HAL
open science

Extraction multimodale de métadonnées de séquences video dans un cadre bayésien

Siwar Baghdadi

To cite this version:

Siwar Baghdadi. Extraction multimodale de métadonnées de séquences video dans un cadre bayésien. Interface homme-machine [cs.HC]. Université Rennes 1, 2010. Français. NNT : . tel-00512706

HAL Id: tel-00512706

https://theses.hal.science/tel-00512706v1

Submitted on 31 Aug 2010

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.

N° d'ordre : **ANNÉE 2010**

THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Traitement du signal

Ecole doctorale MATISSE

présentée par

Siwar Baghdadi

préparée à THOMSON R&D/TEXMEX IRISA

IFSIC

Extraction multimodale de métadonnées de séquences vidéo dans un cadre bayésien

Thèse soutenue à Rennes le 10 Février 2010

devant le jury composé de :

Henri MAÎTRE
Professeur Télécom ParisTech/ *président*
Regine ANDRE-OBRECHT
Professeur Université Paul Sabatier/ *rapporteur*
Denis PELLERIN
Professeur Université Joseph Fourier / *rapporteur*
Claire-Hélène DEMARTY
Ingénieur R&D THOMSON / *examinateur*
Guillaume GRAVIER
HDR CNRS / *examinateur*
Patrick GROS
Directeur de recherche INRIA/ *directeur de thèse*

A mes parents

Table des matières

Table des matières **1**

Introduction **5**

1 État de l'art de la description sémantique de contenus multimédias **9**
1.1 Description sémantique de contenus multimédias 9
1.1.1 Catégories de métadonnées sémantiques 11
1.1.2 Problématiques 12
1.2 Intégration de modalités 13
1.3 Catégories de tâches de description sémantique de contenus multimédias 14
1.3.1 Catégorisation 14
1.3.2 Structuration 14
1.3.3 Détection d'événements 15
1.4 Les approches pour la détection d'événements 15
1.4.1 L'approche syntaxique 16
1.4.2 Les approches basées classification 17
1.4.2.1 Les approches non probabilistes 17
1.4.2.2 Les approches probabilistes 18
1.5 Topologie des approches probabilistes 18
1.5.1 Les modèles de Markov cachés et leurs variantes 19
1.5.2 Les réseaux bayésiens 21
1.6 Conclusion 24

2 Les réseaux bayésiens **25**
2.1 Utilisation des graphes pour la propagation d'information 25
2.1.1 Circulation de l'information dans les graphes 26
2.1.2 Notions de d-séparation dans les graphes 28
2.2 Éléments de la théorie des probabilités 28
2.2.1 Loi de Bayes 29
2.2.2 Indépendance conditionnelle 31
2.3 Les réseaux bayésiens 31
2.3.1 Définitions préliminaires 31
2.3.2 Définition d'un réseau bayésien 32
2.3.3 Paramètres d'un réseau bayésien 32

2.4 Inférence dans les réseaux bayésiens 34
2.4.1 Inférence dans une structure d'arbre : algorithme de propagation de connaissances 34
2.4.2 Inférence dans une structure générique : algorithme du Jtree . . . 36
2.4.3 Inférence approchée 37
2.5 Apprentissage dans les réseaux bayésiens 38
2.5.1 Apprentissage de paramètres 38
2.5.1.1 Apprentissage statistique 38
2.5.1.2 Apprentissage bayésien 39
2.5.2 Apprentissage de la structure 39
2.6 Réseaux bayésiens particuliers 39
2.6.1 Réseau bayésien naïf 39
2.6.2 Réseau bayésien dynamique 40
2.7 Conclusion 41

3 Apprentissage de structure pour l'indexation vidéo 43
3.1 Introduction 43
3.2 Apprentissage de structure 44
3.2.1 Apprentissage de structure par recherche de causalité 44
3.2.2 Apprentissage de structure par optimisation de scores 45
3.2.2.1 Propriétés des fonctions de score 46
3.2.2.2 Score AIC 47
3.2.2.3 Score BIC 47
3.2.2.4 Score d'information mutuelle 48
3.2.2.5 Parcours de l'espace de recherche 48
3.3 Résultats et interprétation 50
3.3.1 Cadre applicatif 50
3.3.2 Construction du modèle 51
3.3.3 Base de données 51
3.3.4 Évaluation 52
3.3.5 Mise en œuvre 52
3.3.5.1 Influence du choix du nœud racine de la structure en arbre 52
3.3.5.2 Influence de l'utilisation de l'ordre dans l'apprentissage de structure par l'algorithme K2 53
3.3.6 Résultats 53
3.3.7 Interprétation des différentes structures récupérées 54
3.4 Conclusion 57

4 Apprentissage de structure pour la classification 59
4.1 Introduction 59
4.2 Apprentissage de structure génératif 59
4.3 Classification par un réseau bayésien naïf augmenté 62
4.3.1 Classification en utilisant un réseau bayésien naïf augmenté de type TAN 62

4.3.2 Classification en utilisant un réseau bayésien naïf augmenté par une structure résultant d'un K2 . . . 63
4.3.3 Résultats et interprétation . . . 64
4.3.3.1 Influence de la complexité de la structure construite par un algorithme K2 . . . 65
4.3.3.2 Comparaison entre les deux approches augmentées . . . 66
4.4 Classification par une approche discriminante . . . 68
4.4.1 Fonction de score . . . 69
4.4.2 Mise en œuvre de la méthode . . . 70
4.4.3 Résultats et interprétation . . . 70
4.5 Classification par l'approche Multinets . . . 72
4.5.1 Processus d'inférence . . . 73
4.5.2 Processus d'apprentissage . . . 73
4.5.3 Résultats et interprétation . . . 74
4.5.3.1 Apport de l'apprentissage de structures pour un réseau Multinets . . . 74
4.5.3.2 Comparaison de réseaux Multinets avec les différentes approches déjà décrites . . . 75
4.6 Influence de la taille de la base d'apprentissage . . . 75
4.7 Conclusion . . . 77

5 Influence de la sélection d'attributs sur l'apprentissage de structure 81
5.1 Sélection automatique d'attributs . . . 81
5.1.1 Méthodes basées *classement* . . . 82
5.1.2 Méthodes tenant compte des corrélations entre attributs . . . 82
5.1.2.1 Méthode de parcours . . . 83
5.1.2.2 Critère . . . 84
5.1.2.3 Méthodes basées *filtrage* . . . 84
5.1.2.4 Méthodes de type *wrapper* . . . 86
5.2 Résultats et interprétation . . . 86
5.2.1 Comparaison des différentes méthodes de sélection d'attributs pour un classifieur donné . . . 86
5.2.2 Influence de la sélection d'attributs sur le réseau bayésien naïf . . . 88
5.2.3 Influence de la sélection d'attributs sur les approches enrichies . . . 89
5.2.3.1 Influence de la sélection d'attributs sur l'approche de type réseau bayésien naïf enrichie par une structure en arbre . . . 89
5.2.3.2 Influence de la sélection d'attributs sur l'approche de type réseau bayésien naïf enrichie par une structure générique . . . 90
5.2.4 Influence de la sélection d'attributs sur les approches Multinets . 92
5.2.5 Influence de la sélection d'attributs sur l'approche générative . . 93
5.2.6 Influence de la sélection d'attributs sur l'approche discriminante 94
5.3 Interprétation des attributs sélectionnés . . . 95

5.4 Conclusion 96

6 Apprentissage de structure dans les réseaux bayésiens dynamiques 97
6.1 Introduction 97
6.2 Les réseaux bayésiens dynamiques 97
6.3 Apprentissage de structure dans les réseaux bayésiens dynamiques . . . 99
6.3.1 Approche utilisant le score BIC 99
6.3.2 Approche augmentée d'apprentissage de structure 100
6.3.3 Approche discriminante 100
6.4 Résultats expérimentaux 101
6.4.1 Comparaison des différentes approches d'apprentissage de structure dynamique 101
6.4.2 Réseau dynamique versus réseau statique 102
6.5 Conclusion 102

7 Conclusion et perspectives 105
7.1 Synthèse des résultats 105
7.2 Travaux futurs 107

Annexe I 111

Annexe II 113

Bibliographie 120

Table des figures 121

Introduction

On est toujours intrigué par la capacité et la facilité avec lesquelles l'être humain peut comprendre, interpréter et s'adapter à de nombreuses situations alors que les machines, même avec la grande puissance de calcul dont elles sont dotées, restent muettes et n'arrivent pas à comprendre leur environnement. On voit par exemple un supporteur du stade rennais somnoler dans son canapé en regardant son équipe préférée jouer lorsque le jeu ne renferme pas d'événements intéressants. Mais dès qu'un but risque d'être marqué, notre supporteur se réveille. Ce moment représente un moment clé pouvant décider de l'issue du match. En effet à cet instant, le supporteur a analysé la vidéo qu'il est en train de regarder. Il déduit à partir de ses expériences passées qu'il y a risque que le score change. Est ce qu'une machine peut atteindre ce niveau de compréhension de l'environnement ? Jusqu'à présent, ce n'est pas encore le cas. Certes, des études et des expériences sont menées dans plusieurs domaines relatifs à la compréhension de la vidéo, mais on est toujours loin du niveau d'analyse accompli par le cerveau humain.

Toutefois, la vidéo et son interprétation deviennent un enjeu financier de plus en plus important, surtout avec la croissance de la quantité et la qualité des vidéos produites chaque jour. Devant cette grande quantité de vidéos, il n'est pas envisageable d'avoir recours à l'annotation humaine pour interpréter et exploiter ces contenus dans leur intégralité. Les moyens financiers sont donc plus que jamais nécessaires pour développer des systèmes capables de remplacer et dépasser l'homme pour des tâches de compréhension sémantique de la vidéo.

Problématique

Avec l'accroissement de la puissance de calcul des ordinateurs, on est de plus en plus capable d'effectuer des traitements sur les vidéos. Ces traitements visent à extraire des attributs bas niveau tels que la quantité de mouvement, le niveau d'une couleur donnée ou la présence d'un visage dans une image de la vidéo. On peut également extraire de l'information à partir du média audio telle que la présence de parole, la présence de silence, le niveau d'énergie.

Toutefois, ces attributs à eux seuls ne sont pas suffisants pour donner un niveau de compréhension de la vidéo comparable à celui de l'être humain. En effet, la détection d'un mouvement fort dans une vidéo peut avoir comme signification par exemple la présence d'un film d'action mais aussi la présence d'une plage de publicité. Toutefois la combinaison du fait que la plage en question renferme, en plus du fort mouvement, un

bout de texte ainsi qu'une grande variabilité sonore est un indicateur supplémentaire du fait de la présence d'une plage de publicité. L'agrégation des différents attributs et leur fusion est donc un moyen pour améliorer le niveau de compréhension des contenus vidéo et ainsi récupérer des index de haut niveau.

Notre problématique est la détection d'événements sémantiques dans un contenu vidéo. Il s'agit, en effet, de reconnaître un certain nombre d'événements de haut niveau compréhensibles par l'humain, à partir de l'agrégation de données ou d'attributs de bas niveau extraits directement à partir de la vidéo.

Tout au long de ce travail, nous utilisons les modèles probabilistes comme cadre de fusion des différents descripteurs. Nous utilisons plus précisément les réseaux bayésiens. Contrairement à la majorité des approches qui existent dans la littérature, nous nous passons de l'hypothèse d'indépendance entre les attributs généralement utilisée faute de connaissance suffisante pour construire correctement la structure du réseau bayésien modélisant l'événement à détecter. Nous proposons, en effet, d'apprendre les interactions qui existent entre les attributs directement à partir de la base d'apprentissage. Nous abordons également l'importance de prendre en compte certaines interactions dans la phase de classification. Un second volet de ce travail est consacré à la pertinence des attributs utilisés et à leur influence sur les résultats de détection. Nous montrons en effet que l'approche que nous proposons permet, lors de la phase d'apprentissage, de déterminer quels sont les attributs pertinents pour la détection de l'événement souhaité.

Méthodologie

Nous avons utilisé dans notre travail une approche probabiliste basée sur les réseaux bayésiens. Ces modèles constituent en effet une extension naturelle des modèles de Markov cachés [1] et des modèles de segments [2] qui ont montré leur aptitude à l'intégration de données provenant de différents supports telles que les attributs issus de l'image et du son. D'autre part, et contrairement à ces modèles, les réseaux bayésiens permettent de relâcher les contraintes sur la structure du modèle utilisé permettant ainsi de modéliser une plus grande classe de systèmes. Cette propriété est d'autant plus importante que les réseaux bayésiens offrent la possibilité d'un apprentissage automatique de cette structure à partir d'une base de données.

Du fait de l'importance de cette propriété, nous avons centré notre travail sur l'apport de l'apprentissage de structure dans un système de détection d'événements.

Ligne directrice du manuscrit

Ce document présente les différents travaux que nous avons menés tout au long de cette thèse sur la problématique de la détection d'événements dans un contenu vidéo. Le premier chapitre propose un bref état de l'art des techniques qui existent dans ce domaine. Il permet également de situer notre travail par rapport à l'existant. Nous présentons dans le deuxième chapitre les éléments théoriques permettant une bonne compréhension des réseaux bayésiens.

Le cœur du document est ensuite organisé en quatre chapitres. Le troisième chapitre présente une première mise en œuvre de l'apprentissage de structure pour l'indexation vidéo. L'aspect classification et son importance pour l'apprentissage de structure sont présentés dans le quatrième chapitre. Le cinquième chapitre présente le volet sélection d'attributs. Nous étudions enfin dans le sixième chapitre l'effet de l'apprentissage de structure sur les réseaux bayésiens dynamiques.

Chapitre 1

État de l'art de la description sémantique de contenus multimédias

Jusqu'à une période récente, il n'y avait guère d'autre choix que le traitement manuel des vidéos numériques chaque fois que l'on voulait structurer ces dernières ou en extraire de l'information, surtout si cette information était de haut niveau sémantique. Toutefois, cette manière de faire se révèle coûteuse et impossible à continuer de nos jours, devant l'explosion de la quantité de données vidéos qui sont désormais produites de partout et accessibles à tout le monde.

La nécessité d'un traitement automatique de la vidéo, limitant l'intervention humaine, s'est donc imposée. La course à la recherche dans le domaine de l'indexation du contenu vidéo est lancée. Diverses méthodes de description de contenu sont apparues dans la littérature cette dernière décennie. Nous en présentons un aperçu tout au long de ce chapitre.

Nous présentons, au début de ce chapitre, le cadre général de la description sémantique de contenu multimédia. Nous passons ensuite en revue les différentes catégories de tâches de descriptions. Nous nous attardons enfin sur la tâche de détection d'événements ainsi que sur les différentes méthodes qui sont utilisées pour réaliser cette tâche.

1.1 Description sémantique de contenus multimédias

Contrairement au domaine du traitement automatique des langues, où le terme indexation désigne l'opération consistant à trouver un ensemble de descripteurs qui représentent au mieux les données, en vidéo le terme indexation est utilisé pour des opérations allant des plus simples, telles que la détection d'une couleur dominante ou d'un niveau de mouvement, aux opérations les plus complexes, mais les plus intuitives pour un être humain, telles que la détection de « touché » dans un match de football américain ou la détection d'une explosion dans un film d'action. Dans le cadre de l'indexation vidéo, les problèmes de détection d'attributs de bas niveau tels que la couleur,

le mouvement ou la texture ont été largement étudiés au cours de ces dernières décennies. Les communautés scientifiques et industrielles disposent actuellement d'un grand éventail de méthodes pour l'extraction de ce type d'information. Toutefois, ce type d'index, dit de bas niveau, reste insuffisant pour fournir à un utilisateur non expert du traitement de la vidéo un cadre lui permettant de manipuler facilement des données vidéos. Ainsi, une application qui fournit tous les segments de vidéo contenant un grand niveau de mouvement ou une couleur dominante particulière ne sera pas d'une grande utilité pour une entreprise voulant récupérer les buts et les actions dans un match de football. Par contre, un système qui fournit tous les paniers marqués dans un match de basketball aura une grande valeur ajoutée pour un industriel travaillant dans l'édition et la diffusion de contenus vidéos.

Ce deuxième cas applicatif se différencie du premier par la nature de l'information à travers laquelle le contenu vidéo est décrit. Ainsi, contrairement à la couleur ou au niveau de mouvement qui sont, somme toute, des informations de bas niveau extraites directement, l'information du *panier marqué* est une information de haut niveau sémantique, compréhensible par l'homme. L'extraction de ces informations dans le contenu vidéo est encore un problème ouvert dans la communauté scientifique. Nous nous sommes intéressés dans le cadre de ce travail au problème de l'extraction d'information de haut niveau pour fournir une description sémantique du contenu vidéo.

La description sémantique d'un contenu multimédia consiste à analyser le contenu et à en extraire une description plus compacte et plus informative, constituée de connaissances de haut niveau. Cette définition du problème exclut donc la description de la vidéo à travers des attributs de bas niveau tels que la couleur, la texture, le niveau d'énergie audio, ou le contenu du texte dans une image. Il s'agit plutôt d'utiliser ces informations pour déduire des connaissances de haut niveau sémantique. Nous pouvons citer par exemple la détection de point gagnant dans un match de tennis, la détection de publicité dans un flux TV, la détection d'un dialogue dans une série télé, la détection d'un but dans un match de football, ou également la détection de la structure d'un journal télévisé. La description sémantique est plus souvent constituée de deux étapes comme illustré par la figure 1.1. La première étape est l'extraction d'un maximum d'attributs ou métadonnées de bas niveau. De nombreux travaux traitent de cette phase du processus de description de contenu. Cette phase fait appel à divers domaines : le traitement de l'image, de la vidéo, du son et/ou des langues naturelles. La deuxième étape consiste à intégrer ces différents attributs pour obtenir une description sémantique sous forme de métadonnées de haut niveau en adéquation avec les attentes de l'utilisateur du système.

Le processus de description sémantique d'un contenu multimédia vise donc à développer des techniques permettant la détection, l'analyse et la reconnaissance d'informations sémantiques ou métadonnées présentes dans le contenu vidéo. Ces dernières constituent un moyen de valorisation du contenu vidéo. Elles permettent en effet l'exploitation du contenu pour faire de la recherche automatique ou de la rééditorialisation par exemple. Nous présentons, dans le paragraphe suivant, les différents types de métadonnées sémantiques qui peuvent être extraites d'un contenu multimédia.

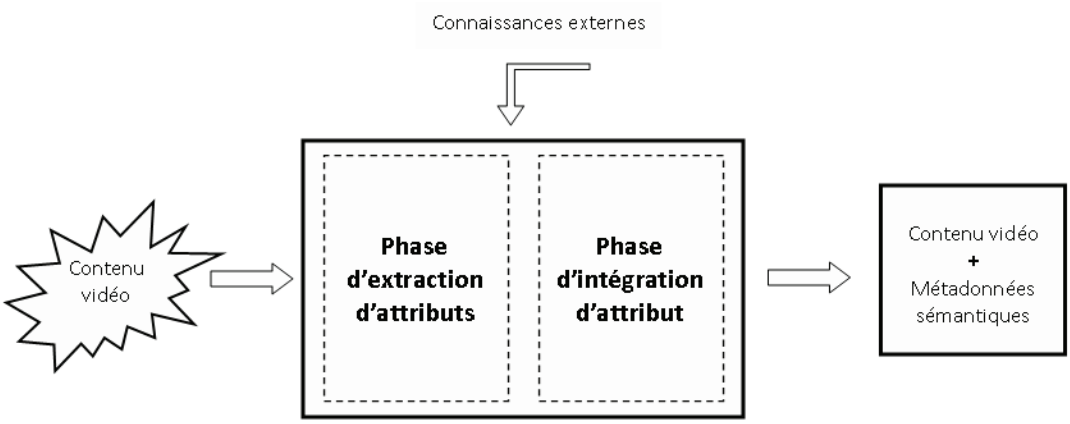

FIG. 1.1 – Schéma d'un système de description de contenu multimédia

1.1.1 Catégories de métadonnées sémantiques

Il existe différent niveaux de métadonnées selon la granularité avec laquelle on observe le document. Dans [3], les auteurs identifient quatre niveaux de métadonnées :

- **genre** : un ensemble de vidéos ayant en commun le même style. Par exemple, *Sport, Film, Information.* Divers travaux permettant d'affecter un genre à une vidéo donnée, à partir d'attributs audio et vidéo, existent dans la littérature [4, 5, 6] ;
- **sous-genre** : un ensemble de vidéos ayant le même genre et qui présentent le même contenu. Par exemple *film d'horreur, vidéo de football, vidéo de tennis,...* Dans [7], les auteurs distinguent 4 genres de sport ; *basketball*, *hockey* ; *football*, et *volleyball* ;
- **unité logique** : une partie continue de la vidéo qui a une signification sémantique. Par exemple *dialogue, ralenti, météo* ;
- **événement nommé** : un segment court de la vidéo ayant un sens et qui ne change pas au cours du temps. On peut citer comme exemples *explosion, but* ou *action* dans un match de football, *point gagnant* dans un match de tennis ou *les cours de la bourse* dans un journal financier.

Il est à noter que les deux premiers niveaux de métadonnées concernent le document. Les deux derniers niveaux concernent seulement une partie du document. De plus en plus de travaux récents en indexation vidéo s'intéressent à ce type de métadonnées. Elles permettent, en effet, un accès non séquentiel à la vidéo, ce qui permet une recherche plus rapide de l'information.

1.1.2 Problématiques

La description sémantique d'un contenu vidéo se heurte principalement à trois types de problèmes : **la multimodalité**, **le caractère temporel de la vidéo** et la **généricité** des méthodes utilisées.

Multimodalité : ce problème provient principalement de la diversité des attributs qui peuvent être extraits de la vidéo (des attributs visuels : histogramme de couleur, vecteurs mouvements, descripteurs de formes), des attributs audio (niveau d'énergie, ZCR, plages de silence), des attributs linguistiques (transcription de la parole dans la bande son). Il est difficile d'intégrer ces attributs de différentes natures dans un même cadre de description de contenu. D'autre part, se pose le problème de la modélisation des différentes corrélations entre les attributs. Comment exploiter cette corrélation pour obtenir une meilleure description sémantique du contenu ? Toujours dans le cadre de la problématique de la multimodalité, la nature différente des médias utilisés fait que les attributs extraits ne sont pas toujours synchronisés et n'ont pas la même fréquence. Des attributs extraits au niveau de l'image telle que la couleur dominante sont extraits à une cadence de 25 ou 30 images par seconde. D'autres attributs visuels peuvent être extraits au niveau du plan. Les attributs audio sont quant à eux extraits le plus généralement à une fréquence de 100 échantillons par seconde. La transcription en texte de la parole aura encore une cadence différente (4 syllabes par seconde). La combinaison de ces attributs pose donc un problème de synchronisation. Faut-il utiliser une fréquence unique ? Dans ce cas une modalité est choisie comme dominante et toutes les autres modalités sont recalées à la fréquence de la modalité dominante. Ce recalage n'est pas toujours évident puisqu'il faut faire des interpolations pour les médias de fréquence plus faible et éliminer de l'information pour les médias de fréquence élevée. Une autre manière de résoudre le problème est de modéliser de manière disjointe les différents médias. Ce type de modélisation implique une prise de décision unimodale, ce qui n'est pas toujours optimal, puisque la corrélation entre les différentes médias est ignorée dans ce cas.

Caractère temporel de la vidéo : le second problème lié à la description de contenus vidéos est lié à la nature temporelle de la vidéo. Il est en effet important de prendre en compte la corrélation temporelle des données vidéos. Une vidéo est composée d'une succession d'images et d'une succession de signaux audio. Ces différents signaux sont corrélés temporellement. Cette corrélation représente un grand potentiel pour effectuer une description efficace du contenu vidéo. Prenons comme exemple le cas de la détection de but dans un match de football. Généralement, quelques instants avant le but, la vidéo présente une succession d'images montrant la zone de but. Au moment de l'événement, on retrouve un plan centré sur la cage de but. Á ce même moment, un grand niveau d'énergie est présent dans le flux audio, du fait de l'excitation de la foule et de celle du présentateur du match. Quelques plans après, le réalisateur insère généralement des plans de ralenti pour rappeler l'action du but. Cette description de l'événement but tient compte de la succession chronologique des attributs dans le temps. Il est donc important que l'approche utilisée prenne également en considération le facteur temps

dans la vidéo.

Généricité : le troisième type de problème relatif à la description de contenus multimédias réside dans la dépendance de ce type d'applications vis-à-vis du domaine de la vidéo traitée. La majorité des approches proposées dans la littérature reste largement dépendantes de la tâche traitée. Ceci vient du fait que les modèles sont construit sur la base d'informations *a priori* relatives à la tâche à traiter. Il est alors nécessaire de fournir des modèles avec une forte capacité d'adaptation au type d'événement à détecter. Il est donc tout à fait souhaitable de pouvoir automatiser la construction du système d'indexation vidéo afin d'avoir recours le moins possible à une intervention humaine toujours coûteuse.

Ces différentes difficultés relatives à la description de contenu sont principalement liées à l'étape d'intégration des différents attributs extraits du flux vidéo. On distingue actuellement différentes stratégies d'intégration selon qu'on tient compte de l'action conjointe des attributs au début du processus ou à la fin. Nous exposons dans la suite les différentes stratégies d'intégration utilisées dans la littérature.

1.2 Intégration de modalités

Dans [8], l'auteur classe les différentes approches d'intégration multimodale en trois catégories et ce selon le niveau auquel on fusionne les données. On distingue alors ainsi :

1. **l'intégration précoce** : durant ce type d'intégration, tous les attributs de différentes modalités sont concaténés en un seul et unique vecteur. La décision est ensuite faite sur ce vecteur multimodal. Ce type d'intégration possède l'avantage de ne pas perdre l'information issue des corrélations entre les attributs des différentes modalités. Toutefois, l'effort d'apprentissage est considérable vu le grand nombre d'attributs mis en œuvre. D'autre part, chaque changement au niveau du comportement d'un attribut implique la mise à jour de tout le système ;
2. **l'intégration tardive** : dans cette approche, la classification est effectuée d'une manière indépendante sur chaque modalité. L'intégration est ensuite faite au niveau des décisions résultantes. Elle est généralement basée sur une approche heuristique à base de règles. Un exemple de ce type d'intégration est l'utilisation de deux HMM, un pour la modélisation du flux visuel et l'autre pour la modélisation du flux audio. Dans [9], les auteurs dressent une comparaison entre les deux types d'intégration présentés ci-dessus. Ils concluent que les résultats ne montrent pas une supériorité nette de l'une ou l'autre des deux approches.
3. **l'intégration intermédiaire** : ce type d'intégration constitue un compromis entre la méthode d'intégration précoce et la méthode tardive. Il se base sur l'utilisation d'un modèle pour représenter l'étape de fusion des différents flux d'attributs. Le modèle a donc comme entrée les flux correspondant aux différentes modalités et comme sortie la décision sur l'existence de la métadonnée ou pas. Le

modèle permet de définir les contraintes de synchronisation entre les flux d'entrée ainsi que les différents niveaux de corrélation qui existent entre les attributs.

L'intégration des modalités est une étape fondamentale vers la production de métadonnées dans les systèmes d'indexation vidéo. Ces systèmes peuvent aussi être catégorisés selon le genre de métadonnées produites.

1.3 Catégories de tâches de description sémantique de contenus multimédias

Pour arriver à une description efficace du contenu multimédia, on distingue principalement trois tâches : catégorisation, structuration et détection d'événements. Ces différentes tâches visent à fournir un ensemble de métadonnées de haut niveau appartenant aux différents types présentés dans le paragraphe 1.1.1. La première tâche fournit généralement des métadonnées appartenant aux deux premiers niveaux de métadonnées. Les deux dernières tâches produisent généralement des métadonnées du troisième et du quatrième niveau.

1.3.1 Catégorisation

La catégorisation consiste à attacher un label à la totalité d'une vidéo pour indiquer son type. On utilise alors des attributs extraits de la vidéo et qui peuvent indiquer la nature de la vidéo en question. Dans [10], les auteurs se basent sur des attributs de bas niveau tels que la couleur, la texture, le mouvement ou encore le niveau audio, pour classer les vidéos de contenus sportifs en différentes catégories telles que : natation, cyclisme, tennis, ou sport à voile. La tâche de catégorisation peut constituer un travail de base pour les autres tâches de description de contenu. Le résultat de la tâche de catégorisation peut, en effet, être utilisé comme une métadonnée pour le choix du traitement à exécuter sur le contenu vidéo dans les phases de structuration ou de détection d'événements.

1.3.2 Structuration

La structuration, assimilée aussi à une segmentation en unités logiques ou sémantiques de la vidéo, consiste à trouver une description de l'organisation temporelle du document vidéo et à localiser les unités correspondant à cette structure. La structuration est dite dense lorsqu'elle décrit toutes les unités de la vidéo, partielle, si elle n'identifie qu'une partie des unités logiques. Dans [11], les auteurs calculent une structuration partielle d'un match de volley en ramenant le problème de structuration à l'extraction de deux scènes majoritaires du match et à la classification des plans selon leur appartenance à l'une ou l'autre des scènes. Dans [12], les auteurs ont proposé une approche multimodale pour effectuer une structuration dense des matchs de tennis, en structurant le contenu sous forme d'« échange », « service manqué », « pause », ou « rediffusion ». La structuration est également utilisée dans [13] pour la description de journaux télévisés.

Dans le cas des contenus qui ne disposent pas d'une structure bien identifiable telle que les vidéos de football, différents travaux ont cherché tout de même à trouver une structure. Dans [14] et [15], les auteurs ont structuré le match de football en phases de jeu/non jeu afin de réduire la quantité de vidéos susceptible d'être vue par un utilisateur.

1.3.3 Détection d'événements

La détection d'événements est une tâche qui a une grande importance dans le contexte de l'indexation vidéo. Cette importance est en effet due à la manière avec laquelle un utilisateur donné perçoit les contenus vidéo. Généralement, on ne se rappelle pas de la totalité d'une vidéo donnée mais seulement des moments importants du contenu tels que le dénouement d'un film, les buts dans un match de football, ou un record battu dans un concours d'athlétisme. Seule une partie de la vidéo semble importante plus que sa structure ou sa catégorie.

La détection d'événements consiste à trouver dans un flux vidéo des moments du document ayant une forte valeur sémantique. Des exemples d'événements peuvent être les dialogues [16], les publicités [17, 18], les événements sportifs dans le basketball [19], dans le football [20], dans le tennis [12, 21]. La détection d'événements s'appuie sur l'organisation temporelle de l'information : la succession des différents plans, l'évolution du mouvement de la caméra... La figure 1.2 présente un exemple d'utilisation de la détection d'événements dans le cadre d'un contenu sportif.

La recherche dans le domaine de la détection d'événements est encore ouverte. Toutefois, nous sommes persuadés que la clé pour le succès de telles techniques est l'utilisation conjointe des différentes modalités présentes dans la vidéo. Il est aussi important de tenir compte du caractère temporel de la vidéo et donc de tenir compte de la corrélation temporelle qui peut exister entre les attributs. Une autre clé de succès est de présenter un système adaptatif, ne nécessitant pas une intervention trop fréquente de la part de l'utilisateur.

Nous approfondissons dans ce qui suit les différents approches de la problématique de détection d'événements.

1.4 Les approches pour la détection d'événements

Le problème de détection d'événements peut être assimilé à un problème de reconnaissances de motifs (de formes). Parmi les types de méthodes utilisées pour la reconnaissance de motifs et identifiés par Jain *et al.* dans [22], on trouve :

- l'approche syntaxique : le motif est reconnu s'il vérifie un ensemble de règles qui combinent l'action des attributs ;
- l'approche par classification : le motif est décrit, il est ensuite reconnu en se basant sur la distribution des motifs dans l'espaces des attributs. Ces méthodes regroupent des classifieurs tels que les réseaux bayésiens, les SVM, les HMM, les arbres de décision et les réseaux de neurones.

Dans les paragraphes suivants, nous détaillons ces deux types d'approches.

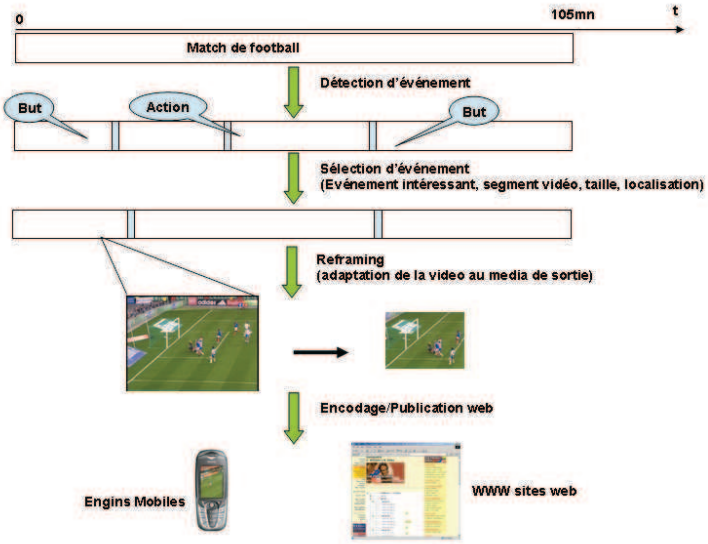

FIG. 1.2 – Schéma de la chaîne de détection d'événements.

1.4.1 L'approche syntaxique

Dans cette approche, la prise de décision s'effectue au travers d'un ensemble de règles explicites définissant les interactions entre les attributs utilisés. Parmi les systèmes utilisant l'approche syntaxique, citons [23, 16, 24, 25, 4, 26]. Dans [27], les auteurs proposent un système pour la détection de paniers dans un match de basket en utilisant le niveau d'excitation de l'audience, le niveau de mouvement et l'apparition de texte. Ces attributs sont fusionnés au travers d'un ensemble de règles. Un exemple de règles est « un niveau d'excitation de la foule est détecté 3 secondes après l'occurrence d'un panier ».

Dans [16], un ensemble de règles est également utilisé pour détecter quatre types de scènes : dialogue, histoire, action et scène générique. Un exemple de ces règles est « si un certain nombre de plans ne présente aucune similarité visuelle, alors ils sont classés dans la classe scène générique », « si les plans présentent des similarités de manière alternée, alors ils sont classés dans la classe dialogue ». Les systèmes basés sur une approche syntaxique donnent des performances acceptables. Ils ont l'avantage d'avoir une syntaxe compréhensible par l'utilisateur ce qui permet une lisibilité assez facile du système. Toutefois, ces approches conduisent à des systèmes peu généralisables. En effet, comme illustré par les deux exemples donnés plus haut dans ce paragraphe, la base de règles qui définit le système est étroitement liée au domaine traité. Ceci rend le système difficilement transposable d'un domaine applicatif à un autre. D'autre part, tout ajout d'attribut nécessite la révision complète de la base de règles pour tenir compte de l'effet de ce nouvel attribut. Un autre inconvénient de ces d'approches réside dans la définition des corrélations entre les attributs pour la construction de la base de règles. La majorité

des travaux utilisant les approches syntaxiques se basent sur des connaissances expertes pour définir les corrélations temporelles qui existent entre les attributs multimodaux. Ces dernières sont rarement disponibles surtout si on dispose d'attributs de bas niveau tels que des histogrammes de couleur ou des flux des mouvements. Ceci constitue un frein au déploiement de ces méthodes. Quelques travaux [28] se sont attaqués au problème de l'apprentissage de la base de règles. Cette démarche s'est avérée très gourmande en termes de puissance de calcul et de données d'apprentissage.

1.4.2 Les approches basées classification

Les approches basées classification reconnaissent les motifs en extrayant de la connaissance d'une base de données contenant des exemples d'occurrence et de non occurrence des événements. On distingue généralement deux sous types d'approches.

1.4.2.1 Les approches non probabilistes

Dans ce type d'approches, on peut inclure les SVM, les arbres de décision, le maximum d'entropie. Un système basé sur le maximum d'entropie est présenté dans [29]. Les auteurs utilisent un classifieur à base de maximum d'entropie pour détecter des évènements dans un match de baseball. Pour chaque plan, un ensemble d'attributs audio visuels sont extraits à partir de la vidéo. L'ensemble de ces attributs forme un vecteur d'attributs donné en entrée du classifieur. La vraisemblance d'une classe est donnée par une fonction exponentielle qui combine l'effet des différents attributs utilisés.

Les méthodes basées sur les SVM ont eu un grand succès ces dernières années. Les SVM se basent sur la recherche de l'hyperplan optimal qui sépare les exemples positifs des exemples négatifs d'une classe donnée. Les techniques d'intégration basées sur les SVM sont considérées comme faisant partie des techniques de fusion précoce. Le point faible de ces approches est que les SVM sont considérées comme une boîte noire à laquelle on peut difficilement ajouter une connaissance *a priori*. D'autre part, il est difficile de gérer le côté temporel de la vidéo. Dans [30], les auteurs utilisent, en plus d'une SVM, un HMM pour gérer le côté temporel du problème d'indexation vidéo. Les attributs de bas niveau sont traités par une SVM pour les classer en différents états. Ces états sont ensuite reliés par un HMM afin de définir la corrélation temporelle qui existe entre les différents états. D'autres types d'intégration temporelle associés aux SVM sont également utilisés dans la littérature. Dans [31], les auteurs utilisent les relations d'Allen [32] pour établir un vocabulaire temporel permettant de relier les différents attributs de moyen niveau reconnus par les SVM.

