADGE. Automated analysis and annotation of basketball video. In *Symposium on Electronic Imaging: Science and Technology: Storage and Retrieval for Image and Video Databases*, pages 176–187, janvier 1997.

-
- N. SAVY. *Probabilités et statistiques pour modéliser et décider*. Ellipses, 2006.
- B. SEEBER, N. YAGER, et A. AMIN. Real-time detection of semi-transparent watermarks in decompressed video. In *Proceedings of the IEEE Workshop on Applications of Computer Vision*, pages 49–54, Los Alamitos, CA, USA, février 2007. IEEE Computer Society.
- A. SHIVADAS et J.M. GAUCH. Real-time commercial recognition using color moments and hashing. In *Proceedings of ACM SIGMM International Workshop on Multimedia Information Retrieval*, octobre 2006.
- M. SLIMANE, G. VENTURINI, J.P. Assdin de BEAUVILLE, T. BROUARD, et A. BRANDEAU. Optimizing hidden markov models with a genetic algorithm. *Artificial evolution, lecture notes in computer sciences*, pages 384–396, 1996.
- A.J. SMOLA et B. SCHÖLKOPF. A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222, août 2004.
- C.G.M. SNOEK et M. WORRING. Multimodal video indexing: a review of the state-of-the-art. *Multimedia Tools and Applications*, 25(1):5–35, 2005.
- B. STENGER, V. RAMESH, N. PARAGIOS, F. COETZEE, et J.M. BUHMANN. Topology free hidden markov models: application to background modeling. In *Proceedings of the Eighth IEEE International Conference on Computer Vision*, pages 294–301, juillet 2001.
- A. STOLCKE et S. OMOHUNDRO. Hidden markov models induction by bayesian model merging. In *Advances in Neural Information Processing Systems 5*, pages 11–18, novembre 1992.
- H. SUNDARAM et S.F. CHANG. Audio scene segmentation using multiple features, models, and time scales. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing 2000*, juin 2000.
- K.K. SUNG et T. POGGIO. Example-based learning for view-based human face detection. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 20(1):39–51, 1998.
- D. SWANBERG, C.F. SHU, et R. JAIN. Knowledge guided parsing in video databases. In *Proceedings of IS&T/SPIE Symposium on Electronic Imaging : Science and Technology*, pages 13–24, mai 1993.
- C.M. TASKIRAN, I. POLLAK, C.A. BOUMAN, et E.J. DELP. Stochastic models of video structure for program genre detection. *Lecture Notes in Computer Sciences : Visual Content Processing and Representation*, 2849, septembre 2003.
- R. THOMSEN. Evolving the topology of hidden markov models using evolutionary algorithms. In *Proceedings of Parallel problem solving from nature VII*, pages 861–870, septembre 2002.
- L. TORGO. A comparative study of reliable error estimators for pruning regression. In *Proceedings of the Iberoamerican conference on Artificial Intelligence*, octobre 1998.
- R. TRONCY. Etude du manuel d'indexation commun à tous les documentalistes. Rapport Technique, INA, avril 2001.
- B.T. TRUONG, S. VENKATESH, et C. DORAI. Automatic genre identification for content-based video categorization. In *Proceedings of 2000 IEEE International Conference on Pattern Recognition*, pages 230–233, septembre 2000.

