



HAL
open science

Recommandation de parcours culturels personnalisés : Étude interdisciplinaire de la proposition automatique de visites

Mayeul Mathias

► **To cite this version:**

Mayeul Mathias. Recommandation de parcours culturels personnalisés : Étude interdisciplinaire de la proposition automatique de visites. Autre [cs.OH]. Université d'Avignon, 2021. Français. NNT : 2021AVIG0293 . tel-03675206

HAL Id: tel-03675206

<https://theses.hal.science/tel-03675206v1>

Submitted on 23 May 2022

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.

THÈSE DE DOCTORAT D'AVIGNON UNIVERSITÉ

SPÉCIALITÉ : INFORMATIQUE

École doctorale 536 « Agrosciences & Sciences »

Laboratoire Informatique d'Avignon (EA 4128)

Recommandation de parcours culturels personnalisés :

Étude interdisciplinaire de la proposition
automatique de visites

Par

Mayeul MATHIAS

Soutenue publiquement le jour/mois/année devant un jury composé de :

M.	Francesco DE PELLEGRINI	Professeur des Universités, LIA, Avignon Université	Examineur
M.	Enrico NATALIZIO	Professeur des Universités, CNRS, LORIA, Université de Lorraine	Examineur
M.	Sébastien MUSTIÈRE	Maître de Conférences HDR, Université Paris-Est, LaSTIG COGIT, IGN ENSG	Rapporteur
M.	Patrice BELLOT	Professeur des Universités, CNRS INS2I, UMR LIS, Aix-Marseille Université	Rapporteur
M.	Juan-Manuel TORRES-MORENO	Maître de Conférences HDR, LIA, Avignon Université	Directeur de thèse
M.	Didier JOSSELIN	Directeur de recherches, CNRS, UMR ESPACE	Co-directeur
M.	Fen ZHOU	Maître de Conférences HDR, CERI-SN, IMT Lille Douai	Co-encadrant

Résumé

CETTE thèse s'intéresse à la recommandation de visites culturelles à travers une approche interdisciplinaire. Ces travaux mêlent des techniques issues de la Recherche Opérationnelle et du traitement automatique de la langue naturelle écrite tout en se basant sur des concepts issus de la sociologie des publics et de la géographie. Nous proposons de nouvelles méthodes autour de l'évaluation des points d'intérêt culturel ainsi que la création automatique de parcours touristiques prenant en compte les envies exprimées par un visiteur. Ces principes sont appliqués à deux échelles et contextes différents, la visite de musées et les parcours culturels dans une ville.

Dans une première partie, nous nous concentrons sur les visites dans les musées d'art en fonction des préférences exprimées par le visiteur et du **prestige des œuvres**. Cette double approche permet de classer les œuvres à la fois en fonction des affinités culturelles du visiteur mais aussi en fonction de leur importance au sein du musée. Cette dernière est calculée en appliquant des algorithmes de résumé automatique de texte aux cartouches officiels du musée qui décrivent les œuvres et permet d'obtenir un profil de visite reflétant la découverte d'un musée en fonction de ses chefs-d'œuvre. Ce profil peut ensuite être modifié en fonction des goûts du visiteur pour obtenir une visite lui correspondant tout en préservant le « point de vue du musée ».

Par la suite, nous assimilons la construction d'un parcours à un problème de routage, visant à trouver un itinéraire parmi les différentes salles et œuvres dont le but est de maximiser la satisfaction du visiteur tout en respectant des contraintes de temps. Deux méthodes sont proposées, un modèle de programmation linéaire en nombres entiers puis une heuristique qui peut être utilisée pour de la proposition de parcours en temps réel, par exemple à leur arrivée dans le musée.

Dans une deuxième partie, nous nous intéressons à la recommandation touristique en ville en établissant des métriques permettant de construire un parcours. À partir d'une étude interdisciplinaire, nous mettons en évidence l'importance de la personnalisation des parcours et identifions un facteur essentiel lors de leur construction en plus des goûts

culturels, la cadence de visite. Une nouvelle méthode de mesure de la qualité d'expérience d'un parcours regroupant ces deux critères est utilisée. Cette dernière unit des méthodes de la littérature pour ce qui est de l'évaluation de l'intérêt culturel et utilise des **actogrammes** comme représentation géographique d'un parcours et ainsi définir une mesure de la cadence de visite.

Par la suite, nous développons un système de recommandation de visites touristiques sous la forme d'un modèle de programmation linéaire en nombre entiers basé sur un formalisme extensible permettant de prendre en compte une grande diversité de contraintes et qui intègre trois critères pour l'évaluation du parcours : d'une part l'intérêt culturel et la cadence de visite, qui dépendent des préférences du touristes sont mesurés à différentes échelles, permettant d'introduire une cohérence dans la construction du parcours ; d'autre part, nous proposons d'intégrer dans la fonction objectif, **l'effet apogée-fin**, une heuristique psychologique célèbre qui a déjà été appliquée dans de nombreux autres domaines.

En nous basant sur des études de cas concrets, nous montrons que l'utilisation conjointe de techniques issues de diverses disciplines permettent d'obtenir de bons résultats, tant au niveau de l'estimation de l'attrait des points d'intérêt que de la construction de parcours touristiques.

Sommaire

1	Introduction générale	1
1.1	Contexte et motivations	2
1.1.1	L'abondance et le paradoxe du choix	2
1.2	Problématiques	3
1.2.1	Les visites muséales	3
1.2.2	Le tourisme urbain	4
1.3	Approches proposées	4
1.3.1	Visite de musée d'art	4
1.3.2	Parcours touristiques en ville	6
1.3.3	Positionnements philosophiques et méthodologiques	8
1.4	Organisation du manuscrit	9
I	Positionnement scientifique	13
2	La recommandation personnalisée	15
2.1	Définitions et objectifs	16
2.1.1	Fonctionnalités	16
2.1.2	Typologie des données	17
2.2	Techniques de recommandation	18
2.2.1	Formulations du problème de recommandation	19
2.2.2	Approches par filtrage collaboratif	20
2.2.3	Approches basées sur le contenu	23
2.2.4	Approches basées sur les connaissances	25
2.3	Conclusion	28

3	L'élaboration de visites culturelles	29
3.1	Introduction	30
3.2	La recommandation de visites muséales	30
3.2.1	La nécessité d'une orientation pendant la visite	30
3.2.2	Systèmes d'aide pour les visites de musées	31
3.3	Le tourisme et les parcours en ville	33
3.3.1	Le tourisme urbain et ses spécificités	33
3.3.2	Le séjour touristique comme un parcours géographique	36
3.3.3	La loi de l'apogée-fin	39
3.4	La recommandation de parcours touristiques	41
3.4.1	Les systèmes de recommandation touristique	41
3.4.2	L'optimisation de voyages touristiques	43
3.5	Conclusion	44
II	Méthodes d'aide à la visite culturelle au musée	47
4	L'analyse de la médiation muséale pour la personnalisation de visites	49
4.1	Introduction	50
4.2	Principes pour l'estimation de l'intérêt des œuvres	50
4.2.1	Processus d'évaluation	51
4.3	Adéquation avec des préférences personnelles	52
4.3.1	Représentation du musée	52
4.3.2	Prédiction de l'intérêt pour le visiteur	52
4.4	Mesure indépendante du prestige des expôts	54
4.4.1	Classement des œuvres et analyse textuelle	54
4.4.2	Adaptation du résumé automatique de texte pour mesure le prestige des expôts	54
4.5	Expériences et résultats	56
4.5.1	Musées étudiés	56
4.5.2	Système de résumé automatique de texte	57
4.5.3	Application	58
4.6	Conclusion	60

5	La recommandation de visites muséales	61
5.1	Introduction	62
5.2	Méthodologie de recommandation	63
5.2.1	Objectif du système	63
5.2.2	Fonctionnement et composants clés	63
5.2.3	Estimation de l'intérêt des œuvres	64
5.2.4	Le musée comme donnée géographique	65
5.3	Formalisation du problème	65
5.3.1	Topologie d'un musée	66
5.3.2	Le problème de la recommandation de visite personnalisée au musée	67
5.3.3	Parcours de visite	67
5.4	Génération de parcours de visite satisfaisant	68
5.4.1	Modèle de programmation linéaire	68
5.4.2	Heuristique gloutonne	72
5.5	Expériences et résultats	75
5.5.1	Configurations des simulations	76
5.5.2	Temps d'exécution	78
5.5.3	Pertinence des parcours	79
5.6	Conclusion	84
III	La ville, un musée à ciel ouvert	85
6	Étude sur la qualité d'expérience des visites culturelles	87
6.1	Introduction	88
6.2	La planification de parcours touristiques personnalisés	89
6.2.1	Le problème de conception de visites touristiques	89
6.2.2	Étude sur les habitudes touristiques	90
6.3	Évaluer la Qualité d'Expérience	96
6.3.1	Formulation générale	97
6.3.2	L'intérêt culturel	97

6.3.3	La cadence de visite	100
6.4	Étude de cas	104
6.5	Conclusion	107
7	Le problème de la recommandation de visite personnalisée en ville	109
7.1	Introduction	110
7.2	La loi de l'apogée-fin dans la recommandation de parcours touristiques personnalisés	111
7.2.1	Définition du problème	111
7.2.2	Vision adoptée pour l'élaboration du parcours	112
7.3	Formalisation du problème	114
7.3.1	L'offre culturelle en tant que donnée géographique et temporelle	114
7.3.2	Proposition d'une modélisation mathématique	116
7.3.3	Flexibilité du système de contraintes	118
7.4	Représentation sous forme de graphe	120
7.4.1	Définition et construction	121
7.4.2	Propriétés et relation avec les parcours	122
7.5	Conclusion	124
8	Un système de recommandation pour la génération de parcours touristiques	125
8.1	Introduction	126
8.2	Architecture et fonctionnement	126
8.3	Estimation de l'attrait des activités	128
8.3.1	Intérêt culturel	128
8.3.2	Cadence de visite	129
8.3.3	Intégration des mesures	131
8.4	Construction du parcours proposé au visiteur	133
8.4.1	Critères d'optimisation	133
8.4.2	Modélisation par recherche de chemin	136
8.5	Expériences et résultats	145
8.5.1	Mesures de l'attrait des activités	145

8.5.2	Scénarios et paramètres	147
8.5.3	Complexité en temps et en espace	149
8.5.4	L'influence des actogrammes sur les recommandations quotidiennes	149
8.5.5	L'effet apogée-fin sur des séjours d'une semaine	152
8.6	Conclusion	154
9	Conclusion et perspectives	157
9.1	Conclusion	158
9.1.1	Visites au musée	158
9.1.2	Parcours en ville	159
9.2	Perspectives	161
9.2.1	Élargissement du cadre méthodologique	161
9.2.2	Visite cohérente au musée	162
9.2.3	Évaluation des critères en ville	162
Annexes		165
A	Le « Festival OFF d'Avignon », contexte et caractérisation du corpus	167
A.1	Présentation du terrain d'étude	168
A.1.1	Popularité lors des éditions passées et orientations des organisateurs	169
A.1.2	Données disponibles	170
A.2	Analyse des données textuelles	170
A.3	Analyse des données structurées	172
A.3.1	Informations spatio-temporelles	172
A.3.2	Attributs des spectacles et genre théâtral	175
A.4	Profils touristiques utilisés en simulation	176
A.4.1	Préférences culturelles et d'organisation du séjour	177
A.4.2	Itinéraire et mode de déplacement	178
Glossaire		181
Liste des figures		184

Liste des tableaux	185
Bibliographie	187
Références	187
Publications personnelles	195

Chapitre 1

Introduction générale

Sommaire

1.1	Contexte et motivations	2
1.1.1	L'abondance et le paradoxe du choix	2
1.2	Problématiques	3
1.2.1	Les visites muséales	3
1.2.2	Le tourisme urbain	4
1.3	Approches proposées	4
1.3.1	Visite de musée d'art	4
1.3.2	Parcours touristiques en ville	6
1.3.3	Positionnements philosophiques et méthodologiques	8
1.4	Organisation du manuscrit	9

1.1 Contexte et motivations

NOUS sommes le 10 juillet, Alice et Bob arrivent en gare d'Avignon. En amateurs de théâtre, ils viennent pour le célèbre « Festival OFF d'Avignon », un des plus grands rassemblements de spectacles vivants du monde. C'est leur première visite de la cité des Papes et, en plus d'assister à quelques pièces de théâtre, ils souhaitent aussi découvrir la ville et ses musées.

1.1.1 L'abondance et le paradoxe du choix

Pour préparer au mieux leur séjour, ils se sont procuré le programme officiel du festival et ont parcouru de nombreux sites Web. Hélas, avec plus de 1 300 pièces à l'affiche, il leur est impossible d'en sélectionner quelques unes à voir absolument simplement en feuilletant le guide papier. Concernant le patrimoine de la ville, les ressources en ligne ne sont guère plus utiles, la plupart se contentant d'exposer une liste d'activités, ordonnées en fonction des avis de précédents visiteurs.

Alice et Bob doivent donc faire un choix, isoler les pièces qu'ils souhaitent ne pas manquer, les musées à visiter et organiser le tout en un emploi du temps, qui tienne compte des représentations, horaires d'ouverture mais leur laissant aussi plusieurs occasions de flâner en ville.

Intuitivement, un plus grand choix est associé à une amélioration du bien être car il satisfait la sensation de liberté, d'autonomie, d'autodétermination et de contrôle sur l'environnement. Ajouter une option ne peut qu'améliorer la situation puisqu'il est toujours possible de l'ignorer si elle n'est pas bénéfique. C'est cette affirmation qu'invalide le paradoxe du choix, décrit par SCHWARTZ (2005). L'abondance de possibilités amènent selon ce dernier deux problèmes majeurs, le coûts de l'opportunité et l'augmentation des attentes.

Lorsqu'ils sont confrontés à l'obligation de choisir une représentation parmi de nombreuses pièces, Alice et Bob ont tendance à envisager des compromis hypothétiques. Ils évaluent chaque spectacle non pas en jugeant de son potentiel mais en termes d'opportunités manquées. Ce ne sont non plus ses atouts qui sont retenus, mais l'ensemble des caractéristiques qu'il n'offre pas par rapport au reste du festival. Ainsi, chaque activité touristique qu'il leur est proposée réhausse leurs attentes sur la qualité du séjour qu'ils élaborent.

En définitive, ces regrets diminuent leur satisfaction vis à vis d'un parcours, pouvant les conduire à une incapacité de prendre une décision, même si elle constitue un séjour qui les ravirait. Pour éviter cette paralysie de l'analyse, PARK et al. (2013) énoncent qu'il

est nécessaire de limiter le choix de l'utilisateur, une tâche pour laquelle un système de recommandation est une solution idéale, car « de tels systèmes présentent un attrait évident dans un environnement où la quantité d'information en ligne dépasse largement la capacité de toute personne à l'étudier » (BURKE 2002).

1.2 Problématiques

Assister le visiteur dans la création d'un parcours touristique apparaît comme une nécessité pour faire face à la surinformation, et l'utilisation de systèmes de recommandation dans ce contexte semble judicieuse. Cependant, ils nécessitent une attention particulière pour adresser les problèmes du démarrage à froid et la sur-spécialisation (AGGARWAL 2016a). Le démarrage à froid est lié au manque de données sur un nouvel utilisateur, le système ne le connaissant pas, il lui est difficile d'effectuer des recommandations personnalisées qui soient pertinentes. La sur-spécialisation quant à elle, désigne la tendance d'un système à ne recommander que des objets similaires au profil de l'utilisateur, limitant la découverte et la nouveauté.

1.2.1 Les visites muséales

Lors des visites muséales, les ouvrages de GILMAN (1916) et MELTON (1935) mettent en évidence le problème de la « fatigue muséale », décrivant une baisse de l'intérêt du visiteur à mesure que la visite progresse. L'objectif pour un système de recommandation est alors d'orienter le visiteur vers les expôts les plus pertinents. Or, étant donné qu'il est peu probable que le visiteur soit connu du système, celui-ci est confronté au problème du démarrage à froid, il lui est impossible d'effectuer des recommandations personnalisées.

Afin de déterminer les goûts et attentes du visiteur, une solution couramment adoptée est de lui soumettre un questionnaire. Cette étape peut être fastidieuse pour l'utilisateur et ne convient pas à un visiteur n'ayant pas de préférence sur le contenu de son parcours ou simplement désireux de découvrir la collection. De plus, une fois le profil du visiteur connu, les propositions du systèmes peuvent être soumises à la sur-spécialisation et donc mener à un parcours peu diversifié.

Ainsi, la problématique est d'éviter le démarrage à froid, sans pour autant imposer un questionnaire avant la visite, mais aussi d'éviter la sur-spécialisation des profils ayant une préférence en y intégrant des œuvres différentes, permettant de découvrir la collection.

1.2.2 Le tourisme urbain

Le tourisme culturel en ville est marqué par des changements de rythme tout au long du séjour. Une étude menée en partenariat avec des chercheurs en sciences sociales (BÉGUIN 2015 ; PICOLOTT 2016) montrent l'importance de moments de repos durant le séjour et la diversité des préférences des visiteurs en terme d'emploi du temps. Par ailleurs, les travaux de WASTIAUX (2009) soulignent l'attrait des touristes pour la découverte de la ville au gré de promenades, qui est donc un aspect important à ne pas négliger.

Un système de recommandation doit donc être capable de proposer un parcours touristique équilibré et adapté au rythme du voyageur. Or, l'offre touristique étant très importante et les profils utilisateurs principalement tournés sur les goûts culturels du visiteur, la sur-spécialisation amène à recommander des parcours uniquement constitués d'activités issus du catalogue (comme les spectacles ou visites de musées) en délaissant les moments de repos. De plus, la flânerie étant imprévisible par nature, il est difficile de l'intégrer à une stratégie de recommandation, ayant pour but de proposer un parcours figé en avance.

Ainsi, la problématique est de s'adapter au rythme du visiteur et diminuer la sur-spécialisation, en proposant des parcours conciliant les temps forts du séjour avec des moments plus calmes, permettant au touriste de se reposer et découvrir la ville.

1.3 Approches proposées

Les travaux réalisés sont fondés sur une démarche interdisciplinaire issue d'une collaboration avec des chercheurs de différents horizons tant en informatique qu'en sciences humaines. Les projets de recherche @MUSE et TCHEK-OFF nous ont permis d'établir de nouvelles hypothèses pour la recommandation de parcours touristiques et de mettre en commun des méthodes et outils pour leur validation sur des cas réels. Notre approche vise notamment à intégrer des concepts issus des sciences humaines pour améliorer l'expérience de visite au musée et en ville.

1.3.1 Visite de musée d'art

Une première partie de nos travaux concerne l'aide à la visite dans les musées d'art dans le but de retarder l'effet de fatigue muséale en améliorant le profil utilisateur construit par les systèmes de recommandation. Cependant, l'utilisation d'un questionnaire pour

identifier les goûts et attentes du visiteur peut être une étape fastidieuse pour celui-ci, d'autant plus s'il n'a pas de préférences précises sur le contenu de sa visite.

La question que nous nous posons ici est « Si un amateur de sculpture visite le Louvre, devrait-il passer devant La Joconde ? » ou plus largement, « Que conseiller à une personne désireuse de découvrir un musée ? ». Pour y répondre, un système de recommandation pourrait prendre en compte lors de la création d'un parcours, la durée de la visite et les éventuelles préférences individuelles mais aussi le prestige de chaque œuvre au sein du musée. En effet, déterminer l'importance de chaque expôt d'une institution permettrait de considérer le « point de vue du musée » et ainsi de proposer des parcours plus diversifiés intégrant les œuvres phares de la collection.

L'idée d'utiliser l'opinion du musée sur sa propre collection rappelle la notion de chefs-d'œuvre, grâce auxquels une institution identifie les expôts qu'elle juge au-dessus des autres, selon des critères qui lui sont propres (BELTING et al. 2000). Or, d'une part, nous avons pu constater que le concept de chef-d'œuvre est rejeté par certains musées qui estiment qu'il n'y a pas de pièce marquant plus que les autres leur collection, rendant l'utilisation de cette ressource limitée à quelques institutions. D'autre part, l'élaboration d'un parcours implique de répartir les expôts entre eux, chefs-d'œuvre ou non, et donc de leur attribuer un score qui puisse être utilisé, afin de sélectionner ceux sur lesquels s'attarder en fonction de la durée de la visite (fixée par le visiteur ou estimée pour prévenir la fatigue muséale).

Établir pour chaque expôt une mesure le différenciant des autres pour le musée nécessite des connaissances sur la collection et les orientations ayant guidé les choix des professionnels de l'institution. De plus, demander de ces derniers d'effectuer cette tâche n'est pas envisageable du fait de la taille de certaines collections et des biais cognitifs exposés plus tôt. Toutefois, nous pouvons constater que les musées mettent en avant certains éléments de leur collection, de part la scénographie des salles mais aussi à travers leurs supports de communication, donnant des indices sur le programme muséographique structurant le discours de l'institution.

Notre hypothèse est qu'il est possible d'identifier ce « point de vue du musée » à partir des textes officiels décrivant les œuvres. Par conséquent, nous proposons d'utiliser des méthodes de traitement automatique de la langue naturelle écrite et notamment le résumé automatique de texte pour attribuer à chaque expôt un score qui traduise son importance par rapport au reste de la collection. Ce procédé permet d'attribuer à chaque expôt un score qui dépend de la description fournie par le musée, en comparaison avec celle des autres œuvres. De ce fait, il offre l'avantage de produire une mesure qui soit indépendante de toute préférence des visiteurs, d'être propre à chaque musée (seuls les expôts de sa collection sont utilisés) et d'exploiter des ressources déjà disponibles issues du travail des

professionnels.

L'objectif est donc d'améliorer l'expérience de visite en proposant des parcours plus variés à partir d'une analyse automatique de la médiation muséale. Lorsque le système de recommandation a connaissance de préférences culturelles du visiteur, il peut départager des expôts jugées similaires et décider de substituer certaines par des œuvres phrases. De plus, il est aussi capable de proposer des parcours à des visiteurs souhaitant découvrir un musée dont la taille ne permet pas une visite complète dans un temps raisonnable.

1.3.2 Parcours touristiques en ville

Dans la deuxième partie de nos travaux, nous nous concentrons sur la recommandation de visites culturelles en ville. Ce sujet, bien que partageant de nombreux points communs avec l'aide à la visite au musée, expose les problématiques spécifiques au tourisme urbain. Parmi celles-ci, nous pouvons citer la place prépondérante de la temporalité, due entre autres, aux horaires des activités et leur espacement géographique, ou encore l'importance de la flânerie, qui est un aspect incontournable mais néanmoins imprévisible de ces parcours à ciel ouvert.

Nous avons mené une étude en partenariat avec des chercheurs en sciences sociales (BÉGUIN 2015 ; PICOLOT 2016). Celle-ci montre effectivement l'importance de moments de repos durant le séjour et la diversité des préférences des visiteurs en terme d'emploi du temps. Un attrait des touristes pour la découverte de la ville au gré de promenades est aussi un aspect important à ne pas négliger. Or, la flânerie étant imprévisible par nature, il est difficile de l'intégrer à une stratégie de recommandation, ayant pour but de proposer un parcours figé en avance.

Nous sommes ainsi amenés à nous demander « Comment organiser un parcours qui allie moments intenses et temps de repos ? ». Pour y parvenir, il semble important que le profil utilisateur d'un système de recommandation décrive non seulement les goûts culturels du visiteur, mais aussi ses préférences sur l'organisation des éléments de son séjour. Ces critères, constituent deux aspects incontournables et complémentaires des motivations du visiteur (BÉGUIN 2015) qu'il faut à la fois arriver à capturer et mettre en relation afin de déterminer ce qui est le plus approprié pour un touriste.

Ce sont particulièrement ses préférences sur l'organisation du séjour qui sont aujourd'hui négligées, alors qu'elles sont capitales pour une meilleure expérience du voyage. Néanmoins, établir le profil détaillé d'un touriste constitue un défi, l'objectif étant de l'assister dans sa recherche d'un agenda qui le satisfasse, sans pour autant que l'étape

de constitution de son profil ne soit trop fastidieuse. Or, décrire les habitudes de mobilité d'un touriste peut être fait au moyen d'actogrammes, outils issus de la géographie temporelle. Ils permettent d'associer à une typologie de visiteur, un emploi du temps représentatif (PICOLOT 2016), qui identifie les périodes propices aux activités et celles où favoriser des temps libres pour du repos ou des promenades en ville. L'utilisation de ces actogrammes amènent la possibilité de juger de la pertinence d'inclure une activité au parcours, en fonction de sa typologie, ses horaires et de l'emploi du temps du visiteur.

Mettre au premier plan l'organisation du parcours nous amène aussi à le considérer non plus comme un ensemble d'activités, mais une série ordonnée de celles-ci. Cette vision permet d'introduire une évaluation des séquences d'activités, vis à vis de l'organisation du séjour mais aussi de leur cohérence culturelle. Dans le premier cas, nous pouvons retranscrire la pertinence d'un temps libre entre deux activités, et donc l'opportunité pour le visiteur, de flâner en ville. Dans le second cas, nous pouvons assurer une similarité des activités culturelles, ou au contraire, favoriser une alternance des styles tout au long du parcours.

L'introduction d'une mesure de la « cadence de visite » ne doit pour autant pas être employée comme une contrainte au parcours, mais en tandem avec les préférences culturelles du visiteur. Deux critères essentiels, qui peuvent amener la recommandation d'un spectacle en dehors des horaires voulues, ou à en retirer un, pour offrir au visiteur le temps d'une promenade dans un parc.

Enfin, l'idée d'une évaluation de l'organisation du parcours et d'une alternance entre des temps forts et moments calmes, fait écho aux théories de KAHNEMAN (2000) sur la construction des souvenirs et du ressenti associés à un évènement. Celui-ci établit la loi de l'apogée-fin, qui désigne le moment le plus intense et la fin d'un évènement, comme les deux facteurs décisifs façonnant le souvenir de ce dernier. Tenir compte de cet effet psychologique, en organisant le séjour de sorte à créer deux journées mémorables pourrait permettre d'améliorer le sentiment d'un voyage exceptionnel, et est un des défis que doit relever l'industrie du tourisme (NAWIJN, MITAS et al. 2013).

Notre hypothèse est donc qu'une évaluation, à la fois multicritère et multiscalaire, permet à un système de recommandation de considérer des profils plus riches et donc, d'améliorer l'expérience de visite. Nous proposons de formaliser et d'intégrer les actogrammes et la loi de l'apogée-fin à la stratégie de recommandation, tout en modélisant le parcours non plus comme un ensemble mais une suite de moments culturels. La mesure de la « cadence de visite » est transposée non pas en contrainte du parcours mais comme un point de vue sur le séjour, qui est mis en relation avec l'intérêt culturel de celui-ci. À l'échelle du séjour, c'est en cherchant à reproduire les effets de la loi de l'apogée-fin que le parcours est construit, tout en considérant l'attrait de chaque activité ainsi que des

séquences qu'elles forment, ceci suivant les préférences du visiteur.

1.3.3 Positionnements philosophiques et méthodologiques

Les approches que nous proposons constituent des changements de paradigmes quant à la façon de penser la recommandation, nécessitant d'établir de nouveaux formalismes. Afin d'évaluer l'impact de ces changements, nous avons choisi de nous limiter à la recommandation de parcours culturels, à partir des envies et attentes exprimées par le visiteur et au moyen d'une modélisation par optimisation linéaire. Par ailleurs, la satisfaction dans le contexte du tourisme étant déjà un sujet complexe (FILEP et al. 2010), l'évaluation des systèmes proposées est basée sur des simulations, réalisées à partir de profils obtenus par des études sur le terrain.

Les problématiques et hypothèses de travail sont le fruit d'un effort interdisciplinaire et, bien que les modélisations proposées aient été voulues les plus génériques possible, nous les avons appliquées sur des terrains d'études spécifiques. Dans la première partie, nous choisissons de nous concentrer sur les musées d'art, conjecturant que la séquentialité du parcours a moins d'impact sur le discours muséal dans ces derniers vis à vis d'autres types d'institutions. En ville, nous étudions les séjours durant le « Festival OFF d'Avignon » (décrit en annexe A), pour lesquels seul l'aspect culturel est évoqué. De plus, l'évènement se déroulant majoritairement en centre ville, qui est rendu exclusivement piéton pour l'occasion, l'aspect multimodal des zones urbaines n'est pas intégré à nos algorithmes.

L'élaboration d'un système de recommandation dépend des données qui constituent le profil utilisateur (AGGARWAL 2016a). Nous choisissons de ne définir ce dernier qu'à partir des envies et attentes exprimés par les visiteurs, sans considérer leurs consommations culturelles antérieures ou leur catégorie socioprofessionnelle. Cette décision a des implications sur les algorithmes, notamment un plus grand respect de la vie privée du visiteur, celui-ci choisissant explicitement le contenu et le moment de ce qu'il partage avec le système. Cependant, cette stratégie peut introduire un décalage entre le visiteur et le système, connu en linguistique comme les notions de signifiant et signifié (SAUSSURE 1916), pour un même terme, le concept qui lui est associé par le visiteur et le système peuvent différer, ce qui n'est pas pris en compte dans nos travaux.

Enfin, nos systèmes de recommandation reposent sur une modélisation par programmation linéaire mixte en nombres entiers (MILP¹) (SCHRIJVER 1998). L'obtention du parcours qui est proposé au visiteur se fait donc en résolvant le problème associé, c'est-à-dire en trouvant une solution optimale à celui-ci. Cette étape est effectuée au moyen de

1. *Mixed Integer Linear Programming*

CPLEX 12.8 (IBM 2017), un outil qui impose une modélisation mono-objectif. Il est aussi à noter que la notion d'optimalité de la solution n'a de sens que dans le cadre du modèle mathématique duquel elle découle. Celui-ci étant une traduction partielle de l'objectif du touriste, tel qu'il le ressent et l'exprime, parler d'optimalité du parcours est un abus de langage qui est employé pour désigner la meilleure visite que le système est capable de proposer compte tenu de ses capacités et sa connaissance de l'individu.

1.4 Organisation du manuscrit

Nos travaux sur l'amélioration de l'expérience des parcours culturels se répartissent donc en deux volets, le premier sur les visites muséales et le second sur les séjours touristiques en ville. Conséquemment, le reste du manuscrit est organisé de la manière suivante :

- **Partie I Positionnement scientifique**

- **Chapitre 2 La recommandation personnalisée**

Ce chapitre rappelle les concepts de base et principales techniques utilisés lors de la création de systèmes de recommandation personnalisée. Nous détaillons les avantages et inconvénients des stratégies de recommandation ainsi que la typologie des données qui sont exploitées par chacune d'elles.

- **Chapitre 3 L'élaboration de visites culturelles**

Ce chapitre passe en revue la littérature autour des visites culturelles au musée puis en ville. Pour chacun de ces domaines, nous présentons dans un premier temps les études issues de sciences sociales sur ces sujets afin d'identifier les enjeux actuels puis les travaux en informatique et notamment les systèmes de recommandation existants.

- **Partie II Méthodes d'aide à la visite culturelle au musée**

- **Chapitre 4 L'analyse de la médiation muséale pour la personnalisation de visites**

Ce chapitre propose une nouvelle méthode pour l'estimation de l'intérêt des œuvres d'un musée. D'une part, celle-ci est basée sur une approche classique de personnalisation à partir des caractéristiques des expôts. D'autre part, nous montrons que l'utilisation de méthodes de traitement automatique de la langue naturelle écrite permet d'identifier un prestige accordé par le musée aux objets de la collection.

– Chapitre 5 **La recommandation de visites muséales**

Ce chapitre décrit une stratégie pour la recommandation de visites muséales à partir des travaux du chapitre 4 sur l'estimation de l'intérêt des œuvres. Dans un premier temps, nous formalisons le problème de recommandation de visite personnalisée au musée (PVRP_m²) et proposons un modèle MILP pour le résoudre. Dans un second temps, nous proposons une heuristique simple pour la résolution du même problème. Enfin, nous étudions l'impact de l'utilisation du prestige des œuvres sur ces deux algorithmes de recommandation.

• Partie III **La ville, un musée à ciel ouvert**

– Chapitre 6 **Étude sur la qualité d'expérience des visites culturelles**

Ce chapitre expose une étude interdisciplinaire sur le tourisme urbain à partir de deux enquêtes sur les pratiques touristiques et styles de visites. Dans ce but, nous formalisons le concept d'actogramme pour représenter l'emploi du temps d'un touriste. De plus, nous proposons de différencier l'adéquation aux préférences culturelles des activités touristiques de leur place dans l'emploi du temps souhaité du visiteur et établissons deux métriques pour refléter ces critères. Enfin, ces nouveaux outils sont appliqués sur des parcours types pour démontrer leur efficacité.

– Chapitre 7 **Le problème de la recommandation de visite personnalisée en ville**

Ce chapitre définit le problème de recommandation de visite personnalisée en ville considérant l'effet de l'apogée-fin (PE-PVRP_c³), une nouvelle stratégie de recommandation pour l'élaboration de parcours touristiques en ville qui est basée sur la loi de l'apogée-fin. Cette nouvelle approche combine différents critères guidant la création du parcours à plusieurs échelles tout en appliquant les métriques définies au chapitre 6. Nous exprimons ensuite ce problème sous la forme d'un objet mathématique extensible qui est comparé aux formulations présentes dans la littérature.

– Chapitre 8 **Un système de recommandation pour la génération de parcours touristiques**

Ce chapitre présente notre système de recommandation répondant aux exigences du PE-PVRP_c. Il intègre les mesures sur l'intérêt culturel et la cadence de

2. *Personalized Visit Recommendation Problem in museum*

3. *Peak-end rule driven Personalized Visit Recommendation Problem in city*

visite afin d'évaluer, non seulement les activités touristiques individuellement mais aussi la cohérence de leurs enchaînements. Ces deux critères sont combinés avec une formalisation innovante de la loi de l'apogée-fin qui est utilisée dans un modèle MILP. Des simulations sur le « Festival OFF d'Avignon » montrent qu'il est possible de proposer un parcours tenant compte des diverses préférences du touriste tout en favorisant l'effet d'apogée-fin pour améliorer le souvenir du séjour.

Pour conclure, le chapitre 9 discute des travaux exposés et suggère les verrous scientifiques restant à explorer tout en proposant des pistes à explorer pour leur résolution.

L'annexe A présente quant à elle, présente le « Festival OFF d'Avignon » et caractérise le corpus des données qui ont été disponibles durant nos travaux.

Partie I

Positionnement scientifique

Chapitre 2

La recommandation personnalisée

Sommaire

2.1	Définitions et objectifs	16
2.1.1	Fonctionnalités	16
2.1.2	Typologie des données	17
2.2	Techniques de recommandation	18
2.2.1	Formulations du problème de recommandation	19
2.2.2	Approches par filtrage collaboratif	20
2.2.3	Approches basées sur le contenu	23
2.2.4	Approches basées sur les connaissances	25
2.3	Conclusion	28

2.1 Définitions et objectifs

Comme nous l'avons vu en introduction, l'ouvrage de SCHWARTZ (2005) met en évidence l'incapacité de l'être humain à prendre des décisions satisfaisantes lorsqu'il est confronté à de nombreuses options. RICCI, ROKACH et SHAPIRA (2011) souligne que les systèmes de recommandation sont reconnus comme efficaces pour résoudre ces problèmes de surcharge d'information et qu'ils constituent depuis plusieurs décennies un domaine de recherche à part entière.

De par leur utilisation importante dans de nombreux domaines, ils ont été spécifiés de multiples manières. Nous choisissons de retenir une définition générale de BURKE (2002) que nous pouvons traduire ainsi :

« Tout système produisant des recommandations individualisées ou permettant de guider l'utilisateur de manière personnalisée vers des solutions utiles ou intéressantes au sein d'un vaste espace de possibilités. » (BURKE – 2002)

Cette définition met en avant deux aspects importants, la personnalisation et l'étendue de l'espace de recherche. Ces concepts imposent de retrouver de l'information qui dépend de l'utilisateur parmi un large éventail de possibilités. Mais avant tout, nous explorons les divers points de vue qui peuvent être pris en compte pour répondre à l'utilisateur de manière personnalisée.

2.1.1 Fonctionnalités

Dans l'ouvrage de AGGARWAL (2016a), on distingue plusieurs fonctionnalités que peut remplir un système de recommandation. La plus évidente est la **pertinence**, un système a pour objectif de recommander des solutions qui soient adéquates à l'utilisateur et sa demande. Mais bien que cet objectif soit primordial, il n'est pas suffisant en soi pour qu'un utilisateur juge une recommandation assez intéressante pour être sélectionnée et d'autres objectifs secondaires doivent être considérés car ayant un impact significatif sur la décision.

La **nouveauté** est en effet aussi un point très important car un système de recommandation n'est réellement utile que lorsqu'il propose des solutions que l'utilisateur ne connaissait pas. Une autre notion proche de celle-ci à considérer est l'**effet de surprise** qui se manifeste lorsque les solutions sont inattendues. Plutôt que de lui proposer quelque chose qu'il ne connaissait pas mais très similaire à ses habitudes, lui recommander un objet réellement surprenant peut lui provoquer de l'intérêt. Un système doit à la fois pro-

poser des éléments qui sont évidents aux yeux de la requête de l'utilisateur mais aussi surprenant, qu'il pourrait assimiler à de la chance.

Les systèmes de recommandation proposent typiquement des listes de solutions, les k objets les plus pertinents (ou *top-k items*). Cependant elles sont toutes similaires les unes aux autres, il y a une possibilité que l'utilisateur n'aime aucune d'entre elles. La **diversité** permet de contrer ce problème en augmentant les chances qu'au moins une solution attire l'utilisateur et a l'avantage qu'il ne soit pas ennuyé par des recommandations trop similaires.

Outre ces objectifs concrets, un certain nombre moins importants peuvent aussi être considérés, tant du point de vue de l'utilisateur que celui du fournisseur du système. L'ergonomie de l'interface peut par exemple inciter l'utilisateur à utiliser le service à nouveau et le détourner des alternatives de la concurrence. Enfin, un système donnant les raisons l'ayant mené à proposer une solution en particulier permet aux utilisateurs de se familiariser avec lui et d'affiner leurs futures recherches. Enfin, comme l'indique BEN TICHA (2015), on attend d'un système de recommandation qu'il ait une bonne couverture, c'est-à-dire la capacité à proposer l'entièreté de son catalogue en faisant varier le profil de l'utilisateur. Pour ce faire, il faut structurer et organiser l'information dont le système dispose.

2.1.2 Typologie des données

Selon RICCI, ROKACH et SHAPIRA (2011), l'information que manipule un système de recommandation peut provenir de sources variées et ces éléments sont répartis en trois catégories : (i) les objets ; (ii) les utilisateurs ; (iii) les relations entre ces deux derniers.

Les **objets** sont les éléments qui sont recommandés par le système et sont caractérisés par leur valeur ou leur utilité qui peut être positive si l'objet est profitable à l'utilisateur, ou négative s'il n'est pas approprié. Par ailleurs, un objet est associé à un coût, qui peut être en partie monétaire si l'article doit être acheté et qui comprend toujours un coût cognitif lié à la recherche effectuée par l'utilisateur et l'analyse de l'information qui lui a été présentée. La pertinence d'une recommandation est donc fonction de l'utilité de l'objet comparée à son coût. Si la valeur perçue par l'utilisateur est supérieure à son coût, alors il sera satisfaisant. Au contraire, si la valeur est dominée par le coût, la recommandation sera perçue comme négative par l'utilisateur.

Estimer la valeur et le coût d'un objet dépend de la complexité de ce dernier, c'est-à-dire les propriétés qui sont considérées par le système pour effectuer les recommandations (BEN TICHA 2015). En fonction de ces caractéristiques, PAZZANI et al. (2007) décrit la

représentation des objets comme étant (i) structurée ; (ii) non-structurée ou (iii) semi-structurée. Lorsqu'un objet est caractérisé par un ensemble d'attributs (dimensions d'un produit, durée d'un film), sa représentation est dite structurée alors que s'il est décrit par un texte libre (résumé d'un livre) elle est dite non-structurée. Enfin, dans de nombreux domaines nous retrouvons l'utilisation conjointe d'attributs et de texte libre et parlons alors de représentation semi-structurée.

RICCI, ROKACH et SHAPIRA (2011) souligne que généralement un système de recommandation ne traite qu'un seul type d'objets et note que les voyages font partie des objets les plus complexes à être considérés pour de la recommandation personnalisée.

Les **utilisateurs** et leur modélisation jouent un rôle central puisqu'afin de personnaliser ses propositions, un système de recommandation doit recueillir et exploiter des informations comme les préférences, besoins et contraintes de l'utilisateur. La nature, la quantité et la représentation de ces informations dépendent du domaine, du degré de personnalisation des interactions homme-machine qui veut être offerte mais aussi de la stratégie de recommandation adoptée par le concepteur du système.