Les réseaux de neurones sont également utilisés dans la détection d'événements [33]. Dans ce travail, l'algorithme utilise des caractéristiques de couleur, de texture et de mouvement. Les blocs qui sont susceptibles d'être en mouvement sont alors détectés et utilisés au niveau d'un réseau de neurones pour vérifier s'ils correspondent à des objets d'intérêt dans une scène de chasse. Toutefois, la modélisation temporelle de la scène est faite dans le cadre de ce travail à travers un ensemble de règles construites à partir de connaissances externes sur les vidéos de chasse.

D'après les exemples que nous avons présentés plus haut, les approches par classification non probabiliste sont souvent couplées avec d'autres traitements pour modéliser les corrélations temporelles des attributs tels que les HMM ou les systèmes à base de règles. Ce couplage est souvent basé sur des connaissances *a priori* introduites par le concepteur du système. Il y a donc une nécessité d'adapter ces connaissances d'un problème à l'autre.

1.4.2.2 Les approches probabilistes

Les approches probabilistes reposent sur la théorie bayésienne. Elles constituent une des approches majeures de la reconnaissance de formes. Ces approches sont généralement utilisées comme une méthode d'intégration intermédiaire des attributs. Les approches probabilistes permettent en effet une modélisation efficace des interactions entre les variables du problème ainsi que la prise en compte de l'aspect temporel du système. Elles supposent que le problème peut être entièrement formulé d'une manière probabiliste et que toutes les probabilités sont disponibles ou peuvent être estimées. Les approches probabilistes sont principalement utilisées dans un souci de modélisation des données. Divers travaux portant sur la description de contenus multimédias et utilisant les approches probabilistes existent dans la littérature. Ils varient selon la structure du modèle probabiliste considéré. Dans [12, 1], les HMM sont utilisés pour détecter des événements dans un match de tennis. Dans [34, 2], les auteurs utilisent les modèles de segments pour cette même tâche. Les réseaux bayésiens et leurs variantes, les réseaux bayésiens dynamiques, ont quant à eux été utilisés pour la détection de divers événements tels que les événements sportifs, les explosions dans les films.

Dans [35], les auteurs comparent un système utilisant une approche syntaxique basée sur un ensemble de règles avec un système basé sur une approche probabiliste utilisant un HMM. L'application proposée est la détection d'événements dans un match de football américain. Les auteurs montrent dans ce travail que le système à base de règles donne des résultats satisfaisants. Toutefois, ces résultats sont surpassés par l'approche statistique. Les auteurs expliquent cette augmentation des performances par le fait que les approches probabilistes ont la capacité de modéliser correctement des systèmes où de multiples attributs sont mis en œuvre. D'autre part, les approches probabilistes gèrent mieux les problèmes où il y a un certain degré d'incertitude.

Nous reviendrons plus en détails dans la section suivante sur les approches probabilistes et leurs utilisations dans le cadre de la détection d'événements.

1.5 Topologie des approches probabilistes

Cette partie est largement inspirée du chapitre élaboré par Gravier [8]. Les approches probabilistes reposent sur l'utilisation d'un modèle stochastique pour représenter les différentes interactions entre les attributs du problème. La complexité de ces modèles dépend de la complexité du problème traité. Selon le modèle utilisé, les approches probabilistes peuvent être utilisées comme une méthode d'intégration précoce ou d'intégration intermédiaire.

1.5.1 Les modèles de Markov cachés et leurs variantes

Les modèles de Markov cachés ou HMM : les HMM sont composés d'un graphe d'états reliés par des arcs dirigés. Une représentation de ce modèle est illustrée sur la figure 1.3. Chaque arc représente une transition possible entre deux états ou la possibilité que le processus reste dans un même état. Dans ces modèles, le temps est considéré comme discret. À chaque instant, le processus fait une transition d'un état à un autre et émet une observation. Les paramètres du modèle sont de trois types : la probabilité d'être dans un état donné à l'instant initial, les probabilités d'observation et les probabilités de transition d'un état à l'autre. Dans les HMM, l'influence du passé se réduit à la seule connaissance de l'état à l'instant précédent. Cependant, des variantes existent pour étendre cette influence, les n-grammes par exemple.

Les HMM ont été utilisés dans plusieurs travaux comme méthode d'intégration précoce pour la détection d'événements. On peut citer parmi ces travaux [36, 37, 38, 5]. Dans [18], un HMM est utilisé pour la détection de plages publicitaires dans un flux vidéo télévisuel. Des attributs audiovisuels ainsi que des attributs spécifiques à la publicité tels que les silences et les images monochromes sont utilisés dans le cadre de ce travail. Les attributs sont concaténés au niveau d'un vecteur descripteur multimédia au sein d'une intégration précoce. Cette méthode a l'avantage de prendre des décisions multimodales. Elle permet également une modélisation temporelle du problème. Les décisions prises à un instant donné sont prises sur la base des valeurs des attributs de tout le voisinage temporel du présent. Toutefois, l'un des inconvénients majeurs de cette méthode reste la synchronisation entre les flux d'attributs. En effet, une synchronisation parfaite est nécessaire pour pouvoir incorporer les descripteurs dans un même vecteur. La majorité des approches utilisant cette méthode travaille à la cadence d'une modalité dominante. Le deuxième inconvénient majeur est la nécessité d'utiliser une grande base d'apprentissage surtout lors de l'utilisation d'un grand nombre d'attributs. Pour réduire la taille de la base, on utilise généralement l'hypothèse d'indépendance entre les attributs ce qui conduit à l'utilisation des modèles de Markov multiflux.

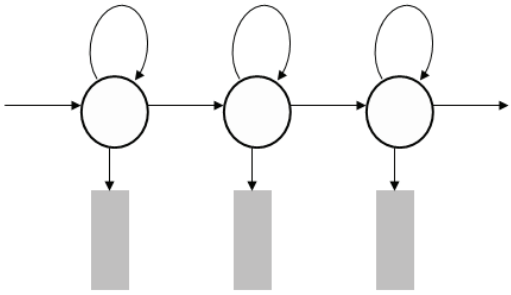

FIG. 1.3 – Modèle de Markov caché.

Modèle multiflux synchrone : dans ce modèle, les attributs sont pris à la même cadence. Ils sont supposés indépendants. Un seul processus caché est mis en œuvre pour la modélisation des flux d'observations. Le modèle multiflux synchrone peut être vu comme un modèle de Markov caché où chaque état possède plusieurs distributions d'observations, une pour chaque flux modélisé. Une illustration de ce modèle se trouve à la figure 1.5. La simplicité de mise en œuvre et d'apprentissage de ce modèle a contribué largement à son utilisation. Parmi les travaux utilisant ce modèle, nous pouvons citer [5, 12, 39].

Modèle produit asynchrone : à la différence du modèle multiflux synchrone, le modèle produit asynchrone permet de définir une corrélation entre les flux de données utilisés dans le modèle. Chaque état du modèle correspond à une combinaison particulière d'états des chaînes de Markov de chaque flux. La probabilité conditionnelle de l'observation d'un modèle produit est la somme pondérée des probabilités conditionnelles pour chaque flux. Les probabilités de transition sont données par le produit des probabilités de transition correspondantes. Le principe de ce modèle est illustré à la figure 1.4. Il a été très utilisé dans le cadre de la reconnaissance audiovisuelle de la parole, mais très peu utilisé en modélisation de contenu multimédia à l'exception de [5].

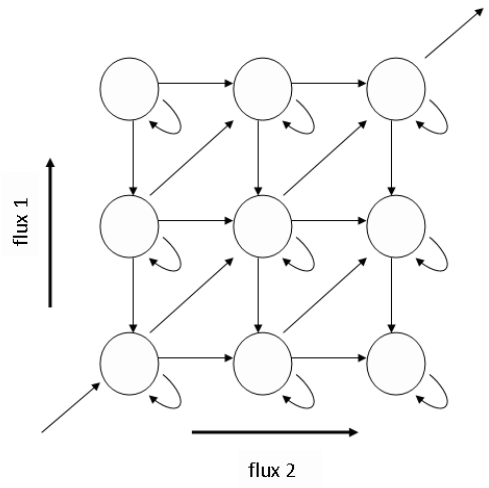

Fig. 1.4 – Modèle multiflux asynchrone.

Les modèles de Markov permettent la modélisation des corrélations temporelles qui existent entre les attributs. D'autre part, ils permettent de modéliser l'asynchronie qui existe entre les différents flux de données. Toutefois, ces modèles ont généralement une structure fixe. Ainsi, lors de la conception du système, c'est au concepteur de choisir le modèle qu'il juge le plus adéquat pour la représentation des données. Cette phase limite la généricité de l'approche.

1.5.2 Les réseaux bayésiens

Un réseau bayésien est un graphe acyclique dirigé. Chaque nœud du graphe correspond à une variable du problème. Les arcs reliant les nœuds peuvent être interprétés comme les relations de corrélation qui existent entre les variables. Ainsi, un arc d'une variable X_1 à une variable X_2 est synonyme d'une dépendance entre les deux variables. Souvent, cette relation de dépendance est interprétée comme une relation de cause à effet. Ainsi, X_1 peut être considéré comme la cause de X_2. Nous explicitons plus en détails dans le chapitre suivant le formalisme des réseaux bayésiens.

D'un point de vue applicatif, le formalisme des réseaux bayésiens a connu un net succès ces dernières années dans divers domaines, allant de la biologie [40] à l'automatique [41] en passant par le domaine de la génétique [42]. Ces modèles ont également séduit la communauté de l'indexation vidéo. Cet engouement de la communauté scientifique à l'égard des réseaux bayésiens vient principalement du grand potentiel de modélisation que présente ce formalisme. Ainsi, contrairement aux modèles stochastiques déjà présentés, les réseaux bayésiens offrent une grande souplesse dans la définition de leur topologie. Cette caractéristique leur permet de définir des interactions complexes entre les différentes variables du système. Des modèles de plus en plus compliqués peuvent être fidèlement représentés à travers les réseaux bayésiens.

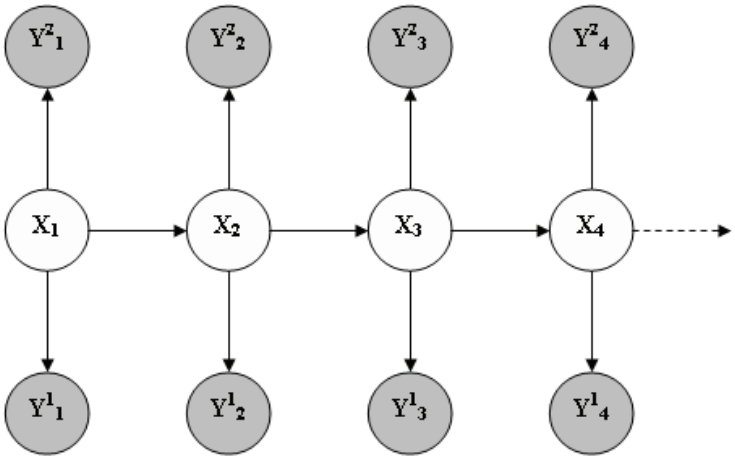

FIG. 1.5 – Représentation d'un modèle multiflux avec le formalisme des réseaux bayésiens. Les états gris correspondent aux variables observées.

Une autre clé du succès des réseaux bayésiens est leur représentation graphique. Cette représentation donne à l'utilisateur du modèle une représentation claire et intuitive des différentes corrélations qui existent entre les éléments du problème. L'utilisateur peut donc intervenir directement sur le modèle, par exemple pour ajouter une information rendue disponible après la phase de construction du modèle. D'autre part, les réseaux bayésiens, permettent, à travers le formalisme des réseaux bayésiens dyna-

miques, la représentation des corrélations temporelles qui existent entre les variables du problème.

Les réseaux bayésiens offrent également un formalisme mathématique unifié pour les divers modèles stochastiques. Dans [43], l'auteur illustre les modèles stochastiques les plus utilisés dans la littérature à travers leurs équivalents dans le formalisme des réseaux bayésiens. Nous présentons ainsi sur la figure 1.5 la représentation d'un HMM multiflux synchrone à travers un réseau bayésien.

Les réseaux bayésiens procurent un cadre pour la représentation de données multimodales. Ils permettent, en effet, de représenter des données discrètes et continues à la fois. Les réseaux bayésiens offrent également une généralisation de nombreux modèles qui ont prouvé être un début de solution pour les problèmes de synchronisation de variables tels que les modèles de Markov asynchrones et les modèles de segments.

En se basant sur la flexibilité des réseaux bayésiens, différentes variantes de modèles de Markov ont été étudiées. Le modèle de Markov couplé a été utilisé dans [44] pour l'analyse de vidéos de sports. Dans les modèles de Markov cachés couplés, l'état d'un flux donné à un instant t dépend de son état à l'instant t-1 et de l'état des autres flux à l'instant t-1. Une représentation de ce modèle à travers les réseaux bayésiens est proposée sur la figure 1.6.

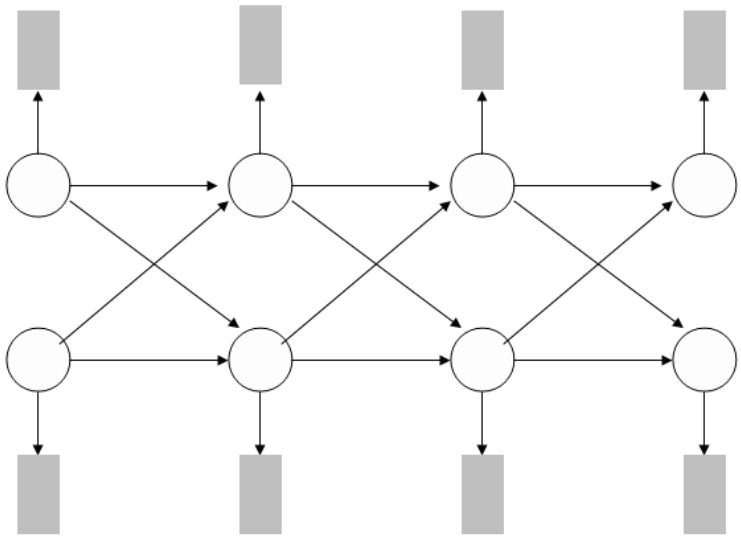

FIG. 1.6 – Représentation du modèle de Markov couplé grâce au formalisme des réseaux bayésiens.

Dans [45] et [46], les auteurs utilisent des réseaux bayésiens pour représenter des concepts sémantiques tels que *Explosion*, *Ciel*. Dans cette représentation, les auteurs utilisent les modalités audio et vidéo. Les interactions entre les attributs sont représentées, pour chaque concept, à travers un réseau bayésien appelé *multiject*. Les connexions qui existent entre les événements sont modélisées par un super réseau bayésien appelé

multinet. Pour construire le modèle, les auteurs ont utilisé des connaissances *a priori* sur les différents événements à détecter. Ces connaissances, qui traduisent l'influence d'un attribut sur l'occurrence d'un événement, résultent en l'existence d'un arc entre le nœud représentant l'événement et celui représentant l'attribut.

Dans [47], les auteurs dressent une comparaison entre les HMM et les réseaux bayésiens dynamiques (DBN) qui seront présentés au début de la section 6. Les deux types de modèles sont utilisés pour la reconnaissance à partir de données multi-sensorielles de l'activité dans un bureau. Les attributs utilisés dans ce travail sont des attributs extraits à partir de l'audio tels que le niveau d'énergie, la fréquence fondamentale et le ZCR (Zero Crossing Rate), des attributs visuels, tels que le taux de pixel de couleur chair, le niveau de mouvement, le taux de pixels considérés comme appartenant à l'arrière plan, le taux de pixels détectés comme appartenant à un visage. Un attribut renseignant sur l'activité souris/clavier a également été utilisé. La comparaison fait apparaître un compromis pour l'utilisation de ces deux modèles. En effet, les HMM s'avèrent plus faciles à apprendre. Le décodage se fait par des algorithmes moins gourmands en capacités de calcul. Toutefois, les réseaux bayésiens permettent une grande diversité de structures régissant les corrélations entre les attributs contrairement aux HMM où la structure est supposée fixe. Les DBN constituent également un cadre unifié pour la représentation de la connaissance contrairement aux HMM où il faut représenter chaque événement à part.

Dans [48], un réseau bayésien dynamique est utilisé pour détecter des événements tels que des *Actions* et des *Buts* dans un match de football. Les auteurs se basent seulement sur des attributs visuels tels que le taux des pixels de couleur de l'herbe, celui des pixels représentant un joueur ou celui des pixels faisant partie de l'arrière plan. Le réseau bayésien utilisé permet de segmenter la vidéo considérée en vues différentes. Dans un second temps, on tient compte de la succession de ces différentes vues pour la détection des événements grâce à ce même réseau dynamique. D'un point de vue construction du modèle, les paramètres du modèle sont appris à partir d'une base de données contenant une vérité terrain comprenant les événements ainsi que les différentes vues définies dans le niveau intermédiaire. Dans ce travail, les auteurs ont fait la supposition que la structure du réseau bayésien est suffisante pour décrire le problème. Cependant, il n'y a aucune garantie que cette hypothèse soit valide.

Dans [49], les auteurs proposent un système d'indexation utilisant un réseau bayésien dynamique permettant de détecter des événements dans des vidéos de formule 1. Ils proposent une méthode de fusion d'attributs basée sur les réseaux bayésiens. Ils ont conclu que différents types de structures peuvent mener à des résultats très différents et que le choix de la structure est crucial pour avoir de bons résultats.

Ainsi, le formalisme des réseaux bayésiens présente un grand potentiel de modélisation. Il offre, en effet, plusieurs avantages : un formalisme clairement défini, permettant de modéliser des variables de différentes natures. Il permet également la représentation d'interactions complexes entre les différentes variables. Ces interactions peuvent être spatiales, sur une même plage temporelle, ou temporelles représentant le cheminement dans le temps de la vidéo. Les réseaux bayésiens représentent donc un cadre intéressant pour la description de contenus multimédias, permettant de traiter le problème

de la multimodalité des données ainsi que la nature temporelle du problème. Toutefois, dans les approches que nous avons présentées, les connexions dans les réseaux bayésiens sont définies manuellement en utilisant des connaissances *a priori* ou d'une manière *ah-doc*. Les réseaux bayésiens offrent, comme propriété supplémentaire, la possibilité d'apprendre les interactions entre les attributs à partir d'une base de données. Très peu de travaux dans la littérature utilisent cette propriété et encore moins en indexation vidéo. C'est pour cette raison que nous explorons dans ce travail cet aspect des réseaux bayésiens afin d'automatiser la tâche de construction du réseau utilisé pour la description d'un contenu multimédia (*cf.* chapitres 3 et 4).

1.6 Conclusion

La description des contenus multimédias est encore un domaine ouvert à la recherche. Il s'agit, en effet, d'extraire un ensemble de métadonnées de haut niveau à forte valeur sémantique. Cette extraction se fait grâce à l'intégration de différents attributs de bas niveau extraits à partir de la vidéo. Nous avons vu que différentes méthodes sont proposées dans la littérature. Les techniques à base d'approches syntaxiques donnent des résultats satisfaisants permettant une intégration des différentes modalités en utilisant une base de règles modélisant les corrélations spatio-temporelles qui existent entre les attributs et les événements à détecter. Les approches non probabilistes présentent un fort pouvoir de classification. Toutefois, ces approches peuvent difficilement modéliser la corrélation temporelle entre les attributs. Cette dernière est généralement gérée par un traitement supplémentaire par des HMM ou par des systèmes à base de connaissances. Les approches probabilistes gèrent, quant à elles, d'une manière naturelle les corrélations à la fois spatiales et temporelles. Ces approches offrent également la possibilité de modéliser conjointement des données de différentes natures. Une grande partie des modèles stochastiques peuvent être représentés dans le cadre des réseaux bayésiens. Ces derniers offrent un cadre souple pour représenter des modèles de plus en plus complexes intégrant des données de différentes natures et présentant une forte corrélation temporelle.

Nous avons toutefois relevé dans le cadre de chapitre que la majorité des approches existantes basées sur le réseaux bayésiens sont dépendantes de connaissances *a priori* pour construire le système de détection. Cette dépendance nécessite de construire un système adapté à chaque métadonnée à extraire et à chaque contenu vidéo traité, ce qui s'avère de plus en plus difficile. Il y a donc besoin d'automatiser cette étape de construction du système pour arriver à s'affranchir de cette contribution humaine.

Nous proposons, dans ce travail, d'apprendre automatiquement cette structure. Nous verrons donc tout au long de ce manuscrit différentes méthodes d'utilisation de cet apprentissage, à commencer par un cas d'utilisation de cet apprentissage dans le cadre d'une application de détection de plages publicitaires.

Chapitre 2

Les réseaux bayésiens

Les réseaux bayésiens sont des modèles de plus en plus utilisés pour représenter l'incertitude des connaissances dans des systèmes complexes. Ils ont émergé de l'association entre la théorie des probabilités et la théorie des graphes pour donner un outil graphique de représentation des probabilités jointes sur un ensemble de variables aléatoires. Nous présentons, dans ce chapitre, quelques aspects de ces deux théories. Nous montrons également comment les modèles graphiques probabilistes et plus particulièrement les réseaux bayésiens permettent une association entre ces deux types de théories. Nous nous attardons ensuite sur les différents aspects d'apprentissage et d'inférence, avant de terminer par la présentation de différents cas particuliers de réseaux bayésiens.

2.1 Utilisation des graphes pour la propagation d'information

Pour présenter la théorie des graphes, nous commençons par un exemple classiquement utilisé dans la littérature et proposé dans [50] :

Monsieur Holmes habite à Los Angeles. Un matin en quittant sa maison, M. Holmes se rend compte que la pelouse de son jardin est mouillée (H). Il se demande s'il a plu dans la nuit, donc si c'est à cause de la pluie (R), ou si c'est parce qu'il a laissé l'arrosage automatique (S) en marche.

Il regarde alors la pelouse de son voisin, Monsieur Watson (W). Il remarque qu'elle aussi est humide. Élémentaire : M. Holmes est presque sûr qu'il a plu la veille.

Une formalisation de cette situation est décrite par le graphe de la figure 2.1.

En effet, lorsque M. Holmes a remarqué que sa pelouse était mouillée, il a fait un raisonnement dans le sens contraire des arcs. En observant H, M. Holmes a augmenté sa certitude par rapport à R et S. L'augmentation de la certitude de R provoque une augmentation de la certitude par rapport à S. M. Holmes vérifie alors l'état de la pelouse de son voisin W qui n'arrose jamais son jardin. En découvrant qu'elle est mouillée elle-aussi, il augmente considérablement la certitude de R sans modifier celle de S.

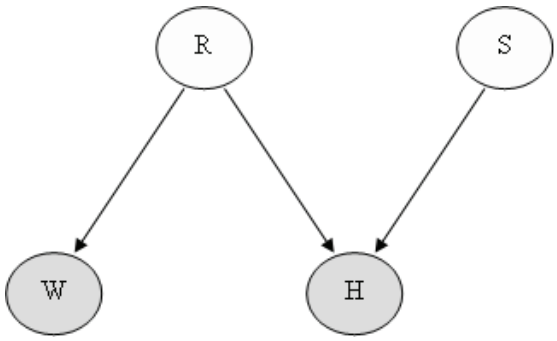

FIG. 2.1 – Modélisation de l'exemple de la pelouse mouillée. La pluie (R) et l'arrosage automatique (S) sont les causes possibles du fait que la pelouse de M. Holmes (H) est mouillée. Seule la pluie est la cause du fait que la pelouse de M. Watson (W) est mouillée.

2.1.1 Circulation de l'information dans les graphes

La représentation en graphe est un moyen de renseigner sur les sens de circulation de l'information entre les différentes variables. Par exemple, s'il existe une relation causale entre deux variables comme c'est le cas entre les variables H et R dans l'exemple de la figure 2.1, toute information sur R peut modifier la connaissance de H et réciproquement toute information sur H peut modifier notre connaissance sur R. Il est essentiel de noter que l'information ne circule pas seulement dans la direction de l'orientation des arcs.

Pour des graphes plus complexes, la circulation de l'information dépend aussi de notre connaissance sur les variables du problème. Le graphe peut en effet contenir des variables dont nous connaissons les valeurs, des variables observées, et des variables dont nous ne connaissons pas la valeur, des variables cachées. Pour déterminer les sens de circulation de l'information, la règle de la balle de Bayes est utilisée. Ce formalisme a été mis en place dans [51]. L'information est alors représentée par une balle qui peut circuler au travers des nœuds comme c'est indiqué dans la figure 2.2. Les nœuds grisés sont les nœuds observés, les autres représentent les nœuds cachés.

L'information peut ainsi circuler dans une connexion en série (c.f. le cas (a) de la figure 2.2) si le nœud représentant la connexion est caché. L'information est toutefois bloquée si le nœud est observé ; en effet dans ce cas c'est l'information du nœud de la connexion lui-même qui devient pertinente et qui est ainsi transmise.

La circulation de l'information est reliée à une notion de séparation dans la théorie des graphes. Ainsi, dans l'exemple de Monsieur Holmes, en ayant l'information sur R, l'information ne circule plus entre H et W. Ces deux variables sont complètement séparée par la variable R. Nous explicitons cette notion de d-séparation dans le paragraphe suivant.

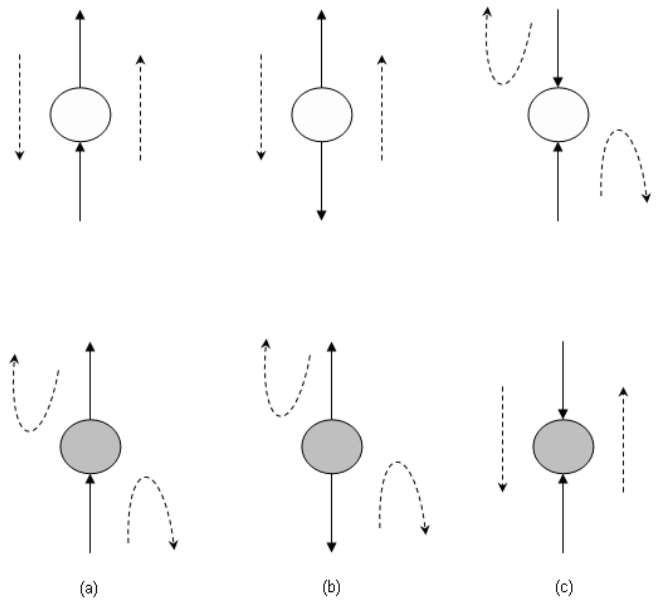

FIG. 2.2 – Règles de passage d'une balle dans un graphe. Les nœuds en gris représentent les nœuds observés, ceux en blancs représentent les noeuds cachés. (a) Connexion en série, (b) Connexion divergente, (c) Connexion convergente.

	Probabilité
R = Vrai	0.4
R = Faux	0.6

TAB. 2.1 – Table de probabilités $P(R)$ de la variable *pluie*(R) : Monsieur Holmes habite dans une zone ou il pleut souvent (40% des jours de l'année).

	Probabilité
S = Vrai	0.4
S = Faux	0.6

TAB. 2.2 – Table de probabilités $P(S)$ de la variable *déclenchement de l'arrosage automatique*(S) : Monsieur Holmes oublie fréquemment de débrancher l'arrosage automatique dans 40% des cas. S'il n'est pas débranché, l'arrossage est déclenché toutes les nuits.

2.1.2 Notions de d-séparation dans les graphes

Définition 2.1 (d-séparation) *Soient (X, Y, Z) des nœuds du graphe $\mathcal{G}$. On dit que X et Y sont d-séparés par Z si, pour tout chemin entre X et Y, il existe une variable W différente de X et Y telle que l'une au moins de deux conditions suivantes soit vérifiée :*

- *la connexion entre X et Y est convergente en W, $W \neq Z$ et W n'est pas une cause directe de Z ;*
- *le chemin qui passe par Z, est soit divergent, soit en série au nœud Z.*

« X est d-séparé de Y par Z »se note par $X \perp_d Y | Z$.

Dans l'exemple que nous avons présenté précédemment, aucune information ne circule entre les variables W et H si on connaît l'état de (R). Si cette dernière est inconnue, les deux variables W et H présentent une corrélation. En effet, connaître l'état de W a permis à Monsieur Holmes de déduire de l'information par rapport à l'état de sa pelouse H.

2.2 Éléments de la théorie des probabilités

Nous transposons, au début cette partie, le formalisme de la théorie des probabilités sur l'exemple de M. Holmes présenté dans la figure 2.1. Nous pouvons traduire nos connaissances subjectives sur l'occurrence des variables du problème en termes de probabilités. Ainsi, en supposant que l'arrosage automatique ne tient pas compte de l'humidité du sol, les événements *pluie* R et *déclenchement de l'arrosage automatique* S sont indépendants. Nous obtenons donc pour ces deux variables les tables de probabilités des tableaux 2.1 et 2.2.

La connaissance selon laquelle l'herbe du jardin de Monsieur Holmes est mouillée ou non se traduit par la table de probabilité 2.3. Cette variable dépend des variables (R) et (S). En effet, la pelouse du jardin de Monsieur Holmes est mouillée seulement s'il a plu la veille ou si Monsieur Holmes a oublié de débrancher l'arrosage automatique.

	S = Vrai		S = Faux	
	R = Vrai	R = Faux	R = Vrai	R = Faux
H = Vrai	1	1	1	0
H = Faux	0	0	0	1

TAB. 2.3 – Table de probabilités $P(H|R,S)$ de la variable état de la pelouse (H) de Monsieur Holmes.

	R = Vrai	R = Faux
W = Vrai	1	0
W = Faux	0	1

TAB. 2.4 – Table de probabilité $P(W|R)$ de la variable état de la pelouse de Monsieur Watson.

Dans notre exemple, la connaissance de l'état de la pelouse de Monsieur Watson W depend uniquement de la pluie R. Cette variable est décrite par la table de probabilité 2.4.

Pour traiter cet exemple, introduisons maintenant la loi de Bayes.

2.2.1 Loi de Bayes

La loi de Bayes lie les probabilités conditionnelles de deux événements. Soit X et Y deux variables aléatoires, telles que $P(X) \neq 0$ et $P(Y) \neq 0$. Alors :

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \tag{2.1}$$

Cette loi est souvent utilisée dans le contexte de l'estimation d'un modèle à partir de données observées. Ainsi, en disposant d'un ensemble de données $\mathcal{D}$, on cherche à estimer le modèle $\mathcal{M}$ qui représente les données. $P(\mathcal{M}|\mathcal{D})$ est la probabilité *a posteriori* du modèle après l'observation des données. La loi de Bayes s'exprime dans ce cas sous la forme :

$$P(\mathcal{M}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{M})P(\mathcal{M})}{P(\mathcal{D})} \tag{2.2}$$

$P(\mathcal{D}|\mathcal{M})$ est la vraisemblance des données par rapport au modèle. $P(\mathcal{M})$ est notre connaissance *a priori* sur le modèle qu'on doit apprendre.

$P(\mathcal{D})$ est notre connaissances *a priori* sur les données. Elle peut se décomposer comme la somme des connaissances sur les données par rapport à tous les modèles.

$$P(\mathcal{D}) = \int_{\mathcal{M}} P(\mathcal{D}|\mathcal{M})P(\mathcal{M})d\mathcal{M} \tag{2.3}$$

L'opération faite dans l'équation 2.3 s'appelle une marginalisation. Nous reprenons l'exemple de M. Holmes. Monsieur Holmes sort de sa maison le matin, il trouve que la pelouse de son jardin est mouillée. Il se demande alors s'il a plu la veille ou s'il

a simplement oublié de débrancher l'arrosage automatique. Cela revient à rechercher la probabilité $P(R = V|H = V)$ pour évaluer s'il a plu la veille, et la probabilité $P(S = V|H = V)$, pour évaluer s'il a oublié de débrancher l'arrosage automatique.

Le calcul de ces deux probabilités conditionnelles se fait en utilisant la loi de Bayes.

$$\begin{aligned} P(R = V|H = V) &= \frac{P(H = V|R = V)P(R = V)}{P(H = V)} \\ &= \frac{P(R = V)}{P(H = V)} \quad car\ P(H = V|R = V) = 1 \end{aligned} \tag{2.4}$$

$$P(S = V|H = V) = \frac{P(H = V|S = V)P(S = V)}{P(H = V)} = \frac{P(S = V)}{P(H = V)}$$

En utilisant le fait que les variables S et R sont indépendantes, on a de plus :

$$\begin{aligned} P(H = V) &= \sum_{R,S} P(H = V, R, S) \\ &= \sum_{R,S} P(H = V|R, S) \times P(R) \times P(S) \\ &= 0,4^2 + 2 \times (0,4 \times 0,6) \\ &= 0,64 \end{aligned} \tag{2.5}$$

On obtient ainsi :

$$P(R = V|H = V) = \frac{0,4}{0,64} = 0,625$$

$$P(S = V|H = V) = 0,625 \tag{2.6}$$

Sans aucune autre information, les deux hypothèses sont donc équiprobables, Monsieur Holmes n'a donc pas d'information sur la cause de l'état de sa pelouse. On se retrouve alors dans la même situation que lors du raisonnement par causalité présenté dans le paragraphe 2.1, où Monsieur Holmes ne peut pas privilégier une hypothèse par rapport à l'autre. Monsieur Holmes a besoin d'une information supplémentaire.

Dans la suite de son raisonnement, Monsieur Holmes a jeté un coup d'œil sur la pelouse de son voisin. Monsieur Holmes compare donc de ce fait les probabilités $P(R = V|H = V, W = V)$ et $P(S = V|H = V, W = V)$. Pour le calcul de $P(R = V|H = V, W = V)$, nous calculons d'abord :

$$\begin{aligned} P(R = V|W = V) &= \frac{P(W = V|R = V) * P(R = V)}{P(W = V)} \\ &= \frac{P(W = V|R = V) * P(R = V)}{P(W = V|R = V)P(R = V) + P(W = V|R = f)P(R = F)} \\ &= 1 \end{aligned} \tag{2.7}$$

Nous pouvons conclure que :

$$P(R = V|H = V, W = V) = 1 \tag{2.8}$$

Les calculs pour obtenir $P(S = V|H = V, W = V)$ sont plus compliqués, nous ne donnons donc que le résultat final :

$$P(S = V|H = V, W = V) = 0.4 \tag{2.9}$$

Donc Monsieur Holmes est certain qu'il a plu la veille.

2.2.2 Indépendance conditionnelle

L'indépendance conditionnelle est un concept qui permet de réduire la complexité de calcul d'une loi jointe entre plusieurs variables. En effet, identifier l'indépendance conditionnelle entre des variables permet de décomposer la probabilité jointe dépendant de toutes les variables, en un produit de probabilités qui ne dépendent que d'un nombre réduit de variables.

Définition 2.2 *Soit X, Y, Z trois variables. X est indépendante de Y conditionnellement à Z (noté $X \perp_{cp} Y|Z$) si, et seulement si, la proposition suivante est vérifiée :*

$$P(X|Y, Z) = P(X|Z)$$

Concrètement cela signifie que dans une tel cas, connaître Y alors qu'on connaît déjà Z n'apporte rien à la connaissance de X.

2.3 Les réseaux bayésiens

2.3.1 Définitions préliminaires

Nous allons poser dans ce paragraphe un certain nombre de définitions relatives aux graphes et qui seront nécessaires pour la compréhension des différentes techniques de manipulation des réseaux bayésiens.

Parents-fils : un nœud X_1 est considéré comme parent d'un nœud X_2, s'il existe un arc dirigé de X_1 vers X_2. X_2 est dans ce cas considéré comme fils du nœud X_1.

Corde : une corde est un arc non dirigé qui n'apparaît pas dans le réseau et qui est ajouté ultérieurement.

Complet : le terme de graphe complet désigne un graphe non dirigé où chaque nœud est connecté à tous les autres nœuds du graphe.

Clique : une clique est un sous-ensemble complet de nœuds.

2.3.2 Définition d'un réseau bayésien

Un réseau bayésien $\mathcal{B}$ est défini par :
- un ensemble de variables aléatoires $X = X_1, ..., X_n$; ces variables forment les nœuds du réseau ;
- un ensemble d'arcs orientés entre ces variables formant un graphe acyclique dirigé ; ces arcs forment la structure $\mathcal{G}$ du réseau ;
- un ensemble de tables de probabilités $\{P(X_i|pa(X_i))\}$. Chaque table donne la probabilité de chaque nœud X_i conditionnellement à l'ensemble $pa(X_i)$ de ses parents dans le graphe $\mathcal{G}$.

Nous avons évoqué, dans les paragraphes précédents, les notions de d-séparation et d'indépendance conditionnelle. Dans [52], les auteurs établissent un théorème fondamental pour le formalisme des réseaux bayésiens.

Théorème 1 *Si les variables X et Y sont d-séparées par Z, $P(X|Z)$ et $P(Y|Z)$ sont conditionnellement indépendantes :*

$$X \perp_d Y|Z => P(X,Y|Z) = P(X|Z) * P(Y|Z) \tag{2.10}$$

Une démonstration de ce théorème peut être trouvée dans [53]. Ce théorème permet d'établir la propriété de décomposabilité de la loi jointe d'un réseau bayésien. Cette propriété s'énonce :

$$P(X_1, ...X_n) = \prod_{i=1..n} P(X_i|pa(X_i)) \tag{2.11}$$

On trouve ainsi une forme compacte de la loi jointe du réseau bayésien. Celle-ci se décompose, en effet, sous la forme d'un produit de termes à chaque nœud du réseau. Chaque terme ne dépend que du nœud auquel il est associé et de l'ensemble de ses parents. Ces termes relatifs à chaque nœud constituent les probabilités conditionnelles de chaque nœud.