- V. VAPNIK. *The nature of statistical learning theory*. Springer-Verlag, 1995.
- V. VAPNIK, S. GOLOWICH, et A. SMOLA. Support vector method for function approximation, regression estimation, and signal processing. *Advances in Neural Information Processing Systems*, 9:281–287, 1997.
- V. VAPNIK et A. LERNER. Pattern recognition using generalized portrait method. *Automation and Remote Control*, 24:774–780, 1963.
- N. VASCONCELOS et A. LIPPMAN. Statistical models of video structure for content analysis and characterization. *IEEE Transactions on Image Processing*, 9(1):3–19, janvier 2000.
- J. VENDRIG et M. WORRING. Systematic evaluation of logical story unit segmentation. *IEEE Transactions on Multimedia*, 4(4):492–499, 2002.
- E. VENEAU, R. RONFARD, et P. BOUTHEMY. From video shot clustering to sequence segmentation. In *Proceedings of International Conference on Pattern Recognition*, pages 254–257, 2000.
- P. VINCENT. *Modèles à noyaux à structure locale*. Thèse de doctorat, Université de Montréal, octobre 2003.
- T. VOLKMER et S.M.M. TAHAGHOGHI. Rmit university at trecvid 2005. In *Proceedings of TRECVID 2005*, 2005.
- A. VULLO. *Learning protein structural representations by specialised recursive connectionist models*. Thèse de doctorat, Université de Florence, 2004.
- H.D. WATCLAR, T. KANADE, M.A. SMITH, et S.M. STEVENS. Intelligent access to digital videos : informedia project. *IEEE Computer*, 29(5):46–52, 1996.
- K. WEBER, S. IKBAL, S. BENGIO, et H. BOURLARD. Robust speech recognition and feature extraction using hmm2. *New computational paradigms for acoustic modeling in speech recognition*, 17:195–211, 2003.
- M. YEUNG, B.L. YEO, et B. LIU. Extracting story units from long programs for video browsing and navigation. In *Proceedings of the 1996 International Conference on Multimedia Computing and Systems*, pages 296–305, 1996.
- D. YOW, B. YEO, M. YEUNG, et G. LIU. Analysis and presentation of soccer highlights from digital video. In *Proceedings of the Second Asian Conference on Computer Vision*, pages 499–503, décembre 1995.
- H.J. ZHANG, S.Y. TAN, S.W. SMOLIAR, et G.Y. HONG. Automatic parsing and indexing of news video. *ACM Journal on Multimedia Systems*, 25(6):256–266, 1995.
- T. ZHANG et C.C. KUO. Audio-guided audiovisual data segmentation, indexing and retrieval. In *Proceedings of SPIE Storage and Retrieval for Image and Video Databases VII*, pages 316–327, 1999.
- Z. ZHAO, S. JIANG, Q. HUANG, et G. ZHU. Highlight summarization in sports video based on replay detection. In *Proceedings of International Conference on Multimedia and Expo*, juillet 2006.

D. ZHONG et S.F. CHANG. Structure analysis of sports video using domain models. In *Proceedings of IEEE International Conference on Multimedia and Expo 2001*, pages 713–716, 2001.

Structuration automatique de flux télévisuels

Dans le cadre du dépôt légal de la télévision, nous proposons une méthode de structuration automatique des flux télévisuels afin de procéder au catalogage des émissions.

La stabilité des grilles de programmes nous permet d'en proposer une modélisation statistique basée sur un modèle de Markov contextuel et un arbre de régression. Entraîné sur les grilles de programmes des années précédentes, ce modèle permet de pallier l'imprécision des guides de programmes (EPG, magazines). En rapprochant ces deux sources d'informations, nous sommes en mesure de prédire les séquences d'émissions les plus probables pour un jour de l'année et d'encadrer la durée des émissions.

A partir de ces grilles de programmes prédites et d'un ensemble de règles indiquant les éléments caractéristiques d'une transition entre deux genres de programmes (images monochromes, silences ou logos), nous sommes en mesure de localiser ces ruptures à l'aide de détections effectuées localement dans le flux.

Mots-clés : flux télévisuel, structuration automatique, macro-segmentation, modèles de Markov contextuels, régression

Automatic Television Stream Structuring

In the setting of the legal deposit of French television, we design an automatic television stream structuring system.

We propose to model the program schedules with a statistical model based on a contextual Markov model and a regression tree. Such a model is justified by the stability of channels' program schedules. Once the model has been trained on past television schedules, the system is able to improve the completeness and the accuracy of program guides (like TV magazines or EPG). The merging of these two sources of information permits to predict all the most probable schedules for a given day and to predict the duration of each telecast.

Rules are then used to characterize the various transitions from a program genre to another by simple detections: logos, monochrome frames, silences. These rules, associated to the predicted program schedules, allow the system to detect the numerous transitions by performing locally detections.

Keywords: television stream, automatic structuring, macro-segmentation, contextual Markov models, regression