L'analyse des usages, c'est-à-dire les **relations** entre les utilisateurs et les objets du système, décrit les réactions entre les utilisateurs et les recommandations du système. Ces dernières sont récoltées pendant les phases importantes d'interactions homme-machine et permettent d'améliorer le fonctionnement de l'algorithme utilisé. Nous pouvons différencier les évaluations *explicites* des évaluations *implicites* d'un utilisateur. En effet le système peut inférer son opinion à partir des actions qu'il entreprend, par exemple lorsqu'il décide d'obtenir plus d'information sur un objet. Les retours explicites quant à eux prennent souvent la forme de commentaires et de votes donnés a posteriori sur les objets qui ont été retenus par l'utilisateur. Tout comme le modèle utilisateur, les relations que collecte le système et leur représentation varient suivant la technique de recommandation qui est utilisée.

2.2 Techniques de recommandation

Les systèmes de recommandation étant utilisés depuis les années 1990 et dans de nombreux domaines, beaucoup de travaux aux techniques variées ont été employées, notamment issues de la recherche d'information. Il est possible de classer ces systèmes selon deux angles : le modèle qu'ils utilisent pour formuler le problème de recommandation et la stratégie mise en place pour le résoudre.

2.2.1 Formulations du problème de recommandation

AGGARWAL (2016a) distingue deux grandes familles pour la modélisation d'un système de recommandation :

- **Problème de prédiction** : le système tente de prédire la note qu'un utilisateur attribuerait à un objet qu'il n'a jamais évalué auparavant. Ce problème est aussi appelé problème de complétion de matrice où le système estime les valeurs manquantes grâce à un modèle d'entraînement.
- **Problème de classement** : le système propose une liste ordonnée en fonction des préférences de l'utilisateur. Plutôt que de prédire la note d'un utilisateur, le système détermine les objets les plus appropriés en fonction de sa requête. Aussi appelé top- k items problem, c'est une formulation très répandue dans la littérature (AGGARWAL 2016a).

SCHAFER et al. (2007) identifie une variante de ces formulations classiques, la **recommandation sous contraintes**, où un ensemble de contraintes restreint les objets que le système peut considérer. Ces contraintes peuvent provenir de l'utilisateur mais aussi être appliquées par le système en fonction de son profil ou d'informations provenant d'autres sources (par exemple les stocks dans la région concernant l'utilisateur).

En plus de la formulation du problème, nous pouvons aussi distinguer les systèmes en fonction de la stratégie de recommandation utilisée, c'est-à-dire la nature des données exploitées et la manière dont elles sont traitées. La classification la plus courante (AGGARWAL 2016a ; RICCI, ROKACH et SHAPIRA 2011 ; BURKE 2002) catégorise ces approches en plusieurs grandes familles dont les principales sont :

- **Le filtrage collaboratif**, où le système recommande à l'utilisateur courant les objets qu'ont apprécié les autres utilisateurs avec lesquels il partage des préférences similaires. Ce type de recommandation utilise exclusivement les relations entre les utilisateurs et les objets et non leurs caractéristiques.
- **La recommandation basée sur le contenu**, qui utilise les propriétés des objets que l'utilisateur a déjà évalués par le passé pour déterminer ceux qui sont susceptibles de l'intéresser. Seuls les votes de l'utilisateur courant sont utilisés pour calculer la similarité entre les objets.
- **La recommandation basée sur les connaissances** se fait en utilisant l'adéquation entre les caractéristiques des objets et les besoins de l'utilisateur qui sont spécifiques au domaine. Ces approches requièrent une connaissance du domaine et des besoins de l'utilisateur mais n'ont pas besoin d'entraînement ou d'évaluations antérieures sur les objets.

Il existe d'autres catégories de techniques de recommandation comme celle basée sur la communauté utilisant les préférences des amis de l'utilisateur ou le filtrage démographique qui construit le profil utilisateur à l'aide de stéréotypes (âge, sexe, profession, etc.) et non de préférences. Enfin, la recommandation hybride qui combine plusieurs des approches que nous avons mentionnées.

La table 2.1 résume les différences entre les trois premiers types d'approches en fonction de la stratégie employée et des données requises. Nous les développons par la suite.

Approche	Stratégie employée	Données exploitées
Collaborative	Recommandations basées sur une approche collaborative qui tire parti des votes et actions de l'utilisateur et de ses pairs.	Votes de l'utilisateur + Votes de la communauté
Basée sur le contenu	Recommandations basées sur le contexte (les attributs) que l'utilisateur a aimé par ses votes ou actions précédents.	Votes de l'utilisateur + Attributs des items
Basée sur la connaissance	Recommandations basées sur les spécifications explicites de l'utilisateur à propos du type de contenu (attributs) qu'il veut.	Spécifications de l'utilisateur + Attributs des items + Connaissance du domaine

TABLE 2.1 – Stratégie et données requises des différentes approches de recommandation (selon AGGARWAL 2016a)

2.2.2 Approches par filtrage collaboratif

Le filtrage collaboratif évalue l'intérêt des objets en utilisant les opinions des autres utilisateurs du système. SCHAFER et al. (2007) compare cette approche au « bouche à oreille » pratiqué depuis toujours par les humains pour se construire un avis. Le système suppose que si des utilisateurs ont les mêmes avis sur un ensemble d'objets, alors ils auront les mêmes avis sur d'autres objets que seulement certains utilisateurs ont notés.

Les évaluations faites par les utilisateurs peuvent être implicites ou explicites et de formes variées comme nous l'avons vu plus tôt mais la plus courante et évidente est la notation à échelle où l'utilisateur indique son niveau de satisfaction d'un objet (AGGARWAL 2016a). Un ensemble de valeurs discrètes ordonnées est proposée à l'utilisateur et celui-ci

sélectionne celle qu'il trouve la plus adéquate à son expérience. Cet intervalle peut être binaire lorsqu'il modélise les évaluations « j'aime / je n'aime pas » ou à 5 niveaux, souvent représentés par des étoiles dans de nombreux services. Dans ce dernier cas, les valeurs sont positives mais il est possible d'utiliser l'ensemble $\{-2, -1, 0, 1, 2\}$ où les deux premières valeurs indiquent une insatisfaction, la valeur centrale un point de vue neutre et les deux valeurs supérieures une opinion positive. C'est le principe utilisé dans l'échelle de LIKERT (1932) très répandue dans les questionnaires de satisfaction.

Méthode de résolution

Les évaluations des utilisateurs peuvent être formalisées par une matrice utilisateurs – objets où chaque valeur est la note attribuée à un objet par un utilisateur. Ainsi, le problème de recommandation s'apparente à celui de la complétion de matrice creuse où les valeurs manquantes sont inférées à partir de celles que le système connaît.

Selon DESROSIERS et al. (2011), les méthodes basées sur les voisins sont les plus populaires pour résoudre le problème de recommandation par filtrage collaboratif. Elles utilisent la totalité ou une partie de la matrice pour prédire une note d'un utilisateur, ce qui peut être fait de deux manières : en se basant sur les utilisateurs ou sur les objets. Dans les deux cas, le système détermine les voisins de l'utilisateur courant par une mesure de similarité puis calcule une note à partir de leur note.

La prédiction basée sur les utilisateurs fonctionne en déterminant les utilisateurs voisins puis en agrégeant leur note sur l'objet dont on veut prédire l'évaluation. La similarité des utilisateurs se base sur les notes qu'ils ont attribuées aux autres objets aussi évalués par l'utilisateur courant. L'estimation de la note se fait ensuite par une moyenne des notes des utilisateurs retenus sur l'objet à estimer. Cette moyenne doit tenir compte de deux éléments, le fait que les utilisateurs ne jugent pas de la même manière les objets et leur ressemblance à l'utilisateur courant. L'estimation est donc formulée par une moyenne des notes des autres utilisateurs sur l'objet, chacune étant ajustée par la moyenne des notes de l'utilisateur qui l'a attribuée, puis pondérée par la similarité avec l'utilisateur à qui le système doit effectuer la recommandation.

La prédiction basée sur les objets a un fonctionnement similaire : le système détermine les objets voisins puis agrège les notes que l'utilisateur courant leur a attribué. Dans ce cas, deux objets sont considérés comme similaires si les autres utilisateurs du système de recommandation leur ont attribué la même note, une fois encore en tenant compte de leur différentes manières de juger. L'estimation de la note de l'objet à évaluer est donc ici une moyenne des notes de l'utilisateur courant sur les autres objets qui sont pondérées par leur similarité avec celui dont le système veut estimer l'utilité.

Ces deux méthodes sont très similaires dans leur fonctionnement, elles utilisent les notes des autres utilisateurs pour trouver des voisins et, à partir de ceux-ci, en déduire celle que l'utilisateur courant donnerait à l'objet. DESROSIERS et al. (2011) souligne qu'en fonction de la méthode choisie, certaines mesures de similarités sont plus efficaces que d'autres et que, dans tous les cas, sélectionner les k plus proches voisins (utilisateurs ou objets) donne toujours des recommandations plus pertinentes pour l'utilisateur.

Avantages et inconvénients du filtrage collaboratif basée sur les voisins

Grâce aux ouvrages de DESROSIERS et al. (2011) et BEN TICHA (2015), nous pouvons identifier les avantages suivants :

- La **généricité** et la **simplicité** : cette approche est simple à implémenter et peut-être appliquée à tout type d'objets, car elle ne requiert aucune connaissance sur les objets à recommander. De plus, elle permet aussi de traiter des objets de divers types, comme le font les sites de commerce en ligne.
- L'**effet de surprise** : la prédiction basée sur les utilisateurs voisins est très efficace pour recommander des objets atypiques qui seront appréciés. Comme nous l'avons vu plus tôt, c'est un des aspect à considérer dans la personnalisation.
- L'**amélioration progressive** : au fur et à mesure que les utilisateurs évaluent les objets, la qualité du système augmente et il devient stable. L'ajout de nouveaux utilisateurs, objets ou notes, ne force qu'une partie de la matrice à être recalculée.

Néanmoins, ces approches sont aussi soumises à certains inconvénients. Les mêmes ouvrages relèvent les suivants :

- Le **démarrage à froid** : ce problème concerne à la fois les nouveaux utilisateurs et les nouveaux objets. Un nouvel utilisateur ne peut recevoir de recommandation que lorsqu'il aura évalué un certain nombre d'objets et un nouvel objet ne pourra être recommandé que lorsqu'il aura été évalué par plusieurs utilisateurs.
- La **sensibilité aux données manquantes** : en raison du fait que les utilisateurs ne notent qu'une faible proportion des objets du système, la matrice utilisateurs – objets a un taux de valeurs manquantes pouvant atteindre 95 % à 99 %. Ces données manquantes rendent complexe la tâche de prédiction puisqu'il y a plus de valeurs à prédire que de valeurs connues.
- Le **problème du mouton gris** : le système a du mal à traiter les utilisateurs aux goûts marginaux puisqu'ils auront peu de voisins similaires (utilisateurs ou objets). Des études montrent qu'il faut un certain nombres d'utilisateurs et de votes pour que les notes prédites soient pertinentes.

2.2.3 Approches basées sur le contenu

Dans un système de recommandation basé sur le contenu (LOPS et al. 2011), les objets ayant des caractéristiques proches de ceux que l'utilisateur a déjà aimés par le passé lui sont conseillés. Le profil de l'utilisateur est souvent constitué de deux profils, un positif et un négatif qui sont construits à partir des objets en fonction des retours d'expériences qu'il a exprimé à leur propos. Le système identifie les propriétés communes entre les objets évalués et les nouvelles possibilités et les classe pour déterminer ceux qu'il proposera. Les techniques employées sont souvent issues de la recherche d'information sans toutefois qu'il y ait de requête utilisateur, puisque ce dernier ne saisit jamais de requête.

Cette méthodologie suppose qu'une description riche et variée soit disponible sur les objets que le système traite. Le modèle vectoriel de SALTON et al. (1975) est couramment utilisé pour représenter les propriétés associées aux objets et s'accommode à la fois des représentations structurées et sous forme textuelle. L'utilisateur est souvent assimilé à un pseudo-objet construit par le système à partir de son historique de consommation mais d'autres approches utilisant des réseaux de neurones ou des arbres de décision ont été proposées (LOPS et al. 2011). L'analyse des usages que nous avons décrite plus tôt montre qu'ils permettent au système d'inférer des préférences de l'utilisateur en attribuant une pondération positive ou négative à un ensemble de caractéristiques extraites des objets qu'il a déjà évalués.

Architecture et fonctionnement général

Une description que donne LOPS et al. (2011) divise l'architecture d'un système de recommandation basé sur le contenu en trois composants, chacun responsable d'une étape spécifique :

- **Analyse du contenu** : lorsque l'information que doit manipuler un système de recommandation n'est pas structurée (comme par exemple du texte libre), un pré-traitement est nécessaire pour extraire les informations pertinentes, les structurer et les représenter de manière adéquate.
- **Apprentissage du profil** : le composant qui en est responsable collecte les données représentatives de l'utilisateur et tente de les généraliser afin de construire le profil utilisateur. Celui-ci est inféré à partir des objets qui ont été évalués, faisant de l'analyse des usages la principale source d'information.
- **Filtrage des objets** : le système suggère des objets pertinents à l'utilisateur courant en exploitant la représentation qu'il en a établi souvent au moyen de mesures de

similarités. Le résultat peut-être binaire, quand le système détermine seulement si l'objet est pertinent ou non, ou continu auquel cas une liste ordonnée selon la pertinence peut-être proposée à l'utilisateur.

Dans les approches basées sur le contenu, les objets sont représentés par un ensemble de propriétés. Or, pour la plupart d'entre eux, LOPS et al. (2011) note que leur description est textuelle et que, par conséquent, il n'existe pas d'attributs bien définis. De ce fait, l'apprentissage du profil utilisateur est compliqué par les ambiguïtés du langage naturel comme la polysémie et la synonymie.

La représentation traditionnelle par mots-clé avec pondération TF-IDF (SPARCK JONES 1972) est couramment utilisée mais transpose aux objets des observations faites sur le texte (notamment quand l'importance des termes est signalée par de multiples occurrences dans un document et peu dans les autres). Les réseaux sémantiques pondérés sont aussi très utilisés, car ils permettent de stocker la signification des mots, ce qui induit un apprentissage plus précis du profil utilisateur (BEN TICHA 2015).

Enfin, les systèmes de recommandation se basant sur ces techniques représentent les objets et les profils utilisateurs par des vecteurs de propriétés pondérées dans un même espace. L'intensité de l'intérêt d'un utilisateur pour un objet est estimée en comparant ces vecteurs par des mesures comme la *similarité cosinus*.

Avantages et inconvénients de la recommandation basée sur le contenu

Les approches basées sur le contenu ont plusieurs avantages sur le filtrage collaboratif, LOPS et al. (2011) donne les suivants :

- **L'autonomie de l'utilisateur** : contrairement au filtrage collaboratif, la recommandation basée sur le contenu traite chaque utilisateur de manière indépendante. Seules les évaluations de l'utilisateur lui-même sont prises en compte pour construire son profil et lui faire des recommandations.
- La **transparence** : l'ensemble des attributs ou des descriptions de chaque objet du système permet d'expliquer comme celui-ci fonctionne. De plus, les recommandations ne peuvent pas être influencées par les autres utilisateurs.
- **L'intégration des nouveaux objets** : sans qu'il soit évalué par aucun utilisateur, un système de recommandation basée sur le contenu peut estimer l'intérêt d'un nouvel objet pour un utilisateur.

Il est à noter que tout comme le filtrage collaboratif, les approches basées sur le contenu ne requièrent pas de connaissance sur le domaine des objets.

Cependant, la recommandation basée sur le contenu présente aussi les inconvénients suivants :

- La **limite de l'analyse du contenu** : la recommandation basée sur le contenu nécessite une représentation des objets qui soit variée, afin de pouvoir les discriminer efficacement. Une connaissance du domaine de recommandation est souvent requise, par exemple pour établir les attributs pertinents de chaque objet.
- La **sur-spécialisation** : par nature, le système ne peut recommander que des objets similaires au profil utilisateur et est donc incapable de lui procurer un effet de surprise. Or comme nous l'avons vu plus tôt, la diversité des propositions est un des objectifs à prendre en compte car important pour l'utilisateur.
- L'**intégration d'un nouvel utilisateur** : pour comprendre les préférences de l'utilisateur, il doit évaluer un certain nombre d'objets afin que le système puisse construire un profil pertinent. Ce problème est aussi connu sous le nom de démarrage à froid pour les utilisateurs.

2.2.4 Approches basées sur les connaissances

Les systèmes de recommandation utilisant une approche basée sur les connaissances (AGGARWAL 2016b) suggèrent des solutions à partir de connaissances sur le degré de pertinence des objets vis à vis des préférences de l'utilisateur. La différence cruciale de ce type de recommandations par rapport aux approches exposées plus tôt est que l'utilisateur doit spécifier ses besoins détaillés sur un domaine complexe et de ce fait, la part de contrôle qu'il a sur le processus de recommandation est bien plus importante. AGGARWAL (2016b) souligne que cette approche est particulièrement utile dans le contexte d'objets qui ne sont pas souvent évalués et indique que c'est le cas des recommandations touristiques.

Le grand degré de personnalisation est obtenu grâce à l'utilisation d'une base de connaissances qui recense les informations pertinentes sur le domaine sous la forme de contraintes ou de mesures de similarité. Ces données sont pour la plupart du temps associées aux objets et à leurs propriétés, mais certains systèmes exploitent aussi des attributs propres à l'utilisateur, en plus des éléments habituellement présents dans sa requête. Dans ces cas, le système doit également encoder des informations sur les relations entre les propriétés des utilisateurs et celles des objets. AGGARWAL (2016b) souligne toutefois que cette pratique est très anecdotique dans les systèmes basés sur les connaissances qui mettent d'avantage l'accent sur les besoins exprimés explicitement par l'utilisateur.

Interactions avec l'utilisateur

Les systèmes de recommandation basés sur les connaissances doivent leur nom au fait qu'ils encodent divers types de connaissances sur le domaine sous la forme de contraintes, règles, métriques de similarité et fonctions d'utilité, qui sont par la suite utilisées durant le processus de recommandation. Ces systèmes peuvent être catégorisés par la méthodologie qu'ils utilisent pour interagir avec l'utilisateur et ce, des deux manières suivantes :

- Les systèmes **basés sur des contraintes** : l'utilisateur donne des prérequis ou contraintes précises sur les attributs des objets. Par la suite, des règles spécifiques au domaine sont utilisées pour trouver des objets correspondant aux besoins de l'utilisateur.
- Les systèmes **fondés sur des cas** : des cas précis sont ici exprimés par l'utilisateur, servant de points d'ancrage ou buts à atteindre pour le système. Ceux-ci sont des avis orientés différenciant les objets positivement ou négativement. Pour ce faire, des mesures de similarité sont définies sur les propriétés des objets.

En fonction de la méthodologie utilisée, les connaissances du domaine sont représentées, soit par les contraintes et leur rapport avec les attributs des objets, soit par le degré d'adéquation entre des caractéristiques et les attributs des objets. Néanmoins dans les deux cas, le système permet à l'utilisateur de modifier ses attentes de manière itérative.

La principale différence entre les deux méthodologies est la manière avec laquelle l'utilisateur spécifie ses requêtes et interagit avec le système pour affiner ses requête jusqu'à ce qu'il soit satisfait du résultat. Dans les systèmes basés sur des contraintes, la requête est modifiée par ajout, suppression, modification ou assouplissement de l'ensemble des prérequis de l'utilisateur. Pour les systèmes fondés sur des cas, soit la cible est modifiée par l'intervention de l'utilisateur, soit les résultats de la recherche sont élagués à l'aide d'avis critiques. Dans ceux-ci, l'utilisateur indique simplement si un attribut spécifique, présent dans les résultats de recherche, doit être mis en avant ou pénalisé.

Bien que la construction de la base de connaissances soit très dépendante du domaine, les interactions entre l'utilisateur et le système peuvent prendre trois formes :

- **Systèmes conversationnels** : les préférences sont déterminées grâce à une boucle de rétroaction dans lequel le système demande les avis de l'utilisateur itérativement.
- **Systèmes basés sur la recherche** : une séquence de question prédéfinies permettent au système d'obtenir les préférences de l'utilisateur.
- **Systèmes basés sur la navigation** : l'utilisateur spécifie un certain nombre de demandes de modification sur la solution actuellement proposée. Grâce à une suite

itérative de changements, il est possible d'arriver à un objet satisfaisant pour l'utilisateur.

Ces différentes formes d'orientation du processus sont plus ou moins adaptées en fonction de la méthodologie d'interaction mais elles peuvent aussi être combinées. Par exemple, un système de recommandation peut commencer par une stratégie de recherche pour établir un profil utilisateur de base et ensuite utiliser une technique basée sur la navigation pour affiner sa connaissance des besoins de l'utilisateur.

Avantages et inconvénients des approches basées sur les connaissances

Les travaux de AGGARWAL (2016b) et BURKE (2002) soulignent que les approches basées sur les connaissances ont des avantages proches de celles basées sur le contenu. Nous pouvons les résumer par les éléments suivants :

- **L'expression de besoins explicites** : la requête est formulée par l'utilisateur directement, ce qui offre une grande transparence du système puisqu'il permet un plus grand contrôle que les autres approches.
- **L'intégration des nouveaux objets** : une fois qu'un objet est ajouté au système, il peut être recommandé à l'utilisateur, puisque les techniques d'estimation de l'intérêt n'utilisent pas les évaluations *a posteriori* des utilisateurs.
- La **sensibilité aux changements** : le système peut prendre en compte les nouvelles préférences de l'utilisateur, puisqu'aucune donnée sur l'historique des recommandations n'est exploitée. Le problème du démarrage à froid est inexistant pour ce type de systèmes contrairement aux approches exposées précédemment.

Cette approche est très efficace lorsqu'il y a peu d'évaluations sur les objets ou quand ils sont complexes, mais elle amène aussi des défauts parmi lesquels :

- La **nécessité d'adaptation au contexte** : mettre en place cette approche requiert une fine connaissance du domaine pour identifier les propriétés importantes des objets, ainsi que les règles qui les gouvernent.
- La **complexité du processus** : proposer une solution satisfaisante est plus long du fait que l'utilisateur doit indiquer explicitement ses besoins. De plus, il peut être amené à effectuer plusieurs itérations avant que le profil utilisateur lui corresponde, ce qui rend l'utilisation plus lourde et chronophage.

La sur-spécialisation est aussi un inconvénient qui peut être présent dans ces systèmes, mais, comme le note AGGARWAL (2016b), pour les méthodologies fondées sur des cas, la

métrique peut combiner une mesure de la similarité et de la distance. Ainsi, ces systèmes peuvent offrir une plus grande diversité, mais il sera tout de même complexe de provoquer un effet de surprise.

2.3 Conclusion

Un système de recommandation a pour objectif d'aider un utilisateur dans son processus de décision, en lui proposant une solution ou en le guidant dans un éventail de possibilités qui est souvent vaste et complexe. De nombreux travaux formulent ce même problème de filtrage de l'information et ce, de diverses manières. Ces techniques ont des forces et des faiblesses complémentaires et choisir la plus appropriée pour un service dépend de la vision du concepteur et des données dont il dispose.

Un autre aspect à prendre en compte est l'explicabilité et la confiance dans les systèmes de recommandation. En plus de l'efficacité, RICCI, ROKACH et SHAPIRA (2011) note qu'un système doit être transparent, « scrutable » et persuasif, afin que les utilisateurs aient confiance. C'est-à-dire qu'il doit être possible d'expliquer son fonctionnement, permettre aux utilisateurs de dire qu'il est erroné et être capable de donner les raisons l'ayant conduit à la solution qui leur est recommandée.

Pour conclure, ce chapitre constitue un aperçu des principes et techniques employés pour effectuer de la recommandation personnalisée. Les ouvrages de RICCI, ROKACH et SHAPIRA (2011) et AGGARWAL (2016a) détaillent les points que nous avons abordés et les formalisent mathématiquement.

Nous avons précisé en introduction que nous souhaitons travailler à partir des préférences exprimées par l'utilisateur et en ce sens, nos travaux suivent les principes de la recommandation basée sur les connaissances.

Chapitre 3

L'élaboration de visites culturelles

Sommaire

3.1	Introduction	30
3.2	La recommandation de visites muséales	30
3.2.1	La nécessité d'une orientation pendant la visite	30
3.2.2	Systèmes d'aide pour les visites de musées	31
3.3	Le tourisme et les parcours en ville	33
3.3.1	Le tourisme urbain et ses spécificités	33
3.3.2	Le séjour touristique comme un parcours géographique	36
3.3.3	La loi de l'apogée-fin	39
3.4	La recommandation de parcours touristiques	41
3.4.1	Les systèmes de recommandation touristique	41
3.4.2	L'optimisation de voyages touristiques	43
3.5	Conclusion	44

3.1 Introduction

De nombreux travaux mettent en évidence le besoin d'assistance dans le domaine du tourisme. Au musée, la préparation de la visite est fortement associée à Internet et l'utilisation du Web est essentielle pour tisser des liens avec de nouveaux publics (POLI 2014 ; ANDREACOLA 2015). Les visiteurs accordent une place grandissante de l'affectif dans leurs attentes et nous pouvons y voir une nécessité de personnalisation de leur expérience pour laquelle un système de recommandation semble être une solution adaptée.

En ville, les enquêtes soulignent que la création d'un système de recommandation répond à un besoin réel de la part des touristes pour palier à l'infobésité (BÉGUIN 2015). Bien que la notion de parcours puisse renvoyer à un itinéraire physique à suivre précisément, ceux-ci ne laissent que peu de place à la flânerie qui est une pratique essentielle à la découverte de la ville. De plus, l'ordre dans lequel un touriste effectue des activités revêt une importance capitale pour la construction du souvenir et la satisfaction qu'il retirera de son voyage.

Ce chapitre présente dans un premier temps les travaux de recherche sur les visites culturelles dans les musées en s'appuyant sur la nécessité de systèmes de recommandation à partir d'études de sociologie. Dans un second temps, nous explorons le concept de tourisme en ville et les travaux provenant de nombreux domaines qui s'y rattachent. Enfin, nous verrons le fonctionnement de différents systèmes de recommandation de parcours touristiques, puis les techniques de recherche opérationnelle utilisées pour modéliser ce problème.

3.2 La recommandation de visites muséales

3.2.1 La nécessité d'une orientation pendant la visite

Les ouvrages de GILMAN (1916) et MELTON (1935) décrivent un important problème lors des visites de musées, la « fatigue muséale » qui se caractérise par une diminution de l'intérêt du visiteur au fur et à mesure de sa visite. Ce phénomène s'exprime par divers comportements de la part du visiteur, celui-ci explore de plus en plus vite la collection, sans faire de pause, et fait preuve d'une sélectivité accrue à l'égard des expôts.

À partir de différentes études de la littérature, DAVEY (2005) identifie deux sources d'explications qui sont les attributs du visiteur et les facteurs environnementaux comme l'arrangement des expôts dans le musée ainsi que leur taille. Les hypothèses concernant

les attributs du visiteur font état de la fatigue physique, qui s'accroît durant la visite (MELTON 1935), de la place de la visite dans leur emploi du temps mais aussi de la capacité d'attention du visiteur. Cette ressource cognitive entraînerait de la fatigue due à une baisse de la capacité d'attention pour les expôts, à chaque étape de la visite. Les études de FALK et al. (1985) montrent que cet effet est perceptible après 30 à 45 min dans le musée. Selon SCHOUTEN (1987), cet effet est partiellement dû à un lien insuffisant entre les œuvres et la perception du monde par le visiteur.

Du point de vue de ROUNDS (2004), la visite d'une collection entière ne permet pas au visiteur de se concentrer sur les œuvres qui l'intéressent. Il note qu'une visite partielle des collections est une stratégie intelligente et efficace pour le visiteur dont le but est d'éveiller et de satisfaire sa curiosité. De plus, le comportement sélectif minimise l'effort de compréhension et donne plus le temps au visiteur d'apprécier les expôts. Nous pouvons noter que l'ouvrage de KAHNEMAN (1973) montre que les personnes ont tendance à se concentrer sur une seule chose à la fois, ce qui appuie la nécessité d'une forme de recommandation des expôts les plus pertinents pour un visiteur. En effet, pour FALK (2009) les visiteurs diffèrent les uns des autres et leur expérience de visite dépend d'aspects qui relèvent de leur identité personnelle.

3.2.2 Systèmes d'aide pour les visites de musées

L'étude de ARDISSONO et al. (2011) détaille des travaux de recherche autour de la personnalisation dans le domaine des visites de musées. Les auteurs font la distinction entre les systèmes, selon plusieurs critères, comme le moment où il interagit avec l'utilisateur (avant ou pendant la visite) ou la méthodologie de construction du profil utilisateur (à partir de préférences, de stéréotypes, etc.). Cette étude est complétée par celle de KOSMOPOULOS et al. (2018) qui se concentre plus sur l'utilisation de la localisation du visiteur dans le musée pour améliorer le profil utilisateur.

Les systèmes de recommandation permettant de planifier une visite avant l'arrivée au musée sont historiquement les premiers à avoir été développés. Aujourd'hui, le système de ROES et al. (2009) est une référence dans la littérature qui propose un parcours en se basant sur une analyse des choix de l'utilisateur pour déterminer ses préférences. Un premier système aide l'utilisateur à découvrir le musée en ligne et permet de lui associer un profil en lui demandant de noter des catégories ou des tableaux qui lui sont présentés. Ces préférences sont ensuite utilisées par un deuxième système construisant un parcours dans le musée qui peut être modifié par l'utilisateur en exprimant des contraintes sur sa durée ou le nombre d'expôts à voir. L'inconvénient majeur de ce système se situe dans l'étape de création du profil utilisateur, car elle demande au futur visiteur de parcourir les

œuvres de la collection sans qu'un pré-filtrage efficace ne soit vraiment appliqué.

Concernant les dispositifs d'aide pendant la visite, nous remarquons une tendance pour les outils suivant la position du visiteur, afin de lui offrir une expérience personnalisée. L'utilisation de codes barres a été une première étape en permettant à un dispositif d'adapter le contenu multimédia (texte et audio) en fonction du profil utilisateur (ROUILLARD 2008). D'autres systèmes s'affranchissent de la nécessité d'une interaction avec l'expôt et exploitent les technologies sans fil pour déterminer l'expôt que le visiteur est en train de regarder. Le problème du démarrage à froid (voir le chapitre 2) est évité en demandant au visiteur de remplir un questionnaire pour déterminer ses préférences et créer un profil utilisateur de base, qui sera enrichi à partir de ses déplacements au sein du musée.

Les travaux de CHIEN et al. (2013) ou de RUOTSALO et al. (2013), par exemple, exploitent ces technologies pour déclencher une interaction avec le visiteur (visualisation du contenu ou écoute de l'enregistrement de l'expôt). La recommandation quant à elle s'effectue à partir du profil du visiteur qui est influencé par les expôts devant lesquels le visiteur s'arrête et consomme du contenu. Une base de connaissances sous la forme d'une ontologie de la collection du musée permet de faire le lien entre les caractéristiques des œuvres et les préférences du visiteur.

Par ailleurs, les travaux de VÉRON et al. (1983) établissent une typologie des visiteurs en observant leur trajectoire et la fréquence de leurs arrêts, ce qui les mène à décrire quatre profils à l'aide de métaphores animales : (i) les fourmis ; (ii) les poissons ; (iii) les sauterelles et (iv) les papillons. Chacun de ces profils de visite correspond à un schéma de mouvement spécifique dans l'espace physique des salles du musée qui traduit leur intérêt pour les expôts devant lesquels ils s'attardent.

Les fourmis passent un temps assez long devant chaque œuvre en restant près de celles-ci, mais évitent les espaces vides dans la salle. Les poissons quant à eux s'arrêtent dans les espaces vides et évitent les zones proches des expôts. Les sauterelles sélectionnent certaines œuvres devant lesquelles elles passent beaucoup de temps mais ignorent le reste de l'exposition. Enfin, les papillons observent presque toute la collection mais accordent des durées très variées à chaque objet. Cette classification est validée par les travaux de ZANCANARO et al. (2007), qui utilisent un algorithme de partitionnement non supervisé sur les trajectoires des visiteurs d'un musée. Ainsi, à partir des mouvements du visiteur dans le musée, il serait donc possible d'estimer son intérêt pour les expôts de la collection.

Le système de LYKOURENTZOU et al. (2013) combine plusieurs attributs dans leur modélisation du visiteur et notamment ces profils de visite pour lesquels ils proposent une fonction indiquant le temps passé par le visiteur devant chaque expôt en fonction de

l'intérêt qui lui est donné par une distribution de probabilités à partir des préférences de l'utilisateur. Le style de visite et les préférences de l'utilisateur sont obtenus en plusieurs étapes (NAUDET et al. 2013), notamment à l'aide d'un jeu en ligne permettant d'associer l'utilisateur à un style visite et un questionnaire pour identifier ses préférences vis à vis du contenu culturel. Les résultats montrent une amélioration de l'expérience de visite en simulations, mais les travaux sur la détermination du style de visite sont encore à un stade expérimental et l'ensemble du système (construction du profil et recommandation) n'a pas encore été évalué. De plus, la quantité de données nécessaires pour mettre en place ce système de recommandation est importante puisqu'il requiert pour chaque expôt, sa position précise dans la salle d'exposition, ainsi que le nombre de visiteurs pouvant l'observer en même temps.

Les outils présentés ici utilisent tous les préférences des visiteurs pour leur fournir des recommandations en adéquation avec leurs attentes ou leur style de visite. L'utilisation des mouvements des visiteurs pour en déduire la suite de leur visite est une piste intéressante, car elle permet de réduire la phase d'initialisation du profil en soumettant un questionnaire moins important mais toujours présent. Il est intéressant de constater que les premiers systèmes dans le domaine (STOCK 1993 ; OBERLANDER et al. 1998) se basaient sur le traitement automatique de la langue naturelle écrite pour produire des pages Web personnalisées. Les descriptifs des expôts ne sont aujourd'hui plus utilisés pour la recommandation personnalisée, mais nous pensons que leur analyse permettrait de résoudre le problème du démarrage à froid pour les visites de musées.

3.3 Le tourisme et les parcours en ville

La seconde partie de cette thèse se focalise sur le tourisme en ville, ce qui constitue un changement d'échelle tant au niveau spatial que temporel. Avant tout, nous décrivons la notion de tourisme et plus particulièrement les aspects qui caractérisent. À partir de ces éléments, nous exposons des concepts issus de la géographie qui permettent de décrire les comportements des touristes pendant leur séjour. Finalement, nous nous concentrons sur les processus de décision afin d'améliorer l'expérience ressentie par le touriste, une fois son séjour terminé.

3.3.1 Le tourisme urbain et ses spécificités

L'ouvrage de GUNN et al. (2002) décrit le tourisme comme un système complexe constitué des groupes suivants : (i) les attractions ; (ii) les services et installations ; (iii) le transport ;

(iv) les informations et l'orientation ; (v) les touristes eux-mêmes. Les attractions constituent le composant le plus important de l'offre touristique ayant pour but de satisfaire le touriste, que ses motivations soient la découverte d'une nouvelle culture, le repos ou l'amusement. Les services et les installations fournissent l'accès à l'hébergement, à la restauration et aux boutiques de souvenirs et sont souvent associés aux attractions dans d'autres modèles. Le transport, bien qu'il soit un service, est défini séparément de par son importance et la complexité à considérer la multimodalité qui est nécessaire pour s'adapter aux différents trajets durant le séjour. Les informations et l'orientation permettent de guider et d'influencer (au moyen de promotions par exemple) le touriste durant son voyage. Enfin, les touristes sont au centre de ce système et se différencient par de nombreux traits, tels leurs centres d'intérêts ou leur culture.

Cette définition du tourisme met en avant la complexité et la diversité des éléments qu'un système de recommandation peut être amené à prendre en compte pour conseiller un utilisateur. Néanmoins, cette description est généraliste et se concentre sur des aspects fonctionnels où le touriste n'est pas mis en avant par rapport aux autres éléments. Or, les recommandations s'adressent aux futurs touristes et, par conséquent, il est important de se centrer sur leurs pratiques.

Du point de vue du visiteur, nous pouvons reprendre la définition du tourisme de l'ouvrage « *Compte satellite du tourisme : recommandations concernant le cadre conceptuel 2008* », qui est la suivante :

« Le tourisme comprend les activités déployées par les personnes au cours de leurs voyages et de leurs séjours dans les lieux situés en dehors de leur environnement habituel pour une période consécutive qui ne dépasse pas une année, à des fins de loisirs, pour affaires et autres motifs non liés à l'exercice d'une activité rémunérée dans le lieu visité. » (*Compte satellite du tourisme – 2017*)

C'est cette notion d'activité de loisir durant un voyage qui nous intéresse principalement pour établir notre stratégie de recommandation. La littérature développée par la suite se concentre donc sur ces aspects des habitudes touristiques en ville.

L'importance de la culture et des promenades

Le mémoire de BÉGUIN (2015) sur le tourisme urbain note qu'il se caractérise par un intérêt renforcé des individus pour les ressources culturelles et patrimoniales de la ville. Grâce à la ville, le visiteur découvre des témoignages de l'histoire qui amènent un folklore et ces sorties culturelles sont vécues comme un moment récréatif permettant de s'évader (LEQUIME 2010). Les travaux de SIMON (2010) sur le même sujet soulignent que bien que le

touriste ait en amont une idée de son voyage, celui-ci n'est pas fixé de manière définitive, le projet initial dessine les contours du séjour et laisse des possibilités d'ajustement plus ou moins ouvertes. Cette improvisation est une des qualités premières de l'espace urbain selon KRACAUER (1995), pour qui la valeur d'une ville se mesure au nombre de lieux qui lui sont dédiés.

La flânerie est un élément important de cette improvisation, c'est au gré du hasard qui le guide qu'un touriste peut être surpris. Selon AMAR et al. (2009), la marche est perçue par le visiteur comme une figure de mobilité traduisant son désir d'autonomie et de mobilité. Les auteurs estiment qu'elle ne vise pas nécessairement à atteindre une destination précise, mais s'apparente à une expérience multisensorielle dans un espace-temps qui est un terrain propice à l'innovation. Il est intéressant de noter que les études de WASTIAUX (2009) révèlent que la promenade en ville est l'activité touristique la plus pratiquée (par 88 % des touristes), faisant de la marche à pied le moyen de transport le plus courant en tourisme urbain.

La marche en ville et l'évaluation du potentiel piétonnier

La marche en ville est donc un aspect important du tourisme urbain, d'une part parce qu'elle est considérée comme une activité touristique à part entière, mais aussi car elle représente l'essentiel des déplacements du visiteur. Or, des travaux initiés par SPECK (2013) autour de la convivialité d'un endroit pour la marche à pied (ou *walkability*) ont initié un nouveau domaine de recherche dont les nombreux travaux montrent la complexité de la tâche de prédiction d'un itinéraire précis à travers une ville.

L'étude de GORRINI et al. (2018) explore cette notion de « marchabilité » dans le contexte du tourisme en ville et considère que le comportement des touristes à pied est principalement influencé par leur exploration de l'environnement urbain. Ils proposent de regrouper les critères d'évaluation du potentiel piétonnier dans les catégories suivantes : (i) l'utilité ; (ii) le confort ; (iii) la sécurité ; (iv) l'attrait et (v) la lisibilité. L'utilité traduit la diversité des bâtiments et services à proximité. Le confort correspond à la qualité de la zone dans laquelle les piétons circulent (présence de bancs, largeur des trottoirs, etc.). La sécurité permet au piéton d'être confiant dans sa progression et des aspects comme la présence d'éclairage, de passages cloutés et l'absence d'obstacle sont des critères mis en avant. L'attractivité est liée à la présence de points d'intérêt pour le touriste et la qualité architecturale du paysage (espaces verts, places, etc.). Enfin, la lisibilité donne au visiteur la possibilité de s'orienter dans l'espace. La mesure de ce potentiel piétonnier peut être effectuée par l'observation de terrain, au moyen de questionnaires ou par analyse du déplacement des visiteurs. GORRINI et al. (2018) note aussi l'existence d'outils collaboratifs

ou commerciaux fournissant ces informations.

Cependant, bien que des mesures du potentiel piétonnier existent, prédire l'itinéraire qu'un touriste empruntera pour se rendre d'un lieu à l'autre (par exemple de son hébergement à une attraction) est un sujet complexe qui dépend de l'objectif du touriste, ainsi que des possibles éléments qui se trouveront sur son trajet (qui pourraient l'inciter à choisir un restaurant plutôt qu'un autre, ou lui donner l'envie de s'asseoir dans un parc...).