La propriété de décomposabilité constitue le fondement du calcul dans les réseaux bayésiens.

2.3.3 Paramètres d'un réseau bayésien

La définition des réseaux bayésiens requiert deux types de paramètres. Le premier type de paramètres, quantitatifs, correspond aux probabilités conditionnelles. Ces probabilités conditionnelles quantifient les corrélations qui existent entre les attributs du problème. Ainsi, pour un nœud X_i ayant pour configuration de parent $pa(X_i) = px_j$ et prenant une valeur x_k, le paramètre associé est θ_{ijk}. Le deuxième type de paramètres, qualitatifs, correspond à la structure du réseau. Cette structure permet de définir les différentes corrélations entre les variables du réseau.

Une illustration de ces paramètres sur l'exemple du problème de M. Holmes est proposée à la figure 2.3.

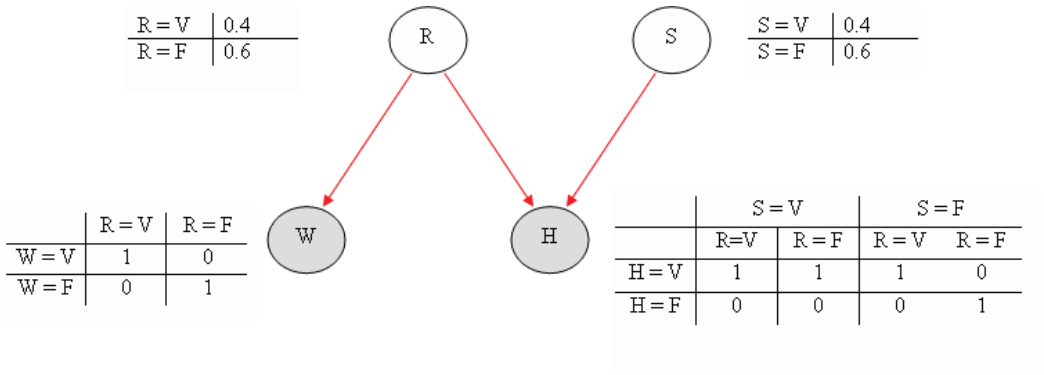

FIG. 2.3 – Exemple de tables de probabilités dans un réseau bayésien à variables binaires.

Deux étapes sont nécessaires pour l'utilisation d'un réseau bayésien (*cf.* figure 2.4). Une première étape dite étape d'apprentissage lors de laquelle il y a construction du réseau à partir de connaissances *a priori* ou en faisant un apprentissage des paramètres du réseau à partir d'une base de données. Une fois que le réseau est construit, vient ensuite l'étape d'utilisation même du réseau. Cette étape s'appelle inférence. Lors de l'inférence, on dispose d'un certain nombre de variables observées et l'on cherche à utiliser ces observations pour déterminer une connaissance sur les variables non observées.

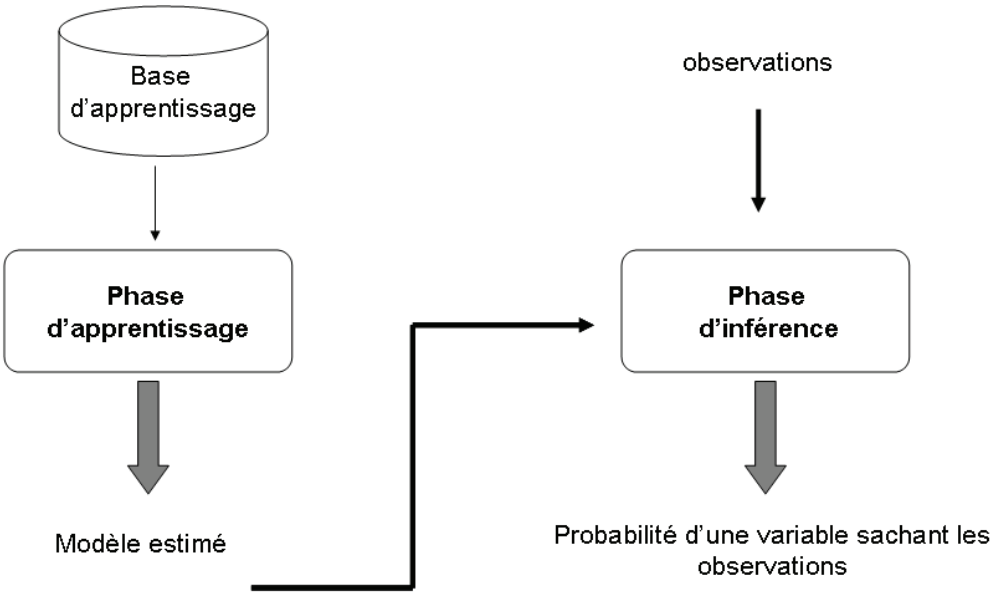

FIG. 2.4 – Processus de fonctionnement d'un réseau bayésien.

Dans les sections suivantes, nous présentons les principes de mise en œuvre de l'étape d'inférence et de l'étape d'apprentissage.

2.4 Inférence dans les réseaux bayésiens

Nous supposons dans cette partie que nous disposons d'un réseau bayésien déjà appris, c'est-à-dire que toutes les probabilités conditionnelles au niveau de chaque nœud sont déjà calculées pour une structure fixe. Nous verrons dans la section suivante comment apprendre ces probabilités ainsi que la structure du modèle. Il ne reste donc qu'à utiliser le réseau bayésien.

Le but de l'utilisation d'un réseau bayésien est, tout en disposant d'un ensemble d'observations e sur un ensemble de variables du problème, de déduire la distribution de probabilité d'une variable X_i du système, $P(X_i|e)$, connaissant cet ensemble d'observations. Les probabilités déjà présentes dans le réseau et qui ont été calculées lors de la phase d'apprentissage ne sont pas toujours celles qui nous intéressent. En effet, les distributions recherchées sont généralement une combinaison des paramètres du réseau bayésien. Le processus de combinaison de ces différentes distributions définit le processus d'inférence. Il est à noter que le problème d'inférence est uniquement un problème de calcul. Il n'y a aucune difficulté théorique derrière. À partir des distributions de probabilité au niveau de chaque nœud calculées pendant la phase d'apprentissage, on peut (en principe) faire le calcul de toutes les distributions de probabilité associées au réseau bayésien étudié.

Il existe diverses méthodes de calcul d'inférence dans les réseaux bayésiens. Ces méthodes peuvent être subdivisées en deux classes différentes : l'inférence approchée et l'inférence exacte. L'inférence approchée est généralement utilisée dans le cas de réseaux bayésiens faisant intervenir un très grand nombre de variables ou de nœuds. Ce type d'inférence se base sur des méthodes de simulations stochastiques pour faire un tirage aléatoire conformément aux probabilités conditionelles du réseau et à l'ensemble des observations disponibles. L'inférence exacte se différencie de la méthode approchée par le fait qu'elle permet un calcul exact de la distribution de la probabilité recherchée. Les méthodes les plus utilisées dans la littérature sont : l'élimination de variables [54], la propagation de connaissances pour les structures en arbre et son extension, le Jtree [55] pour les structures génériques. Nous présentons dans ce qui suit plus en détails ces deux dernières techniques.

2.4.1 Inférence dans une structure d'arbre : algorithme de propagation de connaissances

La propagation de connaissances ou de messages dans un arbre est un algorithme qui a été proposé dans [56]. Cette méthode se base sur la transmission de messages locaux entre nœuds voisins en les faisant transiter à travers les arcs. Chaque nœud communique à son voisinage l'information qu'il a déjà collectée. L'information transite ainsi de proche en proche jusqu'à ce que tous les nœuds mettent à jour leurs probabilités pour tenir compte de l'ensemble d'observations e.

Pour chaque variable X_i, e peut être décomposé en deux sous-ensembles : $e^-_{X_i}$ représentant l'ensemble des variables observées qui sont des descendants de X_i (X_i étant lui même inclus s'il est observé), et $e^+_{X_i}$ représentant toutes les autres variables observées.

L'impact des variables observées sur les connaissances sur X peut être représenté par les valeurs suivantes :

$$\lambda(X_i) = P(e^-_{X_i}|X_i) \tag{2.12}$$

$$\pi(X_i) = P(X_i|e^+_{X_i}) \tag{2.13}$$

X_i pouvant prendre de multiples valeurs discrètes, $\lambda(X_i)$ et $\pi(X_i)$ sont généralement représentés par des vecteurs, où chaque élément est associé à une valeur de X_i :

$$\lambda(X_i) = [\lambda(X_i = x_i^1), ..., \lambda(X_i = x_i^n)] \tag{2.14}$$

$$\pi(X_i) = [\pi(X_i = x_i^1), ..., \pi(X_i = x_i^n)] \tag{2.15}$$

En utilisant 2.12 et 2.13, on obtient la loi *a posteriori*

$$P(X_i|e) = \alpha.\lambda(X_i).\pi(X_i) \tag{2.16}$$

avec $\alpha = \frac{1}{P(e)}$. L'équation 2.16 est le terme qu'il faut calculer pour résoudre le problème de l'inférence. Les valeurs de $\lambda(X_i)$ et $\pi(X_i)$ sont alors propagées à travers le réseau. Nous présentons dans la suite la méthode de calcul de ces deux termes.

Calcul de λ. Si X_i est une variable observée $X_i = x_i^0$, les éléments du vecteur $\lambda(X_i)$ sont calculés de la manière suivante :

$$\begin{aligned} \lambda(X_i) = 0 \quad si \quad x_i \neq {x_i}^0 \\ \lambda(X_i) = 1 \quad si \quad x_i = {x_i}^0 \end{aligned} \tag{2.17}$$

Dans le cas où X_i, n'est pas observée, $\lambda(X_i)$ est calculé en utilisant $\lambda(Y_1), ..., \lambda(Y_m)$, où $Y_1, .., Y_m$ sont les enfants de X_i. soit $e^-_{X_i} = \cup_{j=1}^m e^-_{Y_j}$, en utilisant 2.16, $\lambda(X_i)$ s'écrit :

$$\begin{aligned} \lambda(X_i) &= P(e^-_{X_i}|X_i) \\ &= P(e^-_{Y_1}, ..., e^-_{Y_m}|X_i) \\ &= P(e^-_{Y_1}|X_i)...P(e^-_{Y_m}|X_i) \\ &= \lambda_{Y_1}(X_i)...\lambda_{Y_m}(X_i) \end{aligned} \tag{2.18}$$

Grâce au fait que $e_{Y_1}, ..., e_{Y_m}$ sont conditionnellement indépendants l'un de l'autre et en définissant :

$$\lambda_{Y_j}(X_i) = P(e^-_{Y_j}|X_i) \tag{2.19}$$

Le calcul de $\lambda_{Y_j}(X_i)$ s'effectue de la manière suivante :

$$\begin{aligned}
\lambda_{Y_j}(X_i) &= P(e^-_{Y_j}|X_i) \qquad (2.20)\\
&= \sum_{Y_j} P(e^-_{Y_j}, Y_j|X_i)\\
&= \sum_{Y_j} P(e^-_{Y_j}|Y_j, X_i).P(Y_j|X_i)\\
&= \sum_{Y_j} P(e^-_{Y_j}|Y_j).P(Y_j|X_i)\\
&= \sum_{Y_j} \lambda_{y_j}.P(Y_j|X_i)
\end{aligned}$$

Ceci montre que pour le calcul de $\lambda(X_i)$, seules les valeurs de λ de tous les fils de X_i ainsi que la probabilité conditionnelle des fils de X_i par rapport à X_i lui-même sont nécessaires.

Calcul de π. Le calcul de π passe par le calcul du π du parent U de X_i. En effet :

$$\begin{aligned}
\pi(X_i) &= P(X_i|e^+_{X_i}) \qquad (2.21)\\
&= \sum_{U} P(X_i, U|e^+_{X_i})\\
&= \sum_{U} P(X_i|U, e^+_{X_i}).P(U|e^+_{X_i})\\
&= \sum_{U} P(X_i|U).P(U|e^+_{X_i})\\
&= \sum_{U} P(X_i|U).\pi(U)
\end{aligned}$$

Ce calcul montre que pour obtenir la valeur de $\pi(X_i)$, on a besoin des π du parent de X_i ainsi que la probabilité conditionnelle de X_i connaissant son parent. Le mécanisme de propagation des messages dans un arbre est illustré au niveau de la figure 2.5.

2.4.2 Inférence dans une structure générique : algorithme du Jtree

L'algorithme présenté dans le paragraphe précédent ne fonctionne que dans le cas où on dispose d'une structure en arbre. Dans [57] et [55], les auteurs proposent de transformer toute structure générique en une structure d'arbre appelée arbre de jonction de façon à lui appliquer l'algorithme de propagation de connaissances. Cet arbre de jonction est obtenu par regroupement des nœuds de la structure générique.

D'après Jensen [50], la construction de l'arbre de jonction, correspondant à un réseau bayésien $\mathcal{B}$ ayant pour ensemble de variables $X = \{X_i\}$, passe par les étapes suivantes :

- construire un graphe moral : un graphe non dirigé avec des connexions entre toutes les variables dans l'ensemble $pa(X_i) \cup X_i$ pour toute variable $X_i \in X$.

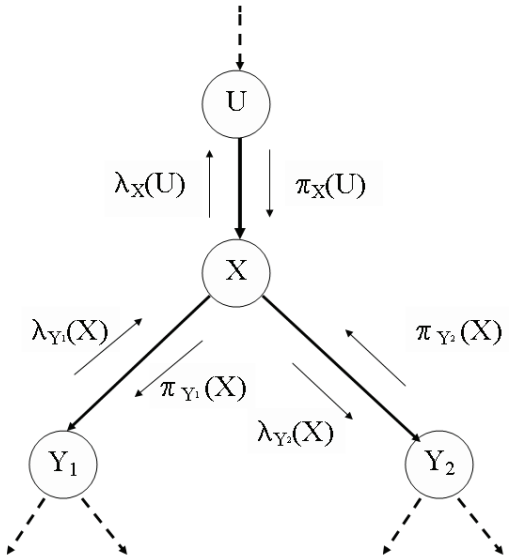

FIG. 2.5 – Mécanisme de propagation de connaissances au niveau d'un nœud dans un réseau bayésien en forme d'arbre.

- trianguler le graphe moral : ajouter des connections jusqu'à ce que tous les cycles ayant plus de trois connexions aient une corde. Les nœuds de l'arbre de jonction sont alors les cliques du graphe triangulé ;
- connecter les cliques du graphe triangulé avec des liens afin de construire l'arbre de jonction ;
- donner à chaque clique une table ne contenant que des unités. Pour chaque variable X_i, trouver une clique contenant $pa(X_i) \cup X_i$ et multiplier sa table par $P(X_i|pa(X_i))$.

L'arbre de jonction ainsi construit représente le réseau bayésien $\mathcal{B}$.

2.4.3 Inférence approchée

L'inférence exacte dans un réseau bayésien est un problème NP-difficile [58]. Une inférence exacte n'est donc pas toujours possible d'un point de vue calculatoire, du fait du très grand nombre de variables. Des méthodes d'inférence approchées ont alors été proposées pour les réseaux très complexes.

Une première catégorie de méthodes approchées [59] opère sur des réseaux approchés du réseau réel. Ces réseaux approchés sont construits en éliminant un certain nombre de connexions jugées faibles du réseau réel.

La seconde catégorie de méthodes se base sur les méthodes stochastiques de simulation. Le principe est d'effectuer un grand nombre de tirages aléatoires conformément aux probabilités conditionnelles du réseau bayésien et d'estimer ensuite la probabilité

recherchée. Parmi cette catégorie de méthodes, on peut citer les méthodes dites de Monte-Carlo ou de *Probabilistic Logic Sampling* [60]. Pour plus de détails concernant les méthodes d'inférence approchées, le lecteur peut se référer à [61] .

2.5 Apprentissage dans les réseaux bayésiens

L'étape d'inférence utilise un modèle complet avec une structure bien défini ainsi que des paramètres représentant les tables de probabilités conditionnelles déjà fixées. Nous nous intéressons, dans cette partie, à la phase d'apprentissage du réseau bayésien.

2.5.1 Apprentissage de paramètres

Les paramètres d'un réseau bayésien sont les probabilités conditionnelles au niveau de chaque nœud. Ils peuvent être fixés manuellement si on les connaît. Toutefois, cela s'avère être très contraignant surtout si le système est composé d'un grand nombre de variables, ce qui se traduit par un grand nombre de paramètres.

On estime donc le plus souvent les paramètres à partir d'une base de données d'apprentissage. Les types d'apprentissage de paramètres les plus connus dans la littérature sont l'apprentissage statistique et l'apprentissage bayésien.

2.5.1.1 Apprentissage statistique

Le but principal de l'apprentissage statistique est la maximisation de la vraisemblance (équation 2.22 des données d'apprentissage $\mathcal{D}$ par rapport à l'ensemble des paramètres θ du modèle.

$$P(\mathcal{D}|\theta) = \prod P(X_i|\theta) \tag{2.22}$$

Cette maximisation est équivalente à la maximisation de la log-vraisemblance :

$$\mathcal{L}(\mathcal{D}|\theta) = \sum \log(P(X_i|\theta)) \tag{2.23}$$

En utilisant la probabilité jointe du réseau bayésien, la vraisemblance se décompose donc comme une somme de termes. L'estimation de paramètres revient alors à la maximisation des termes de cette somme. Notre problème de maximisation initial se décompose donc en des problèmes locaux d'estimation au niveau de chaque nœud du réseau bayésien.

Cet apprentissage s'applique lorsque toutes les variables sont observées. C'est la méthode la plus simple et la plus utilisée dans la littérature. Elle consiste à estimer la probabilité d'un événement par sa fréquence d'apparition dans la base de données. Cette approche est également appelée maximum de vraisemblance MV. Dans ce cas :

$$\begin{aligned} \theta_{ijk}^{MV} &= P(X_i = x_i^k | pa(X_i) = px_j) \\ &= \frac{N_{ijk}}{\sum_k N_{ijk}} \end{aligned} \tag{2.24}$$

où N_{ijk} est le nombre d'occurrences dans la base de données de la configuration « X_i est dans l'état x_i^k et ses parents sont dans la configuration px_j ».

2.5.1.2 Apprentissage bayésien

Le but de l'apprentissage bayésien est la maximisation de $P(\theta|\mathcal{D})$ contrairement à l'apprentissage statistique où il fallait maximiser $P(\mathcal{D}|\theta)$. On cherche donc θ^{MAP} tel que :

$$\begin{aligned} \theta^{MAP} &= \arg\max P(\theta|\mathcal{D}) \\ &= \arg\max P(\mathcal{D}|\theta)P(\theta) \qquad (2.25) \\ & \qquad (2.26) \end{aligned}$$

La quantité $P(\theta|\mathcal{D})$ peut être factorisée comme suit :

$$P(\theta|\mathcal{D}) = \prod_i \prod_j \prod_k P(\theta_{ijk}|\mathcal{D}) \qquad (2.27)$$

Pour le calcul de $P(\theta)$ on suppose que nous avons comme *a priori* sur les paramètres de réseaux une distribution de Dirichlet. Ainsi :

$$P(\theta) = \prod_i \prod_j \prod_k \theta_{ijk}^{\alpha_{ijk}-1} \qquad (2.28)$$

où α_{ijk} peut être vu comme un nombre d'occurrences *a priori* de la configuration « X_i est dans l'état x_k et ses parents sont dans la configuration px_j » si cette dernière n'est pas observée dans la base de données.

Dans [62], l'auteur a démontré que $\theta_{ijk}^{MAP} = P(X_i = x_k|pa(X_i) = px_j)$ s'exprime de la façon suivante :

$$\theta_{ijk}^{MAP} = \frac{N_{ijk} + \alpha_{ijk} - 1}{\sum_k N_{ijk} + \alpha_{ijk} - 1} \qquad (2.29)$$

2.5.2 Apprentissage de la structure

La structure d'un réseau bayésien est un paramètre qualitatif du réseau. L'apprentissage de cette structure constitue une étape clé de notre travail. Plutôt que d'en donner un aperçu dès à présent, nous préférons consacrer une grande partie du chapitre suivant à la présentation des différents aspects de cet apprentissage.

2.6 Réseaux bayésiens particuliers

2.6.1 Réseau bayésien naïf

De tels réseaux ont été largement utilisés dans la littérature pour la classification [63]. Ils se caractérisent, en effet, par la rapidité des opérations d'apprentissage et d'inférence.

Les réseaux bayésiens naïfs correspondent à un type particulier de réseaux bayésiens où l'on suppose que les attributs $X_1, \cdots, X_n$ sont indépendants conditionnellement à la variable classe que nous distinguerons par la notation X_c. Cette hypothèse entraîne la simplification de la loi jointe de l'équation 2.11 sous la forme de l'équation 2.30.

$$P(X_c, X_1, ...X_n) = \prod_{i=1..n} P(X_i|X_c) \tag{2.30}$$

Dans les réseaux bayésiens naïfs, aucune corrélation entre les attributs n'est prise en compte. Toutes les caractéristiques contribuent à la classification de la même manière. Le nœud de classification profite de l'information donnée par chaque caractéristique indépendamment de l'information issue des autres caractéristiques.

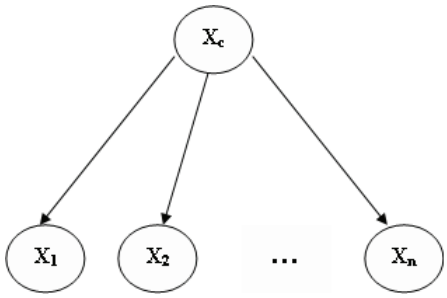

FIG. 2.6 – Réseau bayésien naïf.

2.6.2 Réseau bayésien dynamique

Un réseau bayésien dynamique a pour objectif de modéliser la distribution de probabilité d'une suite de variable sur une séquence dans le temps $T \in \mathbb{N}^*$. Il peut être défini par un couple de réseaux bayésiens$(\mathcal{B}_0, \mathcal{B}trans)$: $\mathcal{B}_0$ a pour but de définir la loi initiale des variables et $\mathcal{B}trans$ définit la loi de transition entre les variables pour deux tranches temporelles successives.

Les réseaux bayésiens dynamiques sont considérés à la fois comme une généralisation et comme un cas particulier des réseaux bayésiens statiques. En effet, en déroulant le réseau sur les T tranches temporelles, on récupère des réseaux bayésiens statiques équivalents au réseau dynamique. La loi jointe du modèle est alors calculée grâce à la formule suivante :

$$P_{\mathcal{B}}(X[0], ..., X[T]) = P_{\mathcal{B}_0} \prod_{t=0}^{T-1} P_{\mathcal{B}trans(X(t+1)|X(t))} \tag{2.31}$$

Dans [43], l'auteur montre que les réseaux bayésiens dynamiques permettent d'unifier de nombreuses approches issues de la modélisation des séries temporelles telles que les modèles de Markov, ou les filtres de Kalman.

2.7 Conclusion

Dans ce chapitre, nous avons proposé une introduction générale du formalisme des réseaux bayésiens. Nous avons également présenté un bref état de l'art des mécanismes d'inférence et d'apprentissage nécessaires à la mise en place et à l'utilisation d'un réseau bayésien.

Comme il a été évoqué dans ce chapitre, les paramètres d'un réseau bayésien sont les probabilités conditionnelles et la structure du réseau. Nous nous attardons dans le chapitre suivant sur l'étude de cette structure et des différents méthodes d'apprentissage utilisées dans la littérature. Nous présenterons également un cas d'utilisation de l'apprentissage de structure et nous étudierons son apport dans un système d'indexation vidéo.

Chapitre 3

Apprentissage de structure pour l'indexation vidéo

3.1 Introduction

Comme il a déjà été expliqué dans le chapitre 2, la conception d'un réseau bayésien passe par l'apprentissage des paramètres du réseau, mais aussi par la mise en place de la structure en elle-même. La structure correspond à l'ensemble de toutes les connexions qui existent entre les variables du système. Ces connexions représentent les corrélations entre les différentes variables. Dans la majorité des modèles d'indexation vidéo basés sur les réseaux bayésiens, les connexions entre les attributs sont fixées par le concepteur du modèle. Ce dernier peut s'appuyer sur des connaissances *a priori* sur le domaine de la vidéo à traiter. Il faut toutefois disposer de ces connaissances *a priori*, et il n'est pas toujours évident de les obtenir. Généralement, lorsqu'on ne dispose pas de connaissances sur le domaine traité, les connexions sont fixées d'une manière arbitraire, en supposant par exemple une indépendance entre les attributs. Toutefois, dans la majorité des cas, cette hypothèse d'indépendance ne tient pas. Par ailleurs, tenir compte des corrélations entre les attributs peut permettre d'augmenter les performances du système. Pour déterminer de façon automatique ces différentes corrélations, nous proposons dans ce chapitre d'utiliser une étape d'apprentissage de la structure elle-même, cette étape est en effet disponible dans le cadre théorique des réseaux bayésiens.

Nous réalisons cet apprentissage de structure pour l'indexation vidéo. Les attributs utilisés étant le plus souvent des attributs de bas niveau, trouver manuellement les différentes corrélations qui existent entre ces attributs n'est pas évident. Le formalisme des réseaux bayésiens permet d'apprendre ces corrélations à partir d'une base d'apprentissage. Cet atout des réseaux bayésiens a été très peu utilisé dans le cadre de l'indexation vidéo.

Dans ce chapitre, nous allons donc introduire l'apprentissage de structure dans les réseaux bayésiens comme méthode pour améliorer les performances du système d'indexation vidéo. Nous allons voir que cette méthode permet d'automatiser la construction du modèle et de nous affranchir de tout besoin d'information *a priori* sur le problème

à résoudre, généralement assez difficile à avoir.

Nous présentons en premier lieu les mécanismes théoriques de l'apprentissage de structure, ainsi que les différentes méthodes existantes permettant de le réaliser. Nous nous attardons sur les méthodes d'apprentissage de structure basées score. Nous présentons enfin un cas d'utilisation de l'apprentissage de structure pour améliorer la performance des systèmes d'indexation vidéo.

3.2 Apprentissage de structure

Comme il a déjà été énoncé dans le chapitre 2, il existe deux types d'apprentissage dans les réseaux bayésiens : un apprentissage quantitatif (c'est-à-dire apprentissage des paramètres) permettant d'apprendre les différentes distributions de probabilités au niveau de chaque nœud et un apprentissage qualitatif permettant d'apprendre la structure, c'est-à-dire l'existence d'arcs entre les nœuds du réseau. Il est évident que chaque fois qu'on change de structure, on doit refaire l'apprentissage des paramètres. L'apprentissage de structure, comme tout type d'apprentissage, utilise une base de données d'apprentissage.

Il existe deux classes d'algorithmes pour faire de l'apprentissage de structure dans les réseaux bayésiens :

- les algorithmes basés sur la recherche de causalité entre les variables ;
- les algorithmes basés sur les scores.

3.2.1 Apprentissage de structure par recherche de causalité

Dans ce type de méthodes, on cherche à déterminer, à partir des données d'apprentissage, les relations d'indépendance qui existent entre deux variables conditionnellement à l'ensemble des variables du problème. Un état de l'art complet des méthodes d'apprentissage de structure basées sur la recherche de causalité est proposé dans [61]. Nous explicitons toutefois les principales méthodes dans ce qui suit.

Deux équipes concurrentes ont travaillé sur le développement de méthodes d'apprentissage de structure par la recherche de causalité : Pearl et Verma avec les algorithmes IC et IC* [64] et Spirtes, Glymour et Scheines pour les algorithmes SGS et PC [65].

L'algorithme PC (pour Peter et Clark) a été introduit dans [65] par Spirtes, Glymour et Scheines. Il utilise des tests statistiques pour évaluer l'indépendance conditionnelle entre les variables du réseau et reconstruire ainsi la structure du réseau bayésien à partir de ces relations d'indépendance. L'algorithme est initialisé par un graphe complètement connecté. Un arc est supprimé si une indépendance conditionnelle entre les deux variables qu'il relie est détectée. Afin de réduire le nombre de tests effectués, ceux-ci sont alors réalisés selon un ordre préalablement établi sur les variables.

Le deuxième type d'algorithme d'apprentissage de structure par recherche de causalité est l'algorithme IC introduit dans [64]. Cet algorithme suit le même principe que l'algorithme PC, à la différence près qu'il est initialisé par une structure vide et qu'on procède par ajout d'arcs au fur et mesure que la structure se construit. Les arcs sont

ajoutés lorsqu'une dépendance entre deux variables conditionnellement à un ensemble de variables est détectée.

D'une façon générale, la recherche des relations d'indépendance dans les méthodes d'apprentissage de structure par recherche de causalité passe par l'utilisation de tests statistiques d'indépendance. On distingue principalement deux tests de ce type :

- le test du χ^2 ;
- le test du rapport de vraisemblance.

Ces deux tests mesurent l'adéquation d'une distribution théorique à une distribution observée. Dans le cadre de leur utilisation pour l'apprentissage de la structure dans les réseaux bayésiens, on teste l'indépendance conditionnelle de deux variables X_i et X_j conditionnellement à un ensemble de variables X_p. On compare ainsi une loi observée à travers les données d'apprentissage à la loi théorique où l'on suppose que les deux variables sont indépendantes conditionnellement à un ensemble de variables parents.

Outre le fait que le nombre de tests d'indépendance, et ainsi la complexité de la méthode, augmentent exponentiellement en fonction du nombre de variables dans le système, cette classe de méthodes d'apprentissage souffre également d'un autre défaut. Elle est en effet basée sur la découverte des indépendances entre les variables et cherche ainsi la structure qui représente le mieux les données. Elle ne peut donc être utilisée si on vise des tâches autres que la description même des liens causaux qui existent entre les variables.

Nous verrons dans le paragraphe suivant, un deuxième type d'apprentissage qui offre, de par sa construction, plus de flexibilité que ce premier type.

3.2.2 Apprentissage de structure par optimisation de scores

Les techniques d'apprentissage de structure basées sur l'optimisation d'un score ont pour origine la théorie de la sélection des modèles stochastiques.

Étant donné un ensemble de modèles, la sélection de modèles par pénalisation consiste à choisir le modèle qui maximise un critère donné. Souvent le critère choisi est composé de la somme de deux termes : un premier terme permettant de mesurer le pouvoir de description des données par le modèle et un deuxième terme mesurant la complexité du modèle. Le deuxième terme est introduit pour permettre une généralisation du modèle et éviter le problème de sur-apprentissage. Ainsi la complexité doit être réduite au maximum tout en gardant un pouvoir de description le plus fiable possible. Dans le cas de l'apprentissage de structure, le modèle à sélectionner est la structure du réseau bayésien. Il s'agit donc de choisir la structure impliquant le moins possible de paramètres et qui permet de décrire au mieux les données d'apprentissage. Cette structure est donc cellequi correspond au critère ou au score optimum.

La flexibilité de ce type de méthodes réside dans le fait que le score n'est pas fixe. Il peut être choisi selon les besoins de l'utilisateur du système.

Pour leur mise en œuvre, un score et une méthode de parcours de l'ensemble des structures sont fixés. Différents scores existent dans la littérature. Dans le cadre de notre apprentissage de structure, nous en présenterons trois : le critère AIC (Akaike Information Criterion), le critère BIC (Bayesian Information Criterion) et le critère

d'information mutuelle. Nous présentons dans la suite ces différents scores, ainsi que les différentes méthodes de parcours de l'espace de structures.

3.2.2.1 Propriétés des fonctions de score

Dans les méthodes d'apprentissage de structure décrites dans cette partie, nous cherchons la structure qui optimise la valeur d'un score donné. Ce score est fixé par l'utilisateur de l'algorithme. Pour permettre une généralisation du modèle, tout en gardant un bonne description des données d'apprentissage, il est nécessaire de tenir compte à la fois de la capacité de représentation et de la complexité de la structure apprise. Pour cela, on introduit ici une mesure de la complexité d'une structure.

Dimension d'un réseau bayésien La dimension d'un réseau bayésien $\mathcal{B}$, $Dim(\mathcal{B})$, est le nombre de paramètres indépendants nécessaires à la description du réseau bayésien. Soit un réseau bayésien $\mathcal{B}$ constitué de n nœuds décrits par les variables $\{X_1, ..., X_n\}$ avec d_i le nombre de configurations possibles de la variable X_i, et D_i le nombre de configurations possibles de l'ensemble des parents $pa(X_i)$ de X_i. Pour représenter la probabilité associée au nœud X_i, $P(X_i|pa(X_i))$, on utilise $dim(X_i) = (d_i - 1).D_i$ paramètres. La dimension totale du réseau bayésien est la somme au niveau de chaque nœud du nombre de paramètres nécessaires à sa description, $dim(X_i)$. La dimension du réseau bayésien $\mathcal{B}$ s'écrit alors sous la forme :

$$Dim(\mathcal{B}) = \sum_{i=1}^{n} dim(X_i, \mathcal{B}) \tag{3.1}$$

$$= \sum_{i=1}^{n} (d_i - 1).D_i \tag{3.2}$$

La dimension d'un réseau bayésien est un indicateur sur sa complexité. Cette quantité sera beaucoup utilisée dans la formulation des scores, pour équilibrer la capacité de description du réseau et sa complexité.

Score décomposable Un score $\mathscr{S}(\mathcal{B})$ d'un réseau bayésien $\mathcal{B}$, est décomposable s'il peut s'écrire comme la somme des scores au niveau de chaque nœud du réseau. Le score $s(X_i, pa(X_i))$ au niveau de chaque nœud est une fonction des paramètres du nœud X_i et des paramètres de l'ensemble de ses parents $pa(X_i)$. Un score décomposable d'un réseau bayésien $\mathcal{B}$ peut alors s'écrire sous la forme :

$$\mathscr{S}(\mathcal{B}) = \sum_{i=1}^{n} s(i, pa(i)) \tag{3.3}$$

Les scores décomposables sont très recherchés pour évaluer la complexité des structures des réseaux bayésiens. En effet, pour évaluer l'ajout d'un nouveau parent à l'ensemble $pa(X_i)$ des parents de X_i, il suffit seulement de recalculer le score au niveau du nœud X_i. Il n'est pas nécessaire de recalculer le score sur tout le réseau, ce qui peut

faire diminuer considérablement la complexité et le temps de calcul des algorithmes d'apprentissage de structure.

Nous décrivons dans la suite trois différentes formules de scores classiquement utilisées dans la littérature.

3.2.2.2 Score AIC

Le score AIC a été proposé dans [66]. Il a comme expression :

$$\mathscr{S}_{AIC}(\mathcal{B}) = \log P(\mathcal{D}|\theta^{MV}, \mathcal{B}) - \frac{1}{2}Dim(\mathcal{B}) \tag{3.4}$$

où $\mathcal{D}$ est l'ensemble de données d'apprentissage et θ^{MV} sont les paramètres des réseaux obtenus par maximum de vraisemblance.

Il peut être interprété comme le fait de choisir le modèle qui décrit au mieux les données en désavantageant les structures les plus complexes. Ce score a l'avantage d'être simple par sa mise en œuvre. Toutefois, il a l'inconvénient de favoriser les modèles complexes si on dispose d'une grande base d'apprentissage. En effet, le premier terme, constitué par la vraisemblance des données par rapport au modèle, augmente lorsqu'on augmente le nombre de données dans la base. Le terme de complexité ne dépend quant à lui que de la dimension du modèle. Donc si le nombre de données augmente considérablement, c'est le premier terme qui prend le dessus dans le score.

3.2.2.3 Score BIC

Le score BIC a été proposé par Schwarz [67]. Il a comme expression :

$$\mathscr{S}_{BIC}(\mathcal{B}) = \log P(\mathcal{D}|\theta^{MV}, \mathcal{B}) - \frac{1}{2}Dim(\mathcal{B}) \log N \tag{3.5}$$

N est le nombre d'exemples disponibles dans la base d'apprentissage $\mathcal{D}$. Le premier terme du score BIC représente la vraisemblance des données par rapport au modèle. Au travers de la maximisation de ce terme, on cherche la structure qui simule le mieux nos données. Le deuxième terme tient compte de la complexité de la structure ainsi que du nombre d'exemples présents dans la base ; donc plus la structure est de grande dimension, plus le deuxième terme diminue le score de cette structure. Le score BIC pénalise ainsi les structures ayant un grand nombre de paramètres même si on dispose d'une base de données d'apprentissage avec un grand nombre d'exemples. D'autre part, le score s_{BIC} est décomposable. Il s'écrit en effet comme la somme de termes calculés de manière indépendante au niveau de chaque nœud.

$$\mathscr{S}_{BIC}(\mathcal{B}) = \sum_{i=1}^{n} s_{BIC}(X_i, pa(X_i), \mathcal{D}) \tag{3.6}$$

Où $s_{BIC}(X_i, pa(X_i), \mathcal{D})$ s'écrit sous la forme :

$$s_{BIC}(X_i, pa(X_i), \mathcal{D}) = \log P(X_i|pa(X_i), \mathcal{B}, \hat{\theta}_i) - \frac{1}{2}dim(X_i, \mathcal{B}) \log N \tag{3.7}$$

où $\hat{\theta}_i$ le paramétre relatif au nœud X.

La propriété de décomposabilité est l'une des raisons de la large utilisation du score BIC en apprentissage de structure.

3.2.2.4 Score d'information mutuelle

Ce score a été proposé par Chow *et al* dans [68]. Il se base sur l'information mutuelle entre deux variables. Le score au niveau de nœud X_i, en supposant que X_j est son parent, s'écrit sous la forme :

$$IM(X_i, X_j) = \sum_{x_i, x_j} P(X_i = x_i, X_j = x_j) \log \frac{P(X_i = x_i, X_j = x_j)}{P(X_i = x_i)P(X_j = x_j)} \tag{3.8}$$

Le score global est égal à la somme sur les scores de tous les nœuds. Ce score est par construction décomposable. Il a été montré dans [69] que l'augmentation du score basé sur l'information mutuelle traduit l'augmentation de la vraisemblance des données par rapport au modèle.

3.2.2.5 Parcours de l'espace de recherche

Dans les méthodes basées sur le calcul d'un score, on choisit au final la structure qui a le plus grand score parmi un ensemble de structure données. Cela suppose qu'on parcoure exhaustivement l'ensemble des structures possibles. Dans [70], l'auteur a prouvé que le nombre de structures possibles pour un réseau bayésien contenant n nœuds est donné par la formule 3.9. Ce nombre a une complexité sur-exponentielle ($R(5) = 29281$ et $R(10) = 4,2 \times 10^{18}$). Le parcours exhaustif de l'ensemble des structures devient donc vite ingérable dès que le nombre de nœuds dans le réseau augmente.

$$R(n) = \sum_{i=1}^{n} (-1)^{i+1} \binom{n}{i} 2^{i(n-1)} R(n-i) \tag{3.9}$$

On doit donc définir des méthodes de parcours non exhaustif de l'ensemble des structures. Le type de parcours variera alors selon l'application envisagée.

Nous présentons à présent les algorithmes de parcours généralement utilisés dans la littérature.

Recherche dans l'espace des arbres. Dans ce type de méthodes, on restreint la recherche de la structure optimale à l'espace des arbres. On utilise l'algorithme MWST (Maximum Weight Spanning Tree [71]) pour trouver l'arbre de recouvrement maximal. Cet algorithme permet de récupérer l'arbre qui passe par tous les nœuds et qui maximise un score donné. Dans [68], les auteurs proposent d'utiliser le score basé sur l'information mutuelle. Chaque connexion entre deux nœuds reçoit donc un poids, correspondant à l'information mutuelle entre les deux variables concernées. L'arbre qui maximise le poids des connexions est ensuite recherché. L'arbre récupéré est un arbre non orienté reliant tous les nœuds. Pour le transformer en arbre orienté, il suffit de choisir un nœud racine

et de diriger chaque arrête à partir de ce nœud. L'arbre ainsi construit constitue la structure du modèle. Dans cet algorithme, on est contraint de choisir un nœud racine. Nous verrons au le paragraphe 3.3.5.1 l'influence du choix de ce nœud racine dans la structure finale.

L'algorithme K2. L'algorithme K2 a été proposé par Cooper *et al.* dans [72]. Les auteurs utilisent le score BIC pour évaluer les structures. Toutefois, n'importe quel score décomposable peut être utilisé. La méthode de parcours de l'algorithme K2 se base sur un ordonnancement de l'ensemble des nœuds à l'initialisation de l'algorithme. Chaque nœud est considéré comme un parent potentiel de tous les nœuds qui le suivent dans l'ordre fixé, par contre il ne peut en aucune manière être parent des nœuds qui le précèdent.

L'algorithme procède à la recherche des parents de chaque nœud en suivant l'ordre de parenté introduit à l'initialisation. Ainsi, le nœud qui arrive le premier dans l'ordre est proposé comme parent potentiel de tous les nœuds. Il est supposé ne pas avoir de parent et correspond donc à la racine de toute la structure. Pour chaque nœud, on dispose d'un ensemble de parents potentiels qui sont les nœuds précédents. On cherche parmi ces nœuds le parent qui augmente le plus le score. Une fois ce nœud trouvé, on l'ajoute à l'ensemble des parents confirmés. On itère ainsi en cherchant à augmenter le score à chaque itération. On arrête l'algorithme lorsqu'aucun parent supplémentaire ne permet d'augmenter le score.

Recherche gloutonne. Cette méthode permet la recherche de la structure dans l'ensemble complet des structures possibles [73]. Elle part d'une structure initiale et, à travers des opérations élémentaires, suppression, inversion ou ajout d'arcs, elle établit un ensemble de structures voisines. La structure voisine de score maximal est alors choisie comme point de départ pour l'itération suivante. Cette méthode se caractérise par sa complexité de calcul puisqu'un grand nombre d'itérations est nécessaire avant de converger. Le nombre d'itérations dépend principalement de la structure d'initialisation. Un autre inconvénient de cette méthode réside dans le fait qu'il est possible de converger vers un minimum local.

Nous avons, jusque là, présenté les aspects théoriques de l'apprentissage de structure dans un réseau bayésien. Nous étudions dans le paragraphe suivant la mise en œuvre pratique de ce type d'apprentissage dans le cadre d'un système d'indexation vidéo. Nous évaluons l'apport que peut apporter un tel apprentissage sur le système en termes de performance et de facilité de mise en œuvre.

3.3 Résultats et interprétation

3.3.1 Cadre applicatif

L'application d'indexation vidéo que nous avons choisie pour cette partie est la détection de plages publicitaires dans un flux vidéo de télévision. Nous explicitons dans ce paragraphe le protocole expérimental de notre application. Nous utilisons six attributs audio-vidéo :

- le niveau audio dans le flux,.
- la cohérence de la couleur avec les plans voisins,.
- l'intensité du mouvement dans le plan,
- la longueur des plans,
- la surface de texte dans une image,
- le nombre de zones de texte présentes dans l'image.

Un ensemble d'attributs est associé à chaque plan de la vidéo. Les détails concernant l'étape d'extraction des attributs sont explicités dans l'annexe 7.2.

Le problème de détection de plages publicitaires peut être modélisé par le réseau bayésien à caractère dynamique de la figure 3.1.

Cette représentation peut être divisée en deux étages.

- Dans un premier étage, on représente les contraintes temporelles du problème de détection de plages publicitaires. Cet étage représente la corrélation entre la variable *Publicité* à un temps donné et ses occurrences dans les temps voisins au temps courant. Cet étage est donc constitué de la variable *Publicité* et de la variable D représentant la durée de la plage de publicité. La variable D est régie par les contraintes suivantes :

si non Publicité	$D = -1$
sinon :	
si début de Publicité	$D =$ durée estimée du segment de la plage de publicité
sinon	$D = D - 1$

  Ainsi au début supposé de chaque segment de publicité, on estime la durée D de la plage de publicité. L'état de la variable *Publicité* ne peut changer vers l'état non *Publicité* qu'à l'expiration de la variable D.
- Le second étage représente l'interaction de la variable *Publicité* avec les variables attributs. Les interactions entre les attributs étant inconnues dans la majorité des cas, l'hypothèse que tous les attributs sont conditionnellement indépendants par rapport à la variable *Publicité* est généralement utilisée, comme cela est illustré dans la figure 3.1. Dans notre cas d'étude, nous supposons une indépendance conditionnelle de l'ensemble d'attributs à un instant donné par rapport aux ensembles d'attributs aux instants voisins. Les interactions temporelles existent donc seulement au niveau de la variable *Publicité* au niveau du premier étage du réseau de la figure 3.1. Nous tenons compte toutefois des corrélations entre les attributs à un instant donné. Nous proposons alors d'utiliser l'apprentissage de structure

pour apprendre automatiquement les différentes interactions entre les attributs.

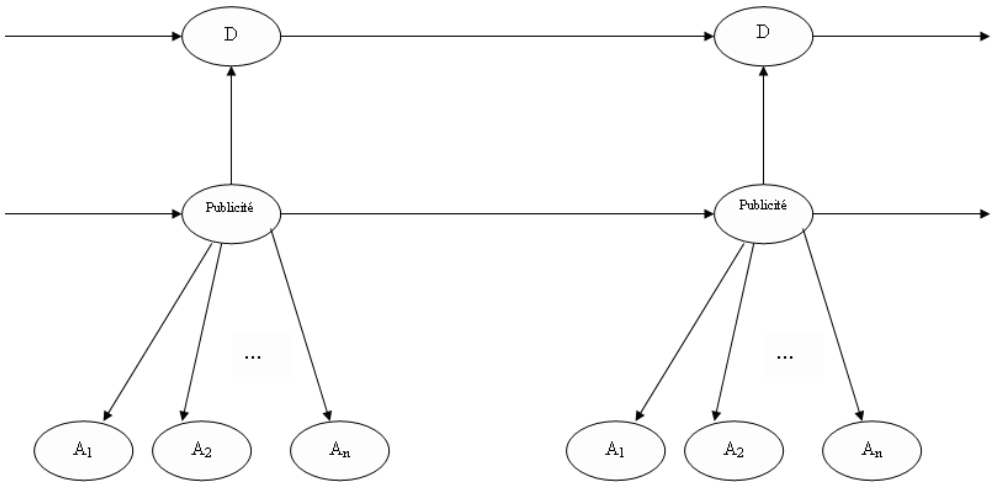

FIG. 3.1 – Réseau bayésien pour la détection de publicité en supposant l'indépendance des attributs A_i.

3.3.2 Construction du modèle

La structure du premier étage, l'étage temporel, est construite manuellement. Nous appliquons l'apprentissage de structure au second étage. Dans la figure 3.1, nous présentons un système sans apprentissage de structure. Ainsi, aucune corrélation entre les attributs n'est prise en considération. Nous proposons de remplacer ce second etage par une structure apprise entre le nœud événement, c'est à dire le nœud *Publicité*, et les autres nœuds représentant les attributs à une plage temporelle donnée.

Puisque dans notre réseau, les deux étages, l'étage temporel et celui représentant les attributs, sont conditionnellement indépendants par rapport à la variable *Publicité*, nous pouvons introduire la structure apprise aisément dans la structure globale.

3.3.3 Base de données

Notre base de données est composée de 36 heures de vidéo extraites de trois chaînes françaises généralistes (France2, TF1 et M6) prises à des tranches horaires variées. Nous disposons d'une vérité de terrain localisant les plages de publicité. Puisque notre base de données n'est pas très grande, nous utilisons un processus tournant de cross-validation pour étudier les performances de notre système. À chaque étape de la cross validation, nous utilisons 32 heures de vidéo pour l'apprentissage de la structure et des paramètres, et 4 heures pour l'inférence.

Nœud racine	Précision	Rappel
Publicité	72	90
Cohérence de la couleur	70	90
Longueur de plan	62	90
Audio	62	90

TAB. 3.1 – Choix du nœud racine dans le cas d'un apprentissage d'une structure en arbre.

3.3.4 Évaluation

Pour chaque plan, nous récupérons la probabilité que le plan soit contenu dans une plage de publicité ou non. Pour évaluer notre système, nous utilisons les mesures de précision et de rappel. Nous rappelons ces deux notions :

$$\text{Précision} = \frac{N_c}{N_c + N_f} \tag{3.10}$$

$$\text{Rappel} = \frac{N_c}{N_c + N_m} \tag{3.11}$$

où N_c est le nombre de plans *événement* correctement détectés, N_m le nombre de plans *événement* non détectés et N_f le nombre de fausses alarmes. N_c+N_f est donc le nombre d'*événements* récupérés par le système. $N_c + N_m$ correspond au nombre d'*événements* dans la base de données.

3.3.5 Mise en œuvre

Les trois méthodes d'apprentissage de structure que nous avons présentées au début de ce chapitre nécessitent des réglages techniques pour leur mise en œuvre. Nous exposons dans les paragraphes qui suivent les différents réglages nécessaires pour obtenir les résultats escomptés. Nous justifions également ces différents choix.

3.3.5.1 Influence du choix du nœud racine de la structure en arbre

La structure recherchée dans ce type d'apprentissage est une structure en arbre. Si l'apprentissage est automatique, il faut cependant préciser le nœud racine de l'arbre. D'une manière intuitive nous avons supposé que le nœud le plus à même d'être le nœud racine est le nœud de la variable *Publicité*. Nous avons toutefois effectué divers tests pour étudier l'effet du changement du nœud racine. Le tableau 3.1 présente les performances des différentes structures obtenues à partir de nœuds racine différents.

Ces résultats montrent que notre choix d'origine, qui est de choisir le nœud de classification comme racine, n'était pas erroné. C'est effectivement cette structure qui donne les meilleurs résultats. En effet, supposer que le nœud de classification est la racine de l'arbre signifie que ce nœud est la cause des attributs de notre problème, ce qui est compatible avec les données dont nous disposons : le fait que le plan appartienne à une plage de publicité ou non, est la cause de la variation des différents autres attributs.

Ordre de parcours	Précision	Rappel
Aléatoire	65	90
Issu de la structure en arbre	78	90

TAB. 3.2 – Influence de l'ordre de parcours des nœuds pour l'apprentissage de structure par l'algorithme K2.

3.3.5.2 Influence de l'utilisation de l'ordre dans l'apprentissage de structure par l'algorithme K2

Comme nous l'avons explicité dans le paragraphe 3.2.2.5 de ce chapitre, l'algorithme K2 requiert un ordre de parenté pour son initialisation. Un nœud X_i ayant un ordre supérieur à celui d'un autre nœud X_j ne peut pas être parent de X_j. En contrepartie, X_j est un parent potentiel de X_i. Cet ordre peut être donné de façon arbitraire. Il peut aussi être fixé en utilisant l'ordre de la structure en arbre.

Les résultats donnés dans le tableau 3.2 montrent l'importance de l'ordre pour l'algorithme d'apprentissage basé sur le K2. Ainsi, l'utilisation d'un ordre issu de la structure en arbre permet d'augmenter les performances du système. Ceci permet de donner une première ossature à la structure recherchée. L'algorithme K2 se base alors sur cette ossature puis la complète afin de maximiser le score BIC. Il est donc important d'utiliser un ordre proche de l'ordre de parenté réel qui existe dans les données.

3.3.6 Résultats

Dans cette partie nous présentons les résultats de l'apprentissage de structure pour la détection de plages publicitaires.

Dans le tableau 3.3, nous comparons les résultats de la détection des plages de publicité en se basant sur un réseau bayésien naïf et les résultats des autres réseaux obtenus par apprentissage de structure en utilisant tout d'abord un parcours dans l'espace des arbre, puis un parcours dans l'espace des structures (K2 et recherche gloutonne). Nous remarquons que les systèmes utilisant les structures apprises donnent de meilleurs résultats que le réseau bayésien naïf dans lequel on suppose l'indépendance conditionnelle des variables attributs par rapport à la variable publicité. Ainsi, la corrélation exploitée dans les structures apprises augmente les performances du système.

Une première conclusion de notre travail est que nous avons réussi à construire de manière automatique la structure du modèle. Nous sommes donc en mesure de nous affranchir d'une connaissance experte du modèle. Une seconde conclusion est, par ailleurs, qu'il existe des corrélations entre attributs et qu'il est important d'en tenir compte pour la modélisation de notre problème. Nous réalisons cela de façon automatique.

Il est intéressant de pousser plus loin notre comparaison des trois méthodes d'apprentissage de structure utilisées. Elles ne donnent en effet pas les mêmes résultats.

- La méthode MSWT donne une structure en arbre (*cf.* figure 3.2), somme toute assez simple, puisque chaque nœud doit avoir un seul parent.
- La structure apprise (*cf.* figure 3.3) à travers le K2 est plus générale. Toutefois

Type de réseau bayésien	Précision	Rappel	Durée d'apprentissage (sec)
Naïf	70	90	–
Structure en arbre	72	90	1,5
Structure obtenue par l'algorithme K2	78	90	7
Structure obtenue par la recherche gloutonne	80	90	480

TAB. 3.3 – Comparaison entre les différentes approches d'apprentissage de structure.

elle requiert un ordre de parenté entre les nœuds que nous pouvons fixer ainsi que nous l'avons mentionné dans le paragraphe 3.3.5.2 à partir de l'ordre donné par la structure en arbre.

– La structure apprise (*cf.* figure 3.4) par recherche gloutonne ne demande pas d'information *a priori* et n'est pas restrictive en termes de structure puisqu'elle donne une structure générale. Toutefois, la complexité et la convergence de cette méthode dépendent à nouveau beaucoup de la structure d'initialisation.

D'après la comparaison des différentes méthodes d'apprentissage de structure, la structure en arbre reste insuffisante pour avoir des résultats satisfaisants pour la détection de publicité, malgré sa simplicité calculatoire. Certes la méthode par la recherche gloutonne donne dans notre cas les meilleurs résultats, mais son coût calculatoire est plus élevé comparé aux deux autres méthodes, ce qui peut constituer un frein à son utilisation. La méthode d'apprentissage de structure par l'algorithme K2 est donc un bon compromis : elle donne des résultats satisfaisants tout en gardant une complexité calculatoire raisonnable.

Dans le paragraphe suivant, nous proposons d'analyser les différentes structures apprises.

3.3.7 Interprétation des différentes structures récupérées

La structure représentée dans la figure 3.2 est la structure en arbre récupérée par l'algorithme MSWT. Nous avons imposé que la racine de l'arbre corresponde au nœud contenant la variable publicité. Dans la structure récupérée, nous remarquons que les fils directs du nœud de classification sont les nœuds *cohérence de couleurs*(CC), *longueur du plan*(SL), *surface des zones de texte*(TS) et *audio*(Au). Ces variables sont effectivement considérées comme les plus pertinentes pour la classification d'un plan en plage de *publicité* ou non.

Nous remarquons aussi que le nœud de la variable *surface des zones de texte* et celui de la variable *nombre de zones de texte* sont connectés dans la structure donnée par le MWST. Ceci peut être expliqué par la forte corrélation qui existe entre ces deux variables. D'autre part, le nœud de la variable *longueur du plan* et celui de la variable *intensité de mouvement* sont également connectés. Cette connexion n'est pas évidente à

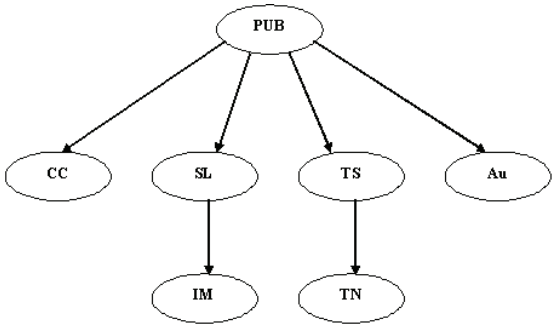

FIG. 3.2 – Structure en arbre construite par l'algorithme MSWT. *plan publicité*(PUB), *cohérence de couleurs*(CC), *intensité de mouvement*(IM), *nombre de zones de texte*(TN) *longueur du plan*(SL), *surface des zones de texte*(TS) et *audio*(Au).

expliquer. Nous pensions, en effet, que la longueur du plan et l'intensité du mouvement n'étaient pas très corrélées. Un plan peut contenir des mouvements rapides donc de grande intensité sans pour autant être de courte durée. L'inverse est tout aussi vrai : un plan de courte durée peut contenir des mouvements lents, ou être sans mouvement, comme dans le cas des plans monochromes délimitant les spots publicitaires par exemple. Toutefois, une analyse plus fine du comportement du détecteur de plan que nous avons utilisé pour cette expérience a montré une tendance du détecteur à sur-segmenter le flux vidéo, chaque fois qu'on est en présence d'un mouvement rapide. La corrélation *longueur de plan*, *intensité de de mouvement* s'explique donc.

La structure de la figure 3.3 est le résultat de l'apprentissage de structure fourni par l'algorithme K2 en utilisant un score BIC. L'ordre de parenté utilisé à l'entrée de cet algorithme est celui de la structure généré par l'algorithme MSWT. La structure récupérée est une structure générique, qui n'a pas de forme particulière, hormis le fait qu'elle respecte l'ordre de parenté introduit au début de l'algorithme.

Nous remarquons que cette structure possède le même squelette que la structure en arbre déjà interprétée dans le paragraphe précédent. Nous ne reviendrons donc pas sur les différentes connexions communes qu'ont ces deux structures. La nouvelle structure issue de l'algorithme K2 a cependant été enrichie par d'autres arcs : un arc entre le nœud de la variable *longueur de plan* et le nœud de la variable *cohérence de la couleur* et un arc entre le nœud de la variable *intensité du mouvement* et la variable *cohérence de la couleur*. On peut attribuer l'augmentation des performances du système basé sur cette structure et illustrée dans le tableau 3.3 à l'ajout de ces arcs. La variable *cohérence de la couleur* mesure en effet la cohérence entre la couleur d'un plan et les plans qui lui sont voisins. Ainsi, si on se trouve au niveau d'un plan de longueur faible, avec une faible cohérence de couleur présentant un gros mouvement, on est probablement au niveau d'un plan de publicité, d'où l'importance de ces connexions rajoutées par l'algorithme K2. Quant à la connexion entre le nœud *Nombre de zones de texte* et le nœud *intensité de mouvement*, elle peut s'expliquer par le fait que lorsqu'on est en

présence d'un mouvement rapide, il y a rarement présence de texte.

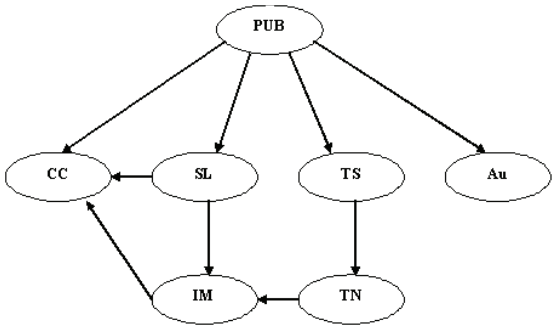

FIG. 3.3 – Structure générique construite par l'algorithme K2 en utilisant le score BIC.

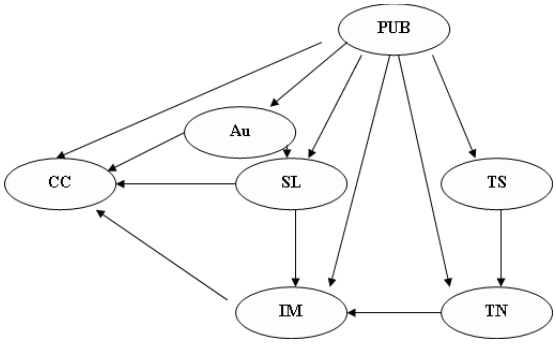

FIG. 3.4 – Structure générique construite par la recherche gloutonne en utilisant un score BIC.

La structure de la figure 3.4 est le résultat de l'apprentissage de structure en utilisant une méthode de recherche gloutonne. Cette structure donne de meilleurs résultats que les autres types d'apprentissage. Il n'y a aucune restriction sur le type de la structure apprise, ni ordre *a priori*. L'apprentissage se fait par ajout ou suppression de l'arc qui permet d'aboutir à la structure avec le meilleur score. Tous les nœuds sont connectés à l'origine au nœud de la variable publicité. Les connexions rajoutées par rapport à la structure obtenue par l'algorithme K2 sont donc la raison de l'augmentation de performances du système de détection de publicité. En effet, les corrélations entre les variables *Publicité (PUB)* et *Intensité de mouvement (IM)* sont prises en compte. D'autre part, la corrélation entre variable représentant l'audio et les variables représentant la *cohérence de couleur (CC)* et la *longueur du plan (SL)* est également prises en compte.

3.4 Conclusion

Dans ce chapitre, nous nous sommes intéressés à l'apprentissage de structure. Nous avons détaillé les différents types d'apprentissage présents dans la littérature. Nous avons également présenté un cas d'utilisation de l'apprentissage de structure pour l'indexation vidéo. Nous avons conclu à travers cet exemple à l'importance des corrélations entre les attributs. Nous avons ainsi utilisé l'apprentissage de structure comme moyen pour incorporer d'une manière automatique les corrélations entre attributs dans un système basé sur les réseaux bayésiens.

Dans notre cadre applicatif, nous avons testé diverses méthodes d'apprentissage de structure basées sur l'optimisation de scores. Nous avons montré que l'algorithme K2 et la méthode utilisant la recherche gloutonne permettent d'avoir des résultats satisfaisants. Toutefois, la complexité pour la recherche gloutonne est nettement supérieure à celle de la recherche par l'algorithme K2, ceci peut constituer un frein à son utilisation tout particulièrement pour les problèmes à grand nombre de variables.

Les scores que nous avons utilisés sont des scores génératifs permettant de rechercher la structure qui représente au mieux les données. Toutefois, la capacité d'une structure à simuler les données ne garantit pas sa capacité à discriminer la classe événement et à classer de nouvelles données. Nous abordons ce problème dans le chapitre suivant. Nous présentons un cas de divergence entre le caractère génératif d'une structure et son caractère discriminant par rapport à la classe événement. Nous présentons également des solutions pour que l'apprentissage de structure permette de récupérer des structures ayant un meilleur pouvoir discriminant.

Chapitre 4

Apprentissage de structure pour la classification

4.1 Introduction

Dans le chapitre précédent, nous avons introduit l'apprentissage de structure comme outil pour aider à la construction d'un système d'indexation vidéo, au travers d'une application de détection de plages publicitaires. Les résultats obtenus ont montré que l'apprentissage de structure permet d'automatiser cette tâche.

Dans l'application de détection de publicité proposée dans le chapitre précédent, on dispose cependant d'un nombre assez restreint d'attributs (six au total), ce qui n'est généralement pas le cas pour la grande majorité des problèmes d'indexation vidéos usuels, surtout si nous tenons compte de l'évolution temporelle des variables.

Nous montrons, dans une première expérience en tout début de ce chapitre qu'effectivement l'application directe de notre méthode générative à des problèmes d'indexation vidéo plus complexes n'est pas immédiate. Nous sommes alors amenés à revoir notre approche de façon à passer d'un problème de simulation des données à un problème de classification. Nous proposons par la suite d'étudier plusieurs approches donnant au nœud de la classe un poids particulier, en commençant par des réseaux bayésiens naïfs augmentés.

4.2 Apprentissage de structure génératif

Nous proposons, ici, d'appliquer la méthode générative d'apprentissage de structure proposée dans le chapitre précédent dans un autre contexte applicatif : celui de la détection des *Actions* dans un match de football. Une *Action* est un moment du jeu où l'une des deux équipes mène le jeu dans la zone de but de l'équipe adverse. Ce jeu risque d'aboutir à un but et se termine généralement par un tir. Ce moment s'accompagne par une forte excitation de la foule et des présentateurs. Sur le plan du montage, il y a généralement une insertion de quelques plans de ralenti quelques secondes après l'*Action*. Pour cette nouvelle application, comme c'est le cas assez généralement dans

les problèmes d'indexation vidéo, nous disposons d'un jeu d'attributs beaucoup plus étendu, constitué de variables extraites du flux audio-vidéo à l'instant courant et de leurs valeurs dans les plans voisins.

Cette application de détection d'*Actions* dans un match de football sera utilisée tout au long de ce chapitre. Nous en proposons une description détaillée dans l'annexe 7.2.

Nous effectuons, dans ce premier paragraphe, une comparaison entre l'approche générative présentée dans le chapitre précédent et une approche basée sur un réseau bayésien naïf présenté dans le paragraphe 2.6.1. L'apprentissage de structure que nous utilisons est basé sur l'algorithme K2 avec utilisation d'un score BIC. La figure 4.1 propose une comparaison entre les courbes précision-rappel de ces deux approches.

Comme nous l'avons déjà fait remarquer, le réseau bayésien naïf est généralement utilisé lorsqu'on ne dispose d'aucune information *a priori* lors de la construction manuelle de la structure du modèle. Cette première expérience montre que les résultats obtenus par l'apprentissage de structure sont de moindre qualité que ceux obtenus par une structure naïve, où tous les attributs sont considérés comme conditionnellement indépendants par rapport à la classe événement, c'est-à-dire la classe *Action*.

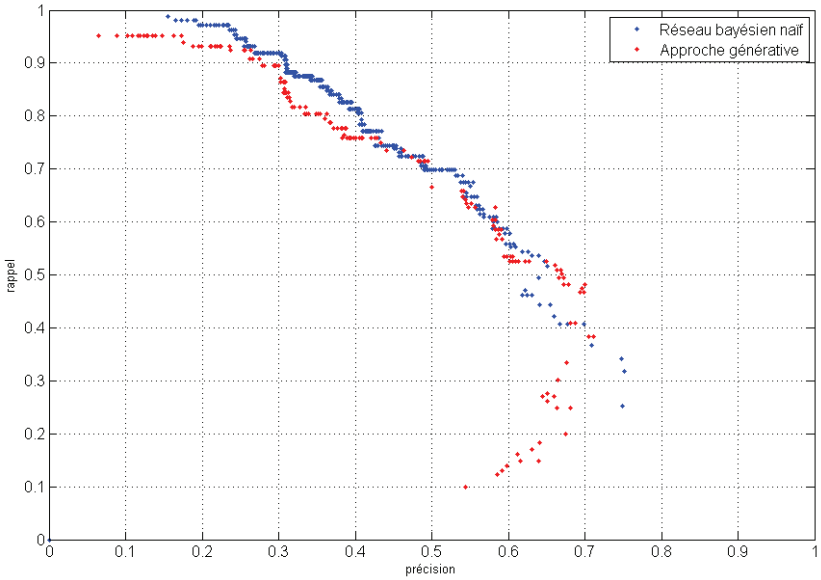

FIG. 4.1 – Comparaison entre les résultats d'un réseau bayésien naïf et l'approche générative.

On peut expliquer ces résultats en se référant à la structure récupérée par l'approche générative (*c.f.* figure 4.2). On remarque que le nœud de classification, X_c, représentant l'événement a été connecté à un faible nombre de variables. Il ne profite donc que de l'information contenue dans ces nœuds, contrairement au réseau bayésien naïf où le nœud de classification est connecté à tous les nœuds attributs.

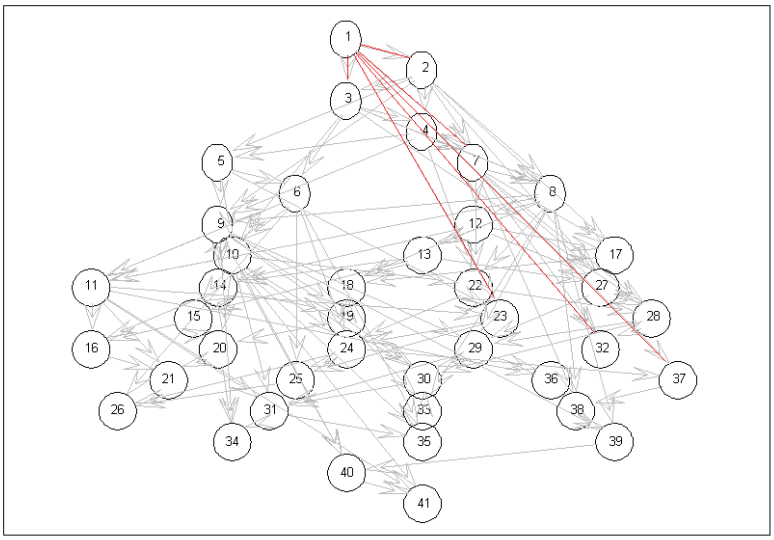

FIG. 4.2 – Structure du réseau bayésien construite par l'approche générative. En rouge les liens directs entre le nœud de la classe et les autres nœuds attributs.

L'analyse du score BIC dont l'expression est décrite dans l'équation 3.5, et qui a servi pour l'apprentissage de la structure dans l'approche générative, montre que nous cherchons à maximiser la log-vraisemblance $\log(P(X_1, ..., X_n, X_c))$. Cette quantité peut s'écrire sous la forme de l'équation 4.1 :

$$\log(P(X_1, ..., X_n, X_c)) = \log(P(X_1, ..., X_n)) + \log(P(X_c|X_1, ..., X_n)) \tag{4.1}$$

Lorsque le nombre de variables augmente, le terme $\log(P(X_1, ..., X_n))$ domine le terme $\log(P(X_c|X_1, ..., X_n))$. En effet, en augmentant le nombre de variables, la probabilité de chaque configuration de l'ensemble de variables $X_1, ..., X_n$ devient de plus en plus petite. Le terme $|\log(P(X_1, ..., X_n))|$ devient alors de plus en plus grand, alors que le terme $\log(P(X_c|X_1, ..., X_n))$ reste à peu près constant. La maximisation de score BIC est donc dominée par le terme $\log(P(X_1, ..., X_n))$ et non pas par le terme $\log(P(X_c|X_1, ..., X_n))$, directement impliqué dans le processus de classification. Ceci ne favorise pas l'utilisation de la structure apprise pour la tâche de classification.

Ainsi, l'approche utilisée jusqu'à présent, c'est-à-dire l'approche générative, cherche à produire un modèle simulant au mieux les données, ce but étant sensiblement différent de notre contexte applicatif général qui est plus proche d'un problème de classification.

Dans ce contexte, un nœud, le nœud de classification, a un rôle particulier qui n'est pas mis en avant dans l'approche générative. Une approche tenant compte de cette différentiation entre les nœuds est indispensable. Pour cela, nous détaillons trois types différents de réseaux. Nous étudions, dans un premier temps, les réseaux bayésiens

naïfs augmentés. Dans un second temps, nous détaillons l'apport d'une « approche discriminante ». Nous étudions, enfin, un dernier type de réseaux qui sont les réseaux bayésiens Multinets et leur apport pour la détection d'événements.

4.3 Classification par un réseau bayésien naïf augmenté

Dans le paragraphe précédent, nous avons montré que les résultats du réseau bayésien naïf, où le nœud de classification tient compte de l'information provenant de tous les attributs, sont équivalents ou surpassent les résultats d'un réseau bayésien dont la structure a été apprise par un score tel que le BIC et où l'on n'a fait aucune restriction sur les connexions entre les nœuds. Toutefois, les réseaux bayésiens naïfs utilisent l'hypothèse simplificatrice que les attributs sont indépendants conditionnellement à la variable classe. Dans la majorité des cas, cette hypothèse n'est pas vérifiée. Les réseaux bayésiens naïfs ne tirent alors pas profit des corrélations qui peuvent exister entre les différentes primitives. D'un point de vue « graphe », cela se traduit par l'absence d'arcs entre les nœuds qui représentent les attributs. Tenir compte de ces corrélations correspond donc à rajouter des arcs entre les attributs. Cet ajout peut être fait de deux façons différentes :

- soit au travers de l'exploitation de connaissance *a priori* sur l'événement à détecter et donc sur nos données. Cette connaissance est toutefois difficile à obtenir ;
- soit par apprentissage de structure permettant d'enrichir la structure du réseau bayésien naïf. Un tel apprentissage peut également permettre d'orienter notre système vers la résolution d'un problème de classification plutôt que vers un problème de description des données.

Par la suite, nous allons présenter deux types de réseaux bayésiens naïfs augmentés. Dans ce type de réseaux la structure est augmentée d'une série d'arcs supplémentaires récupérés lors de la phase d'apprentissage.

4.3.1 Classification en utilisant un réseau bayésien naïf augmenté de type TAN

Dans [74], Friedman *et al.* ont proposé un algorithme basé sur l'apprentissage d'un TAN (Tree Augmented Naive Bayes). L'algorithme apprend, dans ce cas, une structure sous la forme d'un réseau bayésien naïf enrichi par une structure en arbre. L'espace de recherche est cependant restreint à l'ensemble des structures pour lesquelles chaque nœud attribut a un seul parent parmi les autres attributs, en plus du nœud de classification. On retrouve ainsi sous-jacente la structure du réseau bayésien naïf. Elle est cependant augmentée par une structure d'arbre entre les différents attributs. Cette méthode donne un poids particulier au nœud de classification puisque tous les nœuds lui sont connectés. Toutefois, le TAN appartient toujours à la classe des méthodes génératives. En effet, le score utilisé est basé sur l'information mutuelle conditionnelle (*c.f.* équation 4.2) dont toute augmentation implique l'augmentation de la vraisemblance [75]. Pour la mise en œuvre de l'apprentissage de type TAN, les auteurs de [74] utilisent l'algorithme de recherche d'arbre de recouvrement maximal pour construire la structure augmentée. Le

score utilisé dans cet algorithme est l'information mutuelle conditionnelle qui est une extension de l'information mutuelle déjà présentée dans le paragraphe 3.2.2.4. Cette dernière permet de calculer l'information ajoutée par le nœud X_j au nœud X_i, tout en connaissant l'information sur le nœud classe X_c. Elle s'écrit au niveau de chaque nœud X_i sous la forme :

$$IM_c(i,j) = IM(X_i, X_j|X_c) = \sum_{x_i,x_j,x_c} log(\frac{P(X_i = x_i, X_j = x_j|X_c = x_c)}{P(X_j = x_j|X_c = x_c)P(X_i = x_i|X_c = x_c)}) \tag{4.2}$$

Le score final de la structure est la somme des informations mutuelles de tous les nœuds attributs. Le score est ainsi par construction décomposable. L'utilisation de l'information mutuelle conditionnelle comme score sous-entend que l'hypothèse *"le nœud de classification est un parent commun à tous les attributs* $X_1, ..X_n$*"* est vérifiée.

En pratique, une matrice IM_c (information mutuelle conditionnelle) est construite. Chaque terme $IM_c(i,j)$ de la matrice correspond à $IM(X_i, X_j|X_c)$. La structure recherchée dans cette approche est un arbre qui passe par tous les nœuds attributs et qui maximise le score global.