Le tourisme urbain est donc marqué par la découverte du patrimoine culturel, notamment au moyen de la marche, qui constitue à la fois le moyen de transport le plus courant, mais aussi l'activité privilégiée des touristes. Toutefois, planifier l'itinéraire que suivra le visiteur reste une question ouverte qui nécessite de prendre en compte de nombreux paramètres.

3.3.2 Le séjour touristique comme un parcours géographique

Nous pouvons donc voir un parcours touristique comme une combinaison subtile de points d'intérêt (que représentent les attractions, les restaurants, etc.) à atteindre dans un espace de liberté laissant la place à l'improvisation du touriste. Cette description nous amène à considérer la façon avec laquelle un touriste alloue son temps dans l'espace qui l'entoure.

La géographie temporelle ou *Time Geography*

HÄGERSTRAND (1970) décrit un concept d'analyse des activités et déplacements d'une personne par une relation entre l'espace et le temps. L'auteur part du postulat intuitif qu'il existe des limites spatiales et temporelles aux activités humaines. En effet, une personne ne peut physiquement se trouver qu'à un seul endroit à la fois et les activités sont, quant à elles, dispersées et se déroulent à des moments différents. Aujourd'hui, la géographie temporelle (ou *time geography*) est très populaire (ELLEGÅRD 2018). Le modèle qui est proposé reconnaît trois types de contraintes sur les activités humaines :

- les **contraintes de capacité** : elles limitent les activités des individus en raison de leurs capacités physiques, mentales et des ressources dont ils disposent. Des exemples comme les situations de handicap, les ressources financières ou les moyens de transport accessibles en font partie.
- les **contraintes de couplage** : elles définissent où, quand et combien de temps plusieurs individus doivent se rejoindre pour effectuer des activités communes. On peut citer le travail en entreprise ou les sorties entre amis par exemple.

- les **contraintes d'autorité ou de pilotage** : ce sont des relations de pouvoir propres au domaine qui limitent les actions des individus. On peut notamment identifier celles provenant de lois, de règles ou de normes sociales, dans des cas comme les horaires d'ouverture qui restreignent l'accès à un bâtiment ou l'autorité parentale à laquelle est soumise un enfant.

Une activité est souvent disponible pour une durée limitée et dans un nombre restreint de lieux et pour y participer, un individu doit allouer le temps nécessaire pour s'y rendre et la réaliser à un moment où elle est disponible. Les travaux de HÄGERSTRAND (1970) introduisent aussi la visualisation dite en aquarium spatio-temporel qui est illustré par la figure 3.1.

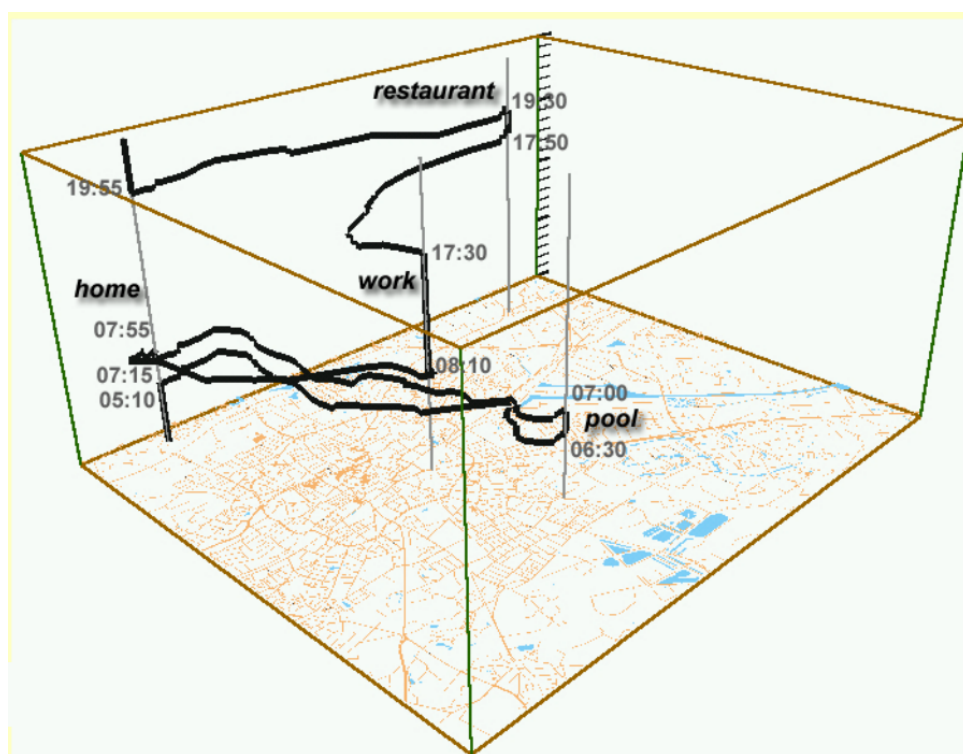


FIGURE 3.1 – Représentation 3D dans l'espace et le temps des activités d'une journée (par KRAAK 1988)

Dans cette représentation, l'espace physique est donné par le plan à deux dimensions tandis que le temps s'écoule selon le 3^e axe. Le comportement d'un individu est retracé par un chemin dont les composantes verticales symbolisent les activités entre les horaires de début et de fin et les segments obliques rendent compte des mouvements entre ces activités (CHARDONNEL et al. 2012). Ces travaux mettent en évidence les relations entre le comportement des individus et les contraintes liés à l'espace et au temps.

Les actogrammes

Les actogrammes, littéralement des programmes d'activités selon THÉVENIN et al. (2007), sont issus de l'étude des populations aux comportements de mobilité variant en fonction de composantes sociales, en plus des contraintes spatio-temporelles que nous venons de voir. Ils rendent compte des enchaînements d'activités et de déplacements et permettent d'identifier des motifs et des séquences de comportements des individus. Leur représentation, illustrée par la figure 3.2, traduit à la fois des actions stationnaires que sont les activités (au sens de HÄGERSTRAND (1970)) et les mouvements, mais au détriment d'une partie de la visualisation spatiale des comportements.

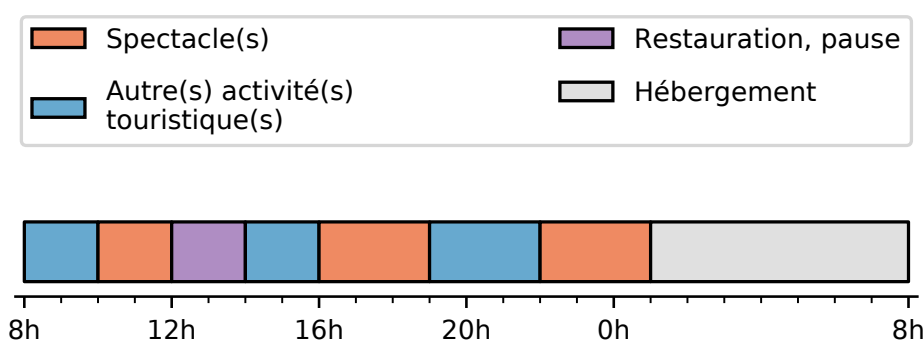


FIGURE 3.2 – Représentation d'un actogramme correspondant à un profil de visiteur du « Festival OFF d'Avignon » (voir le chapitre 6)

Le mémoire de PICOLOT (2016) note que les actogrammes sont construits à partir d'enquêtes et d'analyses multivariées, suite à des hypothèses sur les populations étudiées. Ces travaux présentent aussi l'utilisation de cette méthodologie pour créer des profils types de visiteurs du « Festival OFF d'Avignon ». L'intérêt notable des actogrammes est, d'une part, qu'ils intègrent l'espace et le temps dans le déplacement (on connaît l'activité réalisée, sur une période de temps ou dans des lieux ou des portions d'espace donnés). D'autre part, ils permettent de considérer différents types de touristes, de l'individu aux groupes, en passant par les familles, enrichissant ainsi la définition des profils utilisateur.

Ces actogrammes pourraient aussi être élaborés par des techniques de reconnaissance d'activité, qui permettent d'estimer la séquence des points d'intérêt d'un touriste (MINNEN et al. 2006). De plus, l'utilisation importante des smartphones permet de faciliter la reconnaissance de certaines activités comme le montrent SU et al. (2014).

La géographie temporelle et des actogrammes permettent de décrire le comportement d'un touriste durant son séjour et ainsi identifier les moments qu'il s'octroie pour promener et découvrir la ville. Utiliser ces informations pourrait donner à un système de

recommandation la capacité de proposer les activités pertinentes au moment adéquat en fonction des habitudes du visiteur et des attractions disponibles.

3.3.3 La loi de l'apogée-fin

L'objectif d'un système de recommandation est de fournir à l'utilisateur une proposition qui le satisfasse en fonction de ses besoins. Cette satisfaction dépend de la pertinence de la recommandation, qui sont ici les activités touristiques proposés pendant le séjour. Néanmoins, l'ordre dans lequel ces activités sont effectuées peut influencer le ressenti que le visiteur retirera du séjour.

Se souvenir ou vivre l'expérience

Les travaux de KAHNEMAN (2000) sur les processus de décision décrivent les individus comme constitués de deux entités :

- celle qui **fait l'expérience** (*experiencing self*) : est notre conscience de l'instant présent, qui pense intuitivement et rapidement.
- celle qui **raconte l'expérience** (*remembering self*) : collecte, passe en revue et intègre nos expériences en une histoire. Ce processus de création des récits qui constituent notre mémoire est marqué par de nombreuses modifications et interprétations.

Ces deux concepts amènent à définir deux notions du bien être, celui de l'instant présent et celui que la mémoire nous procure lorsque nous pensons à un évènement passé. Cependant, des études montrent que notre mémoire est construite à partir des moments intenses et significatifs d'une expérience, et ce quelle que soit sa durée. Cet effet, nommé *négligence de la durée* (*duration neglect*), amène les auteurs à proposer la **loi de l'apogée-fin** définie ainsi :

« La loi de l'apogée-fin est une heuristique psychologique dans laquelle les gens jugent une expérience en grande partie en fonction de ce qu'ils ont ressenti à son apogée (c'est-à-dire à son point le plus intense) et à sa fin, plutôt qu'en fonction de la somme totale ou moyenne de chaque moment de l'expérience. » (KAHNEMAN – 2000)

La figure 3.3 illustre cet effet dans le cadre d'un parcours touristique, le visiteur se remémorant l'activité la plus significative (intense), ainsi que la dernière qu'il effectue.

Ainsi, la loi de l'apogée-fin suggère que nos souvenirs sont le fruit d'un processus cognitif permettant de réduire la quantité de détails que notre mémoire doit conserver et ce faisant, de ménager nos ressources cognitives.

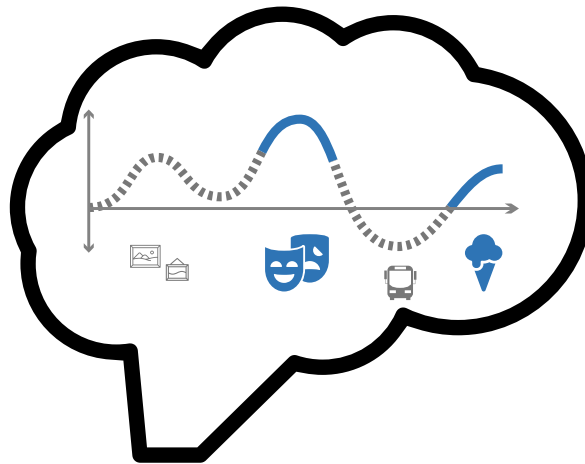


FIGURE 3.3 – Illustration du processus de sélection des expériences selon la loi de l'apogée-fin

Dans son dernier ouvrage, KAHNEMAN (2011) affirme à propos de cette entité qui se souvient que, bien qu'elle se trompe parfois, elle est celle qui gouverne ce que nous apprenons de la vie et la source de nos décisions.

Application au tourisme

L'ouvrage de KAHNEMAN (2011) décrit comment la durée de nos vacances a peu d'impact sur la façon dont nous nous en rappelons. Des vacances de deux semaines ont peu d'avantages par rapport à des vacances d'une semaine parce que de nouveaux souvenirs ne sont pas ajoutés au récit de l'expérience qui deviendra notre souvenir de ce voyage.

En effet, les travaux de KEMP et al. (2008) ont montré qu'allonger la durée des vacances n'améliore pas le ressenti de leur souvenir. Ils observent aussi que la quantité moyenne de bonheur exprimée par les participants au cours de leur séjour est significativement plus faible que celle dont ils se rappellent a posteriori. Ils notent que dans l'ensemble, les souvenirs se fondent sur un échantillon limité de moments qui se sont déroulés au cours de l'évènement passé.

L'étude de NAWIJN, MITAS et al. (2013) concluent que les vacanciers se sentent moins heureux à la fin de leur vacances et souligne que l'industrie du tourisme devrait essayer de créer une apogée durant les vacances, un moment de bien être intense qui serait mémorable.

La loi de l'apogée-fin nous incite à adopter une stratégie particulière lors de la planification d'un séjour touristique. En plus de penser l'ensemble des activités proposées, il faudrait aussi les organiser de sorte à ce que les plus pertinentes soient mémorables. Un système de recommandation tenant compte de cet effet peut améliorer le souvenir du séjour que gardera un touriste.

3.4 La recommandation de parcours touristiques

3.4.1 Les systèmes de recommandation touristique

L'étude de CHAUDHARI et al. (2019) détaille de nombreux systèmes qui sont répartis en fonction du type des objets (hôtels, restaurants, musées, etc.). Néanmoins dans la littérature, les notions d'attractions et de services sont aujourd'hui souvent traités indifféremment par certains systèmes comme celui de WANG et al. (2010) qui intègre monuments et attractions, mais également des hôtels et restaurants. Nous nous focalisons sur ces systèmes génériques ou exclusivement dédiés à la recommandation d'attractions.

L'étude de BORRÀS et al. (2014) distingue quatre fonctionnalités que peuvent offrir les systèmes de recommandation touristique : (i) suggestion d'une destination en forfait touristique ; (ii) recommandation d'activités ; (iii) planification d'itinéraires et (iv) les aspects sociaux.

La suggestion de destination ou de voyages organisés concerne une faible partie des travaux de recherche, le touriste décide souvent lui-même de sa destination et n'a donc pas recours à ce type de systèmes. Les fonctionnalités sociales sont quant à elles très souvent indépendantes des autres fonctionnalités, elles concernent le partage de photos ou de commentaires qui sont utilisés pour effectuer du filtrage collaboratif.

Certains systèmes de recommandation, en plus de suggérer des activités, préparent l'itinéraire que le visiteur devra emprunter tout au long de la journée. Le système de KURATA et al. (2013) en est un très bon exemple, tenant compte de l'aisance de marche et proposant des parcours en fonction des attentes du visiteur qu'il peut affiner en intégrant ou rejetant les propositions du système.

Or, comme nous l'avons vu plus tôt, la marche est elle-même une activité touristique et permet au touriste de découvrir la ville alors que ces systèmes l'utilisent uniquement comme un moyen de transport. Mais bien que la planification de l'itinéraire ne soit pas pertinent aux vues des travaux de recherche en sociologie et géographie présentés plus haut, cet aspect reste néanmoins important. En effet, afin qu'il ne soit pas proposé deux

activités se déroulant en parallèle ou pour lesquelles le temps nécessaire au touriste de les relier ne serait pas suffisant, un système de recommandation doit tenir compte de facteurs comme la durée estimée, les horaires d'ouvertures et les distances séparant les activités en plus du style et de la vitesse de marche du touriste.

La suggestion d'activités reste la principale fonctionnalité des systèmes de recommandation, BORRÀS et al. (2014) notent que cette tâche est particulièrement complexe à cause du large champs des possibles et du nombre de propriétés associés aux objets du système (les types de points d'intérêt qu'il manipule). Les préférences du visiteur, les horaires des activités et la localisation du visiteur sont évidemment très couramment considérés, aux côtés d'autres facteurs comme l'accessibilité des lieux (SANTOS et al. 2017), les conditions méthodologie (LAMSFUS et al. 2009), la popularité ou la diversité des points d'intérêt (BENOURET 2017).

Utiliser ces indicateurs implique la construction d'un profil utilisateur adéquat. BORRÀS et al. (2014) estiment que la plupart des systèmes de recommandation se basent sur des approches hybrides qui utilisent presque toutes une forme de recommandation fondée sur les connaissances. Ces connaissances sont représentées le plus souvent au moyen d'ontologies, des conceptualisations formelles du domaine sur lequel travaille le système. Celles-ci peuvent-être très génériques ou très spécialisées comme dans le système de RUOTSALO et al. (2013) dédié à la visite de musées. Déterminer les points d'intérêt pertinents pour le touriste nécessite de mettre en correspondance le profil ainsi formé et les objets du système. Ce calcul d'adéquation peut être une combinaison des divers facteurs cités plus tôt, comme dans les travaux de MEEHAN et al. (2013) qui proposent l'utilisation d'un réseau de neurones pour établir une pondération des catégories de préférences de l'utilisateur et une logique floue pour représenter l'intérêt des activités et les conditions météorologiques qui influent sur la distance qu'il accepte de parcourir pour y accéder.

L'approche de RODRÍGUEZ et al. (2012) expose un système de recommandation pour les séjours de plusieurs jours dont le modèle se base sur la combinaison de 4 critères : (i) minimiser la distance parcourue ; (ii) minimiser le coût de chaque activité ; (iii) maximiser l'utilité de chaque activité ; (iv) minimiser la différence entre le temps que le touriste s'attend à passer pour chaque type d'activité et celui qu'il y passe réellement. La minimisation de la distance est ici due au fait que ce système est prévu pour une utilisation dans une large zone géographique, dans laquelle le touriste se déplace avec son véhicule personnel. La minimisation du coût total est une somme du coût des activités pondéré par le type de celles-ci (hébergement, restauration et visite). La maximisation de l'utilité des activités est une somme de l'utilité de chaque activité proposée qui est calculée en fonction de son type (monument, promenade, etc.) et de l'importance que lui accorde le touriste. Le dernier critère ajuste le temps passé au cours du séjour sur chaque type d'activité (hébergement,

restaurant et visite) en fonction des préférences du touriste. Ce système permet aussi au visiteur de se détendre : d'une part en laissant la possibilité de fixer des journées libres pendant lesquelles aucune activité n'est prévue ; d'autre part, en exploitant une notion de rythme de séjour, qui permet au touriste d'influer sur la durée à attendre entre deux activités.

3.4.2 L'optimisation de voyages touristiques

Dans la littérature, la recherche de parcours touristique est connue sous le nom du problème de conception de voyage touristique (TTDP¹) qui a été introduit par VANSTEENWEGEN et al. (2007) comme un problème de recherche opérationnelle. Ce problème revient au minimum à donner une liste ordonnée de points d'intérêt qui respecte les contraintes liées au déplacement du visiteur et aux propriétés des activités, tout en correspondant le mieux possible aux préférences du touriste. Que cette planification comporte ou non la planification de l'itinéraire, ce processus de décision doit déterminer une séquence ordonnée et, selon BALAS (2007), peut être formulé comme un problème de voyageur de commerce (TSP²) ou une de ses variantes.

L'étude de GAVALAS et al. (2014) sur les approches pour la résolution du TTDP distingue les systèmes de recommandation proposant un seul parcours pour tout le séjour de ceux qui segmentent la problématique en journées. Cependant, comme nous l'avons vu plus tôt, construire des expériences jugées satisfaisantes nécessite d'avoir un point de vue global afin d'identifier les moments d'apogée et de fin. De ce fait, nous nous concentrons sur les travaux proposant un seul parcours sur tout le séjour. Les auteurs dressent une hiérarchie des approches utilisées pour formuler le TTDP dans le cadre d'une unique proposition pour l'ensemble du séjour, les principales sont les suivantes :

- Le problème de voyageur de commerce collectant des prix (PCTSP³), qui vise à trouver un parcours qui minimise le coût associé au séjour (en terme de distance ou de budget) tout en imposant un seuil de profit que les activités qui le composent doivent atteindre.
- Le problème d'orientation (OP⁴), qui cherche à maximiser le profit total apporté par les activités, tout en maintenant le coût du séjour en dessous d'une valeur donnée.
- Le problème du parcours rentable (PTP⁵), qui recherche un parcours qui maximise

1. *Tourist Trip Design Problem*
2. *Traveling Salesman Problem*
3. *Prize-Collecting TSP*
4. *Orienteering Problem*
5. *Profitable Tour Problem*

le bénéfice des activités du parcours, c'est à dire la différence entre le profit et le coût qui leur sont associés.

Toutes ces formulations sont des variantes du TSP avec profits (FEILLET et al. 2005), une généralisation à deux critères du TSP où il n'est pas nécessaire de visiter tous les sommets et pour lesquels un profil est associé à chacun. L'objectif global est l'optimisation simultanée du profit collecté et des coûts engendrés. Ces deux critères d'optimisation apparaissent soit dans la fonction objectif comme pour le PTP, soit sous forme de contrainte comme pour le PCTSP et l'OP qui transforment respectivement le profit et le coût en une contrainte. GAVALAS et al. (2014) notent que les approches par OP sont les plus utilisées dans la littérature et les plus proches de la définition initiale du TTDP.

Nous pouvons remarquer que les visites muséales sont vues comme des cas particuliers de visites touristiques et sont sujettes aux mêmes formalismes. C'est par exemple le cas du modèle proposé par BERRE et al. (2013) qui est une forme particulière d'un OP (la durée maximale de la visite est donnée en contrainte et le système détermine l'ensemble des expôts maximisant l'intérêt du visiteur).

Enfin, GAVALAS et al. (2014) recensent aussi les algorithmes proposés pour résoudre les différentes formulations exposées qu'ils soient exacts ou des heuristiques. Le système de WÖRNDL et al. (2017), qui adapte l'algorithme de DIJKSTRA (1959) pour proposer un itinéraire au visiteur montre par une expérimentation sur le terrain que les utilisateurs sont satisfaits de ce type de systèmes, bien que ce soit une simple heuristique.

3.5 Conclusion

La visite culturelle, qu'elle se fasse au musée ou en ville, est un sujet qui concerne de nombreux domaines de recherche. Nous avons présenté quelques uns des travaux issus des sciences sociales, sur lesquels nous nous appuyons, qui montrent la complexité des problématiques et donnent des pistes pour l'élaboration de nouvelles approches pour les systèmes de recommandation personnalisés.

Dans les musées, la fatigue muséale qui a été décrite il y a une centaine d'années, est aujourd'hui toujours présentée comme un problème auquel doivent répondre les conservateurs et leurs équipes. Nous avons présenté plusieurs approches, se basant notamment sur des études de sociologie des publics, qui décrivent les déplacements dans les salles d'exposition. Ces comportements peuvent être utilisés pour inférer les préférences du visiteur mais souffrent du problème du démarrage à froid et il pourrait être intéressant d'utiliser d'autres ressources déjà présentes dans les musées pour y répondre comme

les cartels ; ces descriptions textuelles rédigées par les équipes de médiation donnent les informations pertinentes de l'œuvre dans le contexte de l'exposition, c'est-à-dire vis à vis des autres expôts présentés.

En ville, le tourisme se caractérise bien évidemment par des attractions propres aux goûts de chacun, mais aussi par des moments où le visiteur se laisse aller à la flânerie. Celle-ci représente l'activité touristique prédominante qui donne la possibilité de découvrir de nouveaux lieux. De plus, les travaux sur l'évaluation du potentiel piétonnier donnent un aperçu de la complexité à recommander des itinéraires précis à un touriste. D'autres travaux en géographie temporelle dont sont issus les actogrammes permettent de mettre en évidence les conditions spatio-temporelles nécessaires aux activités humaines et d'établir des profils types. Cette notion d'activité regroupe aussi bien les attractions que les autres moments du séjour (hébergement, restaurant, temps libre, etc.) et permet donc de représenter un emploi du temps type dans lequel insérer des propositions.

Les systèmes de recommandation de visites utilisent des formalismes dérivés du TSP bien établis de la recherche opérationnelle, pour suggérer au touriste les activités à effectuer. Ce processus nécessite la construction d'un profil utilisateur qui peut être parfois complexe, car pouvant contenir de nombreuses informations sur le contexte de la visite. Cependant, l'organisation globale du parcours semble être un élément délaissé alors qu'il se pourrait qu'il soit primordial. En effet, les travaux autour de la loi de l'apogée-fin indiquent la nécessité d'organiser les parcours afin que le touriste en retire la meilleure expérience (NAWIJN, MITAS et al. 2013). Cette heuristique psychologique reste néanmoins absente des modèles et algorithmes actuels alors qu'elle pourrait constituer un élément majeur de l'expérience de visite.

Partie II

Méthodes d'aide à la visite culturelle au musée

Chapitre 4

L'analyse de la médiation muséale pour la personnalisation de visites

Sommaire

4.1	Introduction	50
4.2	Principes pour l'estimation de l'intérêt des œuvres	50
4.2.1	Processus d'évaluation	51
4.3	Adéquation avec des préférences personnelles	52
4.3.1	Représentation du musée	52
4.3.2	Prédiction de l'intérêt pour le visiteur	52
4.4	Mesure indépendante du prestige des expôts	54
4.4.1	Classement des œuvres et analyse textuelle	54
4.4.2	Adaptation du résumé automatique de texte pour mesure le prestige des expôts	54
4.5	Expériences et résultats	56
4.5.1	Musées étudiés	56
4.5.2	Système de résumé automatique de texte	57
4.5.3	Application	58
4.6	Conclusion	60

4.1 Introduction

Ce chapitre présente des travaux préliminaires menés sur les musées ayant pour objectif d'estimer l'attrait d'un expôt pour un visiteur. En effet, les travaux de GILMAN (1916) sur la fatigue muséale décrivent une réduction de l'intérêt du visiteur à mesure qu'il progresse dans le musée, se caractérisant par un comportement l'incitant à passer de moins en moins de temps devant les expôts et à sortir du musée prématurément. Un système de recommandation semble tout indiqué dans ce contexte car il permettrait de guider le visiteur vers les salles contenant les expôts qui l'intéressent le plus.

Comme nous l'avons vu dans le chapitre 3, le profil du visiteur peut être déduit en fonction de ses déplacements ou en lui soumettant un questionnaire. Toutefois, ces approches sont chronophages pour le système (recueillir les déplacements) ou pour l'utilisateur (répondre au questionnaire avant la visite) et bénéficieraient de l'apport d'une estimation indépendante pour aiguiller le visiteur.

Nous proposons donc d'établir une nouvelle mesure de la popularité des œuvres qui traduise le point de vue du musée lui-même. À partir des textes de l'institution décrivant la collection, nous estimons l'importance de chaque expôt par rapport aux autres. D'une part, cette tâche n'implique aucune action supplémentaire pour le visiteur mais elle permet aussi d'intégrer au profil une indication sur les éléments phares de la collection.

Attribuer une mesure de l'importance de chaque expôt grâce au texte qui le décrit constitue un défi sur la création d'une nouvelle métrique que nous proposons, comme le **prestige de l'œuvre**. Dans ce chapitre, nous présentons l'utilisation de méthodes issues de la recherche d'information et du traitement automatique de la langue naturelle écrite qui nous permettent d'estimer l'intérêt d'inclure un expôt plutôt qu'un autre lors de la proposition d'un parcours culturel à un visiteur.

4.2 Principes pour l'estimation de l'intérêt des œuvres

Nos travaux se basent sur les attentes et envies exprimées par le visiteur, c'est-à-dire les goûts qu'il juge pertinents et choisit de communiquer à un système de recommandation, afin que celui-ci lui conseille un contenu culturel approprié. Comme nous l'avons vu dans le chapitre 2, cette approche basée sur la connaissance a deux défauts majeurs : (i) la limitation de l'analyse du contenu et (ii) la sur-spécialisation. Pour contrer ces deux inconvénients, nous introduisons une mesure qui se base sur le contenu textuel des expôts dont le but est de compléter les préférences personnelles afin de différencier les œuvres jugées équivalentes.

4.3 Adéquation avec des préférences personnelles

Le point primordial concernant notre stratégie de recommandation est la personnalisation du contenu en fonction des goûts de l'utilisateur. Les expôts sélectionnés doivent donc être en accord avec les affinités culturelles du visiteur, ce qui nécessite de pouvoir représenter et comparer tous ces éléments.

4.3.1 Représentation du musée

En recherche d'information, le modèle vectoriel, introduit par SALTON et al. (1975), est l'approche prédominante de représentation des documents en vue d'une utilisation par des tâches d'accès à l'information (AMINI et al. 2013). L'assimilation d'expôts, ou plus largement d'activités touristiques (ou points d'intérêt culturel), à des documents, se fait couramment par le biais de concepts leur étant attribués comme dans les travaux de ROES et al. (2009) dans les musées, ou de WÖRNDL et al. (2017) pour les visites en ville.

En récoltant les caractéristiques que le conservateur et ses collaborateurs attribuent à chaque expôt, nous pouvons définir un ensemble C_e par expôt e et C l'union de ces ensembles rassemble alors toutes les concepts utilisés pour décrire les œuvres d'un musée.

À partir de ces ensembles, nous pouvons reproduire le modèle vectoriel en projetant les expôts dans un espace vectoriel où chaque dimension représente une caractéristique de l'ensemble C . Ainsi, chaque œuvre est modélisée par un vecteur $v_e = (c_1, \dots, c_n)$ dont chaque élément c_i représente l'adéquation entre l'expôt et la caractéristique correspondante dans C .

Cette représentation permet de traduire une importante variété de concepts. Des attributs simples comme l'auteur d'une œuvre ou la technique picturale (peinture à l'huile, aquarelle...) peuvent se traduire par une valeur binaire tandis que les caractères plus complexes comme le courant artistique (impressionnisme, cubisme...) peuvent être exprimés par une valeur normalisée dans l'intervalle $[-1, +1]$, reflétant l'adéquation entre l'expôt et la caractéristique.

4.3.2 Prédiction de l'intérêt pour le visiteur

En recherche d'information, la pertinence d'un document est calculée en fonction de la requête de l'utilisateur (AMINI et al. 2013), qui dans notre cadre applicatif, traduit les attentes qu'il communique au système de recommandation.

Modélisation des goûts culturels

Une solution évidente est utilisée par d'autres systèmes de recommandation de visites cités plus tôt pour projeter le visiteur dans l'espace vectoriel formé par les caractéristiques C du musée. Nous choisissons ici de laisser au visiteur la possibilité de sélectionner parmi les caractéristiques C du musée celles qui l'importent et de noter son affinité pour chacune d'elle, qu'elle soit positive ou négative.

Cette approche a l'avantage de produire une représentation des préférences du visiteur dans le même espace vectoriel que celui des expôts sans requérir une étape de traduction supplémentaire. De plus, elle permet de générer facilement des profils touristiques réalistes à l'aides d'heuristiques simples et de tirages aléatoires.

Estimation de l'intérêt des expôts

À partir de cette projection dans un espace vectoriel, il nous est possible d'estimer l'intérêt de chaque œuvre en comparant le vecteur qui lui est associé à celui représentant les préférences du visiteur. Cette mesure peut-être faite au moyen de la *similarité cosinus*, très répandue en recherche d'information, qui calcule l'angle formé par les deux vecteurs. Conséquemment, l'équation (4.1) donne pour un expôt e l'intérêt de l'inclure dans un parcours en fonction de la représentation vectorielle v_e de ce dernier ainsi que celle des préférences culturelles du visiteur v_p .

$$\text{interest}(e) = \frac{v_e \cdot v_p}{\|v_e\| \times \|v_p\|} \quad (4.1)$$

Le résultat de cette mesure varie dans l'intervalle $[-1, +1]$, -1 signifiant que le visiteur n'est pas du tout intéressé par l'expôt, 1 qu'il correspond exactement à ses attentes et 0 que le visiteur n'a pas de préférence a priori sur cette œuvre. Notons que cette dernière valeur est la borne inférieure lorsqu'aucun des composants c_i des vecteurs v_e ou v_p n'est négatif (ce qui est le cas pour les attributs à classification binaire).

Cette approche pour l'estimation de l'intérêt d'un visiteur pour les expôts d'une collection présente plusieurs avantages. Contrairement à l'utilisation de simples distributions mathématiques, son utilisation est possible à la fois dans des conditions réelles et lors de simulations. Durant les phases de tests, le vecteur des préférences peut être tiré d'enquêtes ou généré en sélectionnant des caractéristiques du musée, ce que fait un visiteur lorsqu'il est invité à exprimer ses goûts. En conséquence, le système calcule la pertinence de chaque expôt de la même manière, ce qui mène à des simulations plus réalistes, reflétant de véritables profils touristiques.

Néanmoins, si le visiteur déclare être intéressé par un grand nombre de caractéristiques, le système peut avoir du mal à différencier les expôts entre eux. Cette situation peut aussi se rencontrer si l'utilisateur n'a pas de préférences précises ou s'il souhaite découvrir le musée. Ce comportement peut être atténué en utilisant une autre mesure de l'intérêt, qui soit indépendante du visiteur.

4.4 Mesure indépendante du prestige des expôts

Les œuvres sont caractérisés par plusieurs attributs et nous venons de présenter les méthodes de recherche d'information permettant de les discriminer. Pour autant, ce sont des objets semi-structurés puisqu'aussi décrits par du texte libre, issu des équipes de médiation culturelle. Aussi, nous proposons d'adapter le fonctionnement de systèmes de résumé automatique de texte pour établir une mesure du prestige des expôts, à partir des textes qui les dérivent.

4.4.1 Classement des œuvres et analyse textuelle

En parcourant la littérature, nous constatons que le contenu textuel n'est utilisé que pour retrouver des informations structurées, le plus souvent par des méthodes liées au Web Sémantique (HEIDORN et al. 2008 ; BENOURET 2017). Or, le travail des conservateurs de musées et leurs équipes consiste aussi à présenter les expôts de manière précise et cohérente avec le reste de la collection à laquelle elles appartiennent.

Nous souhaitons donc ici exploiter l'important travail que représente la médiation muséale au moyen de méthodes bien établies en traitement automatique de la langue naturelle écrite afin de discriminer les expôts entre eux. En effet, nous pensons que les cartouches d'information rédigés par les équipes retranscrivent leur avis sur la place d'une œuvre parmi les autres et par la même une sorte d'importance qu'elle revêt par rapport aux autres. Une telle analyse textuelle permet de retranscrire le point de vue du musée sur sa propre collection aux visiteurs et donc à un système de recommandation d'intégrer cette vision aux parcours qu'il construit.

4.4.2 Adaptation du résumé automatique de texte pour mesure le prestige des expôts

Cette évaluation par classement à partir du texte peut être accomplie avec des méthodes de résumé automatique de texte. En effet, leur objectif est de proposer le contenu essentiel d'une connaissance particulière qui puisse être un véritable substitut du document d'origine (CLEVELAND et al. 1983).

L'ouvrage de TORRES-MORENO (2014) note qu'il existe plusieurs approches pour la génération de résumé, notamment les méthodes par extraction qui consistent à sélectionner des unités du texte original contenant l'information essentielle pour ensuite les assembler de manière adéquate. Cette sélection impose donc à un algorithme de diviser le texte, puis d'attribuer à chacun des segments textuels une valeur reflétant la quantité d'information qu'il porte.

Nous proposons d'utiliser ce principe pour classer les expôts en fonction de la description produite par les équipes de médiation en considérant le musée comme un corpus de textes où chaque segment est la description d'une œuvre. Un algorithme de résumé par extraction calculera donc pour chaque segment une pondération que nous considérons comme le prestige de l'expôt associé.

La figure 4.2 schématise le procédé utilisé pour obtenir un score pour chacun des expôts à partir de leur description en adaptant le fonctionnement d'un algorithme de résumé automatique de texte.

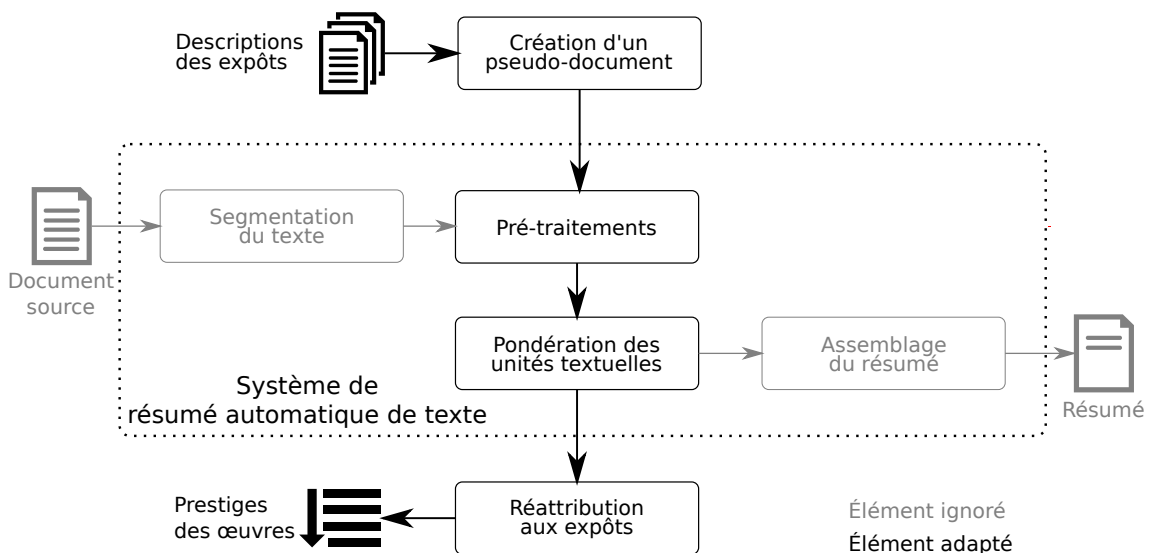


FIGURE 4.2 – Processus de calcul du prestige des expôts par adaptation du résumé automatique de texte

Contrairement à une utilisation classique, nous disposons d'une collection de textes associés aux expôts et non pas un document source unique. L'étape de segmentation est donc assurée par nos soins, la description de chaque expôt est utilisée comme une

seule unité textuelle à classer, bien qu'elle soit elle-même souvent composée de plusieurs phrases. À partir de ce pseudo-document, sont appliqués les pré-traitements habituels (MANNING et al. 1999 ; PORTER 1980), ce qui permet ensuite à l'algorithme d'effectuer une meilleure pondération des segments textuels. Sans continuer le processus de génération du résumé, les résultats sont récupérés et le score associé à chaque unité textuelle est attribué à l'expôt correspondant comme étant le prestige de cette œuvre.

Nous proposons donc de tirer partie des mécanismes utilisés par les algorithmes de résumé automatique de texte pour évaluer les expôts d'un musée. Leur fonctionnement les amène à évaluer l'importance de chaque partie d'un texte, ce que nous exploitons afin de déterminer pour chaque œuvre, une mesure de son importance dans la collection.

4.5 Expériences et résultats

L'application d'un modèle vectoriel et de la *similarité cosinus* est courante en recherche d'information et son application dans le domaine du tourisme est très répandue. Nous nous concentrerons ici à l'évaluation de l'efficacité du classement textuel appliqué au expôts d'un musée.

4.5.1 Musées étudiés

Nous avons appliqué la méthode de classement textuel décrite dans la section précédente à deux musées d'art : le « Musée de l'Orangerie », une galerie d'art de Paris et la « National Gallery », le 7^e musée le plus visité au monde.

Grâce aux sites Web officiels¹ des musées, nous avons collecté l'ensemble des expôts, une description pour chacun d'eux et une liste de chefs-d'œuvre selon le musée. La table 4.1 synthétise les informations caractérisant les corpus caractérisant chaque musée.

Nous pouvons constater tout d'abord la différence dans la taille des collections et la proportion de leurs chefs-d'œuvre. Le « Musée de l'Orangerie » expose une centaine de tableaux et présente 17 d'entre eux comme des chefs-d'œuvre soit 11,8 %. Les équipes de la « National Gallery » en identifie seulement deux fois plus alors qu'elle compte près d'un millier d'expôts, abaissant la proportion de chefs-d'œuvre à 2,9 %.

Les corpus textuels sont aussi très différents d'une institution à l'autre. D'une part les descriptions des expôts sont rédigés en français pour le « Musée de l'Orangerie » et

1. <https://www.musee-orangerie.fr> et <https://www.nationalgallery.org.uk>

	« Musée de l'Orangerie »	« National Gallery »
Nombre d'expôts	144	972
Nombre de chefs-d'œuvre	17	28
Langue des documents	Français	Anglais
Nombre de termes	7 632	94 308
Nombre de termes uniques	2 252	6 845
Taille des descriptions (moyenne — écart-type)	53,0 — 59,3	97,0 — 43,6

TABLE 4.1 – Principaux éléments caractérisant les musées étudiés

en anglais dans la « National Gallery ». D'autre part, le vocabulaire employé est plus varié dans le musée londonien, mais cette augmentation est reflétée par la taille des textes présentés. Cette différence peut être due au fait que la collection de la « National Gallery » regroupe plusieurs courants, tandis que celle du « Musée de l'Orangerie » se focalise autour de quelques peintres. Dans ce dernier, nous pouvons noter que l'écart-type dans la taille des descriptions traduit la présence d'une valeur abérante. En effet, les « Nymphéas » de Monet est une œuvre composée de six panneaux répartis sur deux salles pour lesquels nous n'avons eu accès qu'à une seule description, cependant bien plus fournie que les descriptions des autres expôts de la collection (518 termes dans cette description contre 49 en moyenne dans les autres).

Comme nous l'avons vu dans la section précédente, nous adaptons les algorithmes de résumé automatique de texte en substituant le document d'origine par un pseudo-document contenant déjà les unités textuelles à pondérer. Ce pseudo-document associé à chaque musée est constitué d'autant de segments textuels que d'œuvres et chacun d'eux est composé du nom de l'expôt, sa description et du nom du ou des artistes l'ayant réalisé.

4.5.2 Système de résumé automatique de texte

Nous avons utilisé ENERTEX, un système proposé par FERNÁNDEZ et al. (2007), qui ne requiert aucune connaissance codée par des règles spécifiques, que ce soit à propos du domaine ou de la langue du texte sur lequel l'algorithme est utilisé. Le même protocole a donc pu être appliqué aux deux musées, sans avoir recours à des ressources supplémentaires ou une étape de traduction qui altérerait les résultats.

Cet algorithme utilise un réseau de neurones de HOPFIELD (1982) où chaque neurone représente un terme du document à résumer. Un segment textuel est donc une configuration du réseau dans laquelle l'état de chaque neurone reflète la fréquence ou l'absence du

terme qui lui est associé. La pondération des segments textuels se fait en transposant des principes de mécanique statistique pour calculer l'énergie d'interaction entre les neurones d'une configuration à l'autre. Les auteurs définissent donc une notion d'énergie textuelle, qui représente les liens entre les segments textuels et permet d'identifier ceux qui portent l'information.

Nous pouvons aussi noter que le processus de résumé peut être guidé par une requête qui influence notamment cette phase de pondération que nous exploitons. À l'aide des sites Web et des descriptions récoltées, des muséologues ont proposé une requête pour chaque institution qui est propre au musée, et indépendante de toute préférence d'un visiteur.

Ainsi, nous appliquons les modifications exposées dans la sous-section 4.4.2 au système ENERTEX afin de produire pour chaque expôt, une mesure de son importance vis à vis du reste de la collection.

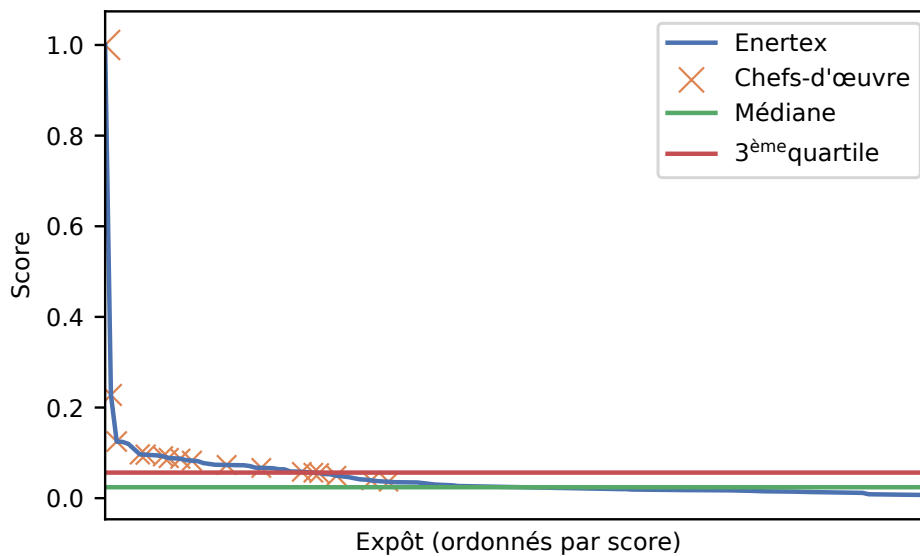
4.5.3 Application

La figure 4.3 montre le prestige calculé par ENERTEX pour chacun des deux musées, ainsi que ceux considérés comme des chefs-d'œuvre par ces institutions.

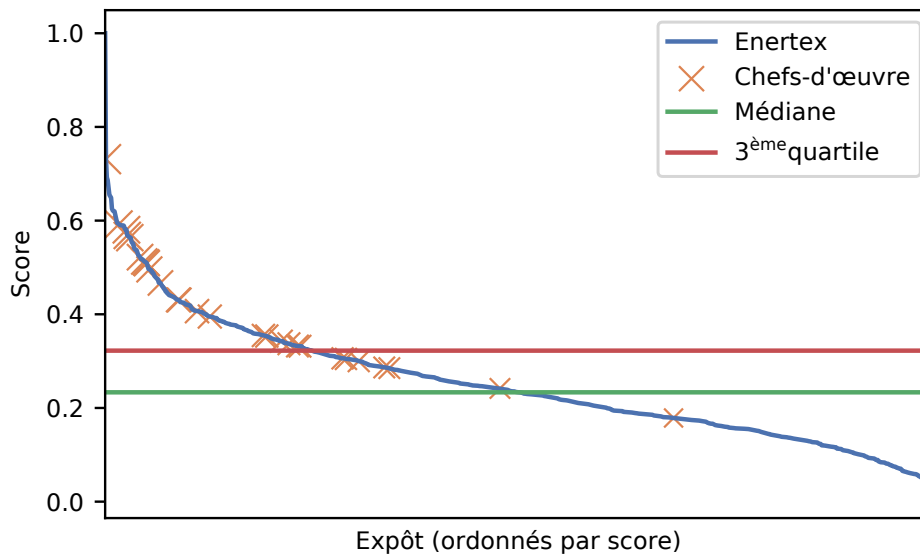
Nous pouvons voir que le classement qui résulte de l'algorithme est en accord avec les informations fournies par les musées. Pour le « Musée de l'Orangerie », les 17 chefs-d'œuvre ont tous un score supérieur à la médiane et 12 d'entre eux (71 %) au troisième quartile. Des résultats similaires sont observables pour la « National Gallery », où un seul chef-d'œuvre a un score inférieur à la médiane et celui de 21 des 28 chefs-d'œuvre (75 %) est plus élevé que le troisième quartile.

La répartition de cette mesure du prestige sur l'ensemble de la collection est quant à elle différente d'un cas à l'autre. Dans la « National Gallery », bien que l'expôt obtenant le meilleur score ne figure pas dans la liste des chefs-d'œuvre, aucune œuvre ne prédomine sur les autres, alors qu'il existe une grande différence (0,77) entre les scores des deux premiers expôts du « Musée de l'Orangerie ».

Dans ce dernier, c'est l'œuvre maîtresse du musée, les « Nymphéas » de Monet, qui obtient ce score disproportionné par rapport au reste de la collection comme nous pouvions nous y attendre au regard de l'atypicité de la description associée. L'impact de l'hétérogénéité dans la taille des données textuelles à disposition se remarque aussi dans la « National Gallery » où le seul chef-d'œuvre ayant un score inférieur à la médiane était accompagné d'une description très sommaire.



(a) « Musée de l'Orangerie »



(b) « National Gallery »

FIGURE 4.3 – Prestige indépendant des expôts calculé en adaptant l'algorithme ENERTEX

Les résultats montrent que le résumé automatique de texte permet d'établir un classement à partir des descriptions des expôts dans lequel les chefs-d'œuvre occupent les premières places. Cette mesure qui estime le prestige des œuvres de la collection d'un musée présente plusieurs avantages, notamment de ne pas nécessiter de travail supplémentaire de la part de l'institution. De plus, le procédé permettant de calculer ces valeurs est très rapide et, étant indépendant des préférences des visiteurs, ce n'est que lorsque la collection est modifiée qu'il est nécessaire de recalculer les scores.

4.6 Conclusion

Guider le visiteur vers les salles contenant les œuvres qui lui correspondent le plus est un défi pour atténuer l'effet de fatigue muséale. L'identification de ces expôts peut se faire à partir des préférences du visiteur, qu'elles soient explicites comme dans le modèle que nous avons exposé, ou inférées comme nous l'avons vu dans le chapitre 3. Toutefois, disposer d'une autre mesure de l'importance des objets permettrait d'améliorer les recommandations, non seulement dans les premières étapes, mais aussi tout au long de la visite.

Exploiter les ressources textuelles mises à disposition par les institutions culturelles au moyen d'outils de traitement automatique de la langue naturelle écrite offre l'opportunité d'utiliser la richesse du travail des conservateurs et de leurs équipes. Elles permettent d'enrichir les recommandations, sans qu'il ne soit nécessaire de demander plus de renseignements aux visiteurs.

Nos résultats montrent que le fonctionnement des algorithmes de résumé automatique de texte peut être adapté, pour extraire de l'information des descriptions provenant des équipes de médiation. Cette information se traduit par une nouvelle mesure du prestige des œuvres, qui permet d'établir un classement dans lequel les chefs-d'œuvre occupent les premières places.

Un système de recommandation pour les visites de musées peut tirer partie de cette estimation du prestige pour proposer des parcours à travers les expôts emblématiques de la collection. Les parcours personnalisés peuvent aussi en bénéficier, permettant de compléter le profil du visiteur, lui permettant de non seulement voir ce qui l'intéresse, mais aussi les incontournables du musée.

Chapitre 5

La recommandation de visites muséales

Sommaire

5.1	Introduction	62
5.2	Méthodologie de recommandation	63
5.2.1	Objectif du système	63
5.2.2	Fonctionnement et composants clés	63
5.2.3	Estimation de l'intérêt des œuvres	64
5.2.4	Le musée comme donnée géographique	65
5.3	Formalisation du problème	65
5.3.1	Topologie d'un musée	66
5.3.2	Le problème de la recommandation de visite personnalisée au musée	67
5.3.3	Parcours de visite	67
5.4	Génération de parcours de visite satisfaisant	68
5.4.1	Modèle de programmation linéaire	68
5.4.2	Heuristique gloutonne	72
5.5	Expériences et résultats	75
5.5.1	Configurations des simulations	76
5.5.2	Temps d'exécution	78
5.5.3	Pertinence des parcours	79
5.6	Conclusion	84

5.1 Introduction

Dans le chapitre précédent, nous avons détaillé des méthodes permettant d'estimer l'intérêt des expôts d'un musée, que ce soit au regard des préférences du visiteur ou en reflétant le point de vue du musée. Ces éléments constituent le matériau de base pour l'élaboration d'un système de recommandation proposant des parcours alliant les goûts culturels spécifiques du visiteur aux œuvres qui font l'identité de l'institution.

D'après SCHOUTEN (1995), une des raisons justifiant que le public se désintéresse des visites de musées est le rôle trop passif des visiteurs qui ne peuvent pas contrôler le rythme auquel le musée leur demande d'acquérir des connaissances. De plus, la plupart des visiteurs ont un budget ou un temps limité à consacrer à chaque site culturel (SERRELL 1997) pendant leur séjour.

Les systèmes de recommandation peuvent en partie répondre à cette problématique, en guidant le visiteur vers les salles qui correspondent le plus à ses attentes, tout en s'assurant que le parcours conseillé n'excédera pas la durée qu'il s'est fixé.

Les travaux présentés ici se concentrent sur la recommandation dans les musées d'art, en émettant l'hypothèse que la temporalité a un impact plus restreint que pour d'autres types d'institutions comme les musées d'histoire. Notre conjecture est que l'ordre dans lequel un visiteur traverse les salles, et donc découvre le musée, n'a pas une importance significative. Le visiteur peut commencer par une salle, puis aller dans une autre, ou inversement, sans que son expérience n'en soit profondément modifiée. Le déplacement au sein d'une salle n'est quant à lui pas modélisé et nous laissons le visiteur découvrir chacune d'entre elle en fonction de son style de visite (VÉRON et al. 1983) et de la scénographie des lieux. Enfin, bien qu'il existe certaines institutions appliquant des contraintes horaires différentes à une partie de leur collection, nous postulons que le visiteur peut accéder à toutes les salles à tout moment, de son arrivée dans le musée et ce jusqu'à son départ.

Nous détaillons ici deux approches pour la proposition de visites culturelles dans un musée. La première est un modèle de MILP, dont l'objectif est de montrer que l'utilisation conjointe des préférences du visiteur et du prestige des expôts par analyse de texte mène à la propositions de meilleurs parcours. Le second algorithme est quant à lui une heuristique proposant des parcours satisfaisants avec une consommation mémoire et un temps de calcul suffisamment courts pour être utilisée dans des situations réelles.

5.2 Méthodologie de recommandation

En premier lieu, nous définissons le problème puis présentons notre système de recommandation par ses composants clés et leurs interactions.

5.2.1 Objectif du système

L'objectif de notre système de recommandation est de proposer une visite d'un musée qui corresponde aux attentes du visiteur tout en prévenant l'apparition de la fatigue muséale. Cette problématique peut être identifiée à un TTDP qui est spécifique aux musées. Nous établissons donc le PVRP_m, qui est formulé par la définition 1 ci-dessous :

Définition 1 : Le problème de recommandation de visite personnalisée au musée – PVRP_m

Au regard des préférences et contraintes du visiteur, recommander une visite personnalisée dans un musée équivaut à proposer un chemin traversant l'institution et passant par les expôts les plus intéressants de la collection tout en prévenant l'effet de fatigue muséale.

Le scénario est ici l'aide à la décision pour le visiteur en amont de sa visite. À partir de ses préférences sur la collection du musée, le système de recommandation doit proposer un parcours dans le musée qui tienne compte de l'effet de fatigue muséale. Ce dernier étant modélisé en fixant une limite sur la durée de la visite, au delà de laquelle le visiteur cherche à sortir de l'institution.

L'objectif du PVRP_m est donc de construire une visite guidant l'utilisateur à travers le musée, en favorisant les parties de la collection qui l'intéressent le plus.

5.2.2 Fonctionnement et composants clés

Notre système de recommandation est constitué de deux modules qui traitent et combinent des informations différentes à propos du musée et du visiteur. La figure 5.1 illustre ces composants ainsi que les différents éléments qu'ils manipulent.

Le premier composant, « Valuation des expôts », estime l'intérêt du visiteur envers les œuvres du musée et a été présenté par la figure 4.2 dans le chapitre 4. Le second, « Générateur de parcours », est responsable de la construction de l'itinéraire qui maximise la

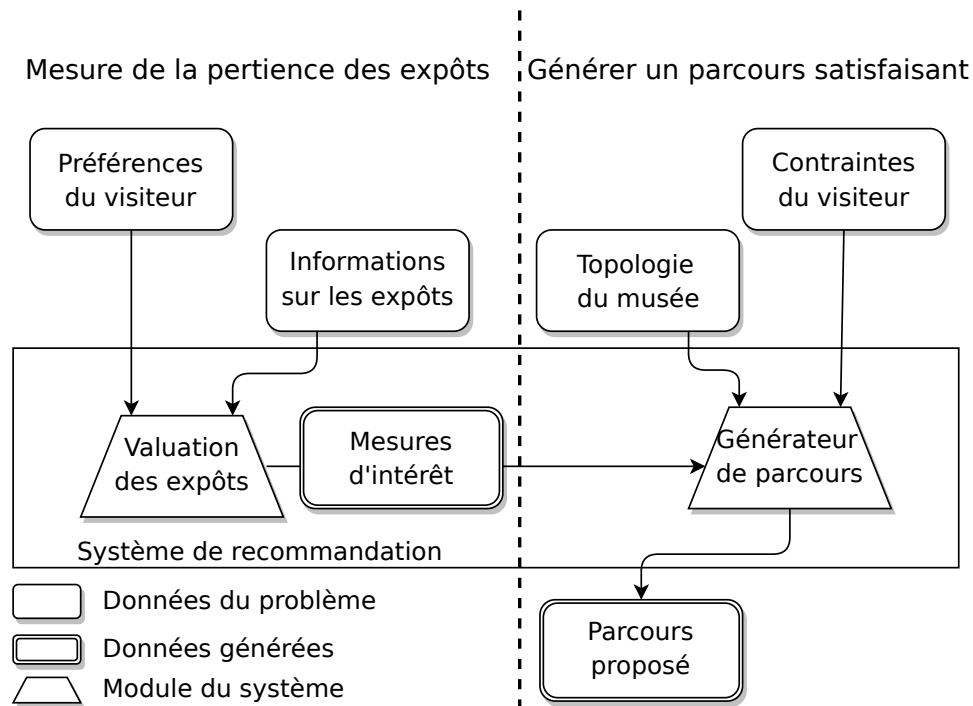


FIGURE 5.1 – Architecture et composants principaux de notre système pour la recommandation de visites personnalisées au musée

satisfaction du visiteur. C'est pour ce dernier que nous développons deux algorithmes dans ce chapitre.

5.2.3 Estimation de l'intérêt des œuvres

La première étape du processus de recommandation consiste à prédire l'intérêt du visiteur pour chaque expôt de la collection du musée.

Dans le chapitre 4, nous avons montré que des techniques bien établies de recherche d'information peuvent être utilisées pour représenter les œuvres d'un musée ainsi que le visiteur à travers ses préférences. De plus, le résumé automatique de texte peut être adapté pour attribuer à chaque expôt un score reflétant son importance au sein du musée, qui est calculé en fonction du contenu textuel décrivant l'œuvre.

La figure 4.1 précédemment exposée décrit le premier composant clé. Pour chaque expôt, il associe une mesure d'importance en utilisant les préférences exprimées par le visiteur quant aux caractéristiques de la collection, mais aussi en fonction de son importance du point de vue du musée. Ce prestige est calculé par une analyse textuelle automatique des cartouches des œuvres exposées.

5.2.4 Le musée comme donnée géographique

Un musée est un ensemble de données géographiques, une série de salles reliés entre elles. Il en résulte une structure représentable par un graphe constitué de sommets (dans notre cas les salles contenant les expôts) et d'arcs (connexions entre deux salles) évalués par des coûts (le temps nécessaire pour aller d'un emplacement à l'autre).

Une visite peut être vue comme une séquence de salles et un ensemble d'expôts à voir dans celles-ci. L'itinéraire final correspond à une façon efficace de visiter le musée en fonction des contraintes et des préférences du visiteur. Ce parcours se caractérise par sa longueur dans le temps, l'ordre des salles et les oeuvres sur lesquelles s'attarder. La figure 5.2 montre un exemple de ces différents niveaux d'information géographique avec lesquels notre système de recommandation doit interagir.

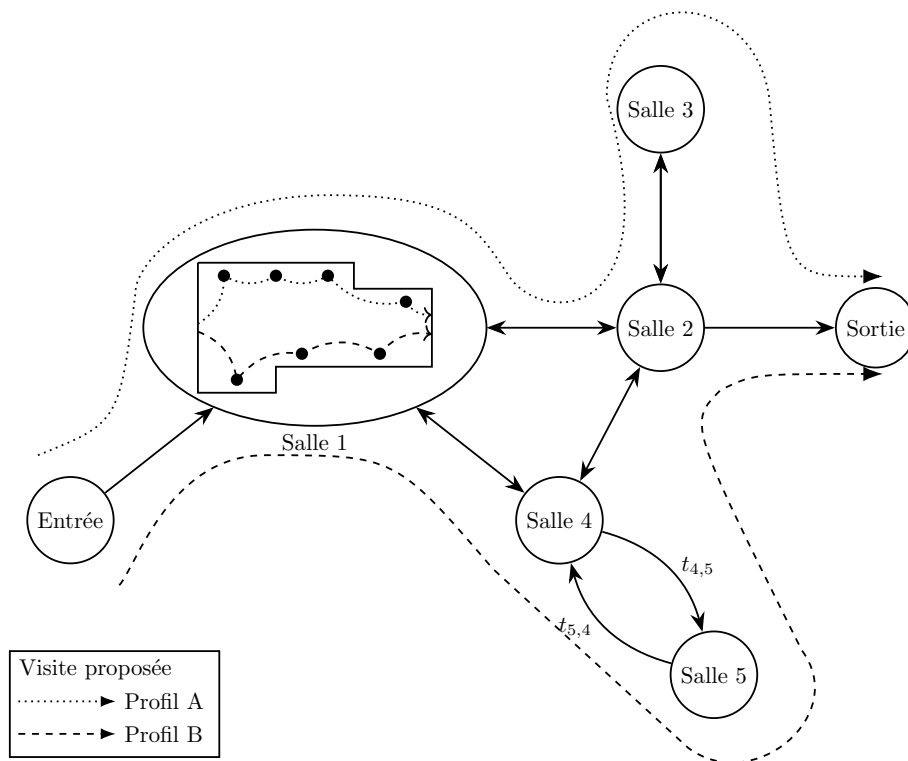


FIGURE 5.2 – Les différents niveaux d'information géographique à modéliser pour un musée et une visite

5.3 Formalisation du problème

Le fonctionnement que nous venons de décrire nécessite tout d'abord de définir les différents éléments qui seront manipulés sous la forme d'objets mathématiques. Le musée

en tant qu'espace à explorer est un prérequis à la formalisation du problème sous forme de graphe, structure servant de base à la proposition de parcours qui en résulte.

5.3.1 Topologie d'un musée

Les parcours proposés étant des chemins traversant l'institution, la connaissance de sa topologie est primordiale pour une résolution correcte du PVRP_m. Nous proposons donc de modéliser le musée comme un 6-uplet $M = (V_E, V_X, V_R, A, E, r)$ où :

V_E est l'ensemble des entrées du musée ;

V_X est l'ensemble des sorties du musée ;

V_R est l'ensemble des salles d'exposition ;

A est l'ensemble des différentes liaisons entre les pièces (entrées, sorties ou salles d'exposition).

Il existe une liaison $(i, j) \in A$ entre deux pièces $i, j \in V_E \cup V_X \cup V_R$ si on peut aller de i à j sans traverser une autre pièce.

E est l'ensemble des expôts du musée ;

r est une fonction $r : \mathcal{D}_r = E \rightarrow V_R$ donnant pour chaque expôt $e \in E$ la salle d'exposition dans laquelle il se situe.

Pour chaque musée M , il est possible de définir un graphe orienté $G_M = (V, A)$ en notant $V = V_E \cup V_X \cup V_R$ l'ensemble des pièces de celui-ci.

Remarque. Dans ce modèle, nous considérons les ensembles V_E , V_X et V_R comme étant disjoints. Ceci impose donc qu'une salle puisse être représentée par plusieurs éléments différents, comme par exemple dans le cas d'un musée où les entrées et sorties sont confondues.

Remarque. Quel que soit le sommet $v \in V$, l'entrée $e \in V_E$ ou la sortie $x \in V_X$, il n'existe aucun arc (v, v) , (v, e) ou (x, v) dans A .

Dans ce modèle, V_E et V_X sont des ensembles pour refléter que des musées, comme la « National Gallery », ont plusieurs entrées et sorties qui sont ouvertes au public. De plus, les liaisons entre les pièces et donc les arcs du graphe sont orientés, car certains musées peuvent imposer un sens de circulation pour diverses raisons (par exemple, réduire la congestion). Nous postulons aussi que, pour tout musée, il existe toujours un chemin entre toute paire de salles d'expositions, c'est-à-dire que le sous-graphe $G = (V_R, A)$ est complet.

5.3.2 Le problème de la recommandation de visite personnalisée au musée

Afin de satisfaire le visiteur, un parcours doit être proposé en accord avec ses préférences et contraintes. Afin de répondre aux objectifs présentés en sous-section 5.2.1, nous exprimons mathématiquement le problème comme un 6-uplet $\text{PVRP}_m = (M, I, R, u, t, T_{\max})$ où :

M est le modèle du musée, tel que formulé plus haut ;

I est l'ensemble des expôts qui doivent être inclus dans la visite ;

R est l'ensemble des expôts qui doivent être exclus de la visite ;

u est la fonction $u : \mathcal{D}_u = E \rightarrow [-1, +1]$ donnant pour chaque expôt du musée, l'intérêt du visiteur pour celui-ci ;

t est la fonction $t : \mathcal{D}_t = V \cup A \cup E \rightarrow \mathbb{R}^+$ donnant pour chaque pièce ou liaison, le temps nécessaire pour la traverser et pour chaque expôt le temps que le visiteur passera à l'observer ;

T_{\max} est le temps maximum que le visiteur veut passer dans le musée, établi de manière cohérente avec t .

Le PVRP_m s'applique donc à un musée M dont le modèle est donné plus haut aux regards des attentes d'un visiteur. Les préférences de ce dernier sont données par la fonction u qui estime l'intérêt qu'il portera sur chaque œuvre du musée, comme nous l'avons vu au chapitre précédent. Les contraintes propres au visiteur sont quant à elles représentées par les ensembles I et R qui doivent être disjoints ($I \cap R = \emptyset$) et permettent à un visiteur ou au musée de forcer certains comportements. Enfin, t et t_{\max} reflètent avant tout l'aspect temporel de la visite, en permettant de ne pas dépasser une limite de temps, qu'elle soit imposée par l'utilisateur ou le musée (afin de respecter les horaires d'ouvertures). Néanmoins, ces temps permettent aussi de modéliser en partie des contraintes d'accessibilité du visiteur : certaines parties du musées peuvent ne pas être adaptées aux personnes à mobilité réduite et utiliser $t(a) = \infty$ permet de représenter l'impossibilité d'accéder à ces espaces.

5.3.3 Parcours de visite

La résolution du PVRP_m mène à la proposition d'un parcours, qui est ici défini comme une séquence ordonnée de paires (v, E_v) où :

v est un pièce du musée (entrée, sortie ou salle d'exposition);

E_v est un sous ensemble de E de tous les expôts qui sont conseillés au visiteur pendant qu'il traverse la salle v (donc $\forall e \in E_v, r(e) = v$).

Toutefois, un parcours $T = ((v_0, E_{v_0}), \dots, (v_n, E_{v_n}))$ ne peut être solution du PVRP_m que lorsque :

- Le parcours commence par une entrée, c'est-à-dire $v_0 \in V_E$;
- Le parcours termine par une sortie, c'est-à-dire $v_n \in V_X$;
- Toutes les successions de salles dans le parcours sont possibles, c'est-à-dire qu'il existe un arc correspondant dans A ;
- Tous les expôts requis sont proposés, c'est-à-dire $\forall e \in I, \exists (v, E_v) \in T | e \in E_v$;
- Aucun des expôts exclus n'est proposé, c'est-à-dire $\forall e \in R, \nexists (v, E_v) \in T | e \in E_v$;
- Le temps total requis pour faire la visite n'est pas supérieur au temps limite T_{\max} .

Cette définition et les contraintes qui y sont associées permettent d'exprimer les visites comme de simples chemins élémentaires¹ dans le graphe du musée ou des constructions plus complexes incluant des cycles.

Ces descriptions mathématiques fournissent un cadre formel pour établir des algorithmes de recommandation de visites muséales personnalisées.

5.4 Génération de parcours de visite satisfaisant

Les méthodes détaillées dans le chapitre précédent fournissent les éléments de base pour notre système de recommandation, une estimation de l'intérêt des expôts. À partir de ceux-ci, nous proposons deux approches pour sélectionner les plus pertinents et les organiser en une visite du musée.

5.4.1 Modèle de programmation linéaire

Le PVRP_m peut être vu comme un problème d'optimisation linéaire et donc formulé sous la forme d'un modèle MILP, qui est détaillé ci-après. Tout comme les travaux de BERRE et al. (2013) et BENOURET (2017), notre modèle suit une logique d'accumulation des éléments intéressants, considérant que la contrainte de temps permet de répondre au problème de fatigue muséale en incitant le système à écarter les expôts les moins intéressants.

1. Au sens de AHUJA et al. (1993), chemin ne passant pas deux fois par un même sommet

Fonction objectif et variables de décision

Le but de notre système est de proposer une visite satisfaisante pour le visiteur en lui conseillant un ensemble d'œuvres à voir pendant son trajet. La fonction objectif recherche donc un chemin à travers le musée qui maximise l'utilité en passant par les salles contenant les œuvres les plus pertinentes pour le visiteur. Cette approche nécessite un ensemble de variables de décisions qui est détaillé par la table 5.1.

Variables binaires associées aux expôts	
$x_e = \begin{cases} 1 & \text{si l'expôt } e \text{ est proposée dans le parcours} \\ 0 & \text{sinon} \end{cases}$	$\forall e \in E$
Variables associées aux arcs	
$y_a = \begin{cases} 1 & \text{si l'arc } a \text{ est traversé pendant la visite} \\ 0 & \text{sinon} \end{cases}$	$\forall a \in A$
f_a	le nombre de pièces traversées lorsqu'on arrive à l'arc a
Variables binaires associées aux salles	
$z_v = \begin{cases} 1 & \text{si la salle associée au sommet } v \text{ est traversée} \\ & \text{pendant la visite, que l'on y propose des expôts ou non} \\ 0 & \text{sinon} \end{cases}$	$\forall v \in V$

TABLE 5.1 – Variables de décisions utilisées pour la résolution du PVRP_m

L'équation (5.1) donne la fonction objectif de notre modèle, qui se traduit par la maximisation de l'intérêt des expôts qui sont proposés pendant la visite.

$$\max \sum_{e \in E} x_e \times u(e) \quad (5.1)$$

Contraintes

Les contraintes nécessaires pour exprimer le PVRP_m sont réparties en trois catégories, les contraintes sur le flot qui assurent la construction d'un chemin réalisable, les contraintes sur les expôts et la contrainte sur la limite de temps.

Contraintes sur le flot

Typiquement, la recherche de chemin dans un graphe est résolu en deux étapes, comme le modèle de BERRE et al. (2013). Dans un premier temps, le système détermine le nombre minimum d'étapes nécessaires pour traverser l'ensemble du musée puis, un second modèle associe à chaque étape la salle à visiter. Or, ces travaux préfigurent la construction de parcours touristiques en ville, pour laquelle la modélisation par étape est impossible à cause des activités se déroulant en parallèle. Par conséquent, nous avons modélisé le PVRP_m par une recherche de flot dans le graphe.

Les contraintes suivantes permettent donc de construire un chemin à travers le musée, en conservant la possibilité d'effectuer des boucles si la topologie l'impose. Pour ce faire, nous notons δ^- et δ^+ les fonctions donnant pour chaque sommet $v \in V$ l'ensemble des arcs dans A entrants et sortants par celui-ci.

$$\sum_{v \in V_E} \sum_{a \in \delta^+(v)} y_a = 1 \quad (5.2a)$$

$$\sum_{v \in V_X} \sum_{a \in \delta^-(v)} y_a = 1 \quad (5.2b)$$

$$\sum_{a \in \delta^-(v)} y_a = \sum_{a \in \delta^+(v)} y_a \quad \forall v \in V_R \quad (5.2c)$$

$$\sum_{a \in \delta^-(v)} y_a + \sum_{a \in \delta^+(v)} y_a \geq z_v \quad \forall v \in V \quad (5.2d)$$

$$f_a \geq y_a \quad \forall a \in A \quad (5.2e)$$

$$f_a \leq |V| \times y_a \quad \forall a \in A \quad (5.2f)$$

$$\sum_{a \in \delta^-(v)} f_a = \sum_{a \in \delta^+(v)} f_a - z_v \quad \forall v \in V_R \quad (5.2g)$$

Ainsi, les contraintes (5.2a) et (5.2b) assurent que le visiteur commence le parcours depuis une seule entrée et termine par une seule sortie respectivement. La contrainte (5.2c) force le visiteur à sortir d'une salle d'exposition s'il y entre, qu'elle soit traversée ou visitée, tandis que la (5.2d) impose qu'une salle doit être traversée si l'un de ses arcs entrants ou sortants est utilisé. L'équation (5.2e) exprime qu'un visiteur doit avoir traversé au moins une pièce avant d'arriver à un arc, alors que (5.2f) impose qu'aucun flot ne se déplace sur un arc s'il n'est pas emprunté dans le parcours. Enfin, l'équation (5.2g) signifie que si un sommet est traversé, alors le débit en arrivant est égal au débit à la sortie moins un.

Contraintes sur les expôts

La recommandation est avant tout un ensemble d'expôts qui sont conseillés au visiteur et c'est ce que traduit la fonction d'objectif de notre modèle.

$$x_e \leq z_v \quad \forall v \in V, \forall e \in \{m \mid r(m) = v\} \quad (5.3a)$$

$$x_e = 1 \quad \forall e \in I \quad (5.3b)$$

$$x_e = 0 \quad \forall e \in R \quad (5.3c)$$

Pour la cohérence de la solution, l'équation (5.3a) assure que si un expôt est inclus dans la visite, alors la salle qui le contient est traversée. Les équation (5.3b) et (5.3c) quant à elles traduisent les contraintes de l'utilisateur sur les expôts qu'il souhaite respectivement inclure et exclure du parcours.

Contrainte sur la limite de temps

$$\sum_{v \in V} z_v \times t(v) + \sum_{a \in A} y_a \times t(a) + \sum_{e \in E} x_e \times t(e) \leq T_{\max} \quad (5.4)$$

Enfin, l'équation (5.4) garantit que le temps passé dans le musée, que ce soit en face des œuvres ou en circulant dans son enceinte, n'excède pas la limite de temps imposé par le visiteur. Cette contrainte est une limite sur le coût (ici temporel) du parcours et, par conséquent, notre modèle formule la problématique par une approche dérivant du OP.

Problème d'optimisation linéaire

Le modèle MILP complet permettant la résolution du PVRP_m est donné par l'équation (5.5).

$$\begin{aligned}
 X^* = \arg \max_X & \sum_{e \in E} x_e \times u(e) \\
 \text{s.t.} & \quad (8.13) \text{ Contraintes sur le flot,} \\
 & \quad (8.13) \text{ Contraintes sur les expôts,} \\
 & \quad (5.4) \text{ Contrainte sur la limite de temps,} \tag{5.5} \\
 & \quad x_e \in \mathbb{B} \forall e \in E, \\
 & \quad y_a \in \mathbb{B} \forall a \in A, \\
 & \quad f_a \in \mathbb{N} \forall a \in A, \\
 & \quad z_v \in \mathbb{B} \forall v \in V
 \end{aligned}$$

À partir d'une solution au problème, le parcours conseillé est obtenu en suivant les valeurs des variables f_a dans l'ordre croissant, tout en ajoutant les salles dont la variable z_v correspondante a pour valeur 1. En suivant la même logique, les expôts dont la variable est à 1 sont ajoutés à la visite, au moment où la salle d'exposition dans laquelle ils sont situés est traversée la première fois. Cette dernière étape permet de créer des parcours contenant des boucles sans toutefois dupliquer les œuvres recommandées.

Cependant, le PVRP_m est un problème NP-difficile (BERRE et al. 2013). Le résoudre requiert de nombreux calculs, rendant l'utilisation de techniques d'optimisation mathématique inadaptée pour de grands musées. En effet, le système peut avoir besoin de plusieurs heures pour proposer une visite, ce qui n'est pas acceptable pour l'utilisateur. Par conséquent, dans la prochaine section, nous proposons une heuristique permettant de calculer des parcours à la volée.

5.4.2 Heuristique gloutonne

Dans le cas d'un système de recommandation proposé directement à l'entrée d'un musée ou dans un office de tourisme, il n'est pas envisageable de faire patienter l'utilisateur plusieurs heures pour lui proposer un parcours. Par conséquent, nous proposons une heuristique dont le but est d'approcher le résultat obtenu par programmation linéaire en des temps suffisamment courts pour qu'un utilisateur ait une réponse en temps réel.

Algorithme 1 Heuristique gloutonne pour le PVRP_m

```

1: function CONSEILLERPARCOURS(PVRPm = (M, I, R, u, t, Tmax))
2:   ▷ Initialisation
3:   etapes ← { (v, Ev) | ∀e ∈ Ev, r(e) = v, ⋃Ev = I }
4:   parcours ← CONSTRUIREPARCOURS(M, etapes)
5:   if TEMPSREQUIS(parcours) > Tmax then
6:     Erreur : Impossible de créer un parcours minimal
7:   end if
8:   ▷ Amélioration du parcours
9:   salles ← Trier VR en fonction de  $f(v) = \frac{\sum_{\substack{e \in E \setminus R \\ r(e)=v}} u(e)}{|\{e | e \in E \setminus R, r(e) = v\}|}$ 
10:  for all salle s d'exposition dans salles do
11:    etapesnouveau ← etapes ∪ {(s, ∅)}
12:    parcoursnouveau ← CONSTRUIREPARCOURS(M, etapesnouveau)
13:    if TEMPSREQUIS(parcoursnouveau) < Tmax then
14:      expots ← Trier les expôts présents dans s en fonction de u
15:      Es ← sous ensemble maximal de expôts sans dépasser Tmax
16:      etapes ← etapesnouveau avec Es
17:      parcours ← parcoursnouveau avec Es
18:    end if
19:  end for
20:  return parcours proposé
21: end function

```

Détails de l'algorithme proposé

L'algorithme 1 détaille une heuristique gloutonne en deux étapes : (i) l'initialisation du parcours ; (ii) son amélioration itérative.

Dans un premier temps, un parcours minimum est construit, validant qu'il existe un chemin dans le musée reliant une entrée à une sortie, que le visiteur pourra effectuer dans le temps imparti, tout en incluant les expôts requis par son profil. La ligne 6 correspond à cette éventualité, qui peut subvenir si le temps maximum imposé par le visiteur est trop faible pour permettre de traverser le musée ou si de trop nombreux expôts sont imposés dans le parcours. Le cas échéant, le système pourrait conseiller à l'utilisateur d'augmenter la durée de sa visite ou de retirer quelques expôts à voir absolument dans le parcours, laissant la décision de les inclure à l'algorithme.

La deuxième phase améliore le parcours itérativement en parcourant les salles d'expositions. Pour ce faire, celles-ci sont dans un premier temps ordonnées en fonction de l'intérêt moyen des expôts (non rejetés par le profil) qu'elle contiennent. Le parcours est

ensuite enrichi en ajoutant les salles, des plus aux moins intéressantes, en incluant le plus d'expôts possibles pour chaque salle. Cela est réalisé en vérifiant tout d'abord qu'il est possible d'ajouter la salle sans dépasser le temps limite, puis en ajoutant pour celle-ci, les expôts des plus aux moins intéressants, toujours en respectant le temps de visite maximal supporté.

Cette heuristique détermine un ensemble de salles ayant le plus grand intérêt, qui forme un parcours réalisable dans la limite de temps imposée. Pour ce faire, la fonction `CONSTRUIREPARCOURS`, donnée par l'algorithme 2, est utilisée pour obtenir un parcours complet passant par toutes les salles. Cette distinction entre les salles que l'on souhaite inclure dans la visite et le parcours qui en découle permet de trier les salles par intérêt et non par proximité car à chaque itération, le parcours résultant pourra changer en fonction de la salle ajoutée, en empruntant un autre chemin changeant l'ordre de la visite.

Algorithme 2 Fonctions `CONSTRUIREPARCOURS` et `TEMPSREQUIS`

```
function CONSTRUIREPARCOURS( $M = (V_E, V_X, V_R, A, E, r)$ , etapes)
    parcours  $\leftarrow$  Trouver le plus court chemin de n'importe qu'elle entrée vers une des
    etapes
    Retirer la salle ajoutée de etapes
    while etapes n'est pas vide do
        chemin  $\leftarrow$  Trouver le plus court chemin de la dernière salle du parcours vers
        une des etapes
        Ajouter chemin à la fin du parcours
        Retirer la salle ajoutée de etapes
    end while
    chemin  $\leftarrow$  Trouver le plus court chemin de la dernière salle du parcours vers n'im-
    porte qu'elle sortie
    Ajouter chemin à la fin du parcours
    return parcours
end function
```

Cette fonction s'appuie sur le calcul des plus courts chemins entre les sommets d'un graphe. Tout d'abord, nous déterminons le plus court chemin reliant une entrée à n'importe qu'elle étape de la visite, permettant de sélectionner l'entrée à utiliser et la première salle à visiter. Itérativement, nous sélectionnons la prochaine étape la plus proche à partir de la dernière salle du parcours déjà établi jusqu'à ce que toutes les salles voulues soient visitées. Enfin, la même logique nous permet de compléter le chemin jusqu'à une sortie du musée.

Le parcours issu de cette fonction est une séquence continue de toutes les salles à traverser, de l'entrée jusqu'à la sortie, que des expôts y soient conseillés ou non. De ce fait, la fonction `TEMPSREQUIS` n'est que la somme des temps requis pour traverser chaque salle et chaque arc les reliant deux à deux, ainsi que celui attribué à chacune des œuvres de la visite.

Analyse de la complexité algorithmique

La complexité algorithmique dans le pire des cas est une mesure courante permettant d'estimer la quantité d'opérations nécessaires pour l'exécution d'un algorithme. La proposition 1 donne celle de notre heuristique.