4.3.2 Classification en utilisant un réseau bayésien naïf augmenté par une structure résultant d'un K2

Dans la structure apprise par l'algorithme TAN présentée ci-dessus, on se restreint à une structure d'arbre pour augmenter le réseau bayésien naïf. Certes, cette structure présente des avantages. Elle garde en effet une complexité assez faible. Toutefois, restreindre le nombre de parents autres que le nœud de classification à exactement un parent pour chaque nœud est une contrainte très forte. La structure ainsi obtenue ne permet pas de représenter le cas où une variable est corrélée avec plusieurs autres variables. Elle ne permet pas non plus de représenter le cas où une variable est conditionnellement indépendante par rapport au nœud de classification de toutes les autres variables. Dans ce cas, le nœud représentant cette variable n'a besoin que du nœud de classification comme parent. L'ajout d'un autre parent ne fait qu'augmenter inutilement la complexité et le nombre de paramètres du réseau.

Pour ces raisons, nous proposons d'utiliser l'algorithme K2 (*cf.* paragraphe 3.2.2.5) pour apprendre la structure qui viendra augmenter la structure du réseau bayésien naïf. Ce choix nous permet de ne plus nous restreindre à une structure d'arbre mais d'avoir une structure plus générique. À l'image de l'information mutuelle qui a été transformée en information mutuelle conditionnelle, nous modifions le score BIC (*cf.* équation 3.5) pour tenir compte du fait que le nœud de classification est un parent commun à tous les nœuds du problème. Le score BIC conditionnel au niveau de chaque nœud X_i s'écrit alors sous la forme proposée dans l'équation 4.3.

$$score^c_{BIC}(X_i, pa(X_i)) = \log(P(X_i|pa(X_i), X_c, \hat{\theta}_i)) - \frac{1}{2}.\dim(X_i, \mathcal{B}).\log N \tag{4.3}$$

Le score de la structure globale reste décomposable. Il est toujours équivalent à la somme des scores au niveau de chaque nœud.

À l'image du score BIC, le score BIC conditionnel est toujours composé de deux termes : un premier terme permettant la maximisation de la vraisemblance que nous avons modifié pour tenir compte du fait que le nœud de classification est un parent commun à tous les nœuds du réseau ; et un second terme permettant de tenir compte de la complexité du réseau construit. Nous verrons, dans la partie 4.3.3.1, comment nous utilisons la pondération entre ces deux termes pour améliorer les performances de classification du système.

4.3.3 Résultats et interprétation

Après avoir décrit les différentes approches d'apprentissage de structure, nous nous intéressons à présent aux performances de ces approches pour la détection d'événements. Nous cherchons, en effet, à mettre en lumière les atouts et les inconvénients de ces méthodes et à trouver ainsi la méthode la plus adaptée à la classification. Nous nous plaçons toujours dans le cadre de notre application de détection d'*Actions* décrite en annexe (*c.f.* annexe 7.2).

Nous comparons à la figure 4.3 les résultats du réseau bayésien naïf et ceux du réseau bayésien augmenté par un arbre de type TAN. Nous remarquons une augmentation des performances du système lorsqu'on utilise le réseau augmenté. En effet, pour une précision de 0.5 nous passons d'un rappel de 0.7 à un rappel de 0.73, et pour une précision de 0.6 nous passons d'un rappel de 0.55 à un rappel de 0.65. L'enrichissement de la structure naïve permet d'augmenter les performances de classification du système. Il permet, comme c'est le cas dans l'approche générative non contrainte présentée dans le chapitre 3, de tenir compte des connexions qui existent entre les différentes variables. Une nouvelle fois, on montre que l'hypothèse d'indépendance entre les attributs ne garantit donc pas des résultats de classification optimaux, bien que cette hypothèse soit utilisée par la majorité des systèmes d'indexation vidéo actuels basés sur les réseaux bayésiens.

Toutefois, contrairement à l'approche générative non restreinte utilisée dans le chapitre 3 et dont les résultats pour notre application ont été présentés dans la figure 4.1, les performances du réseau de type TAN dépassent les performances du réseau naïf. La différence entre l'apprentissage de structure dans l'approche générative et l'approche de type TAN réside, en effet, dans l'importance qu'on accorde au nœud de classification. Dans l'approche générative, tous les nœuds du réseau sont considérés de la même manière. Dans l'approche TAN, le nœud de classification est supposé être une cause commune à tous les nœuds attributs. On se restreint également dans la recherche de structure à l'espace des structures où le nœud de classification est un parent commun à tous les nœuds. Cette supposition permet de faire profiter le nœud de classification de l'information provenant de tous les attributs même si certaines connexions ne sont pas suffisamment fortes pour justifier leur présence dans la structure apprise par l'approche générative non contrainte.

Nous nous intéressons à présent aux résultats de la structure augmentée par une structure générique recherchée par l'algorithme K2. La figure 4.3 montre que les résultats de la structure augmentée par une structure générique ne sont pas toujours supérieurs à

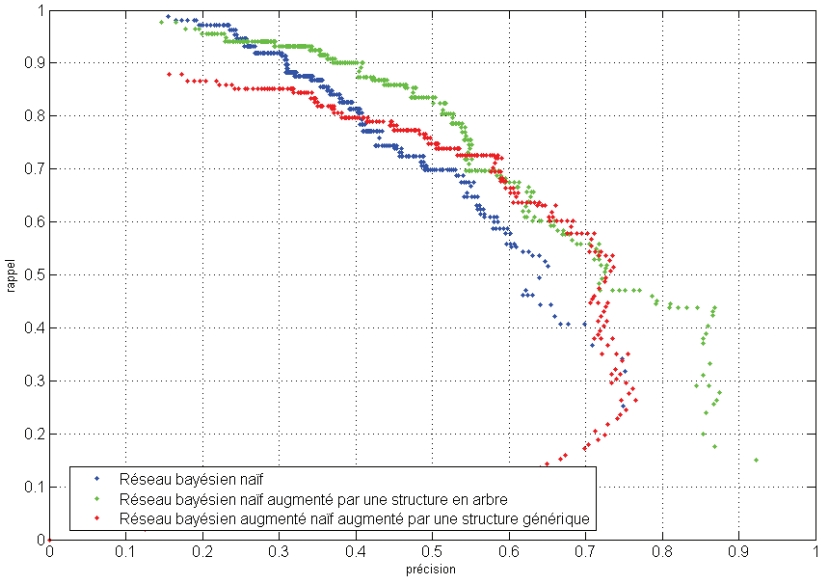

FIG. 4.3 – Comparaison entre un réseau bayésien naïf, un réseau bayésien naïf augmenté par un arbre et un réseau bayésien naïf augmenté par une structure générique.

ceux de la structure naïve. Ils sont aussi inférieurs aux résultats de la structure apprise par l'approche TAN. Ces résultats peuvent être expliqués en faisant une analyse du nombre de paramètres utilisés pour chaque structure. Le tableau 4.1 montre que la structure apprise à travers l'algorithme K2 utilise un nombre nettement plus important de paramètres. On peut soupçonner alors un phénomène de sur-apprentissage dû au grand nombre de paramètres utilisés. L'algorithme est, en effet, basé sur l'enrichissement de la structure naïve par une structure générique construite par l'algorithme K2 avec un score BIC conditionnel (*cf.* équation 4.3). Comme nous l'avons déjà expliqué dans la partie 4.3.2, le score BIC conditionnel est composé de deux termes : un terme d'attache aux données et un second terme permettant de tenir compte de la complexité de la structure. Ceci permet de trouver un compromis entre la simplicité de la structure et la justesse de description des données. Dans notre cadre applicatif, le terme d'attache aux données domine visiblement le score. Nous introduisons dans le paragraphe suivant le score BIC modifié permettant de réguler l'influence du terme d'attache aux données et du terme de complexité.

4.3.3.1 Influence de la complexité de la structure construite par un algorithme K2

Afin de trouver un équilibre entre les deux termes du score BIC, nous utilisons un coefficient de pondération λ au niveau du terme de complexité de la structure. Le score

Type de réseau bayésien	Nombre de paramètres
Naïf	81
Augmenté par un arbre (TAN)	159
Augmenté par une structure générique (K2)	403

TAB. 4.1 – Analyse du nombre de paramètres nécessaires pour chaque type de réseau bayésien.

s'écrit alors sous la forme décrite dans l'équation 4.4.

$$score_{BIC}^{mc}(X_i, pa(X_i)) = \log(P(X_i|pa(X_i), X_c, \hat{\theta}_i), \hat{\theta}_i)) - \lambda. \dim(X_i, \mathcal{B}). \log N \quad (4.4)$$

Nous étudions alors l'influence du paramètre λ, permettant de réguler la pondération entre le terme de vraisemblance et le terme de complexité, sur les résultats du réseau bayésien naïf augmenté par une structure apprise par un K2. Pour évaluer les performances du système, nous utilisons cette fois-ci la F-mesure, notée $F1$. Cette mesure permet d'équilibrer précision et rappel de manière équivalente. Elle s'écrit sous la forme :

$$F1 = \frac{2 * (\text{Précision} * \text{Rappel})}{(\text{Précision} + \text{Rappel})} \quad (4.5)$$

Dans la figure 4.4, nous présentons l'évolution de la F-mesure $F1$ en fonction de l'évolution du paramètre λ. Cette évolution présente deux phases. La première phase correspond à des valeurs de λ faibles : $\lambda < 3$. Durant cette phase, c'est le terme de vraisemblance qui domine le score. La structure construite par l'apprentissage de structure décrit alors avec trop de détails les données d'apprentissage, ce qui conduit à un phénomène de sur-apprentissage. Au fur et à mesure qu'on augmente le paramètre λ, on diminue la complexité de la structure, ce qui se traduit par la diminution du nombre de paramètres du réseau (*cf.* figure. 4.5). L'augmentation de la valeur de λ entraîne donc l'atténuation du phénomène de sur-apprentissage, et l'augmentation des performances de classification. La deuxième phase correspond à des valeurs de λ supérieures à 3.2. Durant cette phase, c'est le terme de complexité du score qui prend le dessus sur le terme de vraisemblance. De moins en moins de connexions sont ajoutées entre les nœuds du réseau. La structure obtenue est alors trop simple pour résoudre notre problème de classification. Il est à noter que pour des valeurs trop élevées de λ, l'apprentissage n'ajoute pas de connexions entre les attributs. Le réseau obtenu correspond alors au réseau bayésien naïf.

4.3.3.2 Comparaison entre les deux approches augmentées

Après avoir exposé deux méthodes d'apprentissage de structure basées sur l'enrichissement du réseau bayésien naïf, une comparaison entre l'efficacité de ces deux méthodes s'impose. La figure 4.6 compare les performances de la structure TAN et de la structure apprise par l'algorithme K2 modifié, avec pour valeur de λ la valeur expérimentale de 3 trouvée précédemment. Nous remarquons que la structure augmentée par une structure

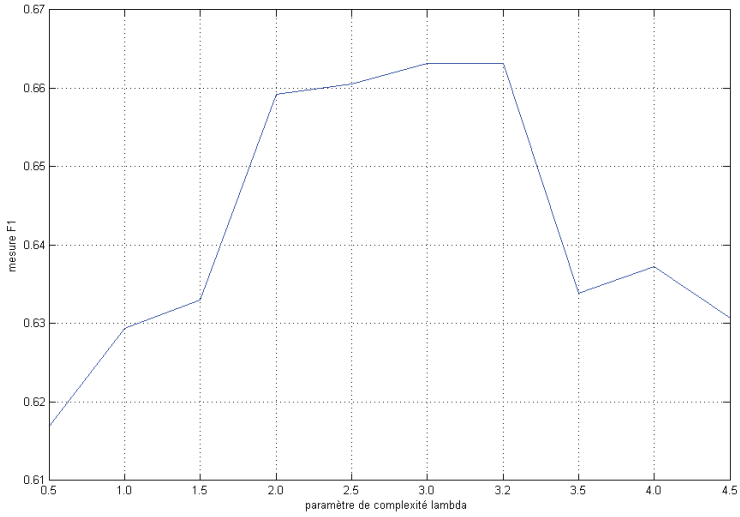

FIG. 4.4 – Influence du paramètre de complexité λ sur les performances de classification d'un réseau bayésien naïf augmenté par une structure apprise par un algorithme K2 (cas de la mesure F1).

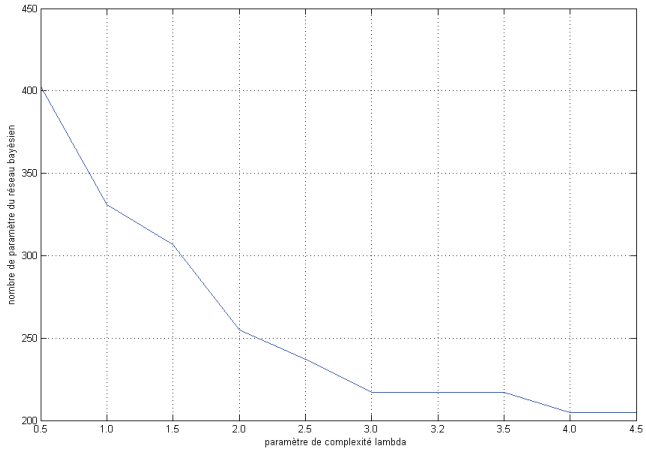

FIG. 4.5 – Evolution du nombre de paramètres du réseau bayésien en fonction du paramètre de complexité λ.

générique donne de meilleurs résultats. Cette augmentation des performances peut être expliquée par le fait que la deuxième méthode ne se restreint pas à un type particulier de structure, tout en restant adaptée à la classification, puisque le nœud de classification est connecté à tous les nœuds attributs.

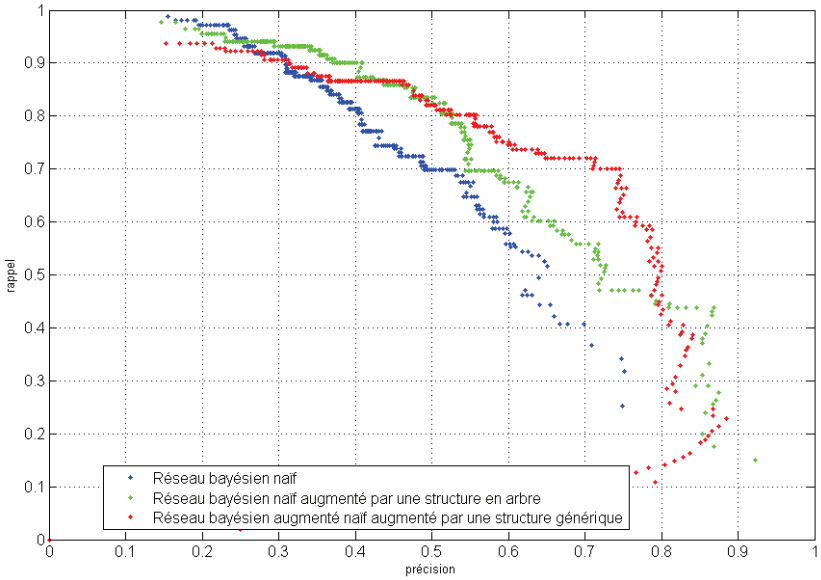

FIG. 4.6 – Comparaison des performances de classification d'un réseau bayésien naïf augmenté par un arbre et d'un réseau bayésien naïf augmenté par une structure générique(K2) en utilisant $\lambda = 3$.

4.4 Classification par une approche discriminante

Jusqu'à présent, toutes les approches que nous avons utilisées cherchent à découvrir la structure qui permet de simuler au mieux les données. Nous avons vu qu'une recherche non restreinte, comme celle utilisée dans le chapitre 3, peut conduire à des structures qui ne sont pas très adaptées au problème de classification. Les approches génératives à recherche restreinte telles que les réseaux bayésiens naïfs augmentés supposent que la classe est la cause de l'occurrence de tous les attributs. Ce type d'approches cherche une structure dans un ensemble qui, vraisemblablement, peut contenir la structure optimale pour la classification. Toutefois, rien ne garantit que la structure découverte soit cette structure optimale.

Plutôt que l'approche générative où l'on cherche à ressembler le plus aux données, nous proposons d'utiliser, dans ce paragraphe, une approche discriminante. Dans ce type d'approche, notre souci principal est d'obtenir une structure qui donne le meilleur

taux de classification. La différence entre une approche discriminante et une approche générative est la différence entre pouvoir reconnaître quelque chose et être capable de le reproduire. Une approche générative converge vers le classifieur qui modélise le mieux la distribution jointe $P(X_c, X_1, ..., X_n)$. Une approche discriminante, quant à elle, permet de mieux approcher la probabilité conditionnelle $P(X_c|X_1, ..., X_n)$. Elle apprend une frontière de façon à minimiser un taux d'erreur de classification.

C'est un choix approprié de la fonction de score qui va nous permettre de donner un caractère discriminant à la structure recherchée.

4.4.1 Fonction de score

Jusqu'à présent, nous avons choisi un score génératif maximisant la vraisemblance des données par rapport au modèle. Nous avons toutefois choisi de restreindre notre recherche de structures dans le sous-ensemble des structures naïves augmentées. Ce choix a été motivé par le fait que les résultats du réseau bayésien naïf étaient meilleurs que les résultats donnés par une structure apprise avec une approche générative non restreinte. Toutefois, notre but principal reste la classification de nos données. Nous cherchons donc une méthode qui donne les meilleurs taux de classification, même si le pouvoir de simulation de la structure apprise est faible.

La classification a pour but de prédire correctement la valeur d'une variable classe X_c à partir d'un vecteur d'attributs $X_1, ..., X_n$. La classe optimale est donc celle qui maximise $P(X_c/X_1, ..., X_n)$. Afin de réaliser l'apprentissage de structure, nous proposons d'utiliser la vraisemblance conditionnelle CLL définie dans l'équation 4.6 comme score pour notre méthode discriminante.

$$score_{dis}(\mathcal{B}|\mathcal{D}) = CLL(\mathcal{B}|\mathcal{D}) = \sum_{taille\ \mathcal{D}} \log(P_{\mathcal{B}}(X_c|X_1, ..., X_n)) \tag{4.6}$$

Il est à noter que la vraisemblance conditionnelle est en fait un terme de la vraisemblance (*c.f.* équation 4.7).

$$LL(\mathcal{B}|\mathcal{D}) = CLL(\mathcal{B}|\mathcal{D}) + \sum_{taille\ \mathcal{D}} \log(P_{\mathcal{B}}(X_1, ..., X_n)) \tag{4.7}$$

où $\mathcal{B}$ est le réseau bayésien contenant la structure à évaluer et $\mathcal{D}$ l'ensemble des données de la base d'apprentissage.

Dans les approches génératives, on cherche justement à maximiser la vraisemblance, ce qui conduit à des structures non adaptées au but de classification. Pour avoir une structure adaptée à cette tâche de classification, nous utilisons uniquement le terme de l'équation 4.6.

Ce nouveau score n'est cependant pas un score décomposable comme c'est le cas des scores utilisés dans les approches précédentes. En effet, la vraisemblance conditionnelle de l'équation 4.6 s'écrit sous la forme suivante :

Structure d'initialisation	Durée moyenne d'apprentissage
Structure naïve	2j : 4h : 50min
Structure du type TAN	1j : 6h : 23min
Structure augmentée par un K2	1j : 4h : 37min

TAB. 4.2 – Influence de la structure d'initialisation sur le temps de recherche de la structure optimale par un critère discriminant.

$$\begin{aligned} CLL(\mathcal{B}|\mathcal{D}) &= \sum_{taille\ \mathcal{D}} \log(P_{\mathcal{B}}(X_c, X_1, ..., X_n)) - \sum_{taille\ \mathcal{D}} \log(P_{\mathcal{B}}(X1, ..., X_n)) \qquad (4.8) \\ &= \sum_{taille\ \mathcal{D}} \log(P_{\mathcal{B}}(X_c, X_1, ..., X_n)) - \sum_{taille\ \mathcal{D}} \log(\sum_{x_c} P_{\mathcal{B}}(X_c = x_c, X_1, ..., X_n)) \end{aligned}$$

Le second terme de l'équation 4.8 ne peut être décomposable en un ensemble de termes n'impliquant qu'un seul nœud et l'ensemble de ses parents. Toute méthode de parcours telle que la méthode de parcours de l'algorithme K2 ou celle de l'algorithme MWST est alors impossible dans ce cas. Puisque l'ajout d'un arc à un nœud entraîne la modification du score global de la structure et non une modification locale au niveau d'un seul nœud, l'utilisation d'un tel score conduit en outre à une grande complexité de calcul.

4.4.2 Mise en œuvre de la méthode

Le score discriminant que nous avons présenté dans le paragraphe précédent est un score non décomposable. Nous ne pouvons donc pas utiliser les méthodes de parcours de structure telles que celles utilisées dans l'algorithme K2 ou le MSWT. Nous utilisons donc la méthode de parcours par la recherche gloutonne présentée dans le paragraphe 3.2.2.5. Ainsi, à partir d'une structure initiale, nous calculons le score pour l'ensemble des structures voisines qui diffèrent de la structure initiale par l'ajout ou la suppression d'un seul arc. Nous utilisons différentes structures pour l'initialisation de l'algorithme de la recherche gloutonne. Nous présentons dans le tableau 4.2, la durée moyenne de l'apprentissage de structure sur les différents jeux de données que nous utilisons. Les structures récupérées à la fin de l'apprentissage sont les mêmes, l'influence de la structure d'initialisation apparaît uniquement dans le temps que prend la recherche pour trouver la structure optimale. Il est donc plus rapide de partir d'une structure qui se rapproche de la structure optimale et qui donne de bonnes performances de classification comme c'est le cas des structures issues des approches augmentées, que de partir de d'une structure en réseau bayésien naïf.

4.4.3 Résultats et interprétation

La figure 4.7 montre la prédominance des résultats de l'approche discriminante par rapport aux approches telles que les réseaux bayésiens naïfs et les réseaux bayésiens

naïfs augmentés par une structure d'arbre. La maximisation d'un critère discriminant s'avère plus efficace pour la tâche de classification.

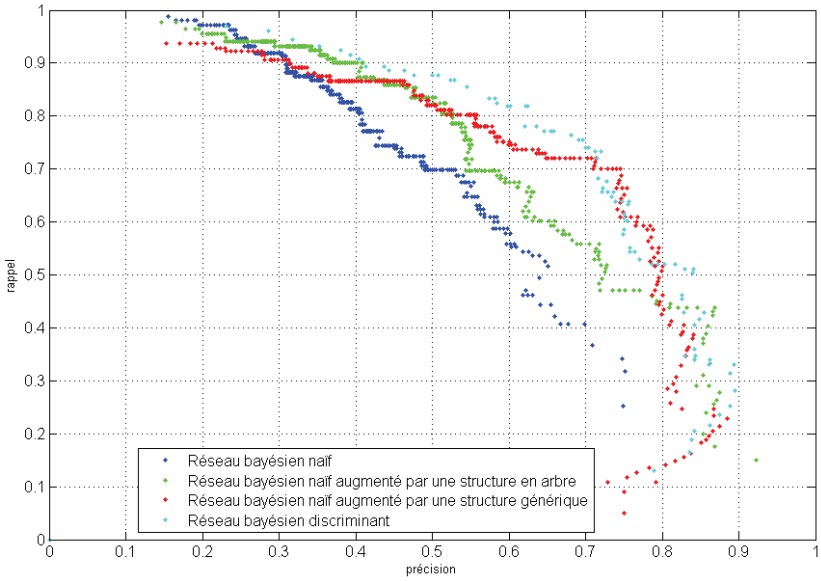

FIG. 4.7 – Comparaison des résultats de classification d'un réseau bayésien discriminant par rapport à ceux d'un réseau bayésien naïf.

Cette amélioration est d'autant plus notable, que nous n'avons utilisé aucune restriction sur la structure recherchée contrairement à l'approche générative, pour laquelle nous avons dû restreindre l'espace de recherche de structures à l'espace des réseaux bayésiens augmentés, pour obtenir des résultats satisfaisants.

Dans la figure 4.8, on représente la structure obtenue par l'apprentissage de structure par un critère discriminant. Le nœud de classification est connecté dans cette structure à un nombre restreint de nœuds attributs. Le reste des nœuds a en effet été rejeté du réseau. Cette constatation permet d'expliquer aussi les bonnes performances de l'apprentissage discriminant dans les réseaux bayésiens. En effet, en rejetant les attributs qui ne sont pas pertinents pour la classification, l'algorithme réduit la taille de la structure qu'il utilise, ce qui rend l'apprentissage de paramètres plus fiable. Il élimine aussi toutes les sources de bruit qui peuvent influer sur l'étape de classification. Nous aborderons plus en détail l'élimination ou la sélection d'attributs dans le chapitre suivant.

Nous notons également que, bien que les résultats des réseaux bayésiens naïfs augmentés par une structure générique sont globalement inférieurs à ceux de l'approche discriminante, ils restent assez proches. Ces résultats satisfaisants peuvent être expliqués par le fait que le paramètre λ a été choisi de façon à maximiser la mesure F1, ce qui revient donc à maximiser le pouvoir de classification. Cela introduit, en effet,

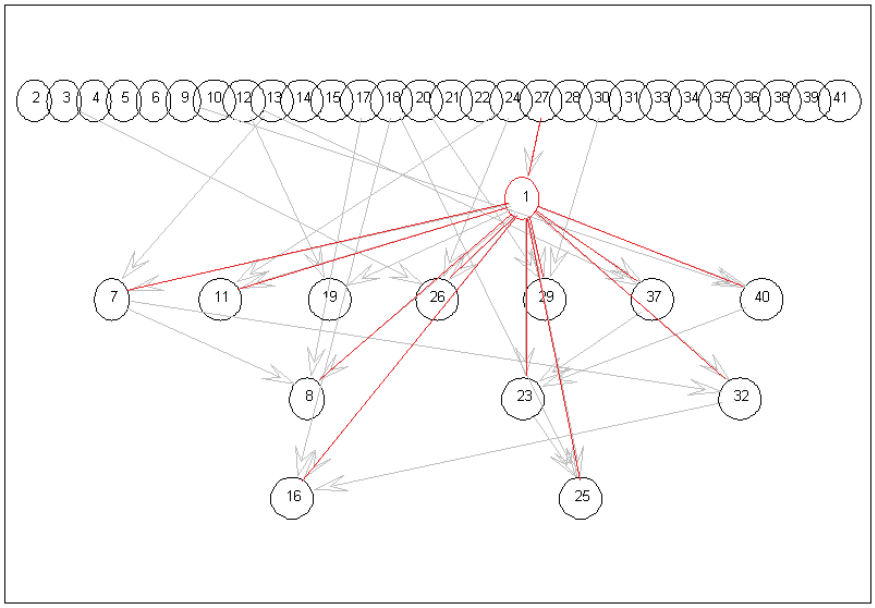

FIG. 4.8 – Structure issue de l'apprentissage de structure par un critère discriminant.

une dimension discriminante à la méthode d'apprentissage par augmentation du réseau bayésien naïf.

4.5 Classification par l'approche Multinets

Jusqu'à présent, nous avons utilisé un modèle unique et donc une structure unique de réseau bayésien pour les différentes instances de la classe X_c. En effet, seuls les paramètres du modèle changent d'une classe à une autre, la structure du réseau reste la même. Toutefois, les relations d'indépendance conditionnelle entre les attributs peuvent changer selon le contexte. Ainsi, deux variables attributs peuvent être indépendantes pour une instance donnée de la variable X_c alors qu'elles sont dépendantes pour une autre instance de la variable classe X_c. Dans le cas d'un modèle unique, les variables sont alors considérées dans tous les cas comme dépendantes et il y a ajout d'un arc entre les deux variables, ce qui n'est pas tout à fait vrai. Il peut donc être préférable, au lieu d'apprendre un modèle unique pour toutes les instances de la variable classe, d'apprendre un modèle pour chaque instance de X_c. L'ensemble de ces modèles ou réseaux bayésiens s'appelle un réseau bayésien Multinet. Ce type de réseau a été présenté dans [76] pour permettre la représentation des indépendances conditionnelles asymétriques pouvant exister entre les variables.

4.5.1 Processus d'inférence

Le processus d'inférence dans les réseaux bayésiens Multinets est différent du processus d'inférence dans un réseau où il y a un modèle unique. La variable classe X_c prend des valeurs dans $c_1, ..., c_m$ où m est la cardinalité de la variable X_c. L'inférence consiste alors à calculer la quantité $P(X_c = c_j|X_1, ..., X_n)$. Dans un calcul d'inférence où il y a un seul modèle, cette quantité s'écrit sous la forme :

$$P(X_c = c_j|X_1, ..., X_n) = \frac{P(X_c = c_j, X_1, ..., X_n)}{P(X_1, ...X_n)} \tag{4.9}$$

Le numérateur de l'équation 4.9 est la vraisemblance du modèle unique. Il se décompose sous la forme du produit de tous les paramètres du modèle. Le calcul du dénominateur passe aussi par le calcul de la vraisemblance du modèle, mais en faisant une marginalisation par rapport à la variable classe X_c. Dans le cas d'un modèle Multinet, on dispose de modèles différents, un pour chaque instance de la classe X_c. La probabilité *a posteriori* pour chaque instance de la classe s'écrit alors sous la forme suivante :

$$P(X_c = c_j|X_1, ..., X_n) = \frac{P(X_c = c_j)P_j(X_1, ..., X_n)}{\sum_i P(X_c = c_i)P_i(X_1, ..., X_n)} \tag{4.10}$$

où $P_j(X_1, ..., X_n)$ correspond à la vraisemblance des données par rapport au modèle local de la classe c_j. Cette quantité correspond donc à une probabilité conditionnelle $P(X_1, ..., X_n|X_c = c_j)$.

Les réseaux bayésiens Multinets peuvent être considérés comme une généralisation des réseaux bayésiens naïfs augmentés. En effet, si les structures représentant les différentes classes sont identiques, l'équation 4.10 s'écrit sous la forme :

$$\begin{aligned} P(X_c = c_j|X_1, ..., X_n) &= \frac{P(X_c = c_j)P(X_1, ..., X_n|X_c = c_j)}{\sum_i P(X_c = c_i)P(X_1, ..., X_n|X_c = c_i)} \\ &= \frac{P(X_c = c_j)\prod_{i=1}^{n} P(X_i|pa(X_i), X_c = c_j)}{\sum_i P(X_c = c_i)\prod_{k=1}^{n} P(X_k|pa(X_k), X_c = c_i)} \end{aligned} \tag{4.11}$$

D'après l'équation 4.11, le nœud représentant la variable X_c est donc un parent commun à tous les nœuds. On se retrouve donc dans le cadre des réseaux bayésiens naïfs augmentés. Les réseaux Multinets peuvent être vus comme un réseau bayésien naïf augmenté par des structures qui diffèrent selon la valeur prise par la classe considérée.

4.5.2 Processus d'apprentissage

Dans les réseaux bayésiens Multinets, un modèle est affecté à chaque instance c_i de la classe X_c. Le nœud de classification n'est donc plus visible au niveau du réseau. Afin de mettre en œuvre l'apprentissage dans un réseau bayésien Multinets, la base d'apprentissage $\mathcal{D}$ est partagée en m bases disjointes $\mathcal{DL}_1, ...\mathcal{DL}_m$ selon l'instance prise par la classe. Chaque base de données $\mathcal{DL}_i$ est utilisée pour apprendre un réseau $\mathcal{B}_i$ ainsi que les paramètres du modèle correspondant à l'instance c_i de la classe X_c. Plusieurs

algorithmes d'apprentissage de structure peuvent être utilisés pour construire les différents modèles. Nous allons étudier dans le paragraphe suivant l'influence du choix de ces algorithmes d'apprentissage sur la qualité de la classification.

4.5.3 Résultats et interprétation

4.5.3.1 Apport de l'apprentissage de structures pour un réseau Multinets

Nous considérons deux types de sous-structures pour les réseaux Multinets : des sous-structures en arbre et des sous-structures génériques. Nous comparons, sur la figure 4.9, les résultats d'un réseau bayésien naïf avec les résultats d'un réseau bayésien Multinets. Pour ce dernier, nous avons appris la structure de la classe *Action* et celle de la classe de rejet en utilisant l'algorithme MWST. Tout comme dans les approches précédentes, l'apprentissage de structure utilisé dans les réseaux Multinets améliore les résultats de classification. En comparaison d'un simple réseau bayésien naïf, ce type d'approche donne encore une fois un rôle particulier au nœud de classification, puisque les structures sont apprises distinctement par rapport à la valeur prise par le nœud de classification.

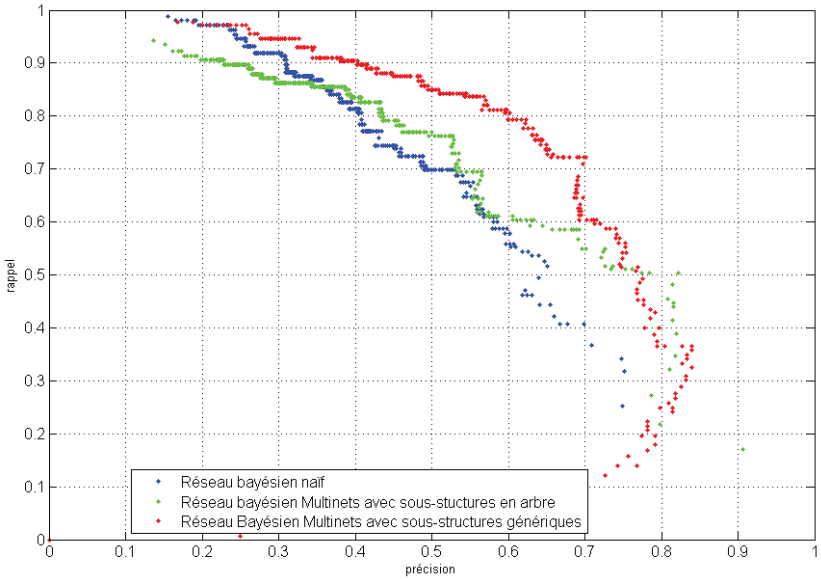

FIG. 4.9 – Influence de l'algorithme d'apprentissage des sous-structures du réseau Multinets.

D'autre part, nous notons la différence de performances entre les réseaux bayésiens Multinets utilisant des sous-structures en arbre et ceux utilisant des sous-structures génériques. Les sous-structures génériques donnent, en effet, de meilleures performances que les réseaux utilisant les sous structures en arbre, à l'image du réseau bayésien naïf

mono-structure augmenté par une structure générique.

4.5.3.2 Comparaison de réseaux Multinets avec les différentes approches déjà décrites

Dans la figure 4.10, nous comparons les résultats de l'approche Multinet avec les différentes approches déjà présentées tout au long de ce chapitre. Nous utilisons dans cette comparaisons les résultats des Multinets à structure générique. Une première lecture de de la figure confirme que l'apprentissage de structure a une grande valeur ajoutée pour les Multinets : leurs performance surpassent en effet celles des réseaux bayésiens naïfs, ainsi que celles des résultats des réseaux bayésiens augmentés par une structure en arbre. La comparaison entre les réseaux bayésiens Multinets et les réseaux bayésiens naïfs augmentés par une structure générique montre par contre que ces deux méthodes donnent des résultats similaires. Utiliser une structure différente pour chaque classe ne constitue donc pas un apport significatif en terme de classification par rapport à une méthode utilisant une seule structure.

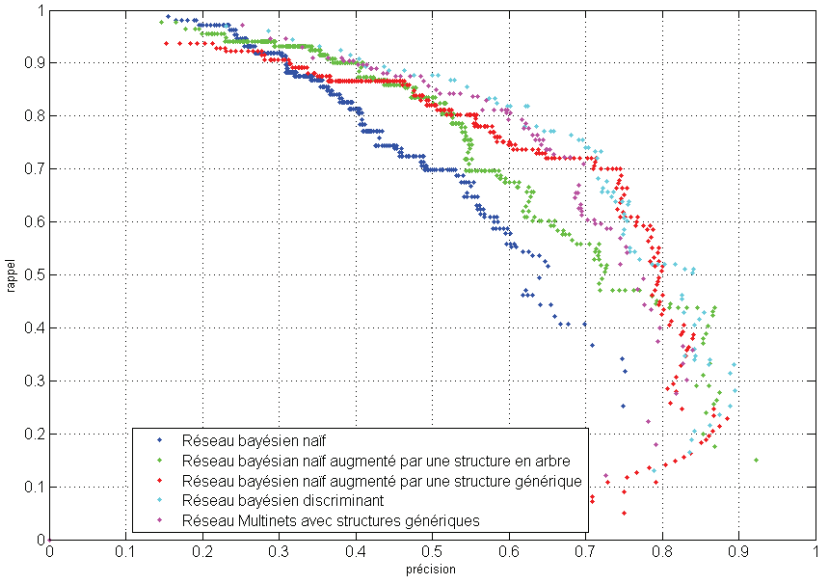

FIG. 4.10 – Comparaison des résultats de classification d'un réseau bayésien Multinets par rapport aux approches décrites précédemment.

4.6 Influence de la taille de la base d'apprentissage

Nous étudions dans cette partie l'influence de la taille de la base d'apprentissage. Nous reprenons donc les approches que nous avons étudiées dans ce chapitre, à savoir :

- réseau bayésien naïf,
- réseau bayésien augmenté par une structure en arbre,
- réseau bayésien augmenté par une structure générique,
- réseau bayésien Multinets,
- réseau bayésien appris par un critère discriminant.