Proposition 1. *La fonction CONSEILLERPARCOURS appartient à la classe des problèmes de complexité $\mathcal{O}(|V|^3 \times \log |V|)$.*

Démonstration. Afin de déterminer les plus courts chemins, nous utilisons l'algorithme de plus court chemin de DIJKSTRA (1959), dont la complexité est $\mathcal{O}(|A| + |V| \times \log |V|)$ selon BARBEHENN (1998). Par conséquent, notre fonction CONSTRUIREPARCOURS a une complexité de l'ordre de $\mathcal{O}(|V| \times |A| + |V|^2 \times \log |V|)$.

Pour conseiller un parcours, nous avons à trier les salles d'exposition du musée. Cette étape peut être faite en $\mathcal{O}(|V_R| \times \log |V_R|)$ avec un algorithme de tri efficace (KNUTH 1997). La fonction TEMPSREQUIS quant à elle est de complexité $\mathcal{O}(|V| + |A| + |E|)$ et, en comparaison avec les autres fonctions, son impact est négligeable.

Nous pouvons donc en déduire que la complexité de l'algorithme CONSEILLERPARCOURS est la suivante :

$$\begin{aligned} &\mathcal{O}(|V| \times |A| + |V|^2 \times \log |V| + |V_R| \times \log |V_R| + |V|^2 \times |A| + |V|^3 \times \log |V|) \\ &\subset \mathcal{O}(|V|^3 \times \log |V|) \end{aligned}$$

□

Avec une complexité polynômiale, notre heuristique rend la recherche de solutions plus rapide qu'un modèle MILP, comme nous le montrons dans la sous-section 5.5.2 ci après. La manipulation des salles d'exposition et non de chaque expôt indépendamment permet de prendre rapidement des décisions, au risque d'inclure des œuvres moins intéressantes qui seraient présentes aux côtés d'autres très pertinentes. Dans la section suivante, nous analysons cette différence de fonctionnement par des simulations sur différents musées.

5.5 Expériences et résultats

L'évaluation d'un système de recommandation culturel est une tâche complexe car il est difficile d'identifier la source principale d'erreur qui peut, par exemple, être due à une

mauvaise estimation de l'intérêt du visiteur, un manque de cohérence de l'ensemble des propositions ou une mauvaise construction du parcours final.

Nous avons décidé ici de réaliser des simulations à partir de différents profils, comme une première étape à une évaluation plus complexe sur le terrain. Nous cherchons ici à montrer l'efficacité d'une heuristique simple par rapport à une modélisation plus complexe.

5.5.1 Configurations des simulations

Nous détaillons ici les deux musées étudiés ainsi que les protocoles suivis pour créer les profils de visite.

Terrain d'étude

Tout comme dans le chapitre 4, nous évaluons notre système de recommandation à partir des deux institutions que sont le « Musée de l'Orangerie » à Paris et la « National Gallery » à Londres. La figure 5.3 montre pour chaque musée, un plan de celui-ci et le graphe associé qui est utilisé par notre système.

Le « Musée de l'Orangerie » expose 144 œuvres par 14 artistes qui sont réparties dans 11 salles. La « National Gallery » regroupe 972 tableaux par 438 artistes sur 65 salles. Cette différence d'échelle permettra d'évaluer la pertinence d'utiliser l'une ou l'autre des méthodes en fonction du temps nécessaire pour proposer une visite.

Profils utilisés

Dans le chapitre précédent, nous avons vu qu'il est possible d'utiliser un espace vectoriel afin de représenter non seulement toutes les œuvres d'un musée, mais aussi les préférences du visiteur. Afin d'utiliser le même protocole pour les deux musées, nous avons choisi de n'utiliser que les peintres comme caractéristiques pour les expôts, étant donné que nous ne disposons pas des mêmes informations dans les deux musées.

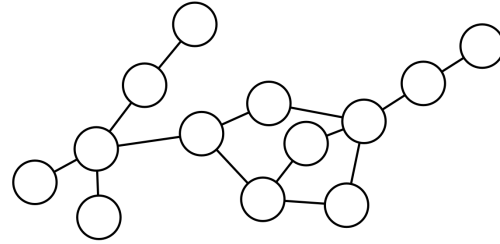
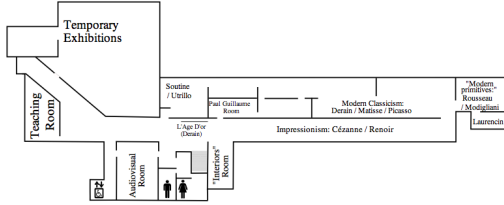
Ainsi, pour calculer l'intérêt des expôts, chaque profil produit différents vecteurs pour les œuvres et les préférences du visiteur :

Sans préférence Le même vecteur $v = (1)$ est utilisé pour chaque expôt ainsi que le profil du visiteur ;

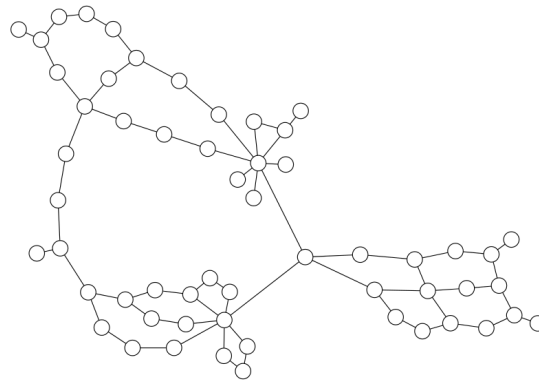
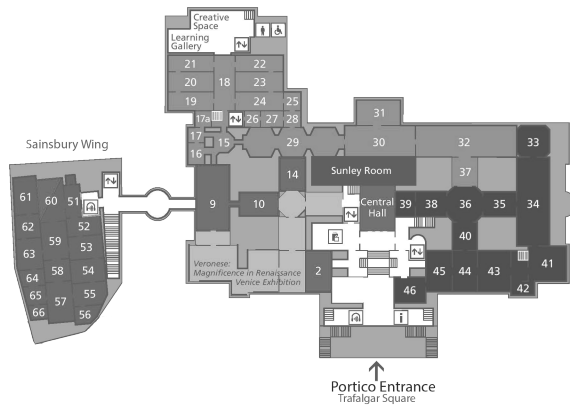
UPPER FLOOR: Water Lilies



LOWER FLOOR: Jean Walter - Paul Guillaume Collection



(a) Cas du « Musée de l'Orangerie »



(b) Cas de la « National Gallery »

FIGURE 5.3 – Plans et graphes associés des différents musées

Point de vue du musée Chaque expôt est associé à un vecteur $v_e = (p_e)$ d'un seul élément p_e correspondant au prestige calculé par résumé automatique de texte tel que décrit dans le chapitre précédent, le profil du visiteur est un vecteur $v_p = (1)$;

Préférences spécifiques Le vecteur $v_e = (a_0, \dots, a_n)$ de chaque expôt est formé avec :

$$a_i = \begin{cases} 1 & \text{Si } a_i \text{ est un artiste ayant réalisé l'œuvre} \\ 0 & \text{Sinon} \end{cases}$$

Celui du visiteur est généré aléatoirement, correspondant à l'ensemble des artistes qui l'intéressent ;

Préférences spécifiques et point de vue du musée Chaque expôt est caractérisé par un vecteur résultant de la concanénation du vecteur des préférences spécifiques et celui du point de vue du musée, soit $v_e = (a_0, \dots, a_n, p_e)$.

Ces différents profils permettent de simuler différentes situations. Le premier est notre cas de référence, il modélise un visiteur intéressé par tous les aspects de la collection. Le second profil ne considère que le prestige des œuvres, pour un visiteur souhaitant découvrir les éléments les plus importants du musée. Le troisième type de profils nous

permet de simuler des préférences spécifiques pour certains artistes. Enfin, le dernier type de profils considère à la fois du prestige des œuvres, tout en conservant les préférences du visiteur.

Lorsque des préférences sont utilisées, c'est-à-dire pour les deux derniers types de profils, nous générons 200 combinaisons d'artistes sélectionnés aléatoirement pour chaque musée (chaque tirage comprenant entre 2 et 5 noms). Ces combinaisons sont ensuite utilisées pour les deux types de profils, afin que les résultats soient cohérents.

Contraintes de temps

Pour chaque musée, nous obtenons un total de 402 profils différents. Pour chacun d'eux, nous utilisons les deux algorithmes de recommandation présentés pour construire 22 parcours en faisant varier la contrainte de temps de la visite entre 30 min et 6 h par palier de 15 min.

Les limites de temps choisies ne reflètent pas des cas d'utilisation en situation réelle. En effet, il est peu probable qu'un visiteur désire effectuer un parcours pendant 6 h sans pause. Néanmoins, cette variation de durée nous permet d'observer comment les méthodes sélectionnent les expôts en fonction du temps dont dispose le visiteur.

5.5.2 Temps d'exécution

Un système de recommandation pour les visites de musées doit être capable de filtrer rapidement l'information pour être utilisé par les visiteurs. En effet, d'après NIELSEN (1993), après 10 s il devient difficile de garder l'attention de l'utilisateur, d'autant plus si le système est incapable d'estimer le temps restant pour accomplir la tâche qu'il exécute. Le temps de calcul est donc un critère important pour mettre en concurrence nos deux approches.

Nous avons effectué toutes les simulations sur une même machine composée de 16 processeurs « Intel Xeon E640 » et 32,9 GiB de mémoire vive. La table 5.2 montre les temps de calcul moyens pour chaque méthode de résolution en fonction de l'institution étudiée.

L'utilisation de l'optimisation linéaire est acceptable pour de petits musées, mais n'est pas envisageable pour de grandes institutions étant donné les temps de calcul considérables. De plus, nous avons constaté que pour la « National Gallery », une simulation pouvait nécessiter jusqu'à 26 GiB en mémoire vive ce qui montre que pour offrir un tel service simultanément à plusieurs centaines d'utilisateurs, un tel musée devrait investir dans une architecture lourde et coûteuse.

Institution	Modèle MILP	Heuristique
« Musée de l'Orangerie »	165 ms	1 ms
« National Gallery »	112 min	10 ms

TABLE 5.2 – Comparaison des temps de calcul des approches proposées

Notre heuristique quant à elle peut être utilisée dans toutes les situations, donnant des réponses en temps réel, permettant à l'utilisateur de rectifier ses choix et ainsi d'affiner son parcours successivement. De plus, nous avons testé cet algorithme sur une machine offrant des capacités similaires aux téléphones d'entrée de gamme² et les temps de calcul concernant la « National Gallery » sont aux alentours de 30 ms.

5.5.3 Pertinence des parcours

Bien que le temps d'exécution soit essentiel, le point le plus important à évaluer selon nous est la capacité du système à conseiller des parcours qui soient intéressants. Pour ce faire, nous évaluons les parcours sur deux aspects que nous pensons complémentaires : (i) l'adéquation avec les préférences de l'utilisateur ; (ii) la capacité à recommander des chefs-d'œuvre. Le défi de ce double objectif est de proposer des parcours incluant des expôts intéressants vis à vis des préférences culturelles du visiteur, tout en intégrant les œuvres signatures du musée.

Métrique d'évaluation

Nous mesurons ces deux aspects avec une unique méthode : pour chaque profil, nous évaluons la proportion des œuvres du parcours par rapport à l'entière collection du musée en pondérant chacune par son importance (équation (5.6)).

$$\text{score} = \frac{100 \times \sum_{e \in E_T} u(e)}{\sum_{e \in E \setminus R} r(e)} \quad (5.6)$$

Dans cette équation, E_T est l'ensemble des expôts proposés dans le parcours et r la

2. Selon <https://www.devicespecifications.com>, avec 1 processeur cadencé à 1 GHz et 512 MiB de mémoire vive

fonction de pondération. Les deux aspects que nous souhaitons évaluer sont exprimés par l'équation (5.7) en donnant à chaque expôt une pondération.

$$r_{\text{interest}}(e) = u(e) \quad (5.7a)$$

$$r_{\text{masterpieces}}(e) = \begin{cases} 1 & \text{Si } e \text{ est un chef-d'œuvre} \\ 0 & \text{Sinon} \end{cases} \quad (5.7b)$$

La première fonction d'évaluation traduit la capacité du système à proposer un parcours satisfaisant en fonction des préférences, tandis que la seconde se concentre sur le nombre de chefs-d'œuvre qui sont conseillés.

Préférences du visiteur

La figure 5.4 présente les résultats évalués par la fonction (5.7a) dans le « Musée de l'Orangerie » en fonction des différents profils de visiteurs.

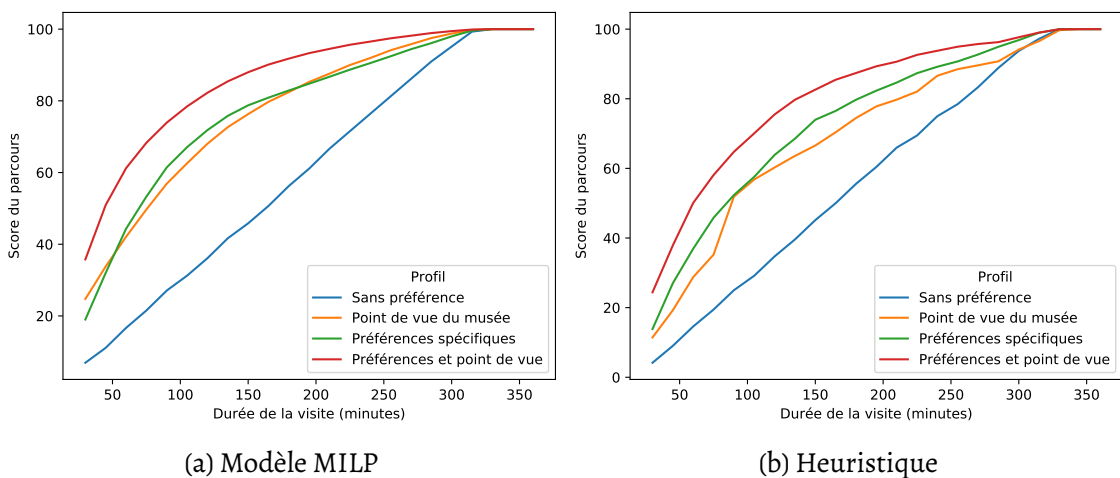


FIGURE 5.4 – Efficacité en fonction de l'intérêt des expôts dans le « Musée de l'Orangerie »

Le profil sans préférence produit une évolution linéaire dans la progression du score, étant donné que tous les expôts ont la même évaluation. Avec l'utilisation des techniques de résumé automatique de texte ou des préférences utilisateur, le système est capable d'identifier rapidement les expôts les plus pertinents pour ces profils. La combinaison à la fois des préférences utilisateur et du prestige des expôts par analyse textuelle produit quant à elle les meilleurs résultats, pour ce qui est de l'identification des œuvres pertinentes.

On observe les mêmes tendances lorsque nous utilisons notre heuristique, avec des résultats similaires et des solutions seulement 10 % moins efficaces.

La figure 5.5 présentent de la même manière les résultats pour la « National Gallery ». Toutefois, à cause des longs temps de calcul, nous n'avons pas simulé les parcours pour les profils incluant les préférences de visite lors de l'utilisation du modèle MILP. Cela s'explique par la différence dans la taille de la collection de la « National Gallery », qui justifie aussi que l'évolution des scores se fasse de manière linéaire, étant donné qu'il faudrait une visite continue sur 2 jours pour parcourir tout le musée. Néanmoins, en observant les résultats de l'heuristique, nous pouvons parvenir aux mêmes conclusions : l'utilisation du point de vue du musée par analyse textuelle améliore les résultats.

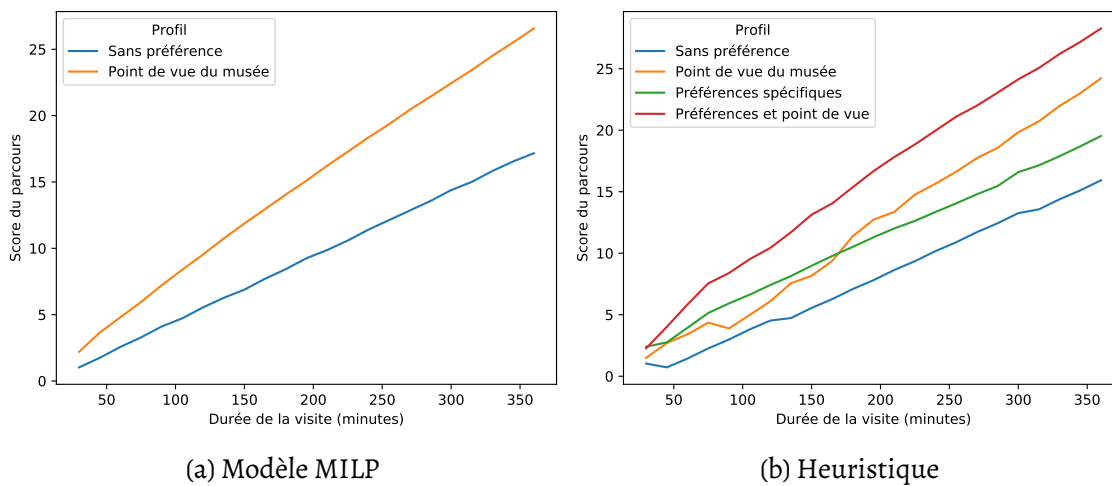


FIGURE 5.5 – Efficacité en fonction de l'intérêt des expôts dans la « National Gallery »

Chefs-d'œuvre

L'originalité de nos travaux se situe principalement dans l'évaluation indépendante du prestige des œuvres. La figure 5.6 montre l'évaluation des parcours pour différentes limites de temps par la fonction d'évaluation (5.7b), retranscrivant le nombre de chefs-d'œuvre inclus dans le parcours.

Quand le visiteur n'a pas de préférence, le système produit des résultats fluctuants. En effet, tous les expôts ayant la même valeur d'utilité, l'augmentation du temps de visite permet de changer de salle contenant plus de tableaux mais moins de chefs-d'œuvre. Les scores donnés par le prestige des œuvres issu de l'analyse des textes obtiennent les meilleurs résultats. Cela est cohérent avec les résultats du chapitre précédent et démontre la capacité du système à inclure le plus de chefs-d'œuvre. Finalement, la combinaison des préférences et de l'analyse de texte donnent des résultats a priori satisfaisants.

La figure 5.7 donnent les résultats pour la « National Gallery ». Des similarités sont observables pour les profils sans préférences, mais l'heuristique offre de meilleurs résultats

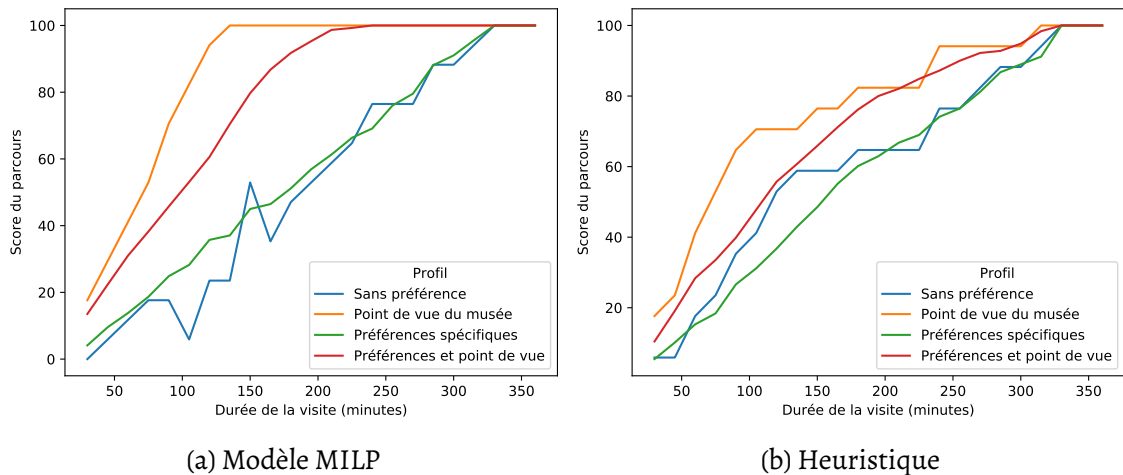


FIGURE 5.6 – Efficacité en fonction du nombre de chefs-d'œuvre dans le parcours pour le « Musée de l'Orangerie »

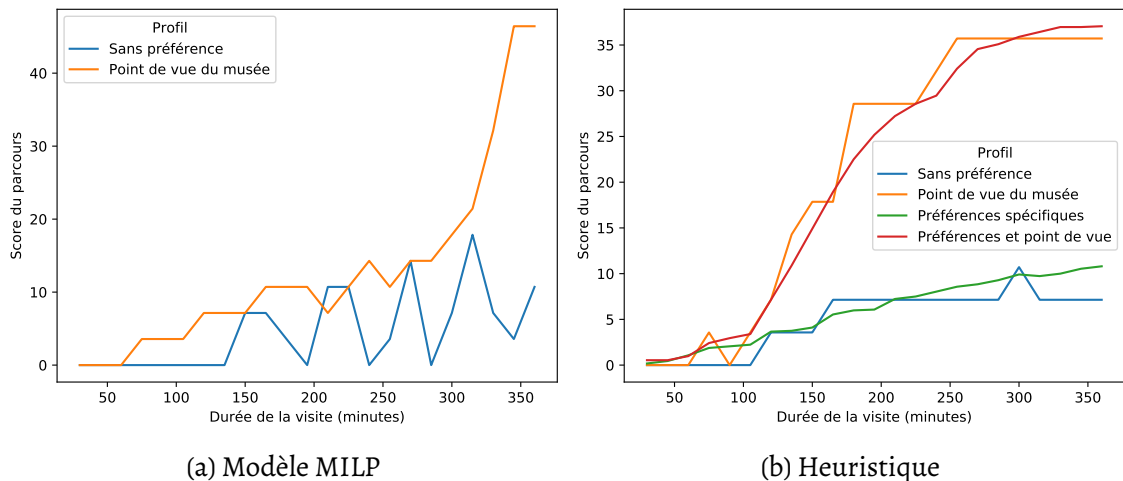


FIGURE 5.7 – Efficacité en fonction du nombre de chefs-d'œuvre dans le parcours pour la « National Gallery »

quand sont considérés les mesures tirées de l'analyse de texte. Les mauvais résultats de cette méthode dans ce musée, comme nous l'avons vu dans le chapitre précédent, sont certainement causés par l'hétérogénéité des textes décrivant les œuvres mises à disposition par le site de l'institution.

L'approche par optimisation linéaire trouve une solution optimale, mais en se basant sur des mesures de mauvaises qualités. La nature gloutonne de notre heuristique, quant à elle, tend à ajouter des expôts par salle. Or, dans la « National Gallery », de nombreux chefs-d'œuvre sont regroupés et lorsqu'une salle est ajoutée au parcours, tous les chefs-d'œuvre sont inclus, bien que l'analyse du prestige des textes ne leur ait pas attribué un haut score.

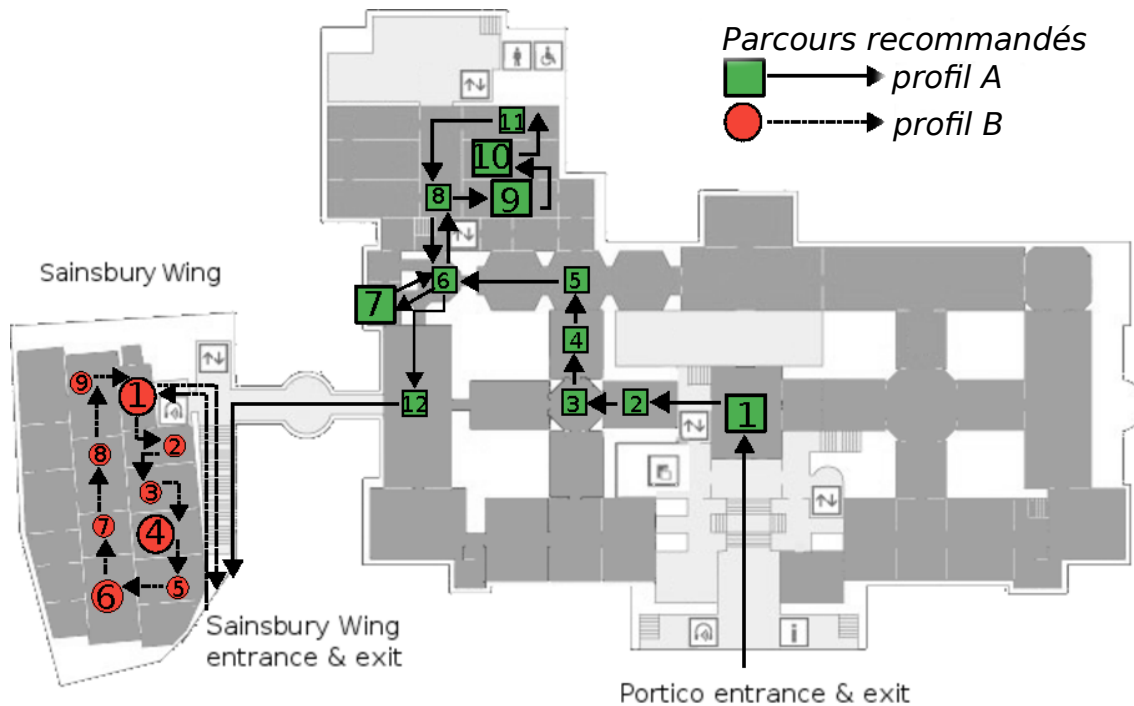


FIGURE 5.8 – Parcours conseillés pour deux profils différents dans la « National Gallery »

Exemples de parcours

Pour illustrer la capacité de notre système de recommandation à personnaliser une visite de musée, nous avons produit des parcours pour deux visiteurs potentiels ayant une même contrainte de temps (1 h et 45 min) mais avec des préférences différentes. Le premier profil (profil *A*) correspond à un visiteur intéressé par la peinture baroque (caractérisée ici par « Rembrandt » et les artistes qui l'ont succédé). Le second (profil *B*) est centré sur les artistes italiens de la Renaissance (ici « Lippi » et « Ucello »).

La figure 5.8 représente les visites proposées pour ces profils par notre heuristique en combinant les préférences personnelles et le prestige des œuvres.

Pour chaque visite, la taille des nœuds est proportionnelle au nombre d'expôts recommandés dans la salle correspondante. Bien que les deux parcours incluent aux alentours de 70 expôts, nous constatons ici que les deux itinéraires sont situés dans deux ailes différentes du bâtiment.

Cet exemple montre que notre système de recommandation peut être utilisé en situation réelle dans de grands musées. Comme il ne faut que quelques millisecondes pour obtenir une réponse, le visiteur peut comparer plusieurs combinaisons de préférences et construire un parcours répondant à ses contraintes et ses goûts culturels.

5.6 Conclusion

Dans cette étude, nous mettons en exergue la difficulté à étudier les attentes et les souhaits des différents publics culturels. Dans le chapitre précédent, nous avons vu que des techniques de résumé automatique de texte peuvent être utilisées pour extraire de précieuses informations données par les questionnaires des musées. Ici, nous montrons qu'il est possible d'exploiter ces informations pour proposer aux visiteurs des parcours différents en fonction de leur appétence culturelle, tout en conservant dans le parcours, l'identité de chaque musée.

L'exploitation des ressources textuelles peut donc permettre de réduire les problèmes liés au démarrage à froid et de sur-spécialisation des systèmes de recommandation. Un parcours peut être enrichi par les expôts phares du musée, diminuant ainsi la nécessité de recourir à un questionnaire approfondi avant la visite.

La modélisation par programmation linéaire reste une étape nécessaire afin d'identifier les problèmes et les leviers fondamentaux, mais il apparaît ici évident qu'elle ne peut être utilisée sur le terrain. Les heuristiques, même très basiques peuvent, de leur côté, apporter une réponse satisfaisante et dans des temps rapides.

Fondamentalement, il apparaît que les points clés permettant à un système de proposer des parcours personnalisés ne se situent pas dans la complexité de l'algorithme de routage, mais bien au niveau des méthodes permettant d'estimer l'intérêt des expôts. Dans cette optique, l'utilisation de savoirs provenant de divers domaines de recherche pour répondre aux problématiques culturelles semble prometteuse.

Dans cette première partie, nous nous sommes intéressés à la recommandation dans un musée, c'est-à-dire un espace clos dont la collection et son organisation sont régies par une équipe réduite. La suite de cette thèse s'intéresse donc à cette même problématique mais à l'échelle d'une ville.

Partie III

La ville, un musée à ciel ouvert

Chapitre 6

Étude sur la qualité d'expérience des visites culturelles

Sommaire

6.1	Introduction	88
6.2	La planification de parcours touristiques personnalisés	89
6.2.1	Le problème de conception de visites touristiques	89
6.2.2	Étude sur les habitudes touristiques	90
6.3	Évaluer la Qualité d'Expérience	96
6.3.1	Formulation générale	97
6.3.2	L'intérêt culturel	97
6.3.3	La cadence de visite	100
6.4	Étude de cas	104
6.5	Conclusion	107

6.1 Introduction

La deuxième partie de nos travaux s'intéresse à l'élaboration de parcours culturels dans le cadre du tourisme urbain. Celui-ci se différencie des visites de musées de part le volume de l'offre touristique et la place prépondérante de la temporalité dans le séjour, deux facteurs nous amenant à considérer différents aspects de l'expérience de visite.

Dans le chapitre 3, nous avons donné un aperçu du large éventail des méthodes utilisées, reflétant les diverses facettes de la problématique. Or, nous constatons que l'évaluation à la fois des points d'intérêt par les algorithmes, mais aussi des systèmes de recommandation eux-même, octroie une place prépondérante de l'intérêt des activités, au détriment de leur organisation durant le séjour. En effet, dans de nombreux travaux, le nombre et la fréquence des activités proposées est un effet secondaire dû à des facteurs comme la distance maximale à parcourir, l'aisance à marcher (KURATA et al. 2013) ou les conditions météorologiques (LAMSFUS et al. 2009).

Le modèle de RODRÍGUEZ et al. (2012) quant à lui, intègre un paramètre permettant d'espacer les activités dans le temps grâce à l'introduction d'un temps mort entre deux éléments du parcours. Cependant, bien qu'il puisse varier en fonction de l'activité, ce critère est modélisé par des contraintes que le système doit donc respecter, diminuant ainsi la flexibilité. En effet, le touriste pourrait être enclin à effectuer deux activités l'une après l'autre si toutes deux sont suffisamment intéressantes. Enfin, les auteurs ne précisent aucune méthode pour l'obtention de ces contraintes de temps, laissant supposer qu'elles sont demandées au visiteur, ce qui peut rendre l'étape de création du profil utilisateur fastidieuse pour ce dernier.

Des études sur les pratiques touristiques (BÉGUIN 2015 ; PICOLOT 2016), menées en collaboration avec des sociologues et géographes, montrent que les visiteurs ne considèrent pas seulement l'attrait culturel des activités constituant leur parcours mais aussi leur répartition durant le séjour, en fonction de leurs perceptions personnelles. Ainsi, pour répondre aux problèmes exposés plus tôt, nous proposons une méthode adaptable et reproductible pour mesurer la qualité de l'expérience d'un parcours touristique personnalisé qui se divise en deux mesures de conformité. La première évalue, en suivant les pratiques de la littérature, l'intérêt culturel des activités proposées pendant le séjour. La seconde mesure, que nous définissons comme la **cadence de visite**, se concentre sur l'organisation dans le temps des activités, tout au long du parcours. Nous utilisons des actogrammes, issus de la géographie temporelle (HÄGERSTRAND 1970) pour représenter un parcours, et le comparer à un cas de référence.

La méthode proposée est flexible, elle ne nécessite que peu d'information sur le par-

cours à évaluer et est agnostique à l’algorithme de recommandation. Les deux mesures proposées sont complémentaires et évaluent un même parcours sous deux angles différents, donnant une vision plus précise de la qualité du séjour recommandé. De plus, il est à la fois possible d’évaluer un parcours vis à vis d’un cas de référence, ou de comparer deux propositions de visite, permettant de déceler les différences et similitudes entre deux stratégies de recommandation.

6.2 La planification de parcours touristiques personnalisés

Dans un premier temps nous exposons le problème d’après la littérature et définissons un modèle général, afin de représenter toute visite guidée. Ensuite, nous présentons deux enquêtes menées sur les pratiques touristiques qui permettent d’identifier quelques facteurs clés quant à la façon de sélectionner et organiser les activités d’un parcours touristique.

6.2.1 Le problème de conception de visites touristiques

La recommandation de parcours touristique personnalisé est un sujet très présent dans la littérature, connu sous le nom du TTDP. À partir des travaux présentés dans le chapitre 3, nous choisissons de le formuler par la définition 2 ci dessous :

Définition 2 : Le problème de conception de voyage touristique – TTDP

Parmi les activités possibles (ou Points d’Intérêt) à l’intérieur d’un territoire et compte tenu des préférences d’un touriste, recommander une visite personnalisée équivaut à proposer une liste d’activités, chacune délimitée dans l’espace et le temps, qui correspondent à ses préférences tout en respectant les contraintes à la fois du touriste et du territoire.

Cette définition mène à l’utilisation des concepts d’activités, de préférences et de contraintes.

- Une activité est toute occupation que peut être amené à effectuer le touriste, comme par exemple visiter un monument, assister à une pièce de théâtre ou aller au restaurant ;
- Les préférences du visiteur concernent tous ses goûts en matière d’activités touristiques, qu’il s’agisse de préférences culturelles, d’affinités avec certains créneaux horaires ou moyens de transport par exemple ;

- Les contraintes sont des limitations imposées par le touriste ou le territoire sur le parcours ou les activités qui le composent (dates, coût maximum ou temps par activité, horaires d'ouverture, lieux de début et de fin).

Ainsi formulé, on peut voir un parcours touristique comme une succession d'activités qui peut par conséquent être modélisé par une séquence ordonnée de 4-uplets (a, s, d, l) avec :

a une activité du parcours ;

s le moment où l'activité débute ;

d la durée de l'activité ;

l une fonction donnant l'emplacement du touriste aux moments s et $s + d$.

L'ordre des éléments de cette séquence est celui des instants de début des activités qui leurs sont associées. De plus il est à noter que cette séquence ne peut contenir aucune paire d'activités se chevauchant dans le temps, ce qui serait équivalent à en effectuer deux en parallèle.

Cependant, la séquence n'est pas tenue d'être continue mais peut décrire un emploi du temps complet du début à la fin du voyage, tout comme un ensemble d'activités à effectuer durant le séjour. Cette flexibilité permet de décrire des parcours pouvant s'étaler sur plusieurs jours sans pour autant planifier chaque minute de la journée. La figure 6.1 montre quelques exemples de parcours touristiques à plusieurs échelles géographiques et avec des niveaux d'information différents.

Les travaux présentés dans ce chapitre ne discutent pas des définitions ou méthodologies utilisées pour collecter ou représenter les activités proposées sur un territoire ou les préférences et contraintes d'un touriste. Mais nous avons constaté que la modélisation présentée ci-dessus est assez robuste pour représenter une activité, peu importe la manière dont elle est définie par un service de recommandation, qu'il s'agisse d'un système automatique ou non. Tous les systèmes de la littérature étudiés (voir la chapitre 3) génèrent des parcours pouvant être représentés par ce modèle, qui permet d'inclure des activités ayant des emplacements de début et de fin différents, bien que cette possibilité ne soit pas exploitée par la grande majorité des travaux existants.

6.2.2 Étude sur les habitudes touristiques

En collaboration avec des laboratoires de recherche en sociologie et en géographie a été réalisé une étude divisée en deux enquêtes portant sur le tourisme à Avignon. La

6.2. La planification de parcours touristiques personnalisés

1. Entrée du Portique 9:20:00 - 20 min	7. Salle 6 10:20:00 - 8 min	13. Salle 11 11:08:00 - 8 min
2. Salle 1 9:40:00 - 8 min	8. Salle 7 10:28:00 - 8 min	14. Salle 8 11:16:00 - 8 min
3. Salle 2 9:48:00 - 8 min	9. Salle 6 10:36:00 - 8 min	15. Salle 6 11:24:00 - 8 min
4. Salle 3 9:56:00 - 8 min	10. Salle 8 10:44:00 - 8 min	16. Salle 12 11:32:00 - 8 min
5. Salle 4 10:04:00 - 8 min	11. Salle 9 10:52:00 - 8 min	17. Sortie aile Sainsbury 11:40:00 - 5 min
6. Salle 5 10:12:00 - 8 min	12. Salle 10 11:00:00 - 8 min	

(a) Parcours continu dans un musée (du chapitre 5)

1. Village du OFF 10:30:00	4. Théâtre du Chêne noir Spectacle à 13:30:00 de 1 h20 min
2. Théâtre du Pixel Spectacle à 11:10:00 de 1 h	5. Théâtre du barouf Spectacle à 16:00:00 de 1 h20 min
3. Restaurant Cul de poule 12:22:00 - 13:00:00	6. Village du OFF 17:34:00

(b) Parcours non continu lors du « Festival OFF d'Avignon » (par PICOLOT (2016))

FIGURE 6.1 – Exemples de parcours touristiques

première de BÉGUIN (2015) était liée aux attentes des touristes et à leur organisation. La deuxième enquête, de PICOLOT (2016), visait à identifier des profils spécifiques à un évènement particulier, le « Festival OFF d'Avignon », en vue de la création d'un système de recommandation présenté dans le chapitre suivant.

Mémoire de BÉGUIN (2015)

L'enquête de BÉGUIN (2015), réalisée en partenariat avec le laboratoire de sociologie, est centrée sur le tourisme à Avignon en dehors de la période du festival et est structurée autour de deux axes principaux : (i) les attentes du visiteur (ii) la manière dont il ou elle organise son parcours. Ces travaux ont notamment donné lieu à la passation d'un questionnaire qui a été réalisée du 8 au 2015-07-17 sur trois sites et auprès de 276 personnes de nationalités et âges différents (voir la table 6.1).

1. Le total est supérieur au nombre de répondants car une partie d'entre eux sélectionnait une combinaison telle que « en couple et en famille » ou « en couple et entre amis ».

		Région	Effectif
		France	58
		Amérique du nord	40
		Royaume-Uni	31
		Allemagne	29
		Pays-Bas	23
		Belgique	18
		Chine et Japon	14
		Australie	10
		Europe de l'est	7
		Suisse	6
		Amérique du sud	5
		Autre	6
		Sans réponse	29
		(b) Par origine géographique	
Âge	Effectif		
Moins de 25 ans	36		
26-35 ans	54		
36-49 ans	39		
50-60 ans	55		
Plus de 60 ans	78		
Sans réponse	14		
		(a) Par tranche d'âge	
Accompagnement	Effectif ^a		
Seul(e)	23		
En couple	154		
En famille	48		
Entre amis	58		
Sans réponse	0		
		(c) Par type d'accompagnement	

TABLE 6.1 – Répartition des répondants de l'enquête de BÉGUIN

En premier lieu, cette enquête a tenté de déterminer les attentes des touristes avant de débiter leur séjour. En demandant aux touristes les lieux qu'ils prévoyaient de visiter, on remarque qu'une grande majorité identifie les monuments les plus célèbres de la ville comme points d'intérêt. À ce propos, l'auteur souligne que « les caractéristiques de la ville et sa mise en tourisme engendrent le rassemblement d'individus au profil culturel plus ou moins similaire », dénotant qu'il est possible d'utiliser la popularité d'un monument comme mesure de son attrait touristique quelque soit le profil du visiteur.

Cependant, il existe tout de même un variété importante de goûts culturels. La figure 6.2 représente les relations entre les différents types de lieux culturels pour lesquels les répondants ont déclaré leur intérêt (la largeur d'une bande entre deux arcs dépend du

nombre de réponses ayant sélectionné les deux catégories associées) et montre que bien que le type « Châteaux, fortifications, palais » soit prédominant, ce qui s'explique par la nature des monuments phares de la ville, on peut constater qu'il y a une part importante de variabilité dans les autres choix sélectionnés par les visiteurs. Cela implique que même si nous utilisons la popularité du monument comme un facteur d'intérêt décisif quel que soit le profil du visiteur, ses préférences personnelles doivent être prises en compte lors de la construction de parcours plus longs.

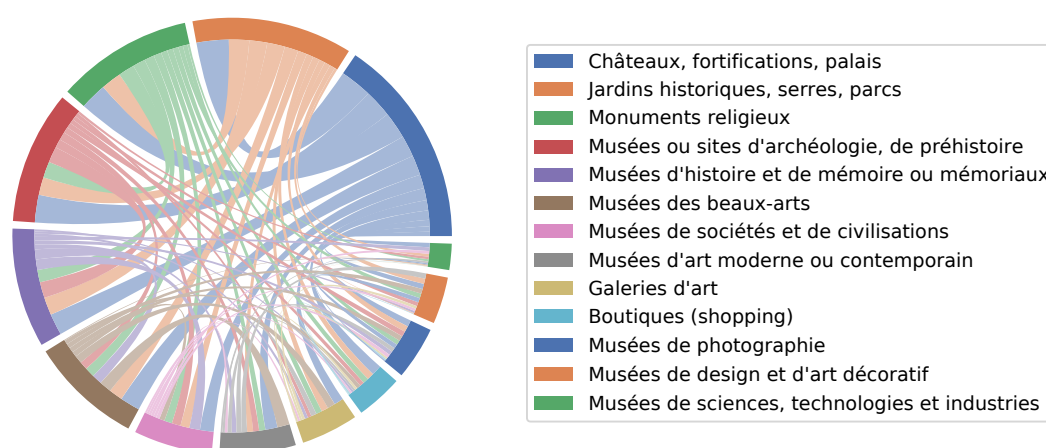


FIGURE 6.2 – Relations entre les intérêts déclarés pour les types de lieux culturels

Le deuxième axe de ce sondage consistait à analyser la manière avec laquelle les touristes construisent leur visite, que ce soit par les outils qu'ils utilisent ou sur l'organisation même de leur parcours. Le résultat fondamental qui nous intéresse ici est que la cadence de la visite est un facteur important lorsque le touriste planifie son séjour. Il a été proposé aux répondants de sélectionner un type d'organisation de leur voyage parmi : (i) découvrir le plus de choses possibles toutes leurs journées ; (ii) alterner visites et moments de repos dans une journée ou (iii) alterner journée intense et journée calme tout au long du séjour. La table 6.2 montrent qu'il existe différents schémas lorsqu'un visiteur organise son emploi du temps et, même si l'on peut identifier certaines tendances en fonction des attentes ou des personnes accompagnantes, nous ne sommes pas en mesure d'établir une règle en fonction de ces facteurs ².

Cette enquête permet d'observer plusieurs tendances lorsqu'un visiteur effectue un séjour touristique. Bien qu'il ne soit pas possible de transposer directement les profils identifiés au public du festival, on peut néanmoins conjecturer que l'intérêt culturel seul ne suffit pas lors du choix des activités d'une visite, mais que la cadence de visite est un

2. Aucune corrélation n'a été trouvée entre les attentes du touriste et les personnes l'accompagnant

Chapitre 6. Étude sur la qualité d'expérience des visites culturelles

Organisation du voyage	Aventure	Amusement	Découverte d'une autre culture	Dépaysement	Repos
Maximiser les activités tous les jours	51,7 %	39,4 %	39,0 %	42,2 %	27,3 %
Alterner les activités et moments de repos durant la journée	29,3 %	42,4 %	37,2 %	37,2 %	56,4 %
Alterner les journées intenses et journées reposantes au long du séjour	19,0 %	18,2 %	23,8 %	20,6 %	16,3 %

(a) En fonction des attentes

Organisation du voyage	Seul	En couple	Entre amis	En famille
Maximiser les activités tous les jours	30,4 %	33,1 %	46,6 %	31,3 %
Alterner les activités et moments de repos durant la journée	43,5 %	42,9 %	29,3 %	45,8 %
Alterner les journées intenses et journées reposantes au long du séjour	26,1 %	24,0 %	24,1 %	22,9 %

(b) En fonction des accompagnants

TABLE 6.2 – Organisation du séjour des visiteurs

facteur à part entière.

Mémoire de PICOLOT (2016)

L'enquête de PICOLOT (2016) s'inscrit dans la dimension géographique de ce projet de recherche et se concentre sur le tourisme au cours du « Festival OFF d'Avignon ». Ses travaux présentent d'une part les concepts et méthodologies liées à la création de profils types et actogrammes et établissent des profils touristiques propres à l'évènement.

Suite à des entretiens personnalisés et la passation d'un questionnaire auprès d'une centaine d'individus, les profils suivants ont été identifiés :

- Personnes à mobilité réduite ;
- Famille avec enfant(s) en bas âge ;
- Famille avec enfant(s) de plus de 5 ans ;
- Amis, couple sans enfant ;
- Personne seule ;
- Professionnel.

Pour chacun d'entre eux, plusieurs déclinaisons ont été proposées donnant notamment lieu à la création de différentes journées types présentées sous la forme d'actogrammes.

Les profils obtenus montrent des schémas d'horaires divers et ouvrent la question de la flexibilité de ces modes d'organisation en fonction du profil touristique.

Nous avons choisi d'utiliser les profils touristiques : (a) « Famille avec enfant(s) de plus de 5 ans », (b) « Amis, couple sans enfant » et (c) « Personne seule ». Pour chacun d'eux, nous avons sélectionné un des actogrammes proposés, qui sont représentés par la figure 6.3. Tous sont représentés sur une échelle de 24 h et constituent des journées typiques pour chaque profil touristique. Les différents segments durent entre 2 et 10 h, ce qui permet de préciser les activités sur une demi-journée tout en laissant suffisamment de liberté pour recommander un ou plusieurs spectacles³. De plus, la variabilité semble suffisante pour s'attendre à une diversité des parcours proposés par un système de recommandation.

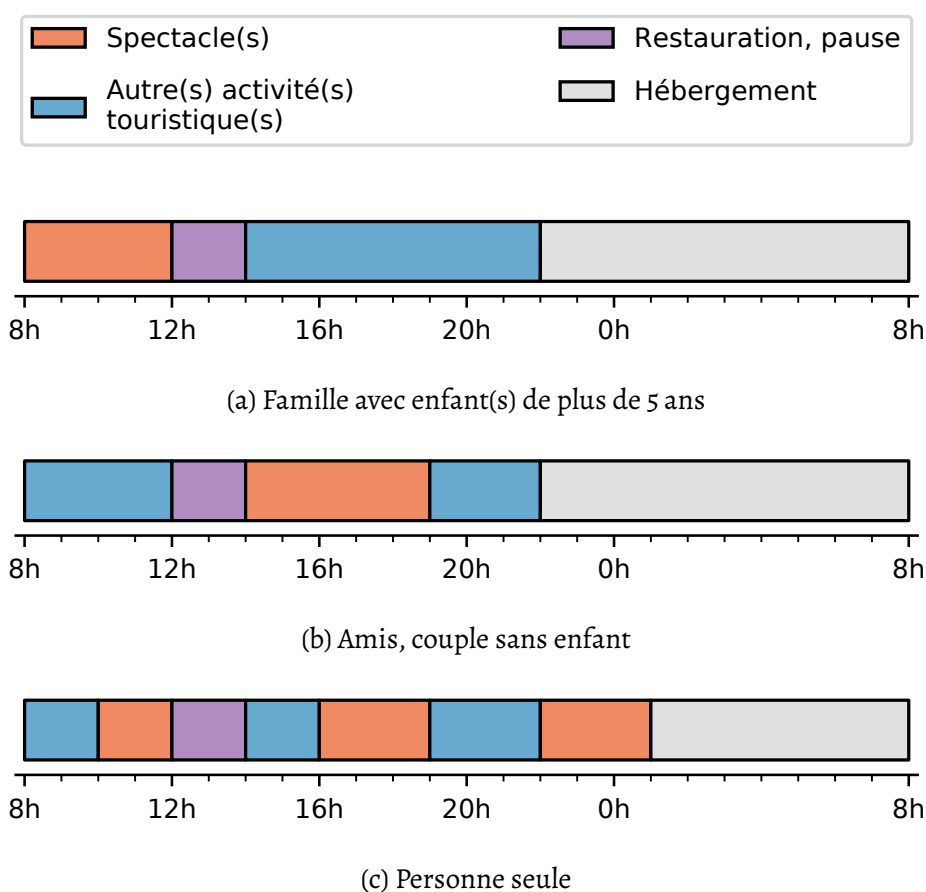


FIGURE 6.3 – Actogrammes des profils touristiques de PICOLOT (2016) pour le « Festival OFF d'Avignon »

Cette enquête met en évidence des conclusions et caractéristiques récurrentes pour ce qui est du processus de choix du contenu culturel des visiteurs. Des pièces de théâtre célèbres ont été réservées en avance, tandis que beaucoup d'autres, intéressant aussi

3. Comme on peut le voir dans l'annexe A, les spectacles durent en moyenne 1 h

Durée du séjour	Proportion des répondants
1 jour	22,6 %
2 à 3 jours	40,8 %
1 semaine	18,3 %
Tout le festival	18,3 %

TABLE 6.3 – Durée du séjour des visiteurs du « Festival OFF d'Avignon »

les visiteurs, n'étaient pas prioritaires. On remarque aussi d'autres éléments qui seront réutilisés par la suite tels que les lieux de départ privilégiés des visiteurs (qui sont la gare et le « Parking des Italiens », visibles sur la figure A.6), la durée moyenne de leur séjour (voir la table 6.3) ou encore que, bien que 83 % des répondants soient déjà venus sur Avignon, une partie importante (32 %) apprécie de déambuler dans la ville, montrant que la flânerie a une place importante durant cet évènement.

Dernièrement, l'étude note que certains répondants ont manifesté l'envie de pouvoir prévoir un emploi du temps de leur visite, ainsi que celle d'obtenir des propositions concernant les pièces de théâtre auxquelles assister. Cela confirme encore l'utilité des travaux autour des systèmes de recommandation culturels.

Pour conclure, ces deux travaux indiquent que la construction d'un parcours touristique doit se faire en fonction de plusieurs critères. Certains de ces facteurs sont liés à l'intérêt culturel de l'activité, sa popularité ou l'adéquation avec les préférences du visiteur. Cependant, l'organisation du parcours, que ce soit dans l'espace ou dans le temps, est déterminante pour le touriste, auquel il faut pouvoir proposer un emploi du temps qui correspond à ses attentes.

6.3 Évaluer la Qualité d'Expérience

Compte tenu des résultats précédents, nous pouvons relier la qualité de l'expérience d'une visite touristique personnalisée à la fois à l'intérêt des activités du parcours et au rythme avec lequel elles prennent place dans le séjour. Dans cette section, nous détaillons d'abord une méthode d'évaluation basée sur ces deux critères, puis nous proposons d'utiliser différents algorithmes de calcul pour mesurer chacun d'eux.

6.3.1 Formulation générale

Afin d'évaluer la qualité de l'expérience, le score final est divisé en deux mesures indépendantes de conformité. La première évalue l'intérêt culturel de la visite et la seconde la cadence de la visite. L'étude suggère que les activités choisies par un touriste dépendent de l'intérêt culturel qu'il y porte, mais aussi de la manière dont elles correspondent aux habitudes d'organisation de son emploi du temps. Par conséquent, ces deux mesures évaluent l'attrait d'un parcours en considérant des points de vue complémentaires et, bien que l'on ne puisse pas affirmer que l'intérêt et la cadence soient strictement indépendants, nous faisons le choix de les mesurer séparément afin de conserver une méthode de calcul à la fois rapide et compréhensible. C'est pourquoi nous proposons d'utiliser une variation de la F-mesure (POWERS 2011), une moyenne harmonique pondérée des deux mesures de conformité. Ainsi, l'équation (6.1) donne un score de satisfaction pour un parcours modélisé comme décrit en sous-section 6.2.1.

$$\text{QoE} = (1 + \beta^2) \times \frac{\text{interest} \times \text{pace}}{(\beta^2 \times \text{interest}) + \text{pace}} \quad (6.1)$$

L'évaluation de la qualité de l'expérience avec cette méthode présente plusieurs avantages. Tout d'abord, elle ne nécessite pas d'intervention humaine pour évaluer un parcours, ce qui permet de calculer rapidement un score pour de nombreuses visites. De plus, la méthode est reproductible : pour un parcours donné, le même score sera toujours obtenu, ce qui rend cette méthode appropriée à la fois pour comparer différents systèmes de recommandation avec un jeu de donnée fixé mais aussi pour évaluer un seul système avec des scénarios différents. Deuxièmement, comme le montre RIJSBERGEN (1979), le facteur β peut être utilisé pour mettre l'accent sur l'intérêt (si $\beta < 1$) ou la cadence (si $\beta > 1$) du parcours évalué. Cela permet d'adapter l'évaluation en fonction du but à atteindre, qu'il soit en rapport avec les priorités d'un organisme de tourisme, de la durée du séjour ou de la flexibilité de l'agenda du touriste.

6.3.2 L'intérêt culturel

La première mesure de conformité est axée sur l'intérêt culturel et ne prend donc en compte que les activités proposées par le circuit. Nous décrivons deux façons d'estimer cet intérêt, selon le contexte et l'information disponible. Tout d'abord, en utilisant la F-mesure pour comparer la visite proposée à un cas de référence, celle-ci reflète la capacité d'un système de recommandation à proposer une visite bien définie, et est compatible avec l'évaluation utilisée par LU et al. (2016) ou LIM (2016), qui essaient de prévoir la visite

précise d'un touriste. La deuxième méthode de calcul ne nécessite pas de tour de référence, mais utilise une fonction de prédiction pour calculer l'intérêt de chaque activité, ce qui permet de noter les systèmes qui utilisent cette approche pour calculer un parcours tels que BERRE et al. (2013) ou les travaux présentés au chapitre 5.

Afin d'évaluer l'intérêt culturel (interest), nous nous basons exclusivement sur les activités qui composent le parcours sans se soucier des horaires ou des localisations où elles prennent place. Ainsi, avec $A_t = \{a \mid (a, s, d, l) \in t\}$ l'ensemble des activités uniques d'un parcours t (tel que défini dans la sous-section 6.2.1), pour faciliter la lecture on notera A_{tour} et $A_{reference}$, où t est respectivement le parcours proposé et de référence.

Comparer la proposition à un parcours de référence en utilisant la F-mesure

Cette méthode permet de formuler l'évaluation en un simple problème de recherche d'information. La précision et le rappel sont calculés par rapport à un ensemble d'activités de référence. Les équations (6.2a) (precision) et (6.2b) (recall) définissent les notions de précision et de rappel pour un circuit touristique personnalisé.

$$\text{interest} = (1 + \beta^2) \times \frac{\text{precision} \times \text{recall}}{(\beta^2 \times \text{precision}) + \text{recall}} \quad (6.2)$$

avec :

$$\text{precision} = \frac{|A_{tour} \cap A_{reference}|}{|A_{tour}|} \quad (6.2a)$$

$$\text{recall} = \frac{|A_{tour} \cap A_{reference}|}{|A_{reference}|} \quad (6.2b)$$

La mesure de l'intérêt (donnée par l'équation (6.2)) évalue le parcours proposé en fonction du parcours de référence (qui n'est pas donné au système de recommandation). Ce dernier peut être obtenu auprès du touriste rétrospectivement ou conçu par un expert. Cette méthode est particulièrement adaptée pour un voyage qui comprendra toutes les activités intéressantes pour un profil donné.

Estimer l'intérêt en utilisant une fonction de prédiction pour chaque activité

Lorsque les parcours satisfaisants sont trop diversifiés et qu'il est impossible d'identifier un parcours de référence, nous proposons ici de nous appuyer sur une fonction de prédiction déterministe, donnant pour chaque activité un score de pertinence indiquant

l'adéquation au goût culturel du touriste. Cela peut être pertinent lorsque les parcours satisfaisants sont trop diversifiés et qu'il est difficile d'identifier un parcours de référence. C'est par exemple le cas pour des séjours de courte durée dans des territoires offrant un nombre important d'activités.

Ainsi, on utilise une fonction de prédiction, notée *relevance*, donnant pour chaque activité du parcours sa pertinence, dans \mathbb{R} , vis à vis des préférences culturelles du touriste. L'équation (6.3) exprime l'intérêt d'un parcours comme le rapport entre la somme des pertinences des activités qui le compose et la somme des pertinences de toutes les activités proposées sur le territoire visité (ensemble noté A).

$$\text{interest} = \frac{\sum_{a \in A_{\text{tour}}} \text{relevance}(a)}{\sum_{a \in A} \text{relevance}(a)} \quad (6.3)$$

Cette mesure de conformité normalise l'intérêt d'un parcours dans l'intervalle $[0, 1]$ (car $A_{\text{tour}} \subset A$) et permet d'évaluer l'intérêt de la visite selon différents critères, avec différentes fonctions de pertinence ou une combinaison pondérée de celles-ci. Comme pour les travaux de MITTAL et al. (2017) ou par analogie à ceux présentés au chapitre 4, on peut envisager une représentation vectorielle à la fois des activités et des préférences culturelles et utiliser une fonction de similarité pour calculer la pertinence de chaque activité (équation (6.4a)). Il est aussi possible de réutiliser la notion de parcours de référence, en vérifiant la présence ou non de chaque activité du parcours évalué (équation (6.4c)). Enfin, comme l'indique l'étude décrite plus tôt, la popularité n'est pas à négliger comme facteur de décision et peut donc être envisagée (équation (6.4b)).

$$\text{relevance}(a) = \text{similarity}(a, \text{preferences}) \quad (6.4a)$$

$$\text{relevance}(a) = \text{popularity}(a) \quad (6.4b)$$

$$\text{relevance}(a) = \begin{cases} 1 & \text{si } a \text{ est dans le parcours de référence,} \\ 0 & \text{sinon} \end{cases} \quad (6.4c)$$

Un avantage en faveur de l'utilisation d'une fonction de pertinence par rapport à la F-mesure est qu'elle permet d'évaluer l'intérêt d'un parcours en fonction de différents critères. La première mesure peut aussi être adaptée à l'utilisation de la popularité en définissant un ensemble d'activités, comme celles étant populaires, et le considérer comme parcours de référence. La détermination de cet ensemble peut se faire manuellement ou

avec un seuil à dépasser à partir d'une mesure. Les deux méthodes de calcul décrites couvrent la majorité des techniques d'évaluation utilisées dans les travaux de l'état de l'art présenté au chapitre 3 et permettent de se concentrer sur l'intérêt culturel seul.

6.3.3 La cadence de visite

La deuxième mesure de conformité est destinée à évaluer le rythme du séjour. Les enquêtes montrent clairement que cet aspect ne peut être négligé lors de l'évaluation d'un parcours et qu'il existe différents types de touristes, avec des modes de visite différents. Dans cette section, nous explorons un concept largement utilisé en géographie pour représenter le voyage d'un touriste et nous proposons une méthode pour évaluer un parcours par rapport à un cas de référence.

Représentation en actogramme

Un parcours peut être représenté par un actogramme décrivant la relation entre le temps et les activités qui le constituent. Cette représentation met en évidence l'utilisation et la répartition dans le temps des activités au sein du parcours, ce qui est approprié pour comparer les différentes cadences de visites.

La modélisation des parcours définie dans la sous-section 6.2.1 permet différentes granularités, que ce soit sur la durée (séjour de quelques heures ou à cheval sur plusieurs jours) ou sur la continuité (de l'horaire exhaustif aux activités clés) du voyage. Cependant, un actogramme est une séquence continue d'actions et de mouvements. Par conséquent, un parcours peut être représentée soit comme un actogramme complet si elle est continue, soit comme un actogramme partiel, si ce n'est pas le cas.

Pour avoir un seul type d'actogrammes sans perdre d'information ou spéculer sur le séjour, il nous faut nous fier uniquement à des actogrammes complets. Les visites continues ne nécessitent donc aucune modification et nous proposons simplement de transformer les autres en ajoutant dans le parcours entre deux activités non consécutives, une activité « nulle » commençant à l'endroit et au moment de la fin de la première et se terminant au début de la seconde.

Ainsi, avec un parcours $t = ((a_0, t_0, d_0, l_0), \dots, (a_n, t_n, d_n, l_n))$, on insère entre chaque paire successive d'éléments i et j pour lesquels $t_i + d_i \neq t_j$ un élément $(\emptyset, t_{i,j}, d_{i,j}, l_{i,j})$ avec :

\emptyset l'activité nulle

$t_{i,j}$ l'instant où l'activité i termine

$$t_{i,j} = t_i + d_i$$

$d_{i,j}$ la durée entre les deux activités

$$d_{i,j} = t_j - t_{i,j}$$

$l_{i,j}$ la fonction donnant la localisation du visiteur

$$l_{i,j} : \{t_{i,j}, t_j\} \rightarrow \{l_i(t_{i,j}), l_j(t_j)\}$$

$$t \mapsto \begin{cases} l_i(t) & \text{si } t = t_{i,j} \\ l_j(t) & \text{si } t = t_j \end{cases}$$

La mesure que nous voulons mettre en place vise à évaluer le rythme du parcours, c'est-à-dire les variations dans les types d'activités et non les activités elles-mêmes. Le type d'une activité est déterminé par l'agent souhaitant évaluer un parcours et n'est donc pas discuté ici, mais nous pouvons noter que cette différenciation peut se faire de plusieurs manières, allant d'un type global (« visite de monument », « restauration », « activité sportive » ou « repos » par exemple), à un niveau de détails plus poussé (le style de monument, de restaurant, etc.).

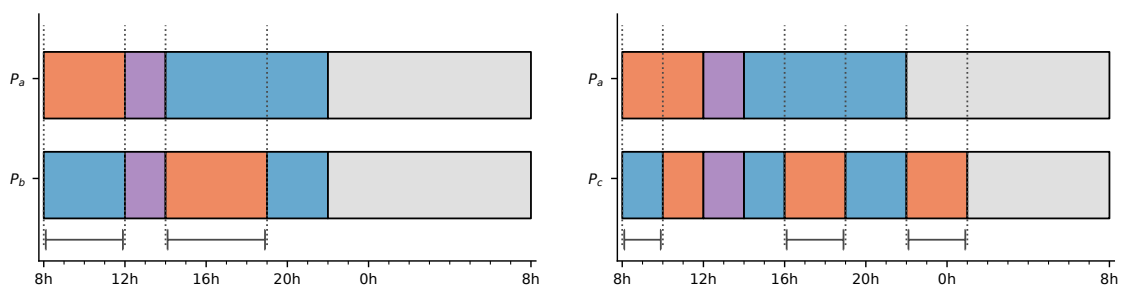
Ainsi, nous créons une nouvelle représentation à partir de la séquence continue où chaque activité est substituée par son type. Ce faisant, il est possible de fusionner des segments consécutifs ayant le même type d'activité. En plus de simplifier l'actogramme final, cela permet d'exiger des parcours de référence plus simple car ils n'ont pas besoin de représenter un programme détaillé. De plus, ces parcours de référence peuvent être obtenus auprès d'experts ou de touristes à la fin de leur séjour.

Cette méthode produit des actogrammes simples et présentant l'occupation du touriste dans le temps. Comme cette opération est effectuée par le processus d'évaluation après la génération des parcours, la cohérence des deux actogrammes est garantie, tant que les mêmes règles sont appliquées lors de la fusion.

Comparaison de deux actogrammes

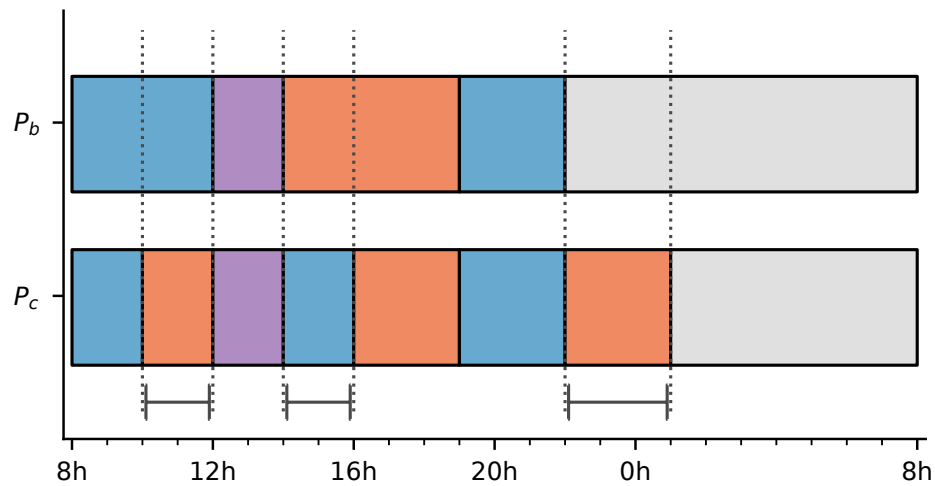
À partir des actogrammes des parcours à évaluer et de référence, il est possible d'évaluer leur conformité du point de vue de l'organisation dans le temps et ainsi obtenir une mesure de la cadence de visite, donnant la cohérence entre deux propositions.

Nous proposons de comparer les représentations obtenues des deux parcours en appliquant à chaque différence un coût de pénalité qui en dépend. Afin de déterminer les différences, nous alignons tout d'abord les deux actogrammes, en ajoutant au début ou à la fin une action nulle, pour obtenir deux actogrammes commençant et finissant au même moment. Ainsi, déterminer les différences revient à trouver les intervalles de temps où les deux actogrammes présentent des types d'activités différents, en veillant à ce que chacune de ces différences ne concerne que deux types d'activités (si un type d'activités dans un actogramme est à cheval sur deux autres types différents dans un autre, cela équivaut à deux différences distinctes) et soit la plus longue possible (les intervalles ne sont pas subdivisés). Plusieurs exemples de ce procédé sont illustrés par la figure 6.4.



(a) Entre « Famille avec enfant(s) de plus de 5 ans » et « Amis, couple sans enfant »

(b) Entre « Famille avec enfant(s) de plus de 5 ans » et « Personne seule »



(c) Entre « Amis, couple sans enfant » et « Personne seule »

FIGURE 6.4 – Différences entre les actogrammes de profils touristiques

Nous obtenons donc un ensemble D de 3-uplets $(r, p, \Delta t)$, chacune représentant une

différence avec :

- r le type d'activité de l'actogramme de référence
- p le type d'activité de l'actogramme proposé à évaluer
- Δt la durée de la différence

À partir de ces différences, l'équation (6.5) évalue la cadence d'un parcours par rapport à une référence comme l'opposé d'un taux de déviation.

$$\text{pace} = 1 - \frac{\sum_{(r,p,\Delta t) \in D} \text{penalty}(r,p) \times \Delta t}{\text{penalty}_{\max} \times T_{\max}} \quad (6.5)$$

Avec :

penalty fonction de pénalité donnant, pour deux types d'activités r et p , la pénalité d'effectuer p plutôt que r

penalty_{\max} la plus grande pénalité applicable

$$\text{penalty}_{\max} = \max \left\{ \text{penalty}(r,p) \mid (r,p,\Delta t) \in D \right\}$$

T_{\max} la durée sur laquelle s'étend les deux actogrammes

$$T_{\max} = \max \left\{ s + d \mid (a,s,d,l) \in \text{tour} \cup \text{reference} \right\} \\ - \min \left\{ s \mid (a,s,d,l) \in \text{tour} \cup \text{reference} \right\}$$

Le taux de déviation tient compte des différences entre les deux actogrammes, chacun étant caractérisé par les deux types d'activités en conflit et une durée. Cette durée est pondérée par un coût donné par la fonction de pénalité, ce qui permet de souligner ou d'ignorer certaines différences. La mesure qui en résulte est une normalisation de la pénalité d'un parcours. En effet, après avoir aligné les deux actogrammes, la somme des durées des différences est inférieure ou égale à la durée totale de ces derniers et comme la pénalité maximale vient pondérer cette durée, le rapport varie dans l'intervalle $[0, 1]$. Il en est donc nécessairement de même pour la mesure finale de conformité.

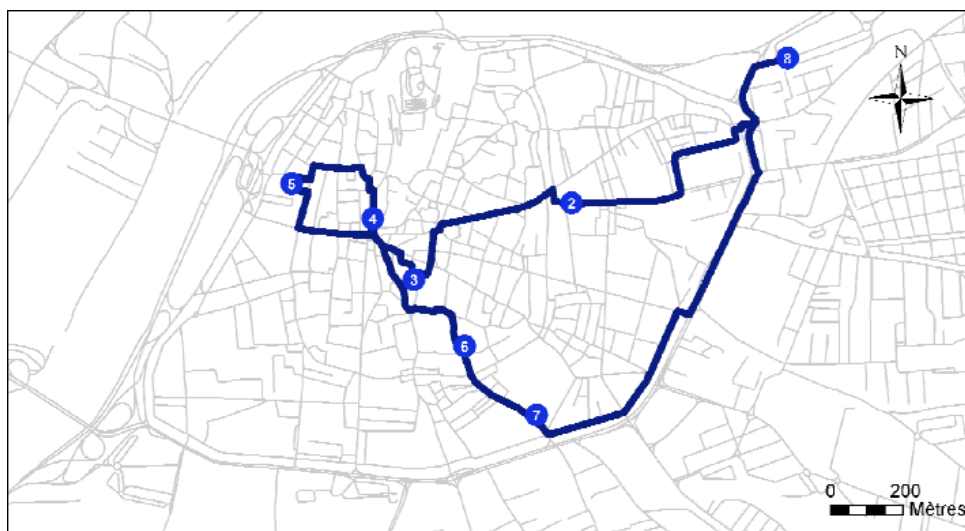
Le type d'une activité peut être une représentation plus ou moins complexe en fonction de la connaissance que l'on a de l'actogramme et de la fonction de pénalité. Cela peut varier d'un attribut unique par type associé à une matrice à une forme plus complexe où chaque

type est représenté par un ensemble d'attribut et la fonction de pénalité une fonction de similarité.

La méthode présentée pour estimer la cadence de visite répond à la nécessité de quantifier la qualité de l'expérience dans les parcours touristiques, comme le montre l'étude portant sur les habitudes touristiques que nous avons résumée plus haut.

6.4 Étude de cas

Afin d'observer comment se comporte cette méthode d'évaluation sur différents types de profils et parcours, nous utilisons des itinéraires proposés par PICOLOT (2016) qui décrivent des journées types de plusieurs profils touristiques (dont deux exemples sont donnés par les figures 6.5a et 6.5b).



Points de passage :

- | | |
|--|---|
| 1. Parking des Italiens
10:30:00 | 5. Théâtre de L'Oulle
Spectacle à 16:45:00 de 1 h35 min |
| 2. Théâtre Cour du Barouf
Spectacle à 11:15:00 de 1 h25 min | 6. Square Pétramale
18:41:00 - 20:20:00 |
| 3. Restaurant Le Potard
12:58:00 - 14:30:00 | 7. Théâtre des Béliers
Spectacle à 20:55:00 de 1 h15 min |
| 4. Terrasse du Venessin
14:35:00 - 16:10:00 | 8. Parking des Italiens
22:39:00 |

(a) Profil « Famille avec enfant(s) de plus de 5 ans »

FIGURE 6.5 – Parcours proposés par PICOLOT



Points de passage :

- | | |
|--|---|
| 1. Parking des Italiens
10:10:00 | 5. Théâtre Cinévox
Spectacle à 16:00:00 de 1 h15 min |
| 2. Théâtre Au magasin
Spectacle à 11:00:00 de 1 h | 6. Restaurant L'essentiel
17:30:00 - 18:36:00 |
| 3. Square Pétramale
12:15:00 - 13:15:00 | 7. Théâtre de Laurette
Spectacle à 19:10:00 de 1 h |
| 4. Palais des Papes
13:24:00 - 15:28:00 | 8. Parking des Italiens
20:43:00 |

(b) Profil « Personne seule »

FIGURE 6.5 – Parcours proposés par PICOLOT

Nous avons évalué chaque profil en fonction des autres avec les deux métriques. Pour l'intérêt culturel, nous avons utilisé l'équation (6.4a) notant la pertinence d'un point d'intérêt en fonction de la similarité avec celles du profil de référence (la meilleure étant retenue). Concernant la cadence de visite, les valeurs données aux pénalités entre les types d'activité sont données par la table 6.4.

Activité proposée	Activité de référence			
	Spectacle(s)	Restauration, pause	Autre(s) activité(s) touristique(s)	Hébergement
Spectacle(s)	0	0,75	0,25	1
Restauration, pause	0,75	0	0,5	0,25
Autre(s) activité(s) touristique(s)	0,25	0,5	0	1
Hébergement	1	0,25	1	0
Aucune	1	0,25	1	0

TABLE 6.4 – Matrice des pénalités appliquées pour l'évaluation de la cadence de visite

En analysant les exemples de parcours de PICOLOT (2016), on peut s'apercevoir que le

parcours proposé pour le profil « Personne seule » n'est pas adapté. En effet, alors que c'est celui qui a une plage de spectacle la plus importante (8 h contre 5 h et 6 h), elle n'est pas utilisée dans son entièreté, car c'est le parcours qui se termine le plus tôt (2 h avant le parcours « Famille avec enfant(s) de plus de 5 ans »). Nous proposons donc un autre parcours « Variation personne seule » décrit par la figure 6.6 qui correspond plus à l'actogramme 6.3c. Il n'y a qu'une seule activité en plus le matin et deux autres qui ont été déplacées afin de mieux correspondre à la journée type.

Points de passage :

- | | |
|--|---|
| 1. Parking des Italiens
9:10:00 | 6. Théâtre Cinévox
Spectacle à 16:00:00 de 1 h15 min |
| 2. Jardin des Doms
9:30:00 - 10:20:00 | 7. Brasserie du Conservatoire
17:35:00 - 22:00:00 |
| 3. Théâtre Au magasin
Spectacle à 11:00:00 de 1 h | 8. Théâtre du Girasole
Spectacle à 22:40:00 de 1 h20 min |
| 4. Square Pétramale
12:15:00 - 13:15:00 | 9. Parking des Italiens
0:30:00 |
| 5. Palais des Papes
13:24:00 - 15:28:00 | |

FIGURE 6.6 – Parcours « Variation personne seule », adapté à l'actogramme correspondant

Les résultats de l'évaluation sont résumés par la table 6.5, qui donne la concordance entre deux profils pour chaque métrique (une valeur proche de 1 indiquant des profils concordants). Il faut noter qu'il y a une différence importante quant aux références utilisées pour les deux métriques. La mesure de l'intérêt s'est faite en fonction des parcours eux-même et non pas des personas issus de la même enquête, car ces derniers ne permettaient pas de valuer toutes les activités des parcours. La mesure de la cadence de visite quant à elle a été réalisée en prenant les actogrammes des profils comme référence (voir la figure 6.3). Or, nous ne disposons que d'un seul parcours par profil alors qu'il y a énormément de pièces de théâtres ou d'évènements envisageables pour chacun d'entre eux (comme le montre l'annexe A).

Pour ce qui est de l'intérêt culturel, les trois cas présentent de bons résultats. Bien évidemment, le parcours évalué par rapport à lui-même obtient la note maximale (puisque chaque activité est au moins parfaitement similaire à elle-même) et les valeurs obtenus en fonction des autres parcours disqualifient celui-ci avec un écart d'au moins 0,7. On remarque que le profil « Variation personne seule » est bien une adaptation du profil « Personne seule », pour lequel l'emploi du temps est plus cohérent car les mesures d'intérêt sont très élevées en comparaison : la similarité est plus importante pour le profil adapté qui s'apparente à un sur-ensemble du profil de base.

Profil évalué	Profil de référence	Intérêt culturel	Cadence de visite
Famille avec enfant(s) de plus de 5 ans	Famille avec enfant(s) de plus de 5 ans	1,0	0,78
	Amis, couple sans enfant	0,3	0,71
	Personne seule	0,25	0,63
	Variation personne seule	0,21	0,63
Amis, couple sans enfants	Famille avec enfant(s) de plus de 5 ans	0,25	0,68
	Amis, couple sans enfant	1,0	0,70
	Personne seule	0,25	0,55
	Variation personne seule	0,21	0,55
Personne seule	Famille avec enfant(s) de plus de 5 ans	0,25	0,69
	Amis, couple sans enfant	0,3	0,66
	Personne seule	1,0	0,56
	Variation personne seule	0,79	0,56
Variation personne seule	Famille avec enfant(s) de plus de 5 ans	0,25	0,63
	Amis, couple sans enfant	0,3	0,62
	Personne seule	0,92	0,68
	Variation personne seule	1,0	0,68

TABLE 6.5 – Évaluation croisée de parcours touristiques

Les résultats évaluant la cadence de visite sont eux moins discriminants. Les deux premiers parcours obtiennent la meilleure note lorsqu'ils sont évalués en fonction de l'actogramme qui leur est associé, mais les valeurs calculées par rapport aux autres profils touristiques restent très proches. Le premier parcours correspondant à une personne seule obtient le moins bon score lorsqu'il est évalué avec l'actogramme lui correspondant, alors que la variation que nous proposons a des caractéristiques similaires aux deux premiers cas. Cela amène à penser que la mauvaise évaluation de la version non modifiée est certainement due à une mauvaise conception du parcours et non à celle de l'actogramme. Comme le détaille l'annexe A, le « Festival OFF d'Avignon » est un évènement proposant une grande variété de spectacles vivants et concevoir un parcours à la main parmi l'immense univers des possibles qui nous est offert est une tâche excessivement ardue.

6.5 Conclusion

Au travers d'une étude interdisciplinaire, nous montrons un résultat fondamental : les préférences sur le type des activités et l'organisation de celles-ci durant le séjour sont deux facteurs importants et complémentaires qui guident les choix des visiteurs. Leurs attentes sur le contenu culturel les regroupent autour des éléments phares de l'offre touristique, mais les goûts personnels doivent être considérés pour compléter les parcours proposés.

Les horaires des activités et le rythme du parcours sont quand à eux tout aussi importants, mais propres à chaque individu.

À partir de ce constat, nous proposons une nouvelle méthode pour évaluer la qualité d'expérience d'un parcours touristique qui se divise en deux mesures de conformité, reflétant nos deux angles d'observation. La première mesure sur l'intérêt culturel se base sur les pratiques largement répandues dans la littérature, s'adaptant aux différentes stratégies de recommandation. La seconde mesure est une nouvelle métrique, se concentrant sur la **cadence de visite** d'un parcours au moyen des actogrammes, qui permettent de représenter les comportements de mobilité des touristes. La méthode qui en résulte est très flexible, elle permet de représenter tout type de parcours et n'impose aucun prérequis au fonctionnement des systèmes de recommandation évalués.

Dans le cadre de la comparaison de différents systèmes, ces deux métriques peuvent être agrégées afin de présenter les résultats d'une évaluation de nombreuses configurations. Néanmoins, le paramètre β doit faire l'objet d'un choix judicieux car, en fonction du nombre de points d'intérêt et de la finesse de définition des actogrammes, les deux mesures de conformité varient dans des plages de valeurs très différentes.

L'enquête sur les habitudes touristiques à Avignon (BÉGUIN 2015) indique que 70 % des participants déclarent utiliser Internet avant leur arrivée, la plupart du temps (74 %) afin de préparer leur visite et identifier les sites touristiques qui les intéressent. Avec les résultats présentés plus tôt, cela souligne l'importance d'une nouvelle génération de systèmes de recommandation, intégrant l'organisation du séjour au cœur de leur algorithme de décision. Dans les chapitres suivants, nous présentons donc un nouveau paradigme pour l'élaboration de parcours touristiques en ville puis, proposons un système de recommandation pour en démontrer les effets.

Chapitre 7

Le problème de la recommandation de visite personnalisée en ville

Sommaire

7.1	Introduction	110
7.2	La loi de l'apogée-fin dans la recommandation de parcours touristiques personnalisés	111
7.2.1	Définition du problème	111
7.2.2	Vision adoptée pour l'élaboration du parcours	112
7.3	Formalisation du problème	114
7.3.1	L'offre culturelle en tant que donnée géographique et temporelle	114
7.3.2	Proposition d'une modélisation mathématique	116
7.3.3	Flexibilité du système de contraintes	118
7.4	Représentation sous forme de graphe	120
7.4.1	Définition et construction	121
7.4.2	Propriétés et relation avec les parcours	122
7.5	Conclusion	124

7.1 Introduction

Avec l'essor du tourisme, l'offre culturelle proposée aux visiteurs est de plus en plus importante. Il apparaît aujourd'hui que la simple présentation des différentes possibilités sous la forme d'une liste ordonnée d'éléments (comme le font TripAdvisor¹ ou les autres géants du domaine), n'est pas une solution satisfaisante, car déléguant à l'utilisateur la tâche de sélectionner et organiser les différentes activités qui constitueront son parcours parmi un ensemble trop important. Cette situation mène au paradoxe du choix, décrit par SCHWARTZ (2005), selon lequel l'abondance de possibilités mène l'utilisateur à prendre une décision qui le satisfera moins, même si cette dernière constitue une bonne option.

Ces dernières décennies ont vu de nouvelles théories sur la façon avec laquelle nous prenons des décisions, notamment l'ouvrage de KAHNEMAN (2000) qui expose une distinction entre la part d'un individu vivant une expérience et celle qui s'en souvient. Cette dualité entre l'*experiencing self* (le « moi qui vit l'instant présent ») et le *remembering self* (le « moi qui se souvient ») pose un problème aux concepteurs de systèmes de recommandation : doivent-ils favoriser le bonheur de l'une ou de l'autre de ces deux entités qui constituent chaque personne ?