Nous réalisons les différents tests sur la moitié de la base de données, ce qui correspond à 109 exemples d'*Action* contre 6300 exemples de la classe de rejet. Nous comparons ainsi les résultats obtenus sur cette demi-base avec les résultats de la base complète que nous avons utilisée tout au long de ce chapitre.

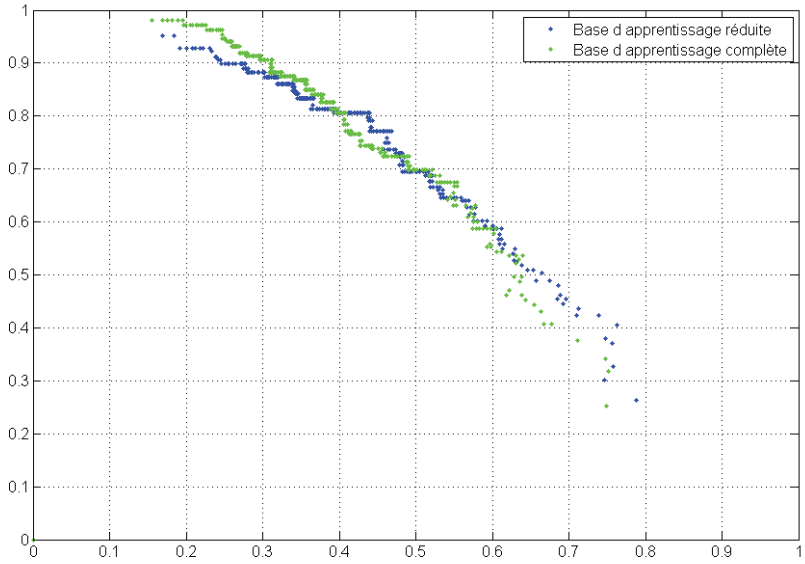

FIG. 4.11 – Influence de la taille de la base sur le réseau bayésien naïf.

Les figures de comparaison montrent que la taille de la base d'apprentissage n'influe pas de la même manière sur les performances des différents systèmes. Ainsi, il s'avère que le réseau bayésien naïf (*cf.* figure 4.11) est le moins sensible à la taille de la base d'apprentissage. Ce résultat était prévisible : ce type de réseau requiert moins de paramètres lors de la phase d'apprentissage ; en outre, il n'y a aucun apprentissage de structure dans cette approche. Les réseaux bayésiens augmentés par un arbre (*cf.* figure 4.12) et par une structure générique (*cf.* figure 4.13) sont quant à eux légèrement affectés par la réduction de la taille de la base d'apprentissage. En effet, le nombre de paramètres dans ces deux réseaux est plus important que celui du réseau bayésien naïf (*cf.* tableau 4.1).

Notre étude montre aussi que les réseaux bayésiens Multinets sont les plus sensibles à la taille de la base d'apprentissage. En effet, en réduisant la taille de la base, les

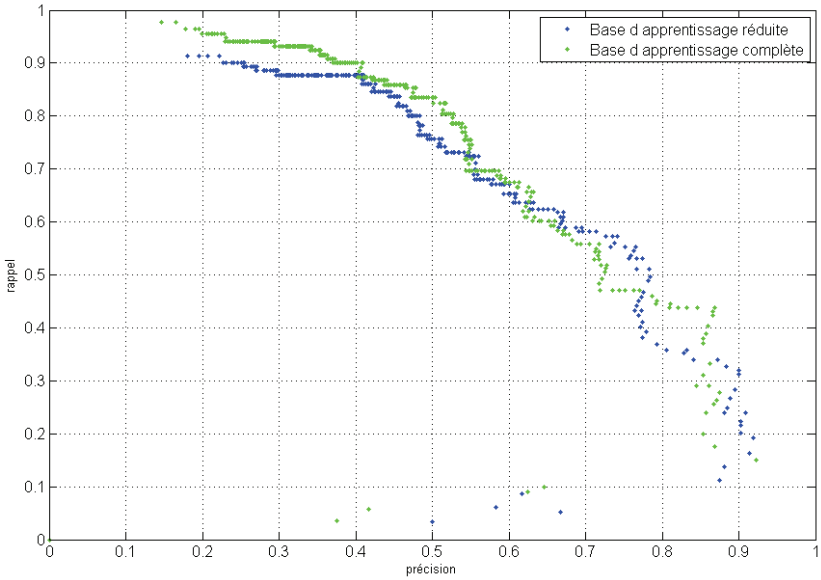

FIG. 4.12 – Influence de la taille de la base de données sur l'apprentissage de structure par augmentation de la structure du réseau bayésien par une structure en arbre.

performances du système basé sur ce type de réseau se dégradent considérablement (*cf.* figure 4.14). L'explication qui peut être donnée à cette dégradation est que, dans les réseaux Multinets, on apprend deux réseaux différents. En particulier, pour la classe *Action*, on apprend la structure du réseau de cette classe à partir d'un nombre très réduit d'exemples, ce qui rend l'apprentissage moins fiable pour ce réseau.

Un dernier volet de cette étude est l'influence de la taille de la base d'apprentissage sur l'approche discriminante. Le score utilisé dans cette approche se base sur le pouvoir de discrimination de la structure récupérée sur les exemples de la base. Ceci explique les résultats illustrés au niveau de la figure 4.15 et qui montrent que l'approche discriminante est plus fiables également en présence d'une plus grande base d'apprentissage.

4.7 Conclusion

Dans ce chapitre, nous avons utilisé diverses approches pour apprendre la structure d'un réseau bayésien dans le but de faire de la classification. Nous avons montré qu'une approche générative non restreinte n'est pas toujours efficace pour faire de la détection d'événements, surtout si on est en présence de plusieurs attributs et lorsque l'événement recherché n'est pas la cause principale du changement d'état des attributs. Nous nous sommes alors dirigés vers des approches qui enrichissent les réseaux bayésiens naïfs. Nous avons montré que l'augmentation de structure par une structure générique donne de

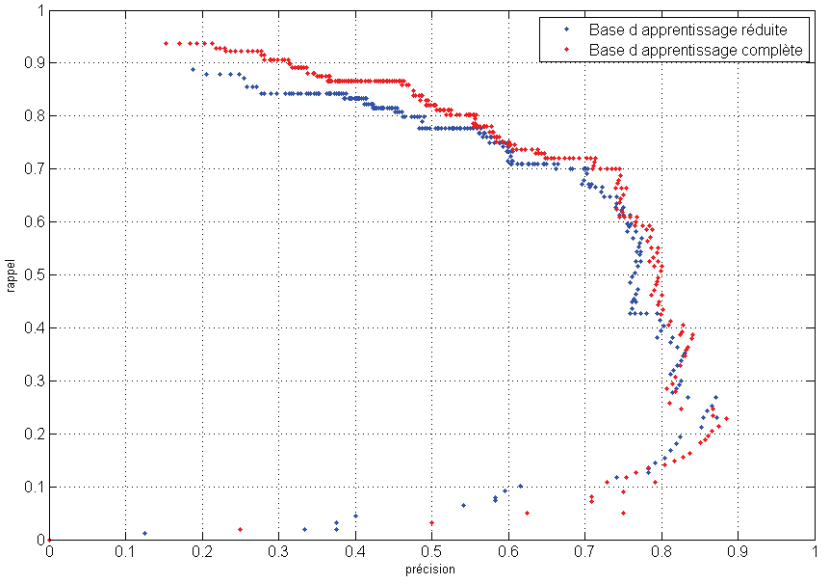

FIG. 4.13 – Influence de la taille de la base de données sur l'apprentissage de structure par augmentation de la structure du réseau bayésien par une structure génerique.

meilleurs résultats qu'une augmentation avec une structure en arbre. Il faut cependant trouver le bon compromis entre la complexité de la structure et son pouvoir génératif pour avoir le meilleur pouvoir de classification possible.

Conscients qu'à travers ce type d'approche on est toujours à la recherche de la structure qui décrit le mieux nos données, alors que notre but principal reste la classification en événement et non événement, nous avons utilisé un score discriminant tenant compte de la vraisemblance conditionnelle. Certes, cette approche est plus complexe d'un point de vue calculatoire, le score utilisé n'étant pas décomposable, mais la structure apprise donne de meilleurs résultats que les approches précédentes se basant sur un score génératif.

Dans un souci d'obtenir un système qui discrimine le plus possible entre les événements et les non-événements, nous avons testé, par la suite, les réseaux bayésiens Multinets qui requièrent un modèle pour chaque classe. Nous avons montré qu'une telle approche donne de meilleurs résultats si les structures recherchées pour les différentes classes ne sont pas contraintes.

Une comparaison globale entre les différentes approches montre que l'approche purement discriminante donne, malgré sa complexité de mise en œuvre, les meilleurs résultats. Les résultats des réseaux bayésiens augmentés par une structure générique et ceux des réseaux bayésiens Multinets, qui comme cela est expliqué dans ce chapitre, sont des approches qui incluent en plus du critère génératif une dimension discriminante à

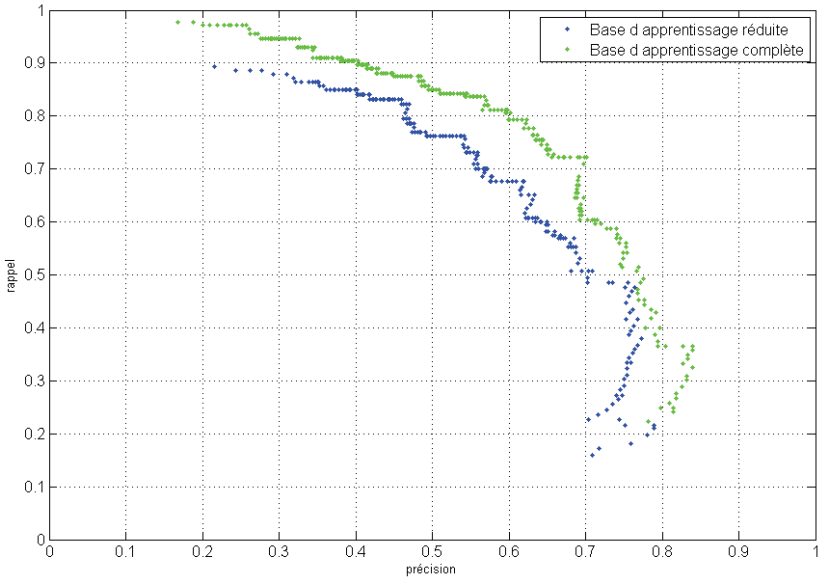

FIG. 4.14 – Influence de la taille de la base de données sur l'apprentissage de structure par l'approche Multinets.

travers le choix du paramètre λ ou à travers la recherche de deux structures différentes pour les deux classes, ces deux approches donc donnent des résultats assez comparables et légèrement inférieurs à une approche purement discriminante. Une approche basée sur un réseau bayésien augmenté par un arbre s'avère enfin la moins efficace pour la classification des événements ; ceci peut être dû à la simplicité de la structure recherchée dans cette méthode.

D'autre part, nous nous sommes intéressés à la stabilité des différentes méthodes vis-à-vis de la taille de la base d'apprentissage. D'après nos expériences, nous avons conclu que le réseau bayésien naïf était le plus stable de toutes les méthodes, de par la réduction de l'effort d'apprentissage qu'il faut faire sur ce type de réseau. Les approches augmentées sont, quant à elles, légèrement influencées par la taille de la base d'apprentissage. L'approche discriminante et l'approche basée sur les Multinets sont de leur coté très influencées par cette réduction de la taille de la base d'apprentissage. Elles sont, en effet, assez gourmandes en données d'apprentissage. L'espace de recherche de structure est, en effet, plus grand dans ces deux types d'approches.

Dans ce chapitre, nous avons utilisé tous les attributs dont nous disposions, sans toutefois, nous soucier du fait que ces attributs soient pertinents ou non pour notre problème. Nous comptions sur une sélection implicite des attributs lors de la phase d'apprentissage de structure. Cette sélection n'a pas eu lieu, hormis lors de la mise en place de l'apprentissage discriminant. Réduire le nombre d'attributs et donc le nombre

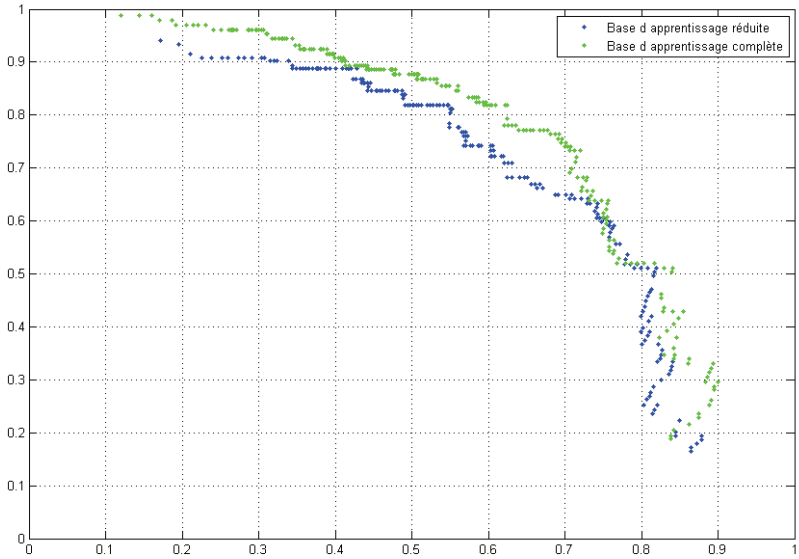

FIG. 4.15 – Influence de la taille de la base de données sur l'apprentissage de structure par l'approche discriminante.

de paramètres ainsi que le nombre d'arcs mis en place dans la structure du réseau, sans porter atteinte aux performances de classification du système, peut être d'une grande aide pour améliorer la robustesse de l'apprentissage de structure. Nous sommes en effet dans le cadre de la détection d'événements rares, pour lesquels nous ne disposons pas de beaucoup d'exemples. Une manière de réduire le nombre d'attributs est d'effectuer une première étape de sélection des attributs. Nous proposons dans le chapitre suivant d'étudier l'effet d'une pré-étape de sélection d'attributs sur l'apprentissage de structure et sur les performances de classification du système.

Chapitre 5

Influence de la sélection d'attributs sur l'apprentissage de structure

La richesse des contenus vidéos fait que l'on peut extraire un grand nombre d'attributs audio-visuels. Le but d'un système de description de contenus multimédias, plus précisément de détection d'événements, est d'intégrer ces différents attributs pour en extraire des métadonnées de haut niveau. Dans les approches où la structure du réseau bayésien est construite manuellement, seule une sélection d'attributs pertinents pour le problème est utilisée par le modèle. Cela suppose donc une connaissance du domaine pour ne choisir que les attributs permettant d'augmenter les performances du système. Dans le chapitre précédent, où nous avons utilisé une approche automatique pour la construction du modèle basée sur l'apprentissage de structure, nous avons utilisé tous les attributs disponibles. Il est toutefois intéressant d'étudier l'effet d'une sélection d'attributs sur les résultats des structures récupérées par le processus d'apprentissage de structure.

Le problème de sélection d'attributs peut être défini comme : *choisir, parmi un ensemble d'attributs candidats, un sous-ensemble d'attributs qui donne les meilleurs résultats pour un système donné de classification.* Des méthodes plus générales qu'une sélection sont également parfois utilisées pour créer de nouveaux attributs à partir de transformations ou de combinaisons des attributs d'origine. Le terme d'**extraction** d'attributs est alors utilisé dans ce cas. Nous nous intéressons toutefois ici au seul problème de **sélection** d'attributs.

Nous étudions dans ce chapitre l'influence de la sélection d'attributs sur les méthodes d'apprentissage de structure dans les réseaux bayésiens que nous avons proposées dans le chapitre 4.

5.1 Sélection automatique d'attributs

La sélection d'attributs peut être définie comme le problème suivant : trouver dans l'ensemble d'attributs X de taille n, $X = \{X_1, ..., X_n\}$, l'ensemble Z de taille $d \leq n$, permettant d'obtenir le meilleur taux de classification.

Dans [77], les auteurs dressent un état de l'art des méthodes classiquement utilisées pour la sélection d'attributs. On distingue ainsi principalement deux types de méthodes de sélection d'attributs. Dans le premier type, on ne tient pas compte des corrélations qui peuvent exister entre les attributs. Il s'agit de méthodes basées sur un classement des attributs. Dans le second type de méthodes, ces corrélations sont bien prises en compte. On distingue alors, dans cette dernière catégorie, deux classes principales : les approches basées *filtrage* et les approches de type *wrapper*. Nous détaillerons dans la suite le principe de ces différentes méthodes de sélection d'attributs.

5.1.1 Méthodes basées *classement*

Pour ce premier ensemble de méthodes, la technique de sélection d'attributs repose sur un classement des attributs en terme de pertinence par rapport à la variable représentant l'événement. Classer les attributs revient à attribuer à chaque attribut un score indépendamment des autres variables attributs. Nous considérons sans perte de généralité que plus le score est élevé plus la variable attribut est pertinente. On choisit ainsi les d attributs qui ont le plus grand score.

Cette méthode de sélection d'attributs se caractérise par une grande simplicité de mise en œuvre. Pour un ensemble $X = \{X_1, ..., X_n\}$, elle ne demande, en effet, que le calcul de n scores. Ces scores ne tiennent compte que de la corrélation qui existe entre chaque variable attribut X_i et la variable classe X_c. Toutefois, elle souffre d'un grand inconvénient : en effet, en ne considérant que la corrélation entre la variable attribut et la variable classe, la sélection d'attributs peut ignorer des attributs peu pertinents individuellement mais dont l'action collective peut s'avérer très utile pour la classification de la variable X_c. Cette sélection d'attributs peut cependant s'avérer suffisante si les attributs ont une action indépendante par rapport à la variable classe.

Dans le cas contraire, il est nécessaire de tenir compte des corrélations existant entre les attributs. Nous passons en revue dans le paragraphe suivant une autre catégorie de méthodes de sélection palliant ce défaut.

5.1.2 Méthodes tenant compte des corrélations entre attributs

Le principe de ce type de méthodes est d'étudier la pertinence d'un sous ensemble d'attributs par rapport à la variable classe ou événement. Cette méthode diffère de l'approche précédente dans le fait que l'on considère l'action de plusieurs attributs à la fois et non pas l'action d'un seul attribut. Pour un sous ensemble d'attributs donné, cette pertinence est évaluée à travers l'utilisation d'un critère ou score. Le parcours exhaustif de tous les sous-ensembles d'attributs, afin de leur attribuer un score, est souvent calculatoirement impossible, on a donc recours généralement à une méthode de parcours non exhaustif de l'ensemble des sous-ensembles d'attributs.

Nous pouvons noter la ressemblance entre la technique de mise en œuvre de la sélection d'attributs et celle d'apprentissage de structure basé sur le score, ceci dans le sens où nous avons besoin dans les deux cas d'une méthode de parcours et d'un critère ou score d'évaluation.

Nous détaillons donc dans la suite les différents éléments nécessaires pour la mise en œuvre des méthodes de sélection d'attributs tenant compte de la corrélation entre les attributs, à savoir :

- la méthode de parcours de l'ensemble des sous-ensembles d'attributs ;
- le critère d'évaluation J pour mesurer la qualité d'un sous-ensemble donné d'attributs.

5.1.2.1 Méthode de parcours

Une méthode de parcours exhaustif de l'ensemble des sous-ensembles d'attributs est toujours envisageable. Toutefois, cela requiert de traiter les $\binom{n}{d}$ sous-ensembles possibles d'attributs de taille d. Ce nombre augmente exponentiellement avec le nombre d'attributs. Un parcours exhaustif devient alors vite impossible d'un point de vue calculatoire.

Différentes méthodes de parcours non exhaustives sont proposées dans la littérature [78] parmi lesquelles, il est possible de citer : la méthode best-first, la méthode branch and bound, le recuit simulé, les algorithmes génétiques et les méthodes séquentielles. Dans [79], les auteurs concluent que la méthode basée sur l'algorithme séquentiel SFFS (*Sequential Forward Floating Selection*) proposée par Pudil *et al.* dans **??** domine les autres méthodes de par ses performances.

L'algorithme SFFS est la combinaison de deux méthodes de parcours de sous-ensembles pour la sélection d'attributs : le SFS (Sequential Forward Selection) et sa contrepartie le SBS (Sequential Backward Selection). Le SFS débute par un sous-ensemble d'attributs vide pour arriver à un sous-ensemble d'attributs de taille d. Il procède par l'ajout à chaque itération de l'attribut qui augmente le plus un critère J. Le SBS est par contre initialisé avec l'ensemble des attributs. Il procède par la suppression des attributs dont l'élimination augmente le plus le critère J et ce jusqu'à atteindre un sous-ensemble d'attributs de taille d. Les deux algorithmes ont toutefois un inconvénient majeur. En effet, un attribut ajouté par le SFS à l'ensemble des attributs sélectionnés ne peut plus être enlevé de l'ensemble (réciproquement dans le cas du SBS, un attribut supprimé de l'ensemble des attributs ne peut plus être récupéré).

Le SFFS est une combinaison des deux algorithmes. Le lecteur en trouvera une description dans l'algorithme 1. Cet algorithme a l'avantage d'être une combinaison des deux méthodes de parcours décrites ci-dessus. On peut ainsi ajouter et éliminer en même temps les attributs. Le but principal est bien sûr d'arriver à un sous-ensemble d'attributs de taille d, qui soit le meilleur possible.

Ayant déjà un sous-ensemble sélectionné, on ajoute l'attribut qui augmente le plus la valeur du critère. Dans une seconde étape, on élimine les attributs dont la suppression conduit à un score supérieur au score obtenu pour la même taille de sous-ensemble. Et on itère jusqu'a ce qu'on arrive à un sous ensemble de taille d.

Au vu de l'état de l'art cité dans [79], pour notre approche de sélection d'attributs, nous avons donc fait le choix d'utiliser la méthode SFFS pour faire le parcours des différents sous-ensembles.

Algorithm 1 SFFS

Entrée : $X = \{X_i | i = 1, ..., n\}$ tous les attributs.
Sortie : $Z^k = \{Z_i | i = 1, ..., k, Z_i \in X\}, k = d$
Initialisation $X_0 = \varnothing; k = 0$
arrêt $k = d$
Etape1(Inclusion)
$Z_i^+ = \arg\max_{Z_i \in X - Z^k} J(Z^k + Z_i)$
$Z^{k+1} = Z_k + Z_i^+;$
$k = k + 1$
Etape2(Exclusion conditionnelle)
$Z^- = \arg\max_{Z_i \in Z^k} J(Z^{k-1})$
si $J(Z^k - Z^-) > J(Z^{k-1})$ *alors*
$Z^{k-1} = Z^k - Z^-;$
$k = k - 1$
retour à l'étape 2
sinon
retour à l'étape 1

5.1.2.2 Critère

La sélection des attributs optimaux se fait au travers de l'optimisation d'un critère que nous notons J dans la suite. Le problème de sélection d'attributs peut donc se formuler sous la forme d'un problème d'optimisation :

$$Y = \arg \max_{Z \subseteq X, card(Z)=d} J(Z) \tag{5.1}$$

Comme nous l'avons évoqué dans l'introduction de ce chapitre, il existe principalement deux classes de méthodes de sélection d'attributs tenant compte des corrélations entre les attributs :

- les méthodes basées *filtrage*,
- les méthodes de type *wrapper*.

Ces deux classes se distinguent par la nature du critère utilisé. Dans les méthodes basées *filtrage*, le score est totalement indépendant de la partie classification qui utilise les attributs sélectionnés. Dans les méthodes de type *wrapper*, la méthode de classification est utilisée pour juger de la pertinence du sous-ensemble d'attributs. Nous proposons dans la suite une présentation de ces deux types de techniques.

5.1.2.3 Méthodes basées *filtrage*

Dans ce premier cas, la sélection d'attributs est faite indépendamment de la méthode de classification qui constitue l'étape suivante dans le traitement des données. Le critère J utilisé est donc complètement indépendant du classifieur utilisé.

Le but principal de la sélection d'attributs est alors la sélection du sous-ensemble d'attributs qui apporte la plus grande quantité d'information. Les méthodes de sélection

d'attributs basées sur les filtrages suivent le schéma de la figure 5.1. Le module de sélection d'attributs a comme entrée l'ensemble des attributs disponibles $X = \{X_1, ..., X_n\}$ et rend en sortie un sous-ensemble de ces attributs. Contrairement aux méthodes de type *wrapper* que nous allons évoquer plus loin, il n'y a aucune interaction entre le module de sélection d'attributs et le module de classification. Les attributs ainsi sélectionnés sont totalement indépendants de la méthode de classification utilisée par la suite.

Cette indépendance entre le module de sélection et le module de classification confère à la sélection d'attributs un caractère de généricité, puisque les attributs sélectionnés sont valables pour tous types de classifieurs. On n'est donc pas dans l'obligation de changer les attributs sélectionnés chaque fois qu'on change de module de classification. Toutefois, cette généricité peut conduire à un système non optimal puisqu'en sélectionnant les attributs indépendamment du module de classification, on n'optimise pas la chaîne de traitement de bout en bout.

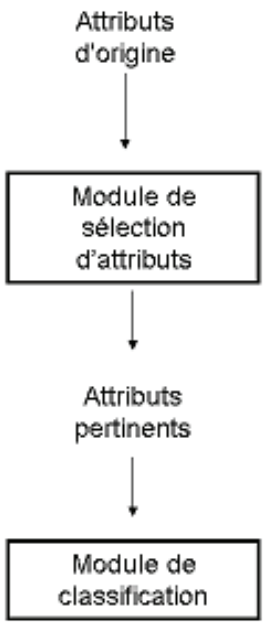

FIG. 5.1 – Principe de la sélection d'attributs basée *filtrage*.

Parmi les scores les plus utilisés dans la sélection d'attributs basée *filtrage*, nous retrouvons le gain d'information $IMG(X_c, Z)$ d'un ensemble Z d'attributs par rapport à la classe X_c (*cf.* équation 5.2). Ce critère mesure la réduction d'incertitude par rapport à la classe X_c quand l'information sur les variables du sous-ensemble Z sont disponibles. Il est constitué d'une différence entre l'entropie $H(X_c)$ représentant l'incertitude par rapport à la classe X_c et de l'entropie conditionnelle $H(X_c|Z)$ représentant l'incertitude

par rapport à la classe X_c en ayant l'information contenue dans le sous-ensemble Z.

$$\begin{aligned} IMG(X_c, Z) &= H(X_c) - H(X_c|Z) \\ avec : H(X_c) &= -\sum P(X_c = c_j) log(P(X_c = c_j)) \\ H(X_c|Z) &= -\sum P(X_c, Z) log(P(X_c, Z)/P(Z)) \end{aligned} \tag{5.2}$$

5.1.2.4 Méthodes de type *wrapper*

Comme nous venons de le voir, les méthodes basées *filtrage* ne tiennent pas compte de la tâche de classification. En effet, elles permettent de sélectionner les attributs qui sont les plus corrélés avec la variable classe sans tenir compte du type de classifieur qui sera utilisé par la suite. Cette indépendance par rapport au module de classification, illustrée dans la figure 5.1, peut conduire à un système non optimal.

Les méthodes de type *wrapper* apportent une alternative aux méthodes basées *filtrage*. On optimise dans ce cas la sélection d'attributs par rapport à la tâche de classification. Le module de classification est utilisé pour mesurer la pertinence des attributs sélectionnés. Le critère J est donc une mesure de la performance du classifieur en présence d'un ensemble d'attributs.

Le classifieur est considéré comme une boîte noire ayant comme entrée un sous-ensemble des attributs et comme sortie une mesure de la qualité de prédiction du classifieur. Le schéma de principe des méthodes de sélection d'attributs de type *wrapper* est illustré dans la figure 5.2. Un bon état de l'art sur ce type de méthodes est disponible dans [78].

Pour le calcul du critère J, nous utilisons un processus de cross-validation. Ainsi, nous faisons l'apprentissage du réseau bayésien ou du classifieur sur une partie des données et nous décodons sur la deuxième partie des données. on répète le processus de façon à ce que toutes les données de la base sont utilisées dans la phase de décodage. Nous mesurons alors le taux de bonne classification sur chacun des sous-ensembles de données issus de la cross-validation. Le critère J global sera la somme des taux de bonne classification sur la totalité des sous-ensembles de données.

En contrepartie de leur avantage de tenir compte de l'étape de classification, les méthodes de type *wrapper* sont cependant très gourmandes en temps de calcul.

5.2 Résultats et interprétation

5.2.1 Comparaison des différentes méthodes de sélection d'attributs pour un classifieur donné

Nous proposons dans ce paragraphe une comparaison des résultats des trois types de méthodes de sélection d'attributs : les méthodes par *classement*, les méthodes basées *filtrage* et celles de type *wrapper*. Pour ce faire, nous choisissons d'utiliser comme classifieur le réseau bayésien augmenté par une structure d'arbre.

Il est évident d'après la figure 5.3 que les résultats de la sélection d'attributs de type *wrapper* surpassent largement ceux de la sélection basée *filtrage* et ceux de la sélection

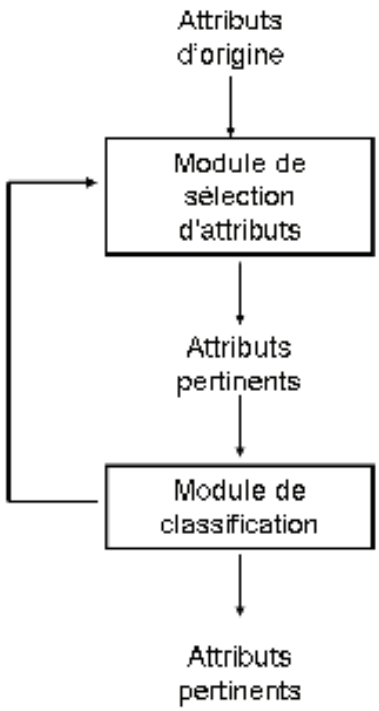

FIG. 5.2 – Schéma de la sélection d'attributs de type *wrapper*.

basée *classement*. En effet, de par sa définition, la méthode de type *wrapper* utilise l'algorithme de classification lui-même, les attributs récupérés s'avèrent donc mieux adaptés à la classification que ceux donnés par les autres approches. D'autre part, la sélection basée *filtrage* donne des résultats légèrement supérieurs à ceux de l'approche basée *classement*. Ceci peut s'expliquer par le fait que cette méthode tient compte de la corrélation qui existe entre les différents attributs du problème, contrairement à la sélection basée *classement*.

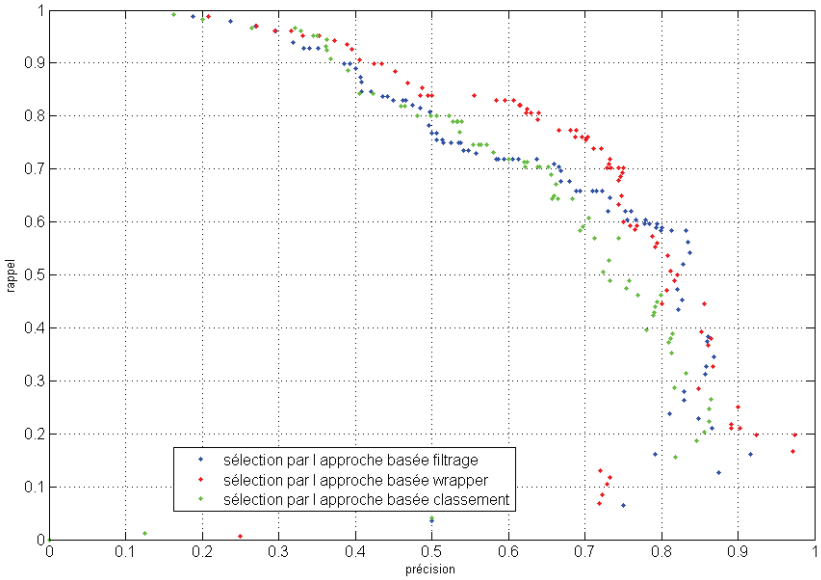

FIG. 5.3 – Comparaison des performances entre les trois méthodes de sélection d'attributs. Classifieur utilisé : réseau bayésien enrichi par une structure en arbre.

Nous utiliserons dans le reste de ce mémoire la méthode de sélection d'attributs de type *wrapper* partout où on ne mentionne pas de façon explicite une autre méthode de sélection.

5.2.2 Influence de la sélection d'attributs sur le réseau bayésien naïf

Une sélection d'attributs de type *wrapper* appliquée à un réseau bayésien naïf permet d'augmenter les performances de classification de ce type de réseau (*c.f.* figure 5.4). En effet, la sélection d'attributs permet de ne garder que les attributs qui maximisent le pouvoir de classification. Il est à noter que l'utilisation de la sélection d'attributs de type *wrapper* introduit un caractère discriminant à la méthode utilisée. En effet, les attributs sélectionnés sont ceux qui donnent la meilleure classification sur l'ensemble d'apprentissage.

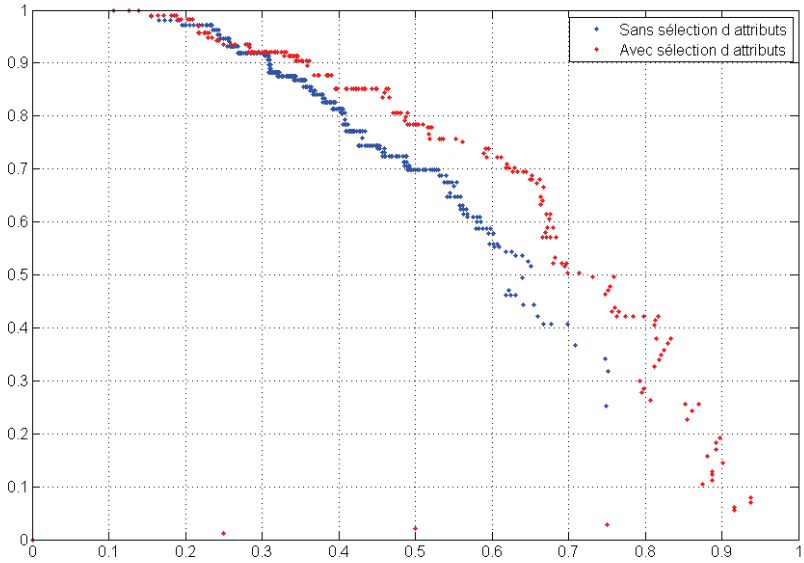

FIG. 5.4 – Influence de la sélection d'attributs sur le réseau bayésien naïf.

5.2.3 Influence de la sélection d'attributs sur les approches enrichies

Comme nous l'avons présenté dans le chapitre 4, l'apprentissage de structure augmentée part d'un réseau bayésien naïf comme structure de base. Tous les attributs sont ainsi connectés au nœud classe. Les connexions inter-attributs sont ensuite ajoutées pour enrichir le réseau. Cet enrichissement de la structure entraîne une meilleure adéquation de la structure aux données. Il conduit toutefois à des distributions de probabilités plus complexes que celles des réseaux bayésiens naïfs. L'existence d'attributs qui ne sont pas pertinents pour la classification augmente la complexité des distributions sans pour autant augmenter le pouvoir de classification du système. Nous testons dans cette partie l'apport de la sélection d'attributs et donc de l'élimination des attributs qui n'ont pas réellement de valeur ajoutée pour notre problème de classification.

5.2.3.1 Influence de la sélection d'attributs sur l'approche de type réseau bayésien naïf enrichie par une structure en arbre

Dans cette partie, nous étudions l'influence de la sélection sur l'apprentissage de structure utilisant l'approche de type TAN (*cf.* paragraphe 4.3.1). Sur la figure 5.5, nous présentons les résultats de l'apprentissage de structure sans sélection d'attributs, avec une sélection d'attributs manuelle et avec une sélection d'attributs de type *wrapper*. Dans le cas de la sélection manuelle, nous avons choisi les attributs en nous basant sur la connaissance dont nous disposons par rapport à la détection d'*Actions* dans un match

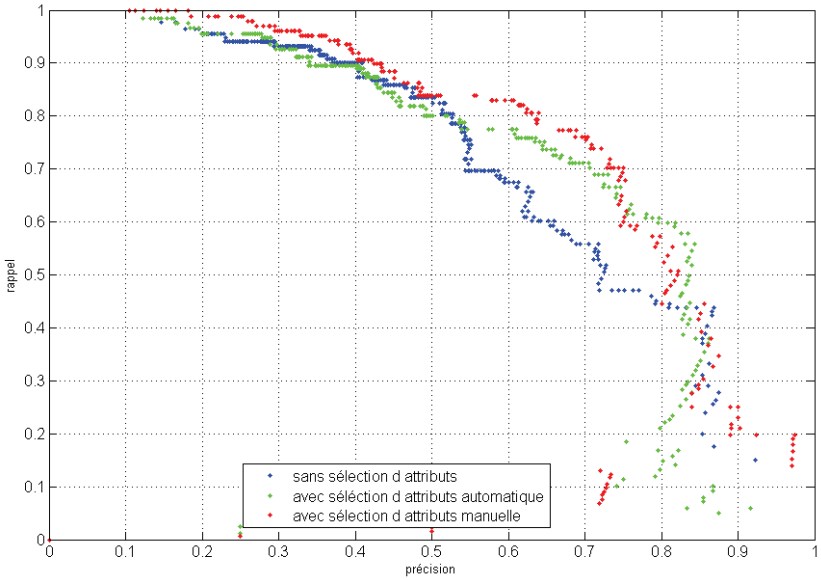

FIG. 5.5 – Influence de la sélection d'attributs pour les réseaux bayésiens naïfs augmentés par une structure d'arbre.

de football. Nous remarquons que la sélection d'attributs augmente considérablement les performances de classification du système. Ainsi, pour une précision de 0.7, nous passons d'un rappel de 0.55 à un rappel de 0.68.