Pour ce qui est des instants heureux, l'accumulation des activités semble être la stratégie à adopter, mais leur évaluation ne peut être faite sur un point de vue seulement culturel. Comme nous l'avons souligné dans le chapitre précédent, les préférences horaires constituent un aspect important des attentes des touristes et semblent pas être corrélées aux goûts culturels.

Enfin, KAHNEMAN (2000) énonce la loi de l'apogée-fin, une heuristique psychologique selon laquelle le jugement d'une personne envers une expérience est soumis à des biais cognitifs qui l'amènent à se remémorer le plus intense et le dernier moment plutôt que la moyenne des sentiments vécus. Tirer partie de cet effet lors de la construction de parcours personnalisés pourrait amener à créer de meilleurs souvenirs de vacances ou d'évènements pour le touriste.

Ce chapitre introduit donc un nouveau problème de recommandation de visite culturelle en s'appuyant sur l'étude développée au chapitre précédent mais aussi sur les principes de la loi de l'apogée-fin.

1. <https://www.tripadvisor.fr>

7.2 La loi de l'apogée-fin dans la recommandation de parcours touristiques personnalisés

Dans la littérature, il n'est nulle part fait mention de l'utilisation de la loi de l'apogée-fin dans la construction de systèmes de recommandation, bien que les travaux de NAWIJN, MITAS et al. (2013) concluent que l'industrie du tourisme devrait s'efforcer à créer un tel effet. Dans cette section, nous proposons donc la définition d'un tel enjeu, puis exposons les pistes que nous envisageons pour y parvenir, en nous basant sur les résultats de l'étude du chapitre 6.

7.2.1 Définition du problème

La problématique est donc de proposer une stratégie de recommandation qui, d'une part, propose des parcours touristiques correspondant aux attentes du visiteur, mais aussi, organise le séjour de sorte que l'expérience qu'il en retire soit la plus satisfaisante au regard de la loi de l'apogée-fin. Nous cherchons donc à résoudre une forme de TTDP, qui considère cet effet en plus des attentes du touriste sur le contenu et l'organisation du parcours. Nous établissons donc le PE-PVRP_c, qui est formulé par la définition 3 ci-dessous :

Définition 3 : Le problème de recommandation de visite personnalisée en ville considérant l'effet de l'apogée-fin – PE-PVRP_c

Un parcours touristique conseillé à un visiteur doit tenir compte de ses préférences et contraintes vis à vis de l'offre culturelle que lui propose un territoire, tout en cherchant à les organiser pour qu'il en retire une meilleure expérience au sens de sa vision du séjour et des principes de l'apogée-fin.

Nous reprenons les concepts de préférences et de contraintes du touriste de la définition 2 sur le TTDP, c'est à dire la différence entre des attentes et des restrictions sur le parcours et y ajoutons un objectif sur l'organisation des activités qui le composent. Ainsi un agent de recommandation de parcours suivant cette définition devrait sélectionner les activités jugées pertinentes au regard des préférences d'un visiteur et les arranger afin qu'elles marquent une apogée et une fin satisfaisante par rapport au reste du séjour, tout en respectant toutes les contraintes imposées.

7.2.2 Vision adoptée pour l'élaboration du parcours

Les principes sur lesquels reposent la conception d'un système de recommandation pour répondre à notre problématique sont variés, comme nous l'avons vu dans le chapitre 2. Ici, nous discutons des lignes de conduite guidant la méthodologie de recommandation que nous avons mise en place. Celles-ci sont réparties en plusieurs catégories, celles ayant trait aux envies du touriste, à la construction et à l'estimation du parcours.

Il est à noter que, si la définition précédente présente le problème de manière généraliste, les choix adoptés pour nos algorithmes de recommandation sont en partie conditionnés par le terrain d'expérimentation que nous avons choisi, à savoir le « Festival OFF d'Avignon » qui est présenté dans l'annexe A.

Les envies du touriste

Le profil touristique est donc composé de contraintes, qui sont des dispositions imposées à la séquence d'activités, et de préférences permettant de guider le système dans ses choix, qui sont elles-mêmes réparties en deux catégories. La première considère les goûts culturels, tel le registre théâtral ou le type de spectacle vivant et la deuxième catégorie se rapporte à l'organisation du séjour, comme les horaires préférés pour voir un spectacle ou au contraire laisser du temps libre au visiteur.

Les préférences culturelles sont exprimées par l'utilisateur en lui offrant la possibilité de sélectionner parmi un ensemble de caractéristiques, celles pour lesquelles il a une affinité particulière, qu'elle soit positive ou négative. Cette méthode est en partie guidée par le corpus de données, mais est suffisamment générique pour être appliquée à de nombreux autres cas comme en témoigne la littérature ou nos travaux sur les musées présentés au chapitre 4.

Les préférences d'organisation énoncent quant à elles au système l'emploi du temps que voudrait respecter un touriste tout au long de son séjour. Il est représenté sous la forme d'un actogramme, qui peut être obtenu à partir d'un ensemble prédéfini en interrogeant le visiteur ou en lui laissant la possibilité de le renseigner lui-même. Cet outil issu de travaux en géographie nous a déjà permis de représenter des parcours et des profils dans le chapitre 6.

À partir de ces éléments, nous pouvons construire un profil touristique qui permettra de traduire les envies du touriste et guider le système dans ses choix lors de la création du parcours.

La construction du parcours

Comme nous l'avons énoncé plus tôt, la finalité d'un système de recommandation est de proposer un parcours au touriste sous la forme d'une séquence d'activités qui lui sont conseillées. Toutefois, il existe une ambiguïté autour du terme *parcours* qui peut être employé aussi bien pour désigner une succession d'étapes à effectuer qu'un itinéraire complet à suivre tout au long du séjour. Or, comme nous l'avons vu dans le chapitre 3, la planification d'un itinéraire complet est une tâche ardue qui constitue un domaine de recherche autour du potentiel piétonnier d'un territoire urbain et n'est pas incluse dans nos travaux.

Nos travaux se concentrant sur la recommandation pendant le « Festival OFF d'Avignon », nous postulons donc que l'utilisateur a déjà fait le choix de sa destination et n'intégrons pas la décision de son lieu de résidence durant le séjour. De surcroît, cette manifestation prend majoritairement place en centre-ville, devenant exclusivement piéton pour l'occasion, rendant les possibilités offertes par la multimodalité moins intéressantes à traiter. Nous nous limiterons donc à la marche à pied en ville et l'utilisation d'un véhicule personnel pour tous les déplacements interurbains.

Il est aussi à noter qu'une grande majorité des spectacles sont jouées plusieurs fois au cours de l'évènement et que chacune de ces représentations se déroulant à une date et une heure précise, nous considérons qu'elles constituent toutes des activités à part entière.

Enfin, notre l'objectif étant de modéliser l'effet de l'apogée-fin, il nous paraît plus simple dans un premier temps de produire un seul parcours continu du début à la fin du séjour, plutôt que plusieurs recommandations quotidiennes.

La qualité du séjour

L'élaboration du parcours se fait à partir du profil touristique et en suivant trois critères opérant à de multiples échelles reflétant chacune un point de vue différent : l'attrait individuel, la pertinence d'une succession et l'existence d'un effet d'apogée-fin.

Premièrement, chaque activité est prise en compte individuellement en jugeant son adéquation à la fois vis à vis des goûts culturels du visiteur, mais aussi de sa place dans l'emploi du temps qu'il souhaite respecter.

Deuxièmement, les paires d'activités successives sont aussi évaluées selon les mêmes facteurs, en suivant de nouvelles méthodes. La mesure de l'efficacité de ces séquences de points d'intérêt culturel permet d'introduire une évaluation de deux notions, la cohérence et la temporalité du parcours. En effet, on peut juger de la pertinence de recommander

une paire d'activité plutôt qu'une autre en fonction de leur similitude en termes de registre théâtral, ce qui permet de grouper les pièces d'un même genre ou au contraire d'offrir la possibilité d'alterner les types de spectacles. Par ailleurs, chaque paire peut aussi être vue comme un temps durant lequel le visiteur est libre d'effectuer tout autre chose. Par conséquent attribuer une valeur à ces moments libres par rapport à son emploi du temps permet de mieux organiser sa journée.

Troisièmement, nous souhaitons appliquer les principes de la loi de l'apogée-fin lors de l'évaluation du parcours, c'est-à-dire en estimant que le bénéfice qu'en retirera le visiteur à son retour dépend fortement du moment le plus intense et de la fin de son séjour. Le calcul de cet effet nécessite de déterminer un découpage en périodes du séjour, de pondérer chacune d'elle pour identifier la plus intense, puis combiner ces valeurs pour obtenir la mesure finale. En outre, les travaux de KEMP et al. (2008) et NAWIJN, MITAS et al. (2013) évaluent la satisfaction des touristes quotidiennement pendant le séjour et ceux de COJUHARENCO et al. (2008) définissent la valeur de l'effet apogée-fin par une moyenne de la plus intense et la dernière expérience du touriste. Néanmoins, ces derniers considèrent des impressions instantanées de satisfaction et non un ressenti sur un intervalle de temps. Notre système de recommandation devra donc définir une mesure associée à chaque période, en fonction des activités recommandées qu'elle englobe.

7.3 Formalisation du problème

Afin de pouvoir adresser le problème de recommandation en ville (PE-PVRP_c) que nous venons de définir, il est nécessaire au préalable de l'exprimer sous la forme d'un objet mathématique utilisable dans le cadre d'une résolution à l'aide de modèles d'optimisation linéaire. Cette description nécessite tout d'abord de préciser l'offre culturelle accessible au visiteur ainsi que de pouvoir comparer les différentes activités qui la composent de diverses manières. Nous pourrions ensuite établir un modèle cohérent et flexible permettant de s'adapter à de nombreux scénarios.

7.3.1 L'offre culturelle en tant que donnée géographique et temporelle

L'offre culturelle mise à disposition du visiteur reflète l'éventail des événements culturels qui lui sont accessibles tout au long de son séjour. Ces éléments peuvent être mis en relation de plusieurs façons, par rapport au visiteur, afin de déterminer les plus intéressants, mais aussi les uns par rapport aux autres. Dans le chapitre précédent, nous avons établi par la définition 2 que ces activités, unités de base d'un parcours touristique,

sont délimitées dans l'espace et dans le temps. Ce sont ces deux dimensions qui nous intéressent dans cette section.

Une des différences essentielles par rapport aux visites de musées est la place prépondérante de la temporalité dans le parcours : choisir de planifier une activité a des conséquences sur celles que l'on peut effectuer avant ou après celle-ci. Les horaires des activités sont bien sûr déterminants mais il ne faut aussi tenir compte de la topologie de l'espace dans lequel le touriste évolue.

Au sens de ALLEN (1984), un évènement est un phénomène ayant un début et une fin qu'il est possible de caractériser par un intervalle temporel. Dans ces mêmes travaux, il définit une logique temporelle sous la forme de treize relations, toutes mutuellement exclusives, entre les intervalles ainsi que des axiomes qui régissent leurs comportements. Par définition, toute activité est donc assimilable à un intervalle temporel et les relations de ALLEN, représentées par la figure 7.1, permettent de les comparer les une par rapport aux autres.

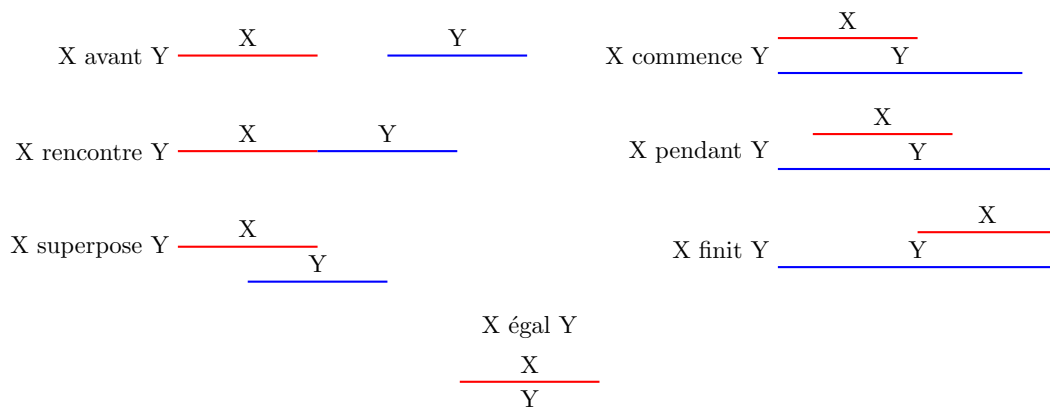


FIGURE 7.1 – Les différentes relations entre deux intervalles temporels (selon ALLEN 1984)

Le second point impactant fortement les choix durant la création d'un parcours est lié à la position des points d'intérêt, la topologie de l'espace dans lequel ils sont situés et les modalités de transport offertes au visiteur. En effet, la localisation des activités joue un rôle essentiel dans les possibilités de parcours car conditionne notamment les modes de transport et les itinéraires envisageables pour se rendre d'un point d'intérêt à l'autre. Ces deux facteurs sont déterminants pour estimer le temps de trajet nécessaire pour rejoindre deux points d'intérêt et donc identifier les séquences impossibles à effectuer. Le problème formulé après ne tient pas compte de ces temps de trajets, qui seront intégrés au système de recommandation présenté au chapitre 8.

7.3.2 Proposition d'une modélisation mathématique

Le PE-PVRP_c, tel que détaillé dans la section 7.2 sélectionne des activités d'une offre culturelle à l'aide de différentes métriques qui sont combinées pour obtenir un parcours touristique plaisant aux yeux du visiteur. Nous pouvons voir ce problème comme la manipulation de différents objets et proposons une décomposition en un 8-uplet PE-PVRP_c = (A, P, C, λ, u, s, e, r).

L'ensemble A traduit l'offre culturelle disponible durant le séjour touristique tandis que P permet de considérer l'effet de l'apogée-fin perçu par le touriste, en répartissant les activités en périodes. Les contraintes regroupées dans C, qui sont détaillées dans la sous-section 7.3.3, offrent une grande adaptabilité aux exigences du visiteur et aux spécificités du territoire. Le vecteur λ influe sur l'organisation du parcours en accentuant l'importance des différentes métriques que sont l'attrait individuel, la pertinence des enchaînements et la loi de l'apogée-fin, ainsi que l'évaluation des deux périodes centrales de cette dernière. Enfin, les fonctions u, s, e et r reflètent les opinions du visiteur et des acteurs culturels à la fois sur les activités prises individuellement, mais aussi les paires qu'elles peuvent former dans le parcours.

Ainsi, nous définissons mathématiquement ce 8-uplet (A, P, C, λ, u, s, e, r) avec :

A l'ensemble fini non vide des activités qu'il est possible d'effectuer.

Comme vu plus haut, les éléments de cet ensemble constituent des intervalles temporels, ce faisant, à partir des relations de ALLEN, on définit < la relation binaire dans A telle que $\forall a, b \in A, a < b$ si et seulement si l'activité a finit avant que l'activité b ne débute.

$$\forall a, b \in A, a < b \iff BEFORE(a, b) \vee MEETS(a, b)$$

∴

$$\forall a \in A, \neg(a < a) \quad \text{antiréflexive}$$

$$\forall a, b \in A, a < b \implies \neg(b < a) \quad \text{asymétrique}$$

$$\forall a, b, c \in A, a < b \wedge b < c \implies a < c \quad \text{transitive}$$

P l'ensemble des périodes de séjour, formant un découpage temporel en groupes d'activités.

Chaque activité appartenant à une seule période, P est donc une partition de A .

$$\begin{aligned} \emptyset &\notin P \\ \bigcup_{p \in P} p &= A \\ \forall p_1, p_2 \in P, p_1 \cap p_2 &= \emptyset \end{aligned}$$

C l'ensemble des patrons de contraintes de groupes d'activités où chaque élément est un 4-uplet $c = (A_c, w_c, \mathcal{R}_c, t_c)$ avec :

A_c l'ensemble des activités partageant la contrainte ($A_c \subseteq A$)

w_c fonction $w_c : \mathcal{D}_{w_c} = A_c \rightarrow \mathbb{R}$ donnant pour chaque activité du groupe le poids associé dans la contrainte

\mathcal{R}_c la relation binaire dans \mathbb{R} associée à la contrainte ($=, \neq, \leq, <, >, \geq$)

t_c seuil de la contrainte dans \mathbb{R} , dont la valeur est cohérente avec la fonction w_c

λ le vecteur de poids associés aux préférences sur l'organisation du parcours :

λ_1 importance de l'utilité des activités au sein du parcours

λ_2 importance de la pertinence des séquences d'activités au sein du parcours

λ_3 importance de la règle de l'apogée-fin au sein du parcours

λ_4 prévalence de l'apogée par rapport à la fin du parcours dans la règle de l'apogée-fin

$$\begin{aligned} \lambda_1, \lambda_2, \lambda_3 &\in \mathbb{N} \\ \sum_{i=1}^3 \lambda_i &> 0 \\ \lambda_4 &\in [0, 1] \end{aligned}$$

u la fonction $u : \mathcal{D}_u = A \rightarrow [-1, +1]$ donnant pour chaque activité $a \in A$ l'utilité d'effectuer cette activité.

s la fonction $s : \mathcal{D}_s \subseteq A \rightarrow [-1, +1]$ dont le domaine de définition (\mathcal{D}_s) est l'ensemble des activités dans A par lesquelles un parcours peut commencer, et pour chaque $a \in \mathcal{D}_s$, $s(a)$ est l'utilité de commencer par a .

e la fonction $e : \mathcal{D}_e \subseteq A \rightarrow [-1, +1]$ dont le domaine de définition (\mathcal{D}_e) est l'ensemble des activités dans A par lesquelles un parcours peut se terminer, et pour chaque $a \in \mathcal{D}_e$, $e(a)$ est l'utilité de terminer par a .

r la fonction $r : \mathcal{D}_r \subseteq A^2 \rightarrow [-1, +1]$ dont le domaine de définition (\mathcal{D}_r) est l'ensemble des couples $(a, b) \in A^2$ tel qu'il est possible d'effectuer l'activité a puis l'activité b et pour lequel la valeur $r(a, b)$ est la pertinence d'effectuer cette succession d'activités. Ainsi on en déduit les propriétés suivantes :

$$\begin{aligned} \forall (a, b) \in \mathcal{D}_r &\implies a \prec b \\ \therefore & \\ \forall a \in A, (a, a) &\notin \mathcal{D}_r \\ \forall (a, b) \in \mathcal{D}_r &\implies (b, a) \notin \mathcal{D}_r \end{aligned}$$

Formuler le PE-PVRP_c de telle manière permet une flexibilité suffisante pour s'adapter à différents types de séjours, car il n'est imposé aucune restriction sur les activités si ce n'est de pouvoir leur attribuer une période et quantifier leur attrait pour le visiteur.

7.3.3 Flexibilité du système de contraintes

Le modèle PE-PVRP_c inclut un élément C permettant de définir des contraintes qui sont propres au profil touristique considéré. Chacun de ces éléments est un patron de contrainte, c'est-à-dire un gabarit qui exprime une contrainte sur les activités retenues par le processus de génération de parcours.

Comme nous l'avons vu plus haut, chaque élément est un 4-uplet $(A_c, w_c, \mathcal{R}_c, t_c)$ qui indique au système de recommandation que le sous-ensemble A_c des activités doit respecter une contrainte. Cette dernière est exprimée arithmétiquement, en soumettant la somme des activités concernées, avec la pondération donnée par w_c , à la valeur t_c par la relation \mathcal{R}_c .

Sans que nous ne présentions un prototype d'interface utilisateur permettant d'exprimer de telles contraintes, voici quelques exemples illustrant les divers scénarios auxquels pourrait être confronté un profil touristique :

Ne pas recommander plusieurs fois la même animation

Dans notre corpus de données, une même pièce de théâtre est jouée plus d'une fois et chacune de ses représentations est une activité différente. Afin de ne pas recommander plusieurs fois un spectacle, il lui est associé à chacun une contrainte de la forme suivante :

A_c l'ensemble des activités issues d'une même pièce;

w_c fonction constante donnant pour chaque activité la valeur 1 ;
 \mathcal{R}_c relation « inférieur ou égal » (\leq);
 t_c de valeur 1.

Cette contrainte implique qu'au plus une représentation d'un spectacle pourra être incluse dans le parcours. En produisant celle-ci pour chaque spectacle du « Festival OFF d'Avignon », nous pouvons nous assurer qu'un parcours ne contiendra pas plus d'une représentation de chaque pièce de théâtre.

Inclure dans le parcours au moins une activité d'un certain genre

À l'inverse, il est possible de forcer le système de recommandation à inclure dans le parcours certaines activités. Si l'on veut qu'il y est au moins n activités, la formulation est :

A_c l'ensemble des activités dont on doit au moins faire une recommandation ;
 w_c fonction constante donnant la valeur 1 ;
 \mathcal{R}_c relation « supérieur ou égal » (\geq);
 t_c de valeur n .

On peut aussi l'étendre à d'autres types d'activités. En imaginant inclure les restaurants dans les activités, on peut ainsi forcer le parcours à inclure une pause déjeuner et/ou un dîner dans le parcours. On remarque aussi qu'il est possible d'inclure exactement toutes les activités identifiées en utilisant $t_c = |A_c|$.

Limiter un type d'activités à une durée totale

Contrairement aux visites muséales, la contrainte de temps n'est ici pas forcément présente. Si néanmoins, on désire limiter la durée d'une visite à une durée T , il est possible de l'exprimer comme suit :

A_c l'ensemble des activités du genre concerné ;
 w_c fonction donnant la durée de chaque activité comme un nombre positif ;
 \mathcal{R}_g relation « inférieur ou égal » (\leq);
 t_c de valeur T , durée totale à ne pas dépasser.

Ne rien effectuer en dehors de certains horaires

Une des situations envisageables est de vouloir limiter les recommandations à une plage horaire bien délimitée. En d'autres termes, interdire les activités en dehors des horaires considérés, ce qui peut s'exprimer avec :

A_c l'ensemble des activités en dehors des horaires désirés ;

w_c fonction constante donnant la valeur 1 ;

\mathcal{R}_c relation d'égalité (=) ;

t_c de valeur 0.

Il est toutefois à noter que l'utilisation des actogrammes comme explicité par la suite permet de considérer des préférences horaires, sans pour autant interdire l'inclusion d'activités en dehors des horaires considérés.

Limites budgétaires

On peut aisément imaginer vouloir fixer un budget limite à ne pas dépasser durant son séjour. En notant d cette limite financière, ces contraintes prennent la forme suivante :

A_c l'ensemble des activités auxquelles il faut appliquer les limites budgétaires ;

w_c fonction donnant le coût de chaque activité ;

\mathcal{R}_c relation « inférieur ou égal » (\leq) ;

t_c de valeur d , budget maximal accordé.

En combinant ces différentes contraintes il est possible de répondre à des scénarios plus complexes, comme par exemple d'inclure tout au long d'un séjour un déjeuner entre 12h et 14h (et uniquement dans cette fenêtre horaire) sans recommander plusieurs fois la même enseigne, tout en ayant un budget à ne pas dépasser.

7.4 Représentation sous forme de graphe

La formulation du PE-PVRP_c qui vient d'être exposée peut être partiellement décrite par un graphe reliant les activités. Cette représentation a différentes propriétés et facilite le positionnement de ce problème vis-à-vis des autres variantes du TTDP.

7.4.1 Définition et construction

Chaque instance possible du problème PE-PVRP_c proposé plus tôt permet de définir un graphe G_c en appliquant la définition 4 énoncée ci-dessous :

Définition 4 : Graphe touristique G_c

Le PE-PVRP_c peut être associé à un graphe orienté G_c dans lequel chaque activité est associée à un unique sommet et un arc relie deux sommets s'il est possible d'effectuer la succession des activités qui y sont associées. De plus, on ajoute deux nœuds start et end ainsi qu'un arc de start vers chaque activité pouvant débiter un parcours puis un arc de end vers chaque activité pouvant le terminer.

En notant la fonction bijective $f_{G_c} : A \rightarrow V \setminus \{start, end\}$ donnant pour chaque activité $a \in A$ l'unique sommet $v \in V$ lui étant associé, on peut exprimer $G_c = (V, E)$ avec :

$$V = \{start, end\} \cup f_{G_c}(A) \quad (7.1a)$$

$$E_s = \left\{ (start, v) \mid \exists a \in \mathcal{D}_s, f_{G_c}(a) = v \right\} \quad (7.1b)$$

$$E_e = \left\{ (u, end) \mid \exists a \in \mathcal{D}_e, f_{G_c}(a) = u \right\} \quad (7.1c)$$

$$E_a = \left\{ (u, v) \mid \exists (a, b) \in \mathcal{D}_r, f_{G_c}(a) = u, f_{G_c}(b) = v \right\} \quad (7.1d)$$

$$E = E_s \cup E_e \cup E_a \quad (7.1e)$$

Remarque. Quel que soit $v \in V$, il n'existe aucun arc (v, v) , $(v, start)$, (end, v) ou $(start, end)$ dans E .

À titre d'exemple, la figure 7.2 expose le graphe associé aux représentations du « Festival OFF d'Avignon » se déroulant le 2015-07-11 entre 15:45:00 et 16:00:00.

Conformément à la définition 4, chaque sommet correspond à une pièce de théâtre se déroulant (au moins en partie) dans l'intervalle de temps et un arc reliant deux nœuds exprime la possibilité d'assister à une pièce finissant dans cet intervalle, puis une autre y débutant. Ceci donne par ailleurs un aperçu de la complexité du problème, considérant qu'il ne s'agit que d'une infime partie de l'évènement.

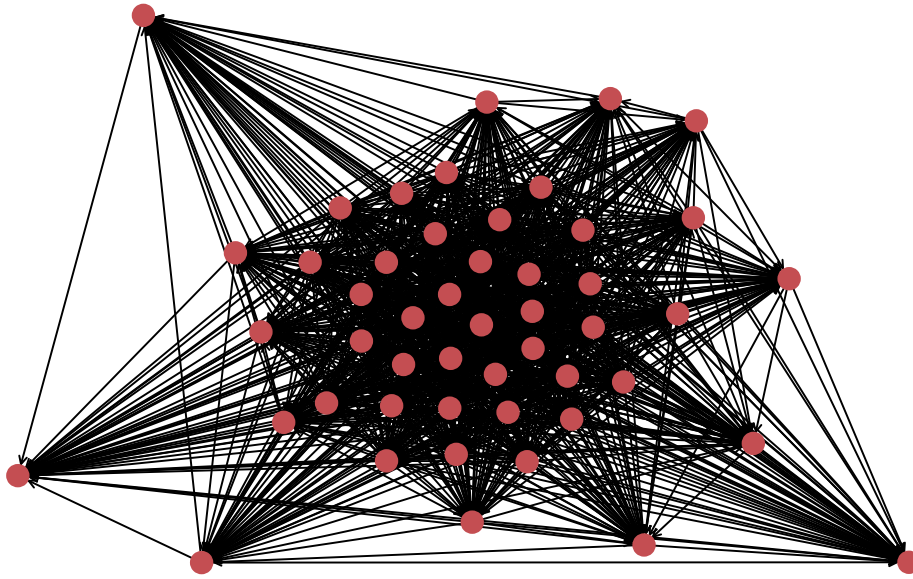


FIGURE 7.2 – Graphe associé à une infime partie du « Festival OFF d'Avignon »

7.4.2 Propriétés et relation avec les parcours

Acyclicité

Étant donné que les arcs entre les différents sommets représentent des liens temporels, on peut intuitivement admettre qu'un tel graphe ne comporte aucun cycle, cette proposition 2 est démontrée ci-après.

Proposition 2. G_c est un graphe orienté acyclique.

Démonstration. Admettons qu'il existe un cycle simple² dans G_c , c'est-à-dire une séquence $c = (v_0, \dots, v_n)$ ayant les propriétés suivantes :

$$n \in \mathbb{N}^+ \quad (7.2a)$$

$$\forall i \in [0, n], v_i \in V \quad (7.2b)$$

$$\forall i, j \in [1, n-1] \mid i \neq j, v_i \neq v_j \quad (7.2c)$$

$$\forall i \in [0, n-1], \exists (v_i, v_{i+1}) \in E \quad (7.2d)$$

$$v_0 = v_n \quad (7.2e)$$

On a les trois cas suivants :

2. Tel que défini dans AHUJA et al. (1993)

Cas $v_0 = start$ donc $v_n = start$ et la suite de la séquence peut être de taille :

$n = 1$ donc par (7.2d) $(start, start) \in E$, ce qui est impossible par (7.1b)

$n > 1$ donc par (7.2d) $(v_{n-1}, start) \in E$, ce qui est impossible par (7.1b)

Cas $v_0 = end$ donc $v_n = end$ et la suite de la séquence peut être de taille :

$n = 1$ donc par (7.2d) $(end, end) \in E$, impossible par (7.1c)

$n > 1$ donc par (7.2d) $(end, v_1) \in E$, ce qui est impossible par (7.1c)

Cas $\exists a \in A \mid f_{G_c}(a) = v_0$ la suite de la séquence peut être de taille :

$n = 1$ donc on a $(v_0, v_0) \in E$ par (7.2d) et par conséquent $(a, a) \in \mathcal{D}_r$ par (7.1d), ce qui est impossible par définition de \mathcal{D}_r

$n > 1$ on a donc les étapes intermédiaires $\tilde{c} = (v_1, \dots, v_{n-1})$ et :

Si $start \in \tilde{c}$, alors $\exists v_i \in \tilde{c} \mid (v_i, start) \in E$ par (7.2d), ce qui est impossible par (7.1b)

Si $end \in \tilde{c}$, alors $\exists v_i \in \tilde{c} \mid (end, v_i) \in E$ par (7.2d), impossible par (7.1c)

Si $\forall i \in [1, n-1], \exists b_i \in A \mid f_{G_c}(b_i) = v_i$, alors par (7.2d) pour tout élément v_i dans \tilde{c} , $(v_i, v_{i+1}) \in E$ et $(v_0, v_1), (v_{n-1}, v_0) \in E$. De plus, $E = E_a$ car $start, end \notin E$. Donc par (7.1d) pour tout i , $(b_i, b_{i+1}) \in \mathcal{D}_r$ et $(a, b_1), (b_{n-1}, a) \in \mathcal{D}_r$.

Par définition de \mathcal{D}_r , on a alors $a \prec b_1, b_{n-1} \prec a$ et pour tout $i, b_i \prec b_{i+1}$. Ainsi, la relation \prec étant transitive, on déduit $b_1 \prec b_{n-1}$. Par cette même propriété, $a \prec b_1, b_1 \prec b_{n-1}$ et $b_{n-1} \prec a$ implique $a \prec a$. Or la relation \prec est antiréflexive, donc c'est impossible.

Il ne peut donc pas y avoir de cycle dans G_c . □

Parcours et chemin dans G_c

Un parcours touriste peut être assimilé à un chemin dans G_c d'origine $start$ et d'extrémité end passant par tous les sommets associés aux activités constituant le séjour. Toutefois, seules les relations temporelles entre les activités étant représentées dans ce graphe, un tel chemin dans G_c ne constitue donc pas toujours une solution valide au problème. En effet, si l'ensemble C des patrons de contraintes n'est pas vide, certains chemins correspondent à des parcours violant les contraintes du profil touristique. Il est à noter que c'est souvent le cas dans notre situation, du fait de la représentation de chaque évènement en de multiple activités, reflétant les horaires auxquels celui-ci est accessible.

Relations avec les formulations existantes

Comme nous l'avons vu dans le chapitre 3, le TTDP, description générale du problème de recommandation touristique, a été traité de multiples manières.

Contrairement aux approches basées sur le PCTSP, l'objectif n'est pas ici de minimiser un coût du trajet mais de maximiser l'intérêt d'un parcours, ce qui rapproche notre formulation des variantes du PTP et du OP. En effet, bien que nous ne considérons pas dans nos travaux de coût associé au déplacement, celui-ci peut être intégré dans la fonction d'utilité u auquel cas résoudre notre problème revient à résoudre le PTP en y ajoutant l'effet de l'apogée-fin. Les variantes basées sur des modèles OP incluent quant à elles une contrainte sur le coût du trajet (exprimé en temps ou en distance) qui est optionnelle dans notre formulation.

7.5 Conclusion

Dans ce chapitre, nous avons défini une problématique constituant un changement de paradigme dans la manière de construire un parcours touristique sur la base de travaux en sociologie, géographie et psychologie. À partir de ce cadre général, nous avons développé des idées maîtresses autour desquelles un système de recommandation devrait s'appuyer pour construire des visites plaisantes. Elles reposent sur une évaluation multicritère et multiscalaire du parcours culturel, afin de proposer des séjours agréables. La recommandation d'une activité se fait en considérant tant l'intérêt culturel que la cadence de la visite, qui sont deux points majeurs mis en avant par le chapitre précédent, mais aussi par une vision séquentielle du parcours, en mesurant la pertinence de chaque étape par rapport à celle qui la précède et la succède. Immanquablement la loi de l'apogée-fin constitue la clé de voûte de ce problème et demande d'être adaptée au contexte de la recommandation touristique afin que le visiteur puisse vivre une expérience satisfaisante, voire mémorable.

À partir de ces concepts, nous avons proposé une formulation du problème qui peut être utilisée pour créer un système de recommandation mais il apparaît que cette approche novatrice de la proposition de visite culturelle requiert de mettre au point un modèle adapté, tâche qui est présentée au chapitre suivant.

Chapitre 8

Un système de recommandation pour la génération de parcours touristiques

Sommaire

8.1	Introduction	126
8.2	Architecture et fonctionnement	126
8.3	Estimation de l'attrait des activités	128
8.3.1	Intérêt culturel	128
8.3.2	Cadence de visite	129
8.3.3	Intégration des mesures	131
8.4	Construction du parcours proposé au visiteur	133
8.4.1	Critères d'optimisation	133
8.4.2	Modélisation par recherche de chemin	136
8.5	Expériences et résultats	145
8.5.1	Mesures de l'attrait des activités	145
8.5.2	Scénarios et paramètres	147
8.5.3	Complexité en temps et en espace	149
8.5.4	L'influence des actogrammes sur les recommandations quotidiennes	149
8.5.5	L'effet apogée-fin sur des séjours d'une semaine	152
8.6	Conclusion	154

8.1 Introduction

L'objectif d'un système de recommandation est de suggérer à l'utilisateur un parcours qui inclut la visite des points d'intérêt jugés les plus pertinents en fonction de ses goûts (GAVALAS et al. 2014). Ces aides à la visite se substituent donc au moins en partie au processus de prise de décision et ont la responsabilité de discerner les activités qui intéresseront le touriste de celles qu'il n'envisagerait pas.

La stratégie actuelle des systèmes de recommandation se concentre sur l'identification et l'estimation des points d'intérêt culturel correspondant aux besoins de l'utilisateur (BENOUARET 2017; CENAMOR et al. 2017) et le développement d'heuristiques permettant de diminuer les temps d'attentes pour avoir une proposition (EXPÓSITO et al. 2019; WÖRNDL et al. 2017). On peut néanmoins identifier deux points à améliorer avec ces approches. D'une part, l'estimation des points d'intérêt n'est souvent faite qu'à partir des goûts culturels du touriste et, d'autre part, l'assomption forte que la stratégie de recommandation est la même quelle que soit la durée du séjour.

Dans le chapitre 7, nous avons formulé le PE-PVRP_c, une nouvelle approche de recommandation de visite culturelle s'appuyant sur l'effet de l'apogée-fin par KAHNEMAN (2000). En conseillant des parcours pendant lesquels une journée intermédiaire et la fin sont remarquables, nous favorisons la construction de meilleurs souvenirs du touriste sur son séjour.

Ce chapitre propose un système de recommandation qui agrège plusieurs métriques pour construire des parcours touristiques correspondant aux attentes du touriste tout en accentuant l'effet de l'apogée-fin dont dépend le ressenti du touriste après son voyage.

8.2 Architecture et fonctionnement

Dans le chapitre 7, nous avons exposé les lignes directrices d'un système de recommandation ayant comme fondations des travaux issus de plusieurs disciplines, tout en considérant leur applicabilité à notre corpus. La figure 8.1 illustre le fonctionnement et les composants principaux du système développé dans ce chapitre.

La première étape de « Création du profil touristique » transforme les informations à propos de l'utilisateur et de l'offre culturelle en une instance du PE-PVRP_c telle que décrite au chapitre précédent. Cette opération nécessite d'utiliser les goûts culturels et les préférences d'organisation du touriste, afin d'estimer l'attrait de chaque activité et paire d'activités selon ces deux métriques. La seconde étape utilise l'ensemble de ces

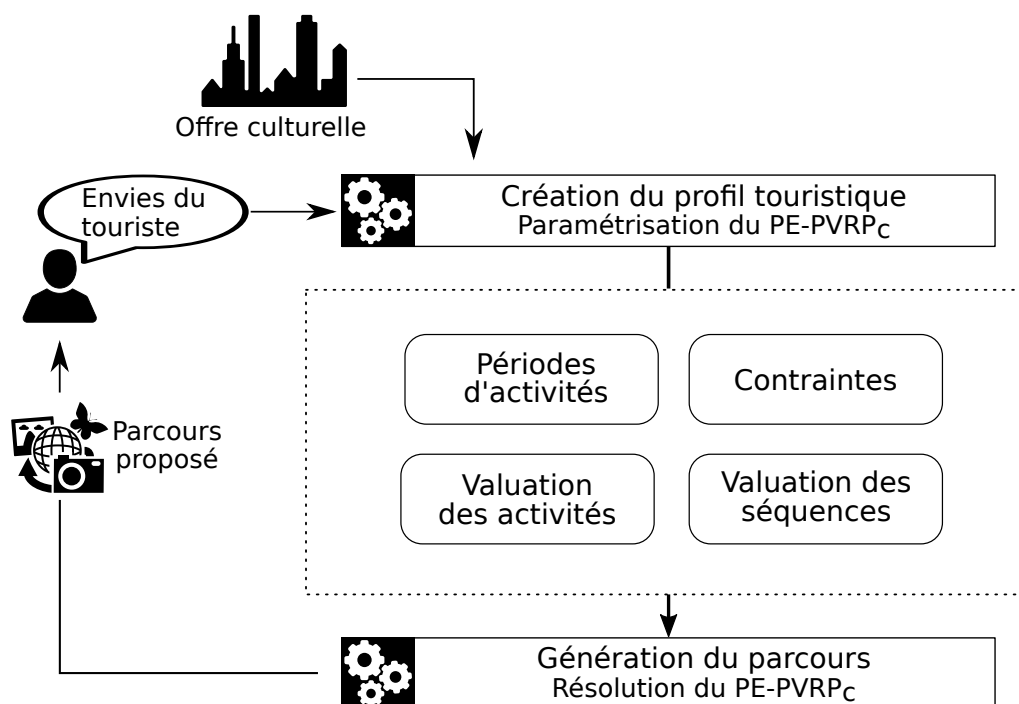


FIGURE 8.1 – Architecture et composants principaux de notre système pour la recommandation de parcours touristiques en ville

éléments afin de formuler le problème par un modèle de MILP qui, une fois résolu, donne un parcours conseillé au visiteur.

Néanmoins, nous avons vu lors de la définition du problème que, pour formuler l'effet de l'apogée-fin, il nous faut répartir les activités en périodes et attribuer à celles-ci une valeur. Les travaux de KEMP et al. (2008) et NAWIJN, MITAS et al. (2013) utilisent un découpage quotidien pour évaluer cet effet sur le tourisme mais le visiteur attribue une note globale à sa journée qu'il est compliquée de prédire. Nous avons donc choisi d'exprimer la valeur de chaque période par la somme des attraits des activités proposées dans le parcours qui la constituent.

Le résultat de notre système est une séquence de points d'intérêt qui est conseillée au touriste et en ce sens, notre approche est assimilée à de la « suggestion d'attractions » au sens de BORRÀS et al. (2014).

Les deux prochaines sections détaillent les deux étapes décrites. Puis, nous évaluons ce modèle en analysant les recommandations qu'il produit sur le « Festival OFF d'Avignon » et les profils touristiques qui sont détaillés en annexe A.

8.3 Estimation de l'attrait des activités

Le processus de « création du profil touristique » est responsable de l'évaluation des éléments dont le parcours peut être constitué, qu'il s'agisse des activités considérées individuellement ou des paires susceptibles d'être formées. Cette section détaille donc les fonctions u et r du PE-PVRP_c qui permettent de valuer chaque noeud et arc du graphe formé par le problème.

De nombreuses approches sont explorées dans la littérature mais, compte tenu du jeu de données que nous utilisons, nous en avons retenu deux : l'intérêt culturel et la cadence de visite. En effet, comme nous l'avons mis en évidence dans le chapitre 6, ces deux métriques sont complémentaires pour une meilleure estimation de l'attrait global d'une activité.

Dans cette section, nous discutons des approches permettant de définir les fonctions u et r du PE-PVRP_c. La première représente l'utilité d'effectuer une activité en elle-même sans tenir compte des éléments qui la précèdent ou la succèdent. La fonction r quant à elle permet de tenir compte de cette séquence d'activité et permet donc d'introduire la temporalité dans le parcours, utilisée ici pour construire une cohérence culturelle, qui peut prendre plusieurs formes.

8.3.1 Intérêt culturel

Tout comme l'approche adoptée dans le chapitre 4 concernant la recommandation de visites muséales, nous utilisons une représentation vectorielle pour comparer chaque activité aux préférences du visiteur. Pour chaque activité $a \in A$, est déterminé un ensemble de caractéristiques C_a et l'union de tous ceux-ci (notée C) permet de créer un espace vectoriel où chaque dimension est une unique caractéristique. Nous pouvons ensuite positionner chaque activité dans cet espace en valuant son adéquation envers chacune des caractéristiques, formant ainsi un vecteur (noté v_a) par activité. Les valeurs représentant l'adéquation entre une caractéristique et une activité peuvent prendre différentes formes, suivant le degré de précision qu'il est possible de leur donner. Par exemple, celles-ci peuvent être binaires pour dénoter la présence ou l'absence d'un trait (par exemple si on peut se restaurer à un point d'intérêt) ou encore varier dans un intervalle, lorsqu'il est possible d'indiquer un degré de compatibilité qui peut être négatif lorsqu'il y a une discordance certaine (par exemple pour différencier une activité reposante d'une activité physique).

Il est ensuite possible de présenter à un touriste tout ou partie de ces caractéristiques

(en fonction de leur nombre, leur capacité à discriminer les points d'intérêt) et l'inviter à identifier celles qui le concernent et leur attribuer une valeur dépendant de son affinité avec cette dernière. À partir de ces valeurs, on peut définir un vecteur de préférences (v_p) qui représente donc l'utilisateur dans le même espace vectoriel. Enfin, l'équation (8.1) montre l'utilisation de la *similarité cosinus* pour estimer l'intérêt d'une activité en accord avec les préférences culturelles du visiteur.

$$\text{activityInterest}_p(a) = \frac{v_a \cdot v_p}{\|v_a\| \cdot \|v_p\|} \quad (8.1)$$

La mesure donne donc des valeurs dans l'intervalle $[-1, 1]$ pour chaque activité représentant son adéquation avec les préférences culturelles du touriste.

Concernant les paires d'activités successives formées par le parcours, bien que nous pourrions elles aussi les comparer avec les préférences du visiteur, nous proposons de nous baser sur une mesure de leur ressemblance, toujours via leur position dans l'espace vectoriel et au moyen d'une *similarité cosinus*. Ainsi cette mesure, exprimée par l'équation (8.2), représente une sorte de cohérence culturelle du parcours, qui tend à regrouper les pièces ayant des caractéristiques similaires.

$$\text{sequenceInterest}(a, b) = \frac{v_a \cdot v_b}{\|v_a\| \cdot \|v_b\|} \quad (8.2)$$

À l'inverse, il serait possible d'utiliser l'opposé de cette mesure de similarité pour inciter le système à créer des parcours alternant les styles culturels des activités.

8.3.2 Cadence de visite

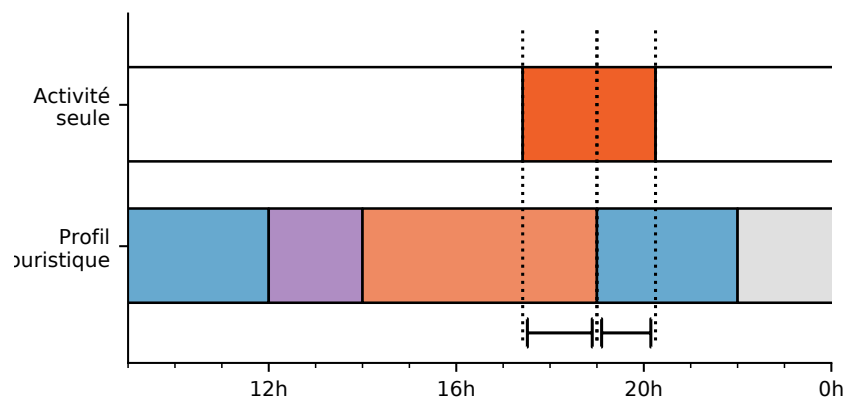
Dans le chapitre 6, nous avons mis en évidence l'importance de prendre en compte le moment où une activité se déroule, ainsi que leur place et l'enchaînement dans le parcours. Les actogrammes, couramment utilisés par les géographes, se révèlent être efficaces pour décrire ces informations spatiales et temporelles d'un itinéraire. De plus, nous avons montré qu'ils peuvent être utilisés pour représenter ces aspects des parcours et mesurer leur distance relative. Ce processus permet une plus grande souplesse dans la définition du profil touristique et englobe de précédentes approches comme celle de RODRÍGUEZ et al. (2012) qui utilise la notion de vitesse de visite permettant d'espacer les activités.

Pour le système que nous proposons, la création du profil touristique revient à associer l'utilisateur à un actogramme qui représentera ses goûts en termes d'organisation du séjour. Cela peut être réalisé par exemple via un questionnaire permettant de déterminer

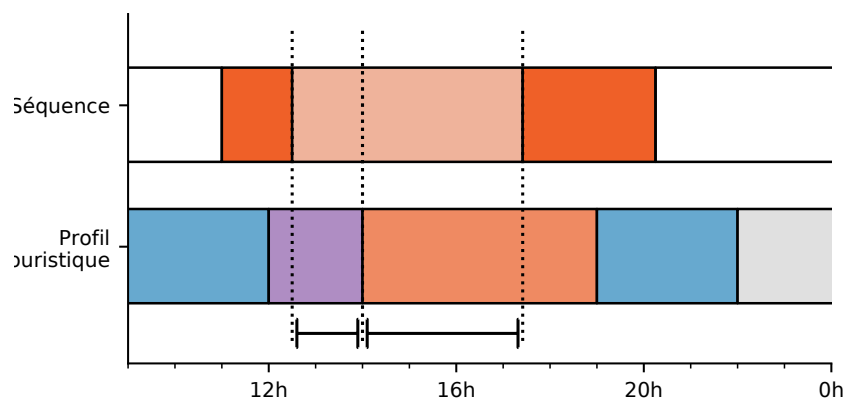
un modèle à utiliser ou en le laissant indiquer lui-même son emploi du temps. Suite à la création du profil touristique, le processus permettant d'identifier les différences entre deux actogrammes présenté dans la sous-section 6.3.3 est réutilisé en substituant l'actogramme à évaluer par un autre, représentant l'élément dont nous souhaitons estimer l'intérêt, qu'il s'agisse d'une activité seule ou de la succession de deux d'entre elles.

Pour ce qui est de l'attrait des activités prises individuellement, l'actogramme est simplement constitué de l'activité elle-même. Quant à l'évaluation des paires potentielles d'activités, un élément fictif est créé, représentant un laps de temps durant lequel aucune recommandation n'est faite, mais qui est précédé et succédé d'activités.

Une fois le processus de comparaison des actogrammes appliqué, on obtient un ensemble de différences vis à vis du profil touristique. Ces dernières couvrant le séjour entier, on retient uniquement celles ayant cours durant l'élément dont il faut estimer la place dans le parcours. la figure 8.2 image le processus de construction et les différences dont il faut tenir compte pour cette première étape d'estimation de l'attrait.



(a) Cas d'une activité seule



(b) Cas d'une séquence d'activités

FIGURE 8.2 – Utilisation des actogrammes comme outil de mesure du confort

Enfin, la mesure finale de l'attrait est calculée en effectuant la moyenne des durées des différences chacune pondérée par un facteur d'adéquation entre l'élément et l'actogramme. L'équation (8.3) donne la proximité entre une activité (ou une paire d'activités successives) et le profil touristique.

$$\text{comfortProximity}(D) = \frac{\sum_{(p,r,\Delta t) \in D} \Delta t \times \text{adequacy}(p, r)}{\sum_{(p,r,\Delta t) \in D} \Delta t} \quad (8.3)$$

Avec :

D l'ensemble des différences obtenu par l'étape précédente ;

adequacy fonction $\text{adequacy} : \mathcal{D}_{\text{adequacy}} \rightarrow [-1, +1]$ donnant pour chaque différence un taux d'adéquation entre l'élément à estimer et l'actogramme du profil.

En conclusion, les deux fonctions calculant la tendance qu'une activité ou une succession d'activités corresponde avec l'organisation du touriste sont respectivement données par les équations (8.4) et (8.5).

$$\text{activityComfort}(a) = \text{comfortProximity}(\text{Act}(a)) \quad (8.4)$$

$$\text{sequenceComfort}(a, b) = \text{comfortProximity}(\text{Act}(a, b)) \quad (8.5)$$

Ainsi, on peut favoriser ou pénaliser une activité en fonction du moment où elle serait proposée dans le parcours. De même, l'absence de recommandation peut-être évaluée ici par la même méthode, mais la vitesse de la visite représentée par le temps libre entre deux activités pourrait être envisagée.

8.3.3 Intégration des mesures

Ces deux différentes métriques permettent d'évaluer l'attrait des activités qui sont susceptibles d'être incluses dans un parcours. Dans ce contexte multicritère, dissocier ces différentes mesures permet de distinguer les parcours optimisant chacune d'elle parmi l'ensemble des possibles. Or, nous travaillons dans le cadre d'optimisation à un seul objectif, donnant lieu à l'obtention d'une unique solution, et donc différencier les solutions issues des différents critères requiert de résoudre le problème en adaptant les facteurs dans chacune des configurations. Par conséquent, les fonctions u et r sont formulées

comme des combinaisons linéaires des deux différentes métriques où un facteur α permet d'insister plus ou moins sur l'intérêt ou la cadence.

Cependant, nous avons choisi de n'exploiter que deux mesures pour estimer l'attrait des différents éléments du parcours, afin de montrer l'impact de chacune d'elle sur la construction d'un parcours mais dans la littérature, de nombreux autres facteurs ont été proposés. La dimension économique du parcours est d'avantage mise en avant par RODRÍGUEZ et al. (2012) qui distinguent l'utilité et le coût de chaque activité. De même, SANTOS et al. (2017) proposent une évaluation de l'accessibilité des activités en fonction des limitations physiques ou psychologiques de l'utilisateur.

L'intégration de ces autres métriques pourrait être faite de la même manière que le confort et l'intérêt, c'est à dire par une combinaison linéaire des différentes mesures, chacune pondérée par un poids reflétant son importance pour le profil touristique. Ainsi, nous aurions des ensembles M_u et M_r , où chaque élément est le couple (f, w) regroupant la fonction f de mesure et le poids w qui lui est associé. Les équations (8.6a) et (8.6b) donnent respectivement les fonctions u et r du modèle PE-PVRP_c proposé.

$$u(a) = \frac{\sum_{(f,w) \in M_u} f(a) \times w}{\sum_{(f,w) \in M_u} w} \quad (8.6a)$$

$$r(a, b) = \frac{\sum_{(f,w) \in M_u} f(a, b) \times w}{\sum_{(f,w) \in M_u} w} \quad (8.6b)$$

Avec dans notre cas :

$$M_u = \{(\text{activityInterest}_p, \alpha), (\text{activityComfort}, 1 - \alpha)\}$$

$$M_r = \{(\text{sequenceInterest}, \alpha), (\text{sequenceComfort}, 1 - \alpha)\}$$

Ces deux fonctions permettent d'estimer l'importance que revêtent les activités aux yeux du touriste et ce, selon différents points de vue et sur plusieurs échelles. Elles constituent donc les fondations sur lesquelles le reste du système doit s'appuyer pour sélectionner et organiser les points d'intérêt, de manière à élaborer un parcours, en considérant au mieux les préférences du touriste.

8.4 Construction du parcours proposé au visiteur

Maintenant que nous avons un moyen de quantifier l'attrait non seulement de chaque activité qu'il est possible de suggérer au visiteur, mais aussi des successions qu'il pourrait être amené à effectuer, il nous faut lui conseiller un parcours dont on puisse retirer une expérience satisfaisante. L'objectif de notre système de recommandation est de construire une visite culturelle conforme aux attentes du touriste d'une part et de la structurer de sorte à se conformer aux principes de la règle de l'apogée-fin. Pour ce faire, nous présentons au préalable la formulation mathématique des différents critères que nous souhaitons prendre en compte pour la création d'un parcours. Nous proposons par la suite un modèle MILP pour la résolution de la problématique utilisant tout ou partie de ces critères.

8.4.1 Critères d'optimisation

Le système de recommandation que nous proposons sélectionne les activités à inclure dans un parcours en articulant sa décision autour de trois métriques distinctes et indépendantes. Chacune de celles-ci juge de la qualité du parcours en reflétant un point de vue différent.

En premier lieu est considérée l'attrait global du parcours en tant qu'un ensemble d'activités déconnectées les unes des autres. La deuxième métrique évalue quant à elle la cohérence des successions d'activités et permet d'intégrer une logique temporelle entre les différents événements, sans qu'ils ne soient examinés individuellement. Enfin, la dernière est une interprétation de l'effet apogée-fin qui favorise les activités durant ces deux périodes décisives.

Afin d'exprimer ces trois différents critères dans un contexte d'optimisation, nous devons utiliser plusieurs variables de décision, propres à chaque PE-PVRP_c, qui sont présentées dans la table 8.1.

La plupart de ces dernières sont binaires et traduisent une caractéristique du parcours telle que la présence ou non d'une activité ou d'une séquence dans la visite recommandée ou bien permettent d'identifier les phases de la règle d'apogée-fin. En outre, afin de pouvoir utiliser cette heuristique psychologique, il nous faut pouvoir attribuer à chaque période une valeur, qui est donnée par l'équation (8.7) et qui correspond à la somme des utilités des activités recommandées au sein de celle-ci.

$$\text{perval}_p = \sum_{a \in p} r_a \times u(a) \quad (8.7)$$

Variables binaires associées aux activités		
$r_a = \begin{cases} 1 & \text{si l'activité } a \text{ est recommandée} \\ 0 & \text{sinon} \end{cases}$		$\forall a \in A$
Variables binaires associées aux séquences d'activités		
$\text{first}_a = \begin{cases} 1 & \text{si le parcours commence par l'activité } a \\ 0 & \text{sinon} \end{cases}$		$\forall a \in \mathcal{D}_s$
$\text{last}_a = \begin{cases} 1 & \text{si le parcours se termine par l'activité } a \\ 0 & \text{sinon} \end{cases}$		$\forall a \in \mathcal{D}_e$
$\text{link}_{a,b} = \begin{cases} 1 & \text{si la séquence } (a, b) \text{ est recommandée} \\ 0 & \text{sinon} \end{cases}$		$\forall a \in \mathcal{D}_r$
Variables binaires associées aux périodes		
$\text{peak}_p = \begin{cases} 1 & \text{si } p \text{ est la période retenue comme étant l'apogée} \\ 0 & \text{sinon} \end{cases}$		$\forall p \in P$
$\text{end}_p = \begin{cases} 1 & \text{si } p \text{ est la dernière période} \\ & \text{contenant une activité recommandée} \\ 0 & \text{sinon} \end{cases}$		
$\text{minor}_p = \begin{cases} 1 & \text{si } p \text{ n'est ni l'apogée ni la fin} \\ & \text{mais contient une activité recommandée} \\ 0 & \text{sinon} \end{cases}$		
Variables continues associées aux périodes		
perval_p	valeur associée à la période p indiquant son importance au sein du parcours proposé	$\forall p \in P$

TABLE 8.1 – Principales variables de décision associées au PE-PVRP_c

Nous faisons le choix d'une formulation simple, qui n'intègre pas les séquences d'activités. Leur inclusion, bien que permettant une meilleure évaluation de l'attrait des périodes, soulève des problèmes quant à la considération des séquences d'activités à cheval sur deux périodes. Les intégrer mènerait à aussi considérer les séquences de périodes, rendant les modèles plus complexes et gourmands en temps et en mémoire.

Chacune des fonctions exprimant un critère différent, on note les ensembles suivants :

$$\begin{aligned}
 X_a &= \{r_a | a \in A\} \\
 X_l &= \{\text{first}_a | a \in \mathcal{D}_s\} \\
 &\quad \cup \{\text{last}_a | a \in \mathcal{D}_e\} \\
 &\quad \cup \{\text{link}_{a,b} | (a, b) \in \mathcal{D}_r\} \\
 X_p &= \{\text{peak}_p, \text{end}_p, \text{minor}_p, \text{perval}_p | p \in P\}
 \end{aligned}$$

Tout d'abord, la fonction *ActivitiesUtility* (8.8a) exprime la métrique concernant l'attrait global des activités prises en compte individuellement. La fonction *LinksRelevance* (8.8b) fait de même pour les séquences d'activités et inclut le début et la fin du parcours. Enfin, la fonction *PeakEndRule* (8.8c) est notre proposition pour traduire l'effet de l'apogée-fin.

$$\text{ActivitiesUtility}(X_a) = \sum_{a \in A} r_a \times u(a) \quad (8.8a)$$

$$\begin{aligned}
 \text{LinksRelevance}(X_l) &= \sum_{a \in \mathcal{D}_s} \text{first}_a \times s(a) \\
 &\quad + \sum_{a \in \mathcal{D}_e} \text{last}_a \times e(a) \\
 &\quad + \sum_{(a,b) \in \mathcal{D}_r} \text{link}_{a,b} \times r(a, b)
 \end{aligned} \quad (8.8b)$$

$$\begin{aligned}
 \text{PeakEndRule}(X_p) &= \sum_{p \in P} \left(\lambda_4 \times \text{peak}_p + (1 - \lambda_4) \times \text{end}_p \right) \times \text{perval}_p \\
 &\quad - \max \left(0, \frac{\sum_{p \in P} \text{perval}_p \times \text{minor}_p}{\sum_{p \in P} \text{minor}_p} \right)
 \end{aligned} \quad (8.8c)$$

Les deux premières expressions manifestent une stratégie d'accumulation globale, l'ajout d'un élément ayant une évaluation positive (que ce soit une activité ou la séquence qui en découle) entraîne inéluctablement une meilleure évaluation du parcours résultant.

Dans la littérature, la loi de l'apogée-fin est souvent représentée par la moyenne des valeurs associées à ces deux moments (COJUHARENCO et al. 2008). Cette formulation permet d'évaluer un parcours en suivant l'effet tel que décrit par KAHNEMAN (2000). Mais, compte tenu que nous sommes ici dans un contexte de construction itérative d'un parcours en suivant plusieurs autres métriques qui suivent une logique d'accumulation, il nous

faut compenser celle-ci.

Notre choix adopté est de distinguer les périodes qui nous intéressent, appelées périodes majeures, que sont l'apogée et la fin des autres périodes dites mineures et de représenter la loi d'apogée-fin comme la différence entre l'intensité moyenne durant les périodes majeures et celle des autres périodes comprenant une activité recommandée.

Par ailleurs, étant donné l'ensemble d'arrivée de la fonction u et la formulation de $perval_p$, il est possible que ces variables prennent une valeur négative et donc que la moyenne des périodes mineures soit elle aussi négative. Or, étant donné que nous choisissons de considérer la différence entre les deux types de périodes, afin de ne pas inciter le système à sélectionner des activités estimées non adéquates au visiteur (et donc avec une valeur négative), ce qui résulterait en une augmentation du score total, nous minorons la moyenne des périodes mineures par 0. Cependant, la moyenne des périodes majeures n'est pas minorée, ce qui permet d'évaluer les situations qui seraient défavorables à l'utilisateur et ainsi les limiter. Cette limite inférieure sur les périodes mineures ne devrait pas biaiser le système, car les études de KEMP et al. (2008) et NAWIJN, MITAS et al. (2013) confirment que les vacances sont des expériences agréables comprenant peu de moments perçus négativement.

Ainsi, le choix d'inclure ou non une activité au parcours n'est altéré que lorsqu'elle se déroule durant une période mineure. D'une part, les deux premières métriques favorisent toute insertion jugée individuellement bénéfique, tandis que la dernière pénalise grandement les activités en dehors des périodes majeures. Il en résulte un équilibre qui est influencé par le paramètre λ propre à chaque profil touristique.

8.4.2 Modélisation par recherche de chemin

Nous proposons ici un modèle qui assimile le problème à une recherche d'un chemin optimal dans un graphe dont la solution recherchée dépend des différents points de vue développés ci-dessus.

Objectif

Telle que formulée précédemment, la fonction *PeakEndRule* n'est pas linéaire et ne peut donc pas être utilisée en l'état dans un modèle MILP. En vue de son emploi, nous ajoutons aux variables détaillées en table 8.1 les variables artificielles de la table 8.2, dont la plupart seront développées plus loin avec les contraintes du modèle. En notant X l'ensemble de toutes les variables de décision, nous reformulons la dernière fonction par :

$$PeakEndRule(X) = \lambda_4 \times peakval + (1 - \lambda_4) \times endval - peravg$$

Où $peakval$ représente la valeur de l'apogée, $endval$ celle de la fin et $peravg$ la moyenne bornée des périodes mineures.

Variables binaires associées aux périodes	
$active_p = \begin{cases} 1 & \text{si } p \text{ contient une activité recommandée} \\ 0 & \text{sinon} \end{cases}$	$\forall p \in P$
$major_p = \begin{cases} 1 & \text{si } p \text{ est la période d'apogée ou de fin} \\ 0 & \text{sinon} \end{cases}$	
$start_p = \begin{cases} 1 & \text{si } p \text{ est la première période} \\ & \text{contenant une activité recommandée} \\ 0 & \text{sinon} \end{cases}$	
Variables continues utilisées pour la fonction objectif	
$peakval$	valeur de la plus intense période
$endval$	valeur de la dernière période
$peravg$	valeur de la moyenne des périodes mineures bornée inférieurement par 0
Variables artificielles pour exprimer $peravg$	
z_{avg}	
$z_{l,p}$	$\forall p \in P$
$z_{r,p}$	$\forall p \in P$

TABLE 8.2 – Variables de décision utilisées pour rendre compte de l'effet apogée-fin

Chaque fonction décrivant un point de vue différent sur le parcours, le PE-PVRP_c donne lieu à un problème multicritère dont les objectifs sont :

$$\begin{aligned} & \max_X ActivitiesUtility(X) \\ & \max_X LinksRelevance(X) \\ & \max_X PeakEndRule(X) \end{aligned}$$

L'ouvrage de EHRGOTT (2005) souligne que les problèmes multiobjectifs ont des solutions formant un front de Pareto qu'ici nous pouvons départager en utilisant les coefficients

de λ . L'auteur détaille plusieurs méthodes pour résoudre ce type de problèmes, dont notamment les approches scalaires qui reviennent à les transformer en problèmes n'ayant qu'une seule fonction objectif. Cette opération est réalisée par une moyenne pondérée, l'évaluation de la distance par rapport à un but ou en convertissant des objectifs du problème en contraintes. La programmation par but n'est ici par envisageable à cause de la complexité à estimer un but pour chaque fonction. L'agrégation par pondération serait ici inefficace car il y a un conflit entre les métriques : améliorer la solution par rapport à *ActivitiesUtility* ou *LinksRelevance* peut la dégrader du point de vue de *PeakEndRule* (si une activité est proposée en dehors des périodes d'apogée et de fin), ce qui traduit un problème non convexe qui requiert de trouver un compromis entre les différents objectifs.

Par conséquent, nous avons décidé d'utiliser la fonction sur l'effet de l'apogée-fin en tant qu'objectif de notre problème et de traduire les deux autres métriques par des contraintes forçant le système à inclure des activités pendant les périodes considérées comme mineures. Nous exprimons ce rapport de force de manière proportionnelle au paramètre λ , mais il nous faut cependant remarquer que bien que les différentes fonctions du PE-PVRP_c ont le même ensemble d'arrivée $[-1, +1]$, les fonctions objectifs manipulent des éléments à des échelles différentes et ont par conséquent des ordres de grandeurs difficilement comparables. En effet, les valeurs de *ActivitiesUtility* et *LinksRelevance* sont proportionnelles au nombre d'activités recommandées dans le parcours, tandis que celles de *PeakEndRule* varient en fonction des activités recommandées lors des deux périodes d'apogée et de fin au regard des activités recommandées lors des autres périodes. Nous notons pour la suite $\tilde{\lambda}$ les poids modifiés qui reflètent ces différences d'échelles.

Finalement, le PE-PVRP_c est donc reformulé en un problème dont la fonction objectif (8.9) est la métrique sur l'effet de l'apogée-fin.

$$\max_X PeakEndRule(X) \quad (8.9)$$

Cette fonction objectif est ensuite assujettie par les contraintes (8.10) traduisant les deux autres métriques. Celles-ci sont pondérées par l'importance qui leur est accordée par le profil touristique tout en tenant compte de la différence d'ordre de grandeur des fonctions objectifs du problème.

$$\tilde{\lambda}_1 \times ActivitiesUtility(X) \geq \tilde{\lambda}_3 \times PeakEndRule(X) \quad (8.10a)$$

$$\tilde{\lambda}_2 \times LinksRelevance(X) \geq \tilde{\lambda}_3 \times PeakEndRule(X) \quad (8.10b)$$

Contraintes

Les contraintes restantes sont réparties en trois catégories : (i) les contraintes de flot assurant la construction d'un parcours réalisable du début à la fin ; (ii) les contraintes sur les groupes d'activités permettant de refléter les exigences spécifiques au profil du visiteur et (iii) les contraintes sur la règle de l'apogée-fin qui rendent possible son utilisation dans la fonction objectif.

Contraintes sur le flot

Le problème étant ici assimilé à la recherche d'un chemin dans un graphe, ces premières contraintes garantissent que le parcours proposé n'entraîne pas l'impossibilité d'assister à toutes les activités recommandées, en tenant compte de leurs horaires, ainsi que des temps de trajet et d'attente de l'utilisateur.

Dans ce but, pour chaque activité $a \in A$, les ensembles d'activités recommandables avant et après sont respectivement donnés par les fonctions δ^- et δ^+ , définies par :

$$\begin{aligned} \delta^- : A &\rightarrow \mathcal{P}(A) \\ a &\mapsto \{b \in A \mid (b, a) \in \mathcal{D}_r\} \\ \delta^+ : A &\rightarrow \mathcal{P}(A) \\ a &\mapsto \{b \in A \mid (a, b) \in \mathcal{D}_r\} \end{aligned}$$

Étant donné que nous avons démontré que le graphe représentant l'espace de recherche est acyclique, la continuité du parcours peut être assurée par l'ensemble de contraintes suivant :

$$\sum_{a \in \mathcal{D}_s} \text{first}_a = 1 \quad (8.11a)$$

$$\sum_{a \in \mathcal{D}_e} \text{last}_a = 1 \quad (8.11b)$$

$$\text{first}_a + \sum_{b \in \delta^-(a)} l_{b,a} = r_a \quad \forall a \in \mathcal{D}_s \quad (8.11c)$$

$$\sum_{b \in \delta^-(a)} l_{b,a} = r_a \quad \forall a \in A \setminus \mathcal{D}_s \quad (8.11d)$$

$$\text{last}_a + \sum_{b \in \delta^+(a)} l_{a,b} = r_a \quad \forall a \in \mathcal{D}_e \quad (8.11e)$$

$$\sum_{b \in \delta^+(a)} l_{a,b} = r_a \quad \forall a \in A \setminus \mathcal{D}_e \quad (8.11f)$$

Les contraintes (8.11a) et (8.11b) assurent respectivement que le parcours débute et se termine par une activité. Les équations (8.11c) et (8.11d) forcent respectivement les activités pouvant et ne pouvant pas débiter un parcours à n'être recommandées que lorsqu'elles sont précédées par une autre activité ou que l'activité est la première du parcours. Il en est de même pour les équations (8.11e) et (8.11f) qui concernent les activités étant suivies par d'autres ou terminant le parcours.

Contraintes sur les groupes

$$\left(\sum_{a \in A_c} w_c(a) \times r_a \right) \mathcal{R}_c t_c(A_c) \quad \forall (A_c, w_c, \mathcal{R}_c, t_c) \in C \quad (8.12)$$

Ces contraintes traduisent les besoins spécifiques du visiteur ainsi que des règles fixées par un organisateur. La sous-section 7.3.3 donne quelques exemples et notamment le patron utilisé pour ne pas recommander plusieurs représentations d'un même spectacle.

Contraintes sur la règle de l'apogée-fin

La fonction objectif que nous avons présentée plus tôt requiert plusieurs constructions intermédiaires afin de correctement formuler la loi de l'apogée-fin. Dans un premier temps, il nous faut exprimer la valeur attribuée aux périodes ainsi que de les discriminer

selon qu'elles sont majeures, mineures ou non impliquées dans le parcours touristique. Par la suite, nous instaurons des règles essentielles quant à la sélection des différentes périodes clés. Cela conduit aux contraintes suivantes :

$$\text{perval}_p = \sum_{a \in p} r_a \times u(a) \quad \forall p \in P \quad (8.13a)$$

$$\text{active}_p \leq \sum_{a \in p} r_a \quad \forall p \in P \quad (8.13b)$$

$$\text{active}_p \geq r_a \quad \forall p \in P, \forall a \in p \quad (8.13c)$$

$$\text{major}_p \leq \text{peak}_p + \text{end}_p \quad (8.13d)$$

$$\text{major}_p \geq \text{peak}_p \quad (8.13e)$$

$$\text{major}_p \geq \text{end}_p \quad (8.13f)$$

$$\text{minor}_p = \text{active}_p - \text{major}_p \quad (8.13g)$$

$$\text{start}_p = \sum_{a \in p} \text{first}_a \quad \forall p \in P \quad (8.13h)$$

$$\text{end}_p = \sum_{a \in p} \text{last}_a \quad \forall p \in P \quad (8.13i)$$

$$\sum_{p \in P} \text{peak}_p = 1 \quad (8.13j)$$

$$\text{peak}_p \leq \text{active}_p \quad \forall p \in P \quad (8.13k)$$

$$\text{peak}_p \leq 1 - \text{end}_p + \text{start}_p \quad \forall p \in P \quad (8.13l)$$

L'équation (8.13a) traduit notre choix précisé auparavant, d'exprimer la valeur émotionnelle associée à chaque période comme la somme des attraits des activités recommandées lors de cette dernière.

Les contraintes (8.13b) et (8.13c) déterminent les périodes actives, c'est à dire comportant au moins une activité recommandée. Une période active est considérée comme majeure si elle correspond à l'apogée ou à la fin, ce qui est formulé par (8.13d), (8.13e) et (8.13f), ou mineure si ce n'est pas le cas (8.13g).

Les contraintes (8.13h) et (8.13i) permettent d'identifier la première et la dernière période, à partir de la première et dernière activité recommandée. De même, (8.13j) assure qu'il existe exactement une apogée et (8.13k) que cette dernière comporte au moins une activité recommandée. Enfin, afin d'accentuer la règle de l'apogée-fin, (8.13l) impose que les périodes d'apogée et de fin sélectionnées soient différentes si le séjour proposé s'étale sur plusieurs périodes.

À partir de ces éléments, nous pouvons exprimer $peakval$, $endval$ et $peravg$ qui sont utilisés pour optimiser le critère lié à la loi de l'apogée-fin. Pour ce faire, il est nécessaire de linéariser la formulation de $peravg$, qui est une moyenne bornée des valeurs des périodes mineures. Dans un premier temps, nous utilisons les variables z_{avg} , $z_{l,p}$ et $z_{r,p}$, qui permettent la décomposition suivante :

$$\begin{aligned}
 peravg &= \max(0, z_{avg}) \\
 z_{avg} &= \frac{\sum_{p \in P} perval_p \times minor_p}{\sum_{p \in P} minor_p} \\
 z_{avg} \times \sum_{p \in P} minor_p &= \sum_{p \in P} perval_p \times minor_p \\
 \sum_{p \in P} \underbrace{z_{avg} \times minor_p}_{z_{l,p}} &= \sum_{p \in P} \underbrace{perval_p \times minor_p}_{z_{r,p}}
 \end{aligned}$$

Les expressions $z_{l,p}$ et $z_{r,p}$ sont donc formulées comme un produit entre une variable binaire et une variable continue. Ceux-ci peuvent être exprimés au moyen de diverses contraintes, qui requièrent de déterminer les bornes supérieure \bar{P} et inférieure \underline{P} des valeurs attribuées aux périodes. Ces dernières sont données par :

$$\begin{aligned}
 \bar{P} &= \max_{p \in P} \left\{ \sum_{a \in p} \max(0, u(a)) \right\} \\
 \underline{P} &= \min_{p \in P} \left\{ \sum_{a \in p} \min(0, u(a)) \right\}
 \end{aligned}$$

Nous pouvons donc, grâce à ces variables, formuler l'effet apogée-fin qui est modélisé par l'ensemble de contrainte suivant :

$$\text{peakval} \geq \text{perval}_p - (1 - \text{peak}_p) \times (\bar{P} - \underline{P}) \quad \forall p \in P \quad (8.13m)$$

$$\text{peakval} \leq \text{perval}_p + (1 - \text{peak}_p) \times (\bar{P} - \underline{P}) \quad \forall p \in P \quad (8.13n)$$

$$\text{endval} \geq \text{perval}_p - (1 - \text{end}_p) \times (\bar{P} - \underline{P}) \quad \forall p \in P \quad (8.13o)$$

$$\text{endval} \leq \text{perval}_p + (1 - \text{end}_p) \times (\bar{P} - \underline{P}) \quad \forall p \in P \quad (8.13p)$$

$$z_{l,p} \geq \text{minor}_p \times \underline{P} \quad \forall p \in P \quad (8.13q)$$

$$z_{l,p} \leq \text{minor}_p \times \bar{P} \quad \forall p \in P \quad (8.13r)$$

$$z_{l,p} \geq z_{\text{avg}} - (1 - \text{minor}_p) \times \bar{P} \quad \forall p \in P \quad (8.13s)$$

$$z_{l,p} \leq z_{\text{avg}} + (1 - \text{minor}_p) \times \bar{P} \quad \forall p \in P \quad (8.13t)$$

$$z_{r,p} \geq \text{minor}_p \times \underline{P} \quad \forall p \in P \quad (8.13u)$$

$$z_{r,p} \leq \text{minor}_p \times \bar{P} \quad \forall p \in P \quad (8.13v)$$

$$z_{r,p} \geq \text{perval}_p - (1 - \text{minor}_p) \times \bar{P} \quad \forall p \in P \quad (8.13w)$$

$$z_{r,p} \leq \text{perval}_p + (1 - \text{minor}_p) \times \bar{P} \quad \forall p \in P \quad (8.13x)$$

$$\sum_{p \in P} z_{l,p} = \sum_{p \in P} z_{r,p} \quad (8.13y)$$

$$\text{peravg} \geq z_{\text{avg}} \quad (8.13z)$$

$$\text{peravg} \geq 0 \quad (8.13aa)$$

La paire de contraintes (8.13m) et (8.13n) attribut à peakval la valeur de la période sélectionnée comme étant l'apogée. Ces contraintes seules ne garantissent pas que la valeur retenue soit celle de l'apogée mais, étant dans un contexte de maximisation et (8.13j) obligeant à retenir une unique période comme étant l'apogée, peakval prendra la valeur voulue. Il en est de même avec (8.13o) et (8.13p) qui fixent la valeur de endval en fonction de end_p .

Les contraintes (8.13q) à (8.13t) et (8.13u) à (8.13x) transposent respectivement les égalités $z_{l,p} = \text{peravg} \times \text{minor}_p$ et $z_{r,p} = \text{perval}_p \times \text{minor}_p$ permettant de formuler la loi de l'apogée-fin par la décomposition exposée plus tôt. Elles sont complétées par l'égalité (8.13y), qui lie toutes les variables $z_{l,p}$ et $z_{r,p}$, forçant ainsi z_{avg} à prendre la valeur souhaitée. Pour finir, la valeur de peravg est donnée par les contraintes (8.13z) et (8.13aa).

Problème d'optimisation linéaire

À partir des variables de décision, de la fonction objectif (8.9) et des contraintes, on obtient le modèle suivant :

$$\begin{aligned}
 X^* = \arg \max_X & \quad \text{PeakEndRule}(X) \\
 \text{s.t.} & \quad (8.10) \quad \text{Contraintes sur les objectifs,} \\
 & \quad (8.11) \quad \text{Contraintes sur le flot,} \\
 & \quad (8.12) \quad \text{Contraintes sur les groupes,} \\
 & \quad (8.13) \quad \text{Contraintes sur la règle de l'apogée-fin,} \\
 r_a \in \mathbb{B} & \quad \forall a \in A, \\
 \text{link}_{a,b} \in \mathbb{B} & \quad \forall (a, b) \in \mathcal{D}_r, \\
 \text{first}_a \in \mathbb{B} & \quad \forall a \in \mathcal{D}_s, \\
 \text{last}_a \in \mathbb{B} & \quad \forall a \in \mathcal{D}_e, \\
 \text{active}_p \in \mathbb{B} & \quad \forall p \in P, \\
 \text{start}_p \in \mathbb{B} & \quad \forall p \in P, \\
 \text{peak}_p \in \mathbb{B} & \quad \forall p \in P, \\
 \text{end}_p \in \mathbb{B} & \quad \forall p \in P, \\
 \text{major}_p \in \mathbb{B} & \quad \forall p \in P, \\
 \text{minor}_p \in \mathbb{B} & \quad \forall p \in P, \\
 \text{perval}_p \in \mathbb{R} & \quad \forall p \in P, \\
 \text{peakval} \in \mathbb{R}, \\
 \text{endval} \in \mathbb{R}, \\
 \text{peravg} \in \mathbb{R}, \\
 z_{\text{avg}} \in \mathbb{R}, \\
 z_{l,p} \in \mathbb{R} & \quad \forall p \in P, \\
 z_{r,p} \in \mathbb{R} & \quad \forall p \in P
 \end{aligned} \tag{8.14}$$

Pour chaque profil touristique, un problème PE-PVRP_c différent est obtenu, qui lorsqu'il est résolu, correspond à un parcours touristique. Ce dernier est obtenu en retenant les activités associées aux variables de décision booléennes r_a , dont la valeur est 1 et en les ordonnant par date et heure de début.

8.5 Expériences et résultats

Nous souhaitons étudier l'impact de l'utilisation de la cadence de visite et l'effet apogée-fin sur la construction de parcours culturels. Or, étant donné qu'il n'existe pas de consensus fort à propos de l'applicabilité de cette théorie dans le domaine du tourisme, une première étape avant une évaluation sur le terrain est d'effectuer des simulations, en variant les profils et paramètres puis, comparer les parcours obtenus.

Afin d'évaluer l'efficacité de notre système de recommandation, nous souhaitons donc vérifier que notre modèle permet de construire des parcours (i) respectant les préférences culturelles et d'organisation du touriste et (ii) accentuant l'effet décrit par la théorie de l'apogée-fin.

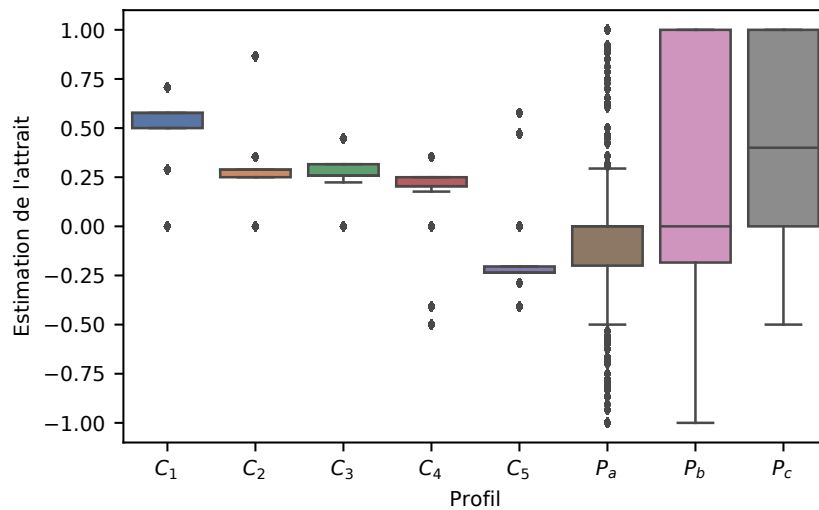
Pour rappel, tous nos travaux sur les visites en ville ont été axés autour de l'édition 2015 du « Festival OFF d'Avignon », un festival de théâtre annuel se déroulant au mois de juillet présenté dans l'annexe A. Cette dernière détaille aussi les 30 profils touristiques utilisés ainsi que les critères à considérer pour déterminer l'ensemble des successions d'activités qu'il est possible d'effectuer.

8.5.1 Mesures de l'attrait des activités