Il apparaît également que les performances de la sélection automatique sont égales à celle de la sélection manuelle. Nous avons donc réussi, au travers de notre système, à automatiser l'étape de sélection d'attributs.

Un autre avantage non négligeable de l'étape de sélection d'attributs est la réduction du nombre d'attributs utilisés, ce qui en terme de puissance de calcul conduit à des algorithmes de classification moins gourmands.

5.2.3.2 Influence de la sélection d'attributs sur l'approche de type réseau bayésien naïf enrichie par une structure générique

Dans cette partie, nous étudions l'influence de la sélection d'attributs sur les approches de type réseau bayésien naïf enrichi par une structure générique. Nous utilisons dans nos tests une sélection d'attributs de type *wrapper*. En comparant la courbe de couleur bleue pour les performances d'un réseau sans sélection d'attributs et la courbe rouge pour les performances d'un réseau avec sélection de la figure 5.6, nous remarquons que la sélection d'attributs n'augmente pas les performances du réseau bayésien enrichi par une structure générique, même si la sélection est réalisée avec une approche de type *wrapper*. Ce résultat est assez décevant, puisque nous nous attendions à ce que

les performances du réseau augmentent avec la baisse du nombre de paramètres et des arcs à apprendre dans la structure.

Toutefois, nous pouvons reprocher à notre mise en œuvre de la sélection d'attributs le fait que nous avons fixé à l'avance le paramètre λ servant pour l'apprentissage de structure du score BIC modifié (*cf.* équation 4.4). Ce paramètre dépend de la complexité de la structure, et donc du nombre de nœuds dans le réseau. Ce nombre de nœuds est variable lors de l'opération de sélection d'attributs. Ceci nécessite donc une adaptation du paramètre λ au nombre de nœuds impliqués dans le réseau au fur et à mesure qu'on sélectionne les attributs. Nous proposons alors d'adapter ce paramètre à partir de sa valeur récupérée de manière expérimentale dans le paragraphe 4.3.3.1.

Ainsi, pour l'adaptation du paramètre λ, nous utilisons la stratégie suivante : nous calculons le rapport α entre le terme de vraisemblance et le terme de complexité du score BIC pour un réseau bayésien naïf. Nous supposons que le rapport $\frac{\alpha}{\lambda}$ est constant pour toutes les tailles de réseaux. Nous avons déjà calculé expérimentalement le paramètre λ pour le cas où le réseau contient tous les nœuds, donc sans aucune sélection d'attributs. Nous pouvons alors calculer les paramètres λ correspondant à chaque nombre de nœuds sélectionnés. Ces nouveaux paramètres λ sont alors utilisés lors de l'apprentissage de structure des réseaux enrichis par une structure générique dans la sélection d'attributs de type *wrapper*.

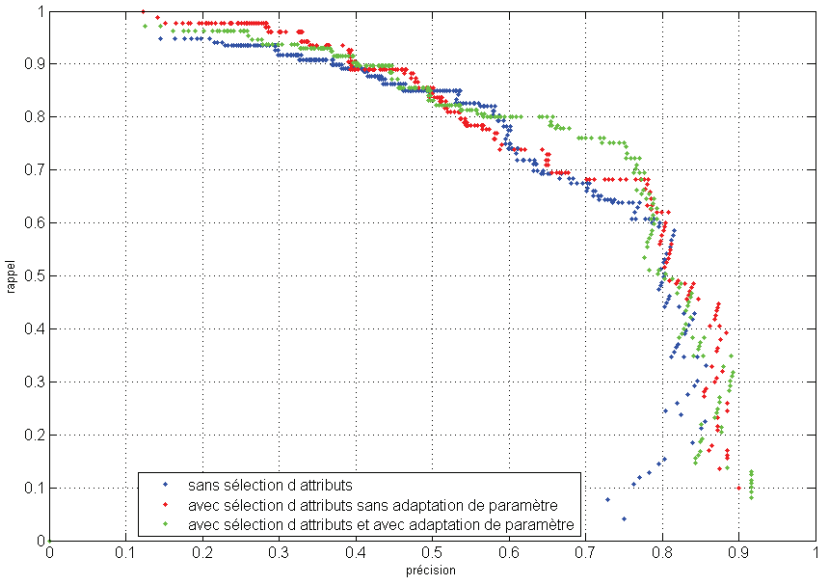

FIG. 5.6 – Comparaison entre les performances des réseaux augmentés par une structure générique sans sélection d'attributs et avec sélection d'attributs avec et sans adaptation du paramètre λ.

Les résultats du réseau bayésien enrichi par une structure générique utilisant des attributs sélectionnés par une approche de type *wrapper* avec adaptation du paramètre

λ sont illustrés sur la figure 5.6. Nous obtenons bien cette fois-ci une augmentation des performances de la détection d'événements en utilisant la sélection d'attributs.

Ceci rejoint la conclusion sur le rôle de la sélection d'attributs pour l'approche d'apprentissage par enrichissement par un arbre. Les approches d'apprentissage par enrichissement de la structure d'un réseau naïf en général sont donc sensibles à la sélection d'attributs. Elles donnent en effet de meilleurs résultats si les attributs qui leur sont présentés sont les plus pertinents possibles.

5.2.4 Influence de la sélection d'attributs sur les approches Multinets

Nous étudions, dans cette partie, l'influence de la sélection d'attributs sur les approches de type Multinets. Nous présentons à la figure 5.7 une comparaison des performances d'un réseau Multinets avec des sous-structures en arbre, précédé ou non d'une étape de sélection d'attributs.

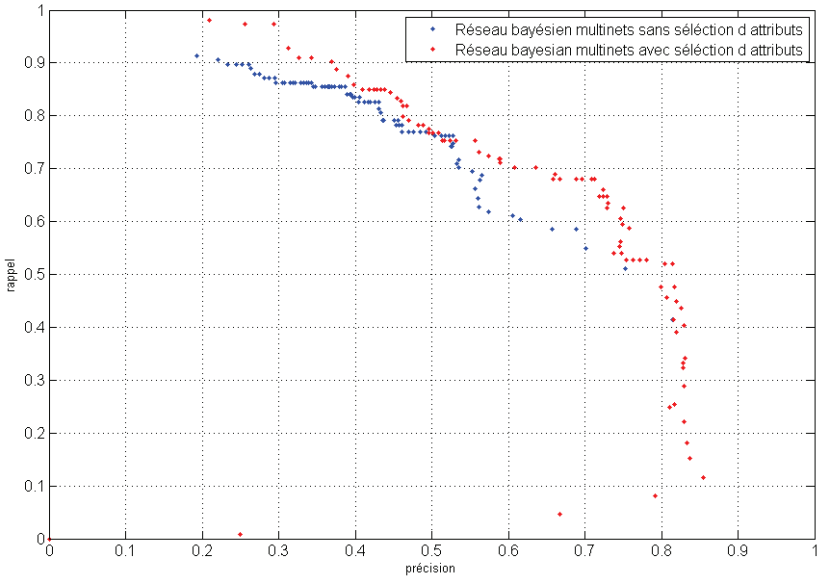

FIG. 5.7 – Comparaison des performances d'un réseau Multinets avec et sans sélection d'attributs.

Dans la figure 5.7, nous remarquons une nette amélioration des résultats des réseaux Multinets lorsqu'on utilise un nombre réduit d'attributs. En effet, en réduisant le nombre d'attributs, l'apprentissage des structures devient plus fiable, particulièrement au niveau de la classe minoritaire où on ne dispose que d'un nombre réduit de paramètres.

5.2.5 Influence de la sélection d'attributs sur l'approche générative

Nous étudions dans ce paragraphe l'influence de la sélection d'attributs sur l'approche générative non restreinte pour l'apprentissage de structure (*cf.* chapitre 3). Nous rappelons que, dans le chapitre 4, nous avons montré que cette approche générative non restreinte ne donne pas de résultats satisfaisants pour la classification des événements rares tels que les *Actions* dans un match de football. Nous avons expliqué ces faibles résultats par le fait que cette approche ne donne pas de valeur particulière au nœud de classification.

La figure 5.8 montre les résultats de l'approche générative avec et sans sélection d'attributs. Dans le cas où la sélection d'attributs a lieu au préalable, il apparaît que les résultats augmentent considérablement. On arrive à un niveau de performance équivalent à celui des approches enrichies.

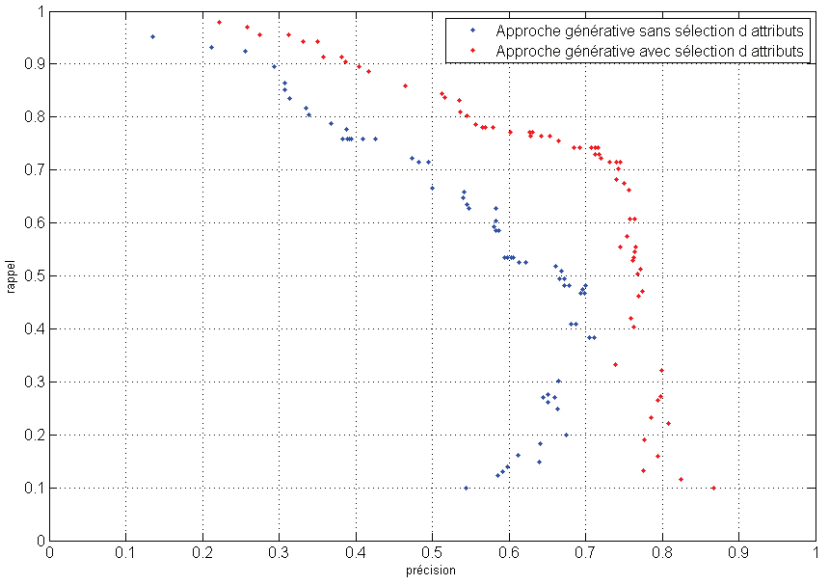

FIG. 5.8 – Comparaison des performances de l'approche générative non restreinte avec et sans sélection d'attributs.

Nous concluons donc que l'approche générative non restreinte est très sensible à la pertinence des attributs que lui sont fournis. Toutefois, si elle est en présence d'attributs pertinents, l'approche donne des résultats satisfaisants. La sélection d'attributs de type *wrapper* ajoute ainsi à l'approche générative non restreinte la dimension qui lui manque par rapport aux autres approches, à savoir : considérer le nœud de classification différemment des autres nœuds. Une étape de sélection d'attributs est indispensable à l'entrée du système de classification se basant sur un apprentissage de structure générique.

5.2.6 Influence de la sélection d'attributs sur l'approche discriminante

Dans les approches précédentes basées sur un score maximisant la vraisemblance, la sélection d'attributs s'est révélée un outil pertinent pour augmenter les performances de ce type de système qui n'incorpore par ailleurs aucun moyen efficace pour distinguer les attributs pertinents pour la tâche de classification. Nous étudions, dans cette partie, l'influence de la sélection d'attributs sur l'apprentissage de structure obtenu par une approche discriminante telle que présentée dans le paragraphe 4.4.

Il est utile de rappeler que nous avons conclu lors de notre étude de l'approche discriminante dans le paragraphe 4.4 du chapitre précédent que cette approche permet de rejeter les attributs dont l'incorporation dans la structure du réseau n'augmente pas les pouvoirs de classification du système. L'apprentissage de structure discriminant peut être apparenté à une approche de sélection d'attributs de type *wrapper*. En effet, lors de la génération de la structure, divers arcs sont éliminés, provoquant l'élimination de certains attributs de la structure.

Dans la figure 5.9, on compare les performances de l'approche discriminante avec ou sans sélection d'attributs. Il est à noter qu'une sélection de type *wrapper* basée sur l'approche discriminante est impossible à mettre en œuvre d'un point de vue calculatoire. Il faut en effet faire une recherche de structure discriminante à chaque fois que l'on désire ajouter un attribut à l'ensemble des attributs pertinents. La sélection d'attributs utilisée dans notre test est donc dans ce cas précis une sélection d'attributs basée *filtrage*. Cette comparaison est effectuée sur deux tailles différentes de sous-ensembles d'attributs. Il apparaît que la sélection d'attributs avec un nombre suffisant d'attributs n'augmente pas les performances de classification. Ceci confirme le rôle de la méthode discriminante comme méthode de sélection d'attributs à part entière.

Ainsi, la méthode basée sur l'apprentissage de structure par un critère discriminant présente un double avantage. Elle permet en effet de construire une structure adaptée à notre besoin de classification et en même temps elle permet de faire la sélection d'attributs pour ne laisser que les attributs les plus significatifs pour la tâche de classification.

Toutefois, l'ajout d'une étape de sélection en entrée de l'approche discriminante permet de réduire la complexité de calcul de cette méthode : le nombre d'attributs diminue ; l'ensemble de recherche de structure optimale diminue également de taille. D'autre part, il y a aussi moins de risque de tomber dans un minimum local pour le calcul de la CLL (*cf.* équation 4.6).

Dans la figure 5.9, nous présentons également les résultats de l'apprentissage discriminant pour un nombre réduit d'attributs. Les performances du système sont alors revues à la baisse. Le système n'arrive plus à trouver toute l'information dans ce nombre réduit d'attributs. Donc, si on désire réduire les variables au niveau de l'apprentissage de structure discriminant, il faut nous assurer que nous présentons au système un nombre suffisant et pertinents de paramètres.

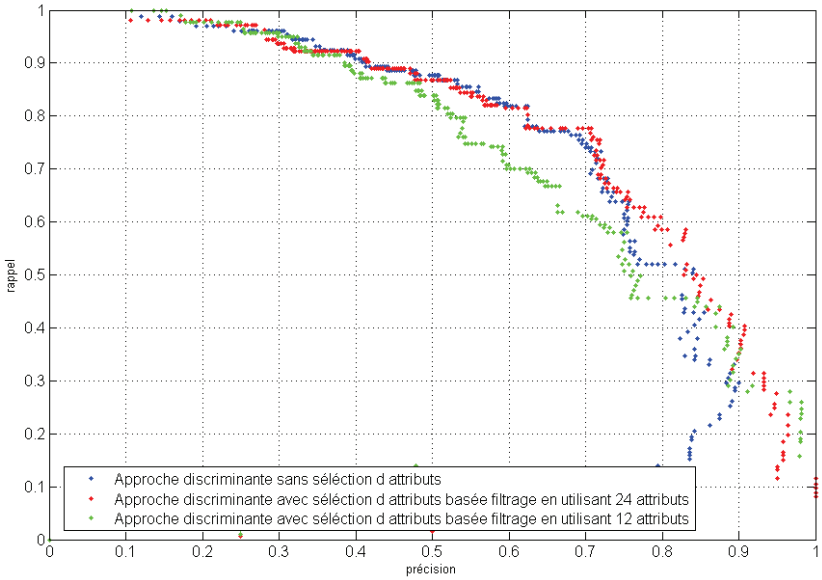

FIG. 5.9 – Influence de la sélection d'attributs sur l'apprentissage de structure par une approche discriminante.

5.3 Interprétation des attributs sélectionnés

Nous proposons dans cette partie de donner une interprétation aux attributs sélectionnés par l'approche de sélection d'attributs. Ainsi, dans le cadre de notre application de détection d'*actions* dans un match de football, un attribut toujours sélectionné par l'algorithme de sélection d'attributs est l'attribut *position sur le terrain*. En effet, cet attribut est très suggestif de l'existence d'une *Action* ou non : une *Action* se produit toujours dans la zone de la cage de but. Le *niveau audio* au niveau du plan courant s'avère aussi un attribut important pour la classification en *Action* ou non. L'algorithme de sélection d'attributs choisit aussi les attributs *Ralenti* des plans qui suivent le plan courant. Dans les attributs sélectionnés, nous retrouvons également les attributs renseignant sur la nature du plan courant, c'est-à-dire si le plan appartient à une zone de jeu ou non et s'il s'agit d'un plan large ou non. Cette sélection d'attributs confirme l'observation qu'une *Action* est une zone de la vidéo avec une augmentation de l'excitation de la foule et une zone de jeu qui se déroule au niveau de la cage de but. Elle est généralement suivie par une plage de ralenti, se trouve au niveau d'une zone non régulière du jeu et se déroule au niveau d'un plan large. Les attributs sélectionnés sont donc cohérents avec une interprétation humaine de ce que sont des *Actions* dans un match de football.

5.4 Conclusion

Nous nous sommes intéressés dans ce chapitre à la sélection d'attributs. Nous avons appliqué la sélection d'attributs comme méthode de pré-traitement avant le module de classification basé sur l'apprentissage de structure dans les réseaux bayésiens. Notre étude a montré que la sélection d'attributs constitue pour les méthodes d'apprentissage basées sur un score génératif un vrai potentiel pour augmenter les performances du système. En utilisant la sélection d'attributs, les tâches de simulation de données et de classification se rapprochent de plus en plus. Éliminer les données non pertinentes constitue une étape de débruitage, ce qui favorise un apprentissage de structure efficace.

Ainsi, l'apprentissage de structure génératif non restreint présenté dans le chapitre 3, qui donne des résultats médiocres pour la détection des *Actions* augmente considérablement ses performances avec une étape de sélection d'attributs. C'est également le cas pour les approches augmentées qui utilisent naturellement tous les attributs disponibles.

L'approche discriminante reste tout de même une exception. En effet, elle donne de bons résultats de classification sans aucun besoin d'une étape de sélection d'attributs. Celle-ci ne permet donc pas d'augmenter les performances du système, contrairement aux autres approches. Cette approche inclut en elle-même un processus de sélection d'attributs qui lui permet de ne garder dans la structure apprise que les attributs qui sont pertinents pour la tâche de classification.

Ainsi, la principale conclusion que nous pouvons tirer de ce chapitre est qu'une étape de sélection d'attributs est indispensable pour garantir des résultats optimaux. Cette étape peut être faite d'une manière disjointe comme c'est le cas des approches génératives ou d'une manière implicite comme c'est le cas pour l'apprentissage de structure basé sur un critère discriminant.

Chapitre 6

Apprentissage de structure dans les réseaux bayésiens dynamiques

6.1 Introduction

Nous avons étudié dans les chapitres précédents l'effet de l'apprentissage de structure sur les réseaux bayésiens statiques. Cet apprentissage s'avère être un moyen pour augmenter les performances des systèmes de détection d'événements. Toutefois, les structures récupérées par cette méthode ne tiennent pas compte de la nature temporelle des données vidéos. Or les réseaux bayésiens permettent de modéliser ce caractère temporel au travers de leur variante dynamique. Nous étudions dans ce chapitre l'apport de l'apprentissage de structure dans les réseaux bayésiens dynamiques.

6.2 Les réseaux bayésiens dynamiques

Les réseaux bayésiens dynamiques sont un cas particulier des réseaux bayésiens. Ils ont été mis en place pour définir un formalisme clair pour la représentation des systèmes évoluant au cours du temps. Le temps est alors discrétisé en unités temporelles. À un instant donné, le réseau représente l'état de l'ensemble des variables du système. Les variables d'une même tranche temporelle sont connectées pour représenter les corrélations entres les variables à un même instant. Les réseaux bayésiens dynamiques modélisent également les corrélations temporelles qui existent entre les variables. Seules les connexions des variables du passé vers le présent sont autorisées. Ces connexions modélisent l'influence du passé sur le présent.

D'un point de vue formalisme, les réseaux bayésiens dynamiques sont définis comme suit :

- un réseau spécifiant les distributions sur l'état initial $\mathcal{B}_0$;
- un réseau de transition définissant les corrélations temporelles entre les différentes tranches temporelles $\mathcal{B}_{trans}$.

Un réseau bayésien dynamique défini par les deux structures $\mathcal{B}_0$ et $\mathcal{B}_{trans}$ correspond à un réseau semi-infini $X[0], ..., X[\infty]$. Toutefois, en pratique, on utilise le réseau

seulement sur une plage temporelle bien précise de longueur T. En déroulant le réseau sur cette plage temporelle, on obtient un réseau bayésien statique. La loi jointe de tout le réseau sera alors :

$$P_{\mathcal{B}}(X[0], ..., X[T]) = P_{\mathcal{B}_0} \prod_{t=0}^{T-1} P_{\mathcal{B}trans(X(t+1)|X(t))} \tag{6.1}$$

La figure 6.3 représente les structures d'un réseau bayésien dynamique ainsi que son équivalent statique déroulé sur une plage temporelle de durée $T = 3$.

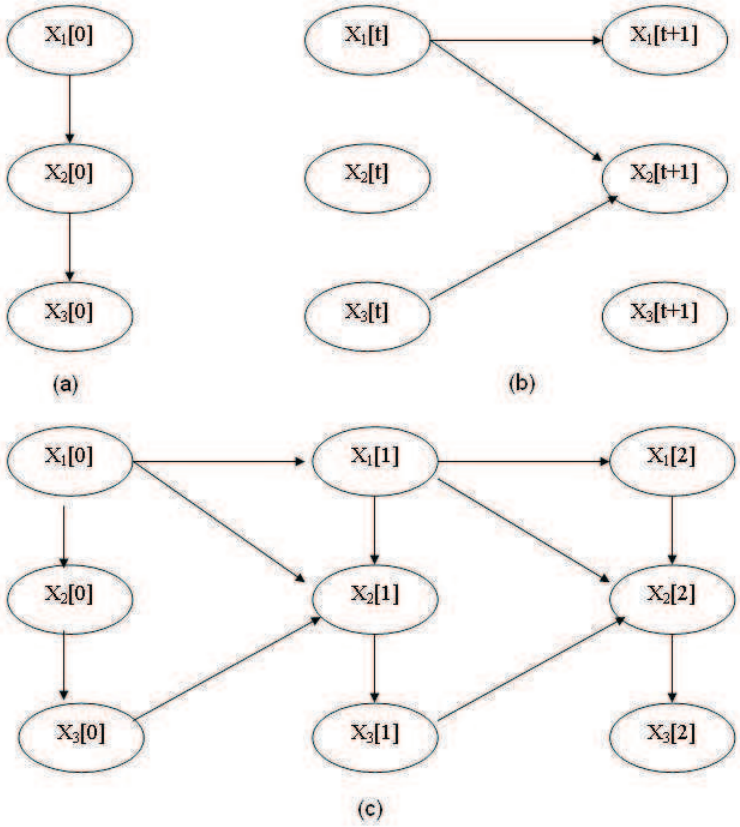

FIG. 6.1 – Structures d'un réseau bayésien dynamique composé de trois variables X_1, X_2, X_3. (a) structure initiale, (b) structure temporelle, (c) réseau bayésien équivalent déroulé sur $T = 3$.

Dans [49], les auteurs utilisent les réseaux bayésiens dynamiques pour détecter des événements dans des vidéos de Formule 1. Dans ce cadre, diverses structures temporelles de réseaux bayésiens ont été testées. Les auteurs ont conclu que la structure de la figure 6.2 donne les meilleurs résultats.

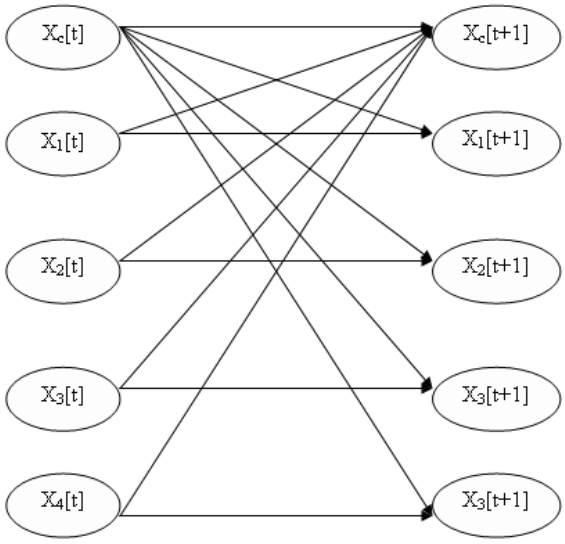

FIG. 6.2 – Structure du réseau bayésien dynamique proposé dans [49].

Ce travail montre l'importance d'un choix adéquat des connexions temporelles dans un réseau bayésien dynamique pour garantir de bonnes performances en terme de détection d'événements vidéo. Toutefois, il n'est pas assuré que la structure présentée dans ce travail soit optimale pour toutes les tâches d'indexation vidéo. D'autre part, la définition de la structure temporelle pour chaque problème n'est pas une tâche évidente. L'apprentissage de structure peut fournir alors une solution.

6.3 Apprentissage de structure dans les réseaux bayésiens dynamiques

Dans cette partie, nous étudions l'apport de l'apprentissage de la structure temporelle. Cet apprentissage permet de mettre en évidence les différentes corrélations qui existent entre les attributs temporels, tout en gardant une structure temporelle proche de la réalité.

Nous appliquons l'apprentissage de structure dans un réseau bayésien dynamique d'une manière similaire à l'apprentissage de structure dans un réseau bayésien statique. Un algorithme d'apprentissage de structure dynamique se base donc sur un score et une méthode de parcours de l'ensemble des structures.

6.3.1 Approche utilisant le score BIC

Le problème de l'apprentissage de structure dans un réseau bayésien dynamique a été abordé dans [80]. À notre connaissance, c'est l'unique travail consacré à ce problème dans

la littérature. Dans ce travail, l'apprentissage est utilisé dans le cadre d'applications de prédiction du comportement des voitures sur une route. Dans ce travail, l'apprentissage de structure de réseaux bayésiens dynamiques est basé sur un score génératif de type BIC. Ce score se décompose en deux scores, un pour la structure du réseau $\mathcal{B}$ et un score pour le réseau $\mathcal{B}_{trans}$ (*cf.* équation 6.2)

$$\mathscr{S}_{BIC}(\mathcal{B}) = \mathscr{S}_{BIC}(\mathcal{B}_0) + \mathscr{S}_{BIC}(\mathcal{B}_{trans}) \tag{6.2}$$

Le score ainsi obtenu est un score décomposable qui garde les mêmes propriétés qu'un score BIC d'un réseau statique. Ce score peut être utilisé dans le cadre d'une recherche de structure optimale en utilisant un algorithme K2 ou un algorithme de recherche gloutonne sur les deux types de réseaux composant le réseau bayésien dynamique.

6.3.2 Approche augmentée d'apprentissage de structure

La méthode présentée dans le paragraphe précédent est une méthode qui favorise la modélisation des données sans se soucier de l'aspect classification. Dans ce paragraphe, nous nous sommes inspirés de la méthode d'apprentissage de structure par enrichissement présentée en 4.3.2 pour construire un algorithme d'apprentissage de structure pour les réseaux bayésiens dynamiques, permettant de répondre de façon plus adaptée à notre problème de classification.

Nous utilisons dans cet algorithme le score BIC modifié (*cf.* équation 4.4). Nous choisissons ainsi pour chaque nœud de la structure *a priori* et chaque nœud de la structure de transition l'ensemble de nœuds parents qui augmentent le plus le score BIC modifié de la structure correspondante. Le score final est ensuite la somme de tous les scores.

6.3.3 Approche discriminante

Dans cette partie, nous utilisons un score discriminant pour l'apprentissage de la structure du réseau bayésien dynamique. Comme nous l'avons déjà précisé dans le chapitre 4, l'utilisation d'un score discriminant introduit une difficulté supplémentaire par rapport à un score génératif. Ceci dans le sens où le score discriminant n'est pas un score décomposable. On ne peut donc pas faire une recherche locale des connexions optimales et utiliser le réseau résultant.

Nous utilisons, dans cette partie, la vraisemblance conditionnelle (*cf.* equation 4.6) comme score discriminant. Une étape d'inférence sur les données d'apprentissage est nécessaire pour calculer le score de chaque structure. Afin de diminuer le nombre de structures candidates, et donc diminuer le nombre d'étapes d'inférence, nous utilisons une recherche gloutonne itérative. Nous décomposons la structure du réseau bayésien en plusieurs sous structures. Nous générons les structures qui correspondent à la perturbation de ces sous-structures par suppression, inversion ou ajout d'un arc. Nous cherchons ensuite la sous-structure qui donne la meilleure CLL, elle sera ensuite insérée dans la structure globale du réseau.

6.4 Résultats expérimentaux

Dans cette partie expérimentale, nous testons l'apprentissage de structure dans les réseaux bayésiens dynamiques sur la base de données de l'application détection d'*Action* dans un contenu de football.

6.4.1 Comparaison des différentes approches d'apprentissage de structure dynamique

Nous comparons dans la figure 6.3 les résultats donnés par les approches d'apprentissage présentées plus haut dans ce chapitre.

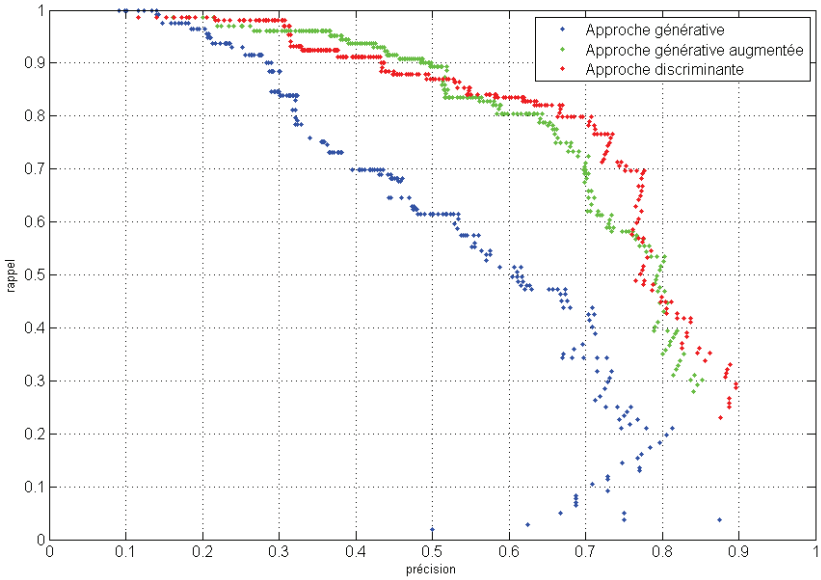

FIG. 6.3 – Résultats des trois approches d'apprentissage de structure dans les réseaux bayésiens dynamiques.

La comparaison entre les trois approches d'apprentissage de structure dans un réseau bayésien dynamique montre que l'approche générative s'avère encore une fois inefficace pour résoudre le problème de détection d'événements.

La recherche de la structure dans un ensemble qui favorise la classification par rapport au pouvoir génératif, tel que c'est fait dans l'approche enrichie, améliore nettement les performances du système. Toutefois, cette méthode possède un inconvénient majeur. Elle se restreint en effet à la recherche des structures dans l'ensemble où le nœud événement est connecté à tous les attributs. Ainsi tous les attributs sont pris en compte sans distinction. L'approche discriminante n'est pas concernée par cet inconvénient de l'approche générative enrichie. Elle n'utilise, en effet, que les attributs qui sont pertinents

pour la classification. Par exemple dans la structure récupérée dans notre exemple, l'attribut position sur le terrain n'est connecté au nœud événement qu'à l'instant courant. Les autres connexions n'ajoutent en effet rien au pouvoir de classification. Elles ne sont donc pas prises en considération.

6.4.2 Réseau dynamique versus réseau statique

La comparaison des résultats du réseau bayésien dynamique et ceux issues du réseau bayésien statique (*cf.* figure 6.4), montre que l'utilisation du réseau dynamique n'entraîne pas un changement significatif au niveau des résultats de classification. Il est toutefois important de noter que dans le cadre de notre problème de détection d'*Action* dans un match de football, l'influence des attributs se limite au voisinage de l'événement. Ainsi, en prenant en compte les attributs du voisinage de l'événement comme nous l'avons fait dans le cadre des réseaux statiques, nous sommes parvenus à exploiter toute l'information présente dans les données. L'information provenant de tranches temporelles distantes et qui circule à travers les réseaux bayésiens n'est donc pas importante dans le cadre de notre exemple.

Toutefois, avec des résultats comparables aux réseaux bayésiens statiques, les réseaux bayésiens dynamiques génèrent des structures plus compréhensibles, vu qu'ils respectent la causalité temporelle inhérente à la vidéo. Un autre avantage des réseaux bayésiens dynamiques par rapport aux réseaux bayésiens statiques est que l'organisation spécifique des données des réseaux dynamiques permet une incorporation plus facile de connaissances *a priori* tel qu'un modèle de durée par exemple tel que celui utilisé au chapitre 3.

6.5 Conclusion

Nous avons étudié dans ce chapitre l'apport de l'apprentissage de structure dans les réseaux bayésiens dynamiques. Nous avons mis en place deux techniques d'apprentissage de structure. La première technique se base sur l'enrichissement de la structure naïve. La deuxième se base sur l'utilisation d'un score discriminant. Nos expériences ont montré que cette dernière approche donne de meilleurs résultats tout en permettant de sélectionner les connexions les plus pertinentes. D'autre part, nous montrons que l'apprentissage de structure permet d'avoir des résultats équivalents aux résultats de la structure du réseau statique. Toutefois, la structure du réseau bayésien dynamique a l'avantage de pouvoir mieux incorporer des connaissances *a priori* vu qu'elle modélise le caractère temporel de la vidéo.

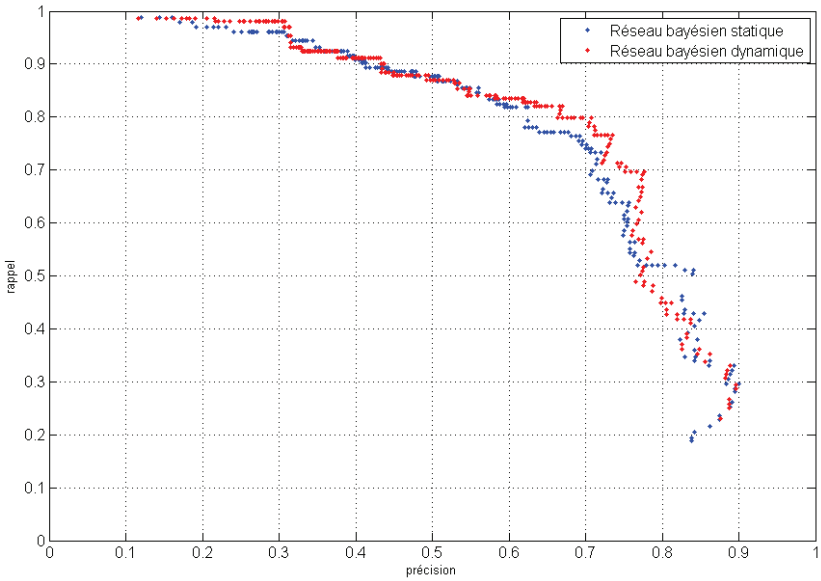

FIG. 6.4 – Effet de l'apprentissage de structure sur les résultats des réseaux bayésiens statiques et des réseaux bayésiens dynamiques.

Chapitre 7

Conclusion et perspectives

Nous nous sommes intéressés au cours de ce travail au problème de la détection d'événements dans un contenu vidéo. Nous avons proposé un cadre statistique générique pour la détection d'événements que nous avons appliqués à la détection de *plages publicitaires* et d'*actions* dans un match de football. Dans cette conclusion, nous proposons de dresser un bilan des différents résultats obtenus dans cette thèse. Nous présentons ensuite différentes perspectives que nous trouvons pertinentes comme suite pour ce travail.

7.1 Synthèse des résultats

Le problème de description de contenus multimédias est un problème de plus en plus étudié dans la littérature. De même dans le milieu industriel des solutions à ce problème sont de plus en plus demandées, et avec des contraintes supplémentaires inhérentes au monde industriel, qui sont la recherche de l'automatisation du processus. Or cette automatisation ne va pas sans une certaine généricité du système au regard des différents contenus et applications.

Les approches probabilistes et tout particulièrement les réseaux bayésiens offrent un grand potentiel de modélisation des données issues de la vidéo. Ils offrent également des propriétés intéressantes peu étudiées jusqu'à présent et qui conduisent à plus de généricité dans la modélisation d'un tel système. C'est pour cette raison que nous avons, dans ce travail, proposé une approche à base de réseau bayésiens. Comme conclusion de nos travaux, nous avons en effet abouti à différents aspects de généricité des modèles bayésiens que nous rappelons à présent.

Généricité par rapport à la construction de la structure Les systèmes utilisant les réseaux bayésiens sont généralement limités par la nécessité de fournir une structure du modèle de réseau bayésien à utiliser. Cette structure est construite à partir de connaissances expertes, qui sont par ailleurs rarement disponibles, ou, à défaut à partir d'une structure que le concepteur juge proche du but de son application. Ceci demande donc une intervention humaine lors de la construction du modèle. D'autre part, on n'a aucune garantie que la structure utilisée est la plus adaptée pour le problème visé.

Nous avons donc étudié l'apprentissage structure comme moyen permettant d'automatiser la construction du réseau bayésien. Nous avons, dans un premier temps, démontré la validité de notre approche dans le cadre de l'application de détection de plages publicitaires. Nous avons montré que l'apprentissage de structure permet la construction automatique du modèle. Le modèle ainsi obtenu aboutit à une fusion automatique des attributs. L'apprentissage de structure permet donc de s'affranchir de l'étape de définition des différentes corrélations entre attributs, nécessitant souvent une connaissance poussée du système et aussi une connaissance de la mise œuvre des réseaux bayésiens. Il n'est toutefois pas exclu d'ajouter une quelconque information *a priori* disponible. Nous avons ainsi pu introduire l'information de durée en ajoutant un réseau bayésien en parallèle au réseau de données permettant, ainsi, de simuler l'information de durée des plages publicitaires.

Toujours dans un souci de généricité, nous avons été amenés à étudier l'influence du critère de score sur les résultats de la détection d'événements. Nous avons, en effet, démontré dans le cadre de l'application détection d'*action* dans un match de football, les limites de l'apprentissage de structure en utilisant un score génératif. Ce dernier favorise en effet les structures les plus vraisemblantes vis-à-vis des données. Or plus le nombre de variables augmente plus le pouvoir génératif et le caractère discriminant divergent. Les arcs qui représentent les corrélations entre les variables permettant de discriminer la variable événement sont souvent ignorés au profit de corrélations plus fortes au sens de la vraisemblance par rapport aux données. Nous avons alors introduit deux types d'approches pour distinguer le nœud de classification dans le processus d'apprentissage de structure. La première méthode consiste à considérer que le nœud de classification est un parent commun aux nœuds attributs. Ceci conduit à une structure de réseau naïf augmentée. La seconde manière de procéder consiste à prendre en compte le pouvoir de classification du réseau bayésien résultant au niveau du critère d'apprentissage de structure. D'après nos résultats présentés dans le chapitre 4, ces deux méthodes se sont avérées efficaces pour augmenter considérablement les performances de détection du système. Cet apprentissage de structure permet donc une intégration automatique des attributs avec comme objectif principal l'augmentation des performances de classification.

Génericité grâce à la sélection d'attributs Nous avons également étudié l'effet de la sélection d'attributs comme deuxième volet vers l'automatisation du système de détection. Nous avons montré que la sélection d'attributs représente un complément pour l'automatisation et donc de la généricité des systèmes de description de contenus multimédias. Elle permet, en effet, de réduire le nombre de paramètres des réseaux bayésiens utilisés, ce qui résulte en un apprentissage plus fiable, tout particulièrement dans le cadre de la détection d'événements, où la base d'apprentissage comporte généralement un faible nombre d'exemples. La sélection d'attributs permet aussi de rendre la structure résultante plus visible vu le nombre réduit de nœuds utilisés. Dans le cadre de nos expériences, nous avons ainsi montré que la sélection d'attributs permet d'améliorer les performances de l'apprentissage de structure pour les structures naïves augmentées qui utilisent par construction tous les attributs du problème. Les approches généra-

tives sont également sensibles à la sélection d'attributs. L'association apprentissage de structure génératif et sélection d'attributs basée sur les « wrappers » permet d'augmenter les performances de classification, dans la mesure où seuls les attributs pertinents pour la tâche de classification sont utilisés dans l'apprentissage. En ce qui concerne l'apprentissage de structure par un critère discriminant, nous avons montré que cette approche incluait déjà une étape de sélection d'attributs. Dans ce contexte, nous avons abouti à la conclusion que pour ce cas particulier la sélection d'attributs n'apportait pas d'amélioration de performances au final.

Généricité grâce à la modélisation temporelle Dans une dernière étape, nous avons étudié l'apport de l'apprentissage de structure pour un réseau bayésien dynamique. Nous avons montré que, même pour ce type d'approche qui permet de modéliser le caractère temporel des données, un apprentissage tenant compte de la classification est nécessaire. L'avantage de faire de l'apprentissage de structure sur un réseau bayésien dynamique par rapport à un réseau statique est d'aboutir à une structure plus lisible. Il est donc plus facile d'ajouter des connaissances externes telles que les connaissances sur la durée des événements par exemples.

Généricité par rapport aux domaines d'application Nous avons testé notre système dans le cadre de la détection de publicités et la détection d'*actions* dans les matchs de football. Nous avons travaillé à l'automatisation du processus de construction du modèle et à l'introduction du minimum de connaissances *a priori*. Notre méthode reste valable pour de nombreux types d'applications de détection d'événements. Il suffit de disposer pour cela d'un ensemble d'attributs et aussi d'une base d'exemples. L'association entre sélection d'attributs et apprentissage de structure permet d'utiliser les attributs pertinents pour la détection et de construire automatiquement les différentes corrélations entre les composantes du système. Ceci permet de fournir le modèle de détection d'événements souhaité. Il est donc envisageable de l'utiliser pour la détection d'autres types d'événements dans un match de football ou d'autres types de contenus sportifs. On pourrait également l'employer pour des applications de détections d'événements très différents comme par exemple la détection d'événements dans les films par exemple, pour une application de contrôle parental pour la détection de scènes violentes.

7.2 Travaux futurs

Ce travail ouvre sur de nombreuses perspectives. D'un point de vue purement applicatif, on peut aisément imaginer améliorer les performances du système grâce à l'ajout de nouveaux attributs. On peut, par exemple, détecter un certain nombre de mots clés dans le flux de parole du présentateur tels que les mots : *but, action, gardien, raté*. On peut également ajouter l'information sur la position du ballon sur le terrain. Il est, en effet, toujours intéressant de rajouter des attributs dont nous savons déjà qu'ils vont être utiles pour la tâche de détection d'événements. Cela constitue une source d'information *a priori* que l'on peut ajouter au système, et qui ne nuit en aucun cas à sa généricité.

Une autre connaissance *a priori* que l'on peut ajouter facilement au réseau bayésien dynamique construit dans le chapitre 6 est un modèle sur la durée des *actions*, durée qui reste relativement similaire d'une action à une autre. Cette connaissance doit contribuer également à l'amélioration des performances de notre système pour la détection des actions.

Toujours sur le plan applicatif, il serait intéressant de continuer à valider notre démarche et de tester la généricité de notre approche sur d'autres applications de détections d'événements telles que celles que nous avons proposées dans la section précédente.

Sur le plan plus théorique de l'étude de l'apprentissage de structure, il serait intéressant de tester l'approche discriminante sur les modèles de type Multinets, pour lesquels, par manque de temps, nous n'avons pu tester que l'apprentissage génératif. À la différence de l'apprentissage avec le score génératif où chaque modèle est appris d'une manière indépendante, nous ne pouvons pas envisager cette manière de faire dans le cas du modèle discriminant. En effet, dans le calcul de la probabilité *a posteriori* nécessaire au calcul du score discriminant, les deux modèles sont mis à contribution. La recherche gloutonne devra donc dans ce cas se faire d'une manière conjointe sur les deux modèles de la classe et de sa négation.

D'autre part, dans les systèmes vidéos, les corrélations entre les attributs et les événements peuvent être expliqués par les relations de cause à effet qui existent entre les deux types de variables. Elles peuvent être également expliquées par l'existence d'autres concepts qui font office de couche de liaison entre les attributs et les événements. Une perspective intéressante pour ce travail sera donc de tenir compte de la présence de ce genre de concepts dans l'apprentissage de structure. Une première manière de traiter ce problème consiste à identifier manuellement les concepts et à établir une base de données les incluant. Ils seraient alors traités dans la phase d'apprentissage comme le reste des variables du système. La deuxième manière de traiter le problème, qui présente à notre avis un plus grand défi scientifique, serait de découvrir les concepts recherchés au niveau de la phase d'apprentissage de structure. Découvrir de tels concepts, en permettant une meilleure modélisation de certains systèmes, serait un pas supplémentaire vers plus de généricité, puisque des modèles plus complexes pourrait être pris en compte et faire l'objet d'apprentissage.

Une autre perspective intéressante de notre travail et que nous n'avons pas pu traiter dans le cadre de ce travail est la problématique des données asynchrones, problématique qui résulte bien souvent de la multimodalité des contenus. Nous avons considéré dans le cadre de ce travail que les données étaient complètement synchrones. Elles sont toutes prises à la fréquence du plan. Il serait donc intéressant de traiter le problème où les données sont asynchrones. Nous pouvons envisager dans un premier temps une combinaison des réseaux bayésiens avec les modèles de segments pour résoudre le problème d'asynchronie. L'apprentissage de structure reste au niveau des réseaux bayésiens, le modèle de segments permettant de traiter le problème de l'asychronie des données.

Et enfin comme dernière perspective, nous proposons de poursuivre notre étude des réseaux bayésiens dynamiques qui reste succincte. Nous pourrions même aller plus loin. En effet, l'utilisation des réseaux bayésiens dynamiques impose une observation continue de données vidéos et une connaissance précise de l'état de la vidéo à chaque instant.

Toutefois, cette condition n'est pas toujours nécessaire dans le cadre de la description de contenus vidéos et plus précisément dans le cadre de la détection d'événements. En effet, dans ce type d'applications, l'intérêt porte essentiellement sur une partie de la vidéo et pas sur la totalité. Un modèle intéressant, est le modèle basé sur les réseaux bayésiens temporels [81]. Dans ce type de modèles, un nœud est associé à chaque variable. La valeur de la variable n'est plus l'occurrence ou pas d'un événement, mais son temps d'occurrence. Cette manière de faire permet de modéliser des corrélations temporelles à long terme ce qui était difficile dans le cadre des réseaux bayésiens classiques. Cela suppose, en effet, l'utilisation d'un réseau de taille importante qui pose généralement des problèmes de calcul. D'autre part, avec ce type de modèle temporel, le problème de synchronisation ne se pose plus. En effet, il n'y a plus besoin d'échantillonnage pour connaître l'état du système à un instant donné, mais il s'agit plutôt de connaître les instants d'occurrence des événements les uns par rapport aux autres. Ces nouveaux réseaux bayésiens temporels mériteraient donc qu'on s'y attarde.

Avec cette liste, bien sûr non exhaustive de perspectives nouvelles, nul doute qu'il reste à faire dans l'étude des réseaux bayésiens.

Application « détection de plage publicitaire »

L'objectif de cette application est la détection de plages publicitaires dans un flux vidéo télévisé. Nous considérons comme plages publicitaires, les annonces pour les produits commerciaux, ainsi que les annonces pour les programmes télévisés.

Base de données Nous disposons d'une base de données composée de 36 heures de vidéo enregistrées à partir de trois chaînes nationales françaises :

- 24 heures de France2,
- 4 heures de TF1,
- 8 heures de M6.

Les vidéos sont prises à des plages horaires différentes.

Une décomposition en plan est faite sur la vidéo. Pour chaque plan, on extrait une image clé, considérée comme l'image la plus représentative de la vidéo. On extrait cinq attributs visuels et un attribut audio. Nous détaillons dans ce qui suit ces différents attributs.

Longueur de plan : Pour rendre les plages publicitaires plus attirantes, les concepteurs de ces dernières y introduisent une grande activité sous forme de couleurs attirantes, de mouvements rapides, ou de changements rapides de plan. L'un des attributs permettant de mesurer cette activité est donc la longueur des plans vidéos.

Surface et nombre de blocs de texte : Pour chaque plan, on cherche à détecter la présence de zones de texte dans son image clé. On cherche dans une première étape les zones à fort gradient et dont l'orientation est susceptible de correspondre à du texte. Des opérations morphologiques sont ensuite effectuées pour faire ressortir ressortir les zones de l'image susceptibles de constituer une suite de lettres. La sortie du détecteur est un ensemble de boîtes englobantes contenant les éventuelles zones de texte. Les deux attributs basés sur le texte que nous considérons sont la surface globale des blocs de texte ainsi que leur nombre.

Intensité de mouvement : In autre moyen de caractériser l'activité dans un plan donné consiste à mesurer la quantité de mouvement dans ce plan. Ainsi, nous mesurons

pour chaque image du plan, le vecteur mouvement associé à l'image. Nous mesurons en suite l'intensité moyenne du vecteur mouvement sur les différentes images du plan.

Similarité visuelle : Les plages publicitaires présentent généralement une grande diversité visuelle. Nous utilisons donc un descripteur renseignant sur la similarité visuelle du plan courant par rapport aux plans voisins. La similarité visuelle entre deux plans est mesurée à travers la distance entre les histogrammes couleurs des images clés des deux plans considérés. Nous considérons ensuite la distance moyenne du plan courant par rapport aux deux plans qui le précédent et aux deux plans qui le suivent.

Audio : Nous utilisons dans le cadre de notre travail le LSTER (*Low short time energy ratio*) [82]. Cet attribut permet de mesurer la variation de l'énergie audio à court terme sur un segment audio donné. Les auteurs de [82] montrent que cet attribut renseigne sur la présence de la parole dans le flux audio. Nous considérons une moyenne de cet attribut sur la durée d'un plan donné.

Application « détection d'événement dans un match de football »

Nous nous intéressons dans cette application à la détection d'*actions* dans un match de football.

Action : Une *action* est un moment du jeu où une équipe risque de marquer un but. La détection de ce type d'événements présente un grand intérêt industriel. En effet, les actions sont parmi les événements qui intéressent le plus les spectateurs. Elles constituent, alors, une grande proportion du résumé d'un match donné. D'autre part, détecter les actions permet généralement de détecter les buts en même temps. Il ne reste donc qu'à filtrer ceux-ci parmi les actions. Un autre intérêt de la détection des *actions* est qu'elle permet de renseigner sur le niveau offensif du match.

Succession des plans dans un segment vidéo d'une *action*.

L'occurrence d'une action a généralement une incidence sur le contenu que le réalisateur décide de montrer dans les minutes qui suivent l'*action*. Par exemple comme l'indique la figure 7.2, lors d'une *action*, le plan est généralement un plan large contenant la cage de but. Il est accompagné par une grande excitation de la foule et du commentateur du match. Le plan suivant l'*action* est généralement centré sur le visage de la personne qui a mené l'*action*. Enfin, pour accentuer l'*action*, le réalisateur insère souvent des plans de ralenti montrant les différentes phases de l'*action*.

Base de données Notre base de données est constituée de sept matchs enregistrés lors de la coupe du monde 2006. Cette base correspond à un volume horaire de 14 heures. Nous utilisons 8 attributs, 7 attributs visuels et un attribut audio. Nous présentons dans

ce qui suit une description de ces attributs.

Plan large Ce type de plans montre une vue d'ensemble du terrain. Ce plan inclus une grande partie du terrain ainsi qu'une partie des gradins du stade. Pour détecter ce type de plan, on cherche si le masque de la couleur dominante dans l'image clé du plan présente une zone connexe de taille plus grande qu'un seuil donné.

Couleur verte : Cet attribut indique la présence de la couleur verte du terrain. Cet attribut permet de renseigner sur la présence d'une partie du terrain dans l'image.

Zone de jeu régulière : Les matchs de football présentent généralement de longues phases de jeu souvent monotones où les joueurs ne font qu'échanger le ballon. Le réalisateur utilise alors, à ce moment du jeu, une succession de plan large/plan autre, afin de rompre avec la monotonie du jeu. L'attribut que nous utilisons résulte de la détection de cette succession de plans en se basant sur un ensemble de règles.

Attribut visage : Pour chaque plan , nous cherchons l'existence d'un visage donc la surface dépasse un certain seuil. Cet attribut permet de renseigner sur la présence d'un gros plan sur un joueur donné.

Plan de ralenti : Le réalisateur insère fréquemment des plans de ralenti afin de rappeler l'événement qui vient de se passer. Ces plans de ralenti sont généralement délimités par deux plans de transitions caractéristiques.

Plan de transition : Lors des phases de jeu où il y a un événement marquant, le réalisateur insère souvent des effets de transition et d'animation pour augmenter l'attractivité de l'événement. Notre détecteur de plan permet de récupérer ces transition au fur et mesure que la découpe en plans est faite.

Position de la zone de but : Cet attribut indique la présence ou pas de la zone de but dans l'image. Pour chaque plan détecté comme plan large, on extrait le masque de la couleur dominante, celle du terrain. Nous cherchons, alors, sur cette image masque, les éventuelles lignes en utilisant une transformée de Hough. Le plan est considéré comme proche de la zone de but si plus de trois lignes parallèles sont détectées.

Acclamation de la foule : Ce descripteur indique les plans qui sont susceptible de contenir une acclamation de la foule renseignant sur l'occurrence d'un événement marquant. Nous utilisons pour avoir cet attribut, les outils Spro et Audioseg développés par G. Gravier.

Bibliographie

[1] Ewa Kijak. *Structuration multimodale des vidéos de sports par modèles stochastiques.* Thèse de doctorat, Université de Rennes 1, 2003.

[2] Manolis Delakis. *Multimodal Tennis Video Structure Analysis with Segment Models.* Thèse de doctorat, Université de Rennes 1, 2006.

[3] Cees G. M. Snoek and Marcel Worring. Multimodal video indexing : A review of the state-of-the-art. *Multimedia Tools and Applications*, 25(1) :5–35, janvier 2005.

[4] Rainer Lienhart, Wolfgang Effelsberg, Stephan Fischer. Automatic recognition of film genres. In *ACM Multimedia*, pages 295–304, San Francisco, United States, 1995.

[5] H. L. Wang , J. Huang, Z. Liu, Y. Wang, Y. Chen, E. K. Wong. Integration of multimodal features for video scene classification based on HMM. In *IEEE Workshop on Multimedia Signal Processing*, pages 53–58, 1999.

[6] R. S. Jasinschi, N. Dimitrova, T. Mcgee, L. Agnihotri, J. Zimmerman, D. Li Philips, and D. Li. Integrated multimedia processing for topic segmentation and classification. In *IEEE International Conference on Image Processing*, volume 3, pages 366–369, Greece, 2001.

[7] Emile Sahouria, Avideh Zakhor. Content analysis of video using principal components. In *IEEE Transactions on Circuits and Systems for Video Technology*, volume 9, pages 1290–1298, 1998.

[8] G.Gravier. *Description multimodale des documents multimédia*, chapter 7. L'indexation multimédia : description et recherche automatiques. lavoisier, second edition, may 2007.

[9] Cees G. M. Snoek, Marcel Worring, and Arnold W. M. Smeulders. Early versus late fusion in semantic video analysis. In *Proceedings of ACM international conference on Multimedia*, pages 399–402, Singapore, 2005.

[10] Kieron Messer, William Christmas, Josef Kittler. Automatic sports classification. In *International Conference on Pattern Recognition*, volume 2, pages 1005 – 1008, Los Alamitos, USA, 2002. IEEE Computer Society.

[11] H. Lu and Y-P Tan. Sports video analysis and structuring. *IEEE Workshop on Multimedia Signal Processing*, October 2001.

[12] E. Kijak, G. Gravier, L. Oise, P. Gros. Audiovisual integration for tennis broadcast structuring. *Multimedia Tools and Application*, 30(3) :289–311, 2006.

[13] Hong-Jiang Zhang , Shuang Yeo Tan, Stephen W. Smoliar,Gong Yihong. Automatic parsing and indexing of news video. *Multimedia Systems*, 2(6) :256–266, 1995.

[14] A. Ekin, M. Tekalp. Generic play-break event detection for summarization and hierarchical sports video analysis. In *ICME '03 : Proceedings of the 2003 International Conference on Multimedia and Expo*, pages 169–172, Washington, DC, USA, 2003.

[15] Lexing Xie, Peng Xu, Shih-Fu Chang, Ajay Divakaran, Huifang Sun. Structure analysis of soccer video with domain knowledge and hidden Markov models. *Pattern Recognition Letters*, 25(7) :767–775, 2004.

[16] C. Saraceno, R. Leonardi. Identification of story units in audio-visual sequences by joint audio and video processing. In *International Conference on Image Processing*, volume 1, pages 363–367, Chicago, USA, 1998.

[17] Hong-Jiang Zhang Xian-Sheng Hua, Lie Lu. Robust learning-based TV commercial detection. In *IEEE International Conference on Multimedia and Expo*, pages 4 –7, 2005.

[18] M. Mizutani, S. Ebadollahi, S. Chang. Commercial detection in heterogeneous video streams using fused multi-modal and temporal features. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages 157–160, Philadelphia, USA, 2005.

[19] Xingquan Zhu, Xindong Wu, Ahmed K. Elmagarmid, Zhe Feng, and Lide Wu. Video data mining : semantic indexing and event detection from the association perspective. *IEEE Transactions on Knowledge and Data Engineering*, 17(5) :665–677, 2005.

[20] Xiaofeng Tong, Qingshan Liu, Hangging Lu. Semantic units based events detection in soccer videos. In *International Conference on Image Processing*, volume 3, pages 1621– 1624, Singapore, 2004.

[21] Ming-Chun Tien, Yi-Tang Wang, Chen-Wei Chou, Kuei-Yi Hsieh, Wei-Ta Chu, Ja-Ling Wu. Event detection in tennis matches based on video data mining. In *IEEE International Conference on Multimedia and Expo*, pages 1477–1480, Hanover, Germany, 2008.

[22] Ramesh Jain, Arun Hampapur. Metadata in video databases. *SIGMOD Record*, 23(4) :27–33, 1994.

[23] Vasanth Tovinkere and Richard J. Qian. Detecting semantic events in soccer games : Towards a complete solution. *IEEE International Conference on Multimedia and Expo,*, 0 :212, 2001.

[24] Rainer Lienhart, Silvia Pfeiffer, Wolfgang Effelsberg. Scene determination based on video and audio features. In *Multimedia Tools and Applications*, pages 685–690, 1998.

[25] Jeho Nam, Masoud Alghoniemy, and Ahmed H. Tewfik. Audio-visual content-based violent scene characterization. In *IEEE International Conference on Image Processing*, pages 353–357, 1998.

[26] Di Zhong, Shih-fu Chang. Structure analysis of sports video using domain models. *IEEE ICME*, pages 22–25, 2001.

[27] Surya Nepal, Uma Srinivasan, Graham Reynolds. Automatic detection of 'goal' segments in basketball videos. In *MULTIMEDIA '01 : Proceedings of the ninth ACM international conference on Multimedia*, pages 261–269, Ottawa, Canada, 2001. ACM.

[28] Leonid I. Perlovsky. Conundrum of combinatorial complexity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(6) :666–670, 1998.

[29] Yihong Gong, Mei Han, Wei Hua, and Wei Xu. Maximum entropy model-based baseball highlight detection and classification. *Computer Vision and Image Understanding*, 96(2) :181–199, 2004.

[30] Tae Meon Bae, Cheon Seog Kim, Sung Ho Jin, Ki Hyun Kim, and Yong Man Ro. Semantic event detection in structured video using hybrid HMM/SVM. In *international conference on image and video retrieval*, pages 113–122, Singapore, 2005.

[31] Cees G. M. Snoek and Marcel Worring. Multimedia event-based video indexing using time intervals. *IEEE Transactions on Multimedia*, 7(4) :638– 647, 2005.

[32] James F. Allen. Maintaining knowledge about temporal intervals. *Communication of the ACM*, 26(11) :832–843, 1983.

[33] Niels Haering, Richard J. Qian, M. Ibrahim Sezan. A semantic event-detection approach and its application to detecting hunts in wildlife video. *IEEE Transactions on Circuits and Systems for Video Technology*, 10(6) :857–868, 2000.

[34] Manolis Delakis, Guillaume Gravier, and Patrick Gros. Audiovisual integration with segment models for tennis video parsing. *Computer Vision and Image Understanding*, 111(2) :142–154, 2008.

[35] B. Li and I. Sezan. Semantic sports video analysis : approaches and new applications. In *on Proceedings International Image Processing Conference*, volume 1, pages 17–20, 2003.

[36] A. Aydin Alatan, Ali N. Akansu, and Wayne Wolf. Multi-modal dialog scene detection using hidden markov models for content-based multimedia indexing. *Multimedia Tools Applications*, 14(2) :137–151, 2001.

[37] Nevenka Dimitrova, Lalitha Agnihotri, and Gang Wei. Video classification based on HMM using text and faces. In *European Signal Processing Conference*, 2000.

[38] Stefan Eickeler, Stefan Müller, and Stefan M Uller. Content-based video indexing of TV broadcast news using hidden markov models. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 2997–3000, Phoenix, USA, 1999.

[39] G. Potamianos, C. Neti, G. Gravier, A. Garg, and A. W. Senior. Recent advances in the automatic recognition of audiovisual speech. *Proceedings of the IEEE*, 91(9) :1306–1326, 2003.

[40] N. Friedman. Inferring cellular networks using probabilistic graphical models. *Science*, 303(5659) :799–805, February 2004.

[41] Jan Lunze. Fault diagnosis of discretely controlled continuous systems by means of discrete-event models. *Revue Discrete Event Dynamic Systems*, 18(2) :181–210, 2008.

[42] A. V. Werhli, M. Grzegorczyk, D. Husmeier. Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical gaussian models and bayesian networks. *Bioinformatics*, July 2006.

[43] Kevin P. Murphy. *Dynamic Bayesian Networks : Representation, Inference and Learning.* Thèse de doctorat, UC Berkeley, Computer Science Division, 2002.

[44] Ziyou Xiong. Audio-visual sports highlights extraction using coupled hidden markov models. *Pattern Analysis and Application*, 8(1) :62–71, 2005.

[45] Milind R. Naphade, Igor Kozintsev, Thomas Huang. Probabilistic semantic video indexing. In *Proceedings of Neural Information Processing Systems*, pages 967–973, 2000.

[46] Milind R. Naphade, Thomas S. Huang. Semantic video indexing using a probabilistic framework. *International Conference on Pattern Recognition,*, 3 :3083, 2000.

[47] Eric Horvitz Nuria Oliver. A comparison of HMMs and dynamic Bayesian networks for recognizing office activities. *lecture notes in computer science*, 3538 :199–209, 2005.

[48] Fei Wang, Yu-Fei Ma, Hong-Jiang Zhang, Jin-Tao Li. Dynamic Bayesian network based event detection for soccer highlight extraction. In *International Conference on Image Processing*, pages 633–636, singapore, 2004.

[49] M. Petkovic, V. Mihajlovic, W. Jonker, and S. Djordjevic-Kajan. Multi-modal extraction of highlights from TV formula 1 programs. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, pages 817–820, 2002.

[50] Finn Verner Jensen. *An introduction to Bayesian Networks.* UCL Press, 1996. ISBN : 1-85728-332-5.

[51] Ross D. Shachter. Bayes-ball : The rational pastime for determining irrelevance and requisite information in belief networks and influence diagrams. In *Uncertainty in Artificial Intelligence*, 1998.

[52] T.S. Verma, J. Pearl. Causal networks : Semantics and expressiveness. In *Uncertainty in Artificial Intelligence*, pages 352–359, 1988.

[53] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference.* Morgan Kaufmann, Santa Mateo, CA, USA, September 1988.

[54] Nevin L. Zhang and David Poole. Exploiting causal independence in bayesian network inference. *Journal of Artificial Intelligence Research*, 5 :301–328, 1996.

[55] F. V. Jensen, S. L. Lauritzen, K. G. Olesen. Bayesian updating in causal probabilistic networks by local computations. *Computational Statistics Quaterly*, 4(5) :269–282, 1990.

[56] J. H. Kim and J. Pearl. A computational model for combined causal and diagnostic reasoning in inference systems. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, Karlsruhe, Allemagne, 1983.

[57] S. L. Lauritzen, D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Readings in uncertain reasoning.*

[58] G. F. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42(2-3) :393–405, 1990.

[59] Uffe Kjaerulff. Approximation of bayesian networks through edge removals. Technical report, Aalborg University, 1993.

[60] M. Henrion. Some practical issues in constructing belief networks. In *Proceedings of the 3rd Annual Conference on Uncertainty in Artificial Intelligence (UAI-87)*, pages 161–174, New York, NY, 1987.

[61] Patrick Naim, Pierre-Henri Wuillemin, Philippe Leray, and Olivier Pourret. *Réseaux bayésiens.* Eyrolles, 3 edition, 11 2007. isbn : 978-2-212-11972-5.

[62] David Heckerman. A tutorial on learning with bayesian networks. Technical report, Learning in Graphical Models, 1996.

[63] P. langley, W. Iba, K. Thompson. An analysis od bayesian classifiers. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, pages 223–228, San Jose, CA, 1992.

[64] J.Pearl, T.S. Verma. A statistical semantics for causation (corr : 93v3 p59). *Statistics and Computing*, 2(2) :91–95, 1992.

[65] P. Spirtes, C. Glymour, R. Scheines. *Causation, Predection and Search.* 1993.

[66] H. Akaike. Statistical predictor identification. *Ann. Inst. Stat. Math.*, 22 :203–217, 1970.

[67] Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2) :461–464, 1978.

[68] C. K. Chow and C. N. Liu. Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, IT-14(3) :462–467, May 1968.

[69] Jeff Bilmes. Dynamic bayesian multinets. In *UAI '00 : Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, pages 38–45, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc.

[70] R. W. Robinson. Counting unlabeled acyclic digraphs. *Combinatorial Mathematics V*, 622 :28–43, 1977.

[71] J. B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. In *Proceedings of the American Mathematical Society 7*, pages 48–50, 1956.

[72] Gregory F. Cooper and Tom Dietterich. A bayesian method for the induction of probabilistic networks from data. In *Machine Learning*, pages 309–347, 1992.

[73] D. Chickering, D. Geiger, D. Heckerman. Learning bayesian networks : Search methods and experimental results. In *Proceedings of the fifth Conference on Artificial Intelligence and Statistics*, pages 112–128, 1995.

[74] N. Friedman, D. Geiger, and M. Goldszmid. Bayesian network classifiers. *Machine Learning*, 29(2) :131–163, 1997.

[75] Jeff Bilmes. Dynamic bayesian multinets. In *UAI '00 : Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, pages 38–45, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc.

[76] Dan Geiger, David Heckerman. Knowledge representation and inference in similarity networks and bayesian multinets. *Artificial Intelligence*, 82(1-2) :45–74, 1996.

[77] Isabelle Guyon, André Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3 :1157–1182, March 2003.

[78] Ron Kohavi and George H. John. Wrappers for feature subset selection. *Artif. Intell.*, 97(1-2) :273–324, 1997.

[79] Anil Jain and Douglas Zongker. Feature selection : Evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19 :153–158, 1997.

[80] Nir Friedman, Kevin Murphy, and Stuart Russell. Learning the structure of dynamic probabilistic networks. pages 139–147. Morgan Kaufmann, 1998.

[81] Eugene Santos. Unifying time and uncertainty for diagnosis. *Journal of Experimental and Theoretical Artificial Intelligence*, 8 :75–94, 1996.

[82] Lie Lu, Hong-Jiang Zhang, Hao Jiang. Content analysis for audio classification and segmentation. *IEEE Transactions on Speech and Audio Processing*, 10 :504–516, 2002.

Table des figures

1.1 Schéma d'un système de description de contenu multimédia 11
1.2 Schéma de la chaîne de l'utilisation de la détection d'événements. 16
1.3 Modèle de Markov caché. 19
1.4 Modèle produit asynchrone. 20
1.5 Représentation d'un modèle multiflux avec le formalisme des réseaux bayésiens. 21
1.6 Représentation du modèle de Markov couplé grâce au formalisme des réseaux bayésiens. 22

2.1 Modélisation de l'exemple de la pelouse mouillée. 26
2.2 Règles de passage d'une balle dans un graphe. 27
2.3 Exemple de tables de probabilités dans un réseau bayésien à variables binaires. 33
2.4 Processus de fonctionnement d'un réseau bayésien. 33
2.5 Mécanisme de propagation de connaissances au niveau d'un nœud dans un réseau bayésien en forme d'arbre 37
2.6 Réseau bayésien naïf. 40

3.1 Réseau bayésien pour la détection de publicité en supposant l'indépendance des attributs. 51
3.2 Structure en arbre construite par l'algorithme MSWT 55
3.3 Structure générique construite par l'algorithme K2 en utilisant le score BIC. 56
3.4 Structure générique construite par la recherche gloutonne en utilisant un score BIC. 56

4.1 Comparaison entre les résultats d'un réseau bayésien naïf et l'approche générative. 60
4.2 Structure du réseau bayésien construite par l'approche générative. . . . 61
4.3 Comparaison entre un réseau bayésien naïf, un réseau bayésien naïf augmenté par un arbre et un réseau bayésien naïf augmenté par une structure générique. 65

4.4 Influence du paramètre de complexité λ sur les performances de classification d'un réseau bayésien naïf augmenté par une structure apprise par un algorithme K2 (cas de la mesure F1). 67
4.5 Évolution du nombre de paramètres du réseau bayésien en fonction du paramètre de complexité λ. 67
4.6 Comparaison des performances de classification d'un réseau bayésien naïf augmenté par un arbre et d'un réseau bayésien naïf augmenté par une structure générique(K2) en utilisant $\lambda = 3$. 68
4.7 Comparaison des résultats de classification d'un réseau bayésien discriminant par rapport à ceux d'un réseau bayésien naïf. 71
4.8 Structure issue de l'apprentissage de structure par un critère discriminant. 72
4.9 Influence de l'algorithme d'apprentissage des sous-structures du réseau Multinets. 74
4.10 Comparaison des résultats de classification d'un réseau bayésien Multinets par rapport aux approches décrites précédemment. 75
4.11 Influence de la taille de la base de données d'apprentissage sur le réseau bayésien naïf. 76
4.12 Influence de la taille de la base de données sur l'apprentissage de structure par augmentation de la structure du réseau bayésien par une structure en arbre. 77
4.13 Influence de la taille de la base de données sur l'apprentissage de structure par augmentation de la structure du réseau bayésien par une structure génerique. 78
4.14 Influence de la taille de la base de données sur l'apprentissage de structure par l'approche Multinets. 79
4.15 Influence de la taille de la base de données sur l'apprentissage de structure par l'approche discriminante. 80

5.1 Principe de la sélection d'attributs basée *filtrage*. 85
5.2 Schéma de la sélection d'attributs de type *wrapper*. 87
5.3 Comparaison de performances entre les trois méthodes de sélection d'attributs. Classifieur utilisé : réseau bayésien enrichi par une structure en arbre. 88
5.4 Influence de la sélection d'attributs sur le réseau bayésien naïf. 89
5.5 Influence de la sélection d'attributs pour les réseaux bayésiens naïfs augmentés par une structure d'arbre. 90
5.6 Comparaison entre les performances des réseaux augmentés par une structure générique sans sélection d'attributs et avec sélection d'attributs avec et sans adaptation du paramètre λ. 91
5.7 Comparaison entre les performances des réseaux Multinets avec et sans sélection d'attributs. 92
5.8 Comparaison des performances de l'approche générative non restreinte avec et sans sélection d'attributs. 93

5.9 Influence de la sélection d'attributs sur l'apprentissage de structure par une approche discriminante. 95

6.1 Structures d'un réseau bayésien dynamique 98
6.2 Structure du réseau bayésien dynamique proposé dans [49] 99
6.3 Résultats des trois approches d'apprentissage de structure dans les réseaux bayésiens dynamiques. 101
6.4 Effet de l'apprentissage de structure sur les résultats des réseaux bayésiens statiques et des réseaux bayésiens dynamiques. 103

Résumé

Le domaine de la description de contenus multimédias est un domaine relativement récent qui a pris une grand importance dans le monde industriel et celui de la recherche, vu l'augmentation considérable de la production de contenus. Un besoin grandissant de systèmes capables de fournir une description sémantique est plus ce que jamais à l'ordre du jour. Dans ce domaine, les réseaux bayésiens ont été largement utilisés pour modéliser les données vidéos, afin d'en extraire des métadonnées sémantiques. Toutefois, les systèmes basés sur les réseaux bayésiens nécessitent qu'on fixe préalablement leur structure. Cette opération se fait, généralement, soit en utilisant des connaissances *a priori*, ce qui résulte en un système peu généralisable, soit en utilisant l'hypothèse d'indépendance des flux de données, ce qui résulte en un système peu optimal. Motivés par la nécessité de fournir des systèmes génériques capables de s'adapter à la grande diversité des applications envisageables, nous utilisons l'apprentissage de structure pour construire automatique le réseau bayésien. En apprenant la structure automatiquement à partir d'une base de données, nous n'avons plus besoin de connaissances externes ou de faire des suppositions, souvent peu réalistes, pour la mise en place de la structure du réseau bayésien utilisé. Différentes techniques d'apprentissage de structure ont été utilisés. Nous concluons à la nécessité d'adapter l'apprentissage de structure dans les réseaux bayésiens statiques et dynamiques à la classification. En associant Apprentissage de structure et sélection d'attributs, nous obtenons un cadre permettant de construire automatiquement des systèmes de descriptions de contenus sans être dépendants de connaissances externes.

Abstract

The description of multimedia contents field is a relatively recent one which takes a large importance in both industrial and research world, considering the massive increase of content production. A growing need for systems able to provide a semantic description is more than ever within the order of the day. In this domain, Bayesian networks are largely used to model the video data in order to extract semantic metadata. However, the Bayesian networks based systems require a beforehand fixed structure. This operation is done, generally, whether using "a priori" knowledge, which results in a not very generalizable system, or by using the assumption of independence of the data flows, which results in a not very optimal system. Moved by the need for providing generic systems capable of adapting themselves to the great diversity of applications, we use training of structure to automatically build the Bayesian network. By automatically learning the structure from a database, we no longer need external knowledge or not very realistic assumptions to build the structure of the used Bayesian network. Various structure training techniques were used. We conclude with the need to adapt training of structure in the static and dynamic Bayesian network in classification. By associating Training of structure and selection of attributes, we obtain a framework allowing to automatically building content description systems without being dependent on external knowledge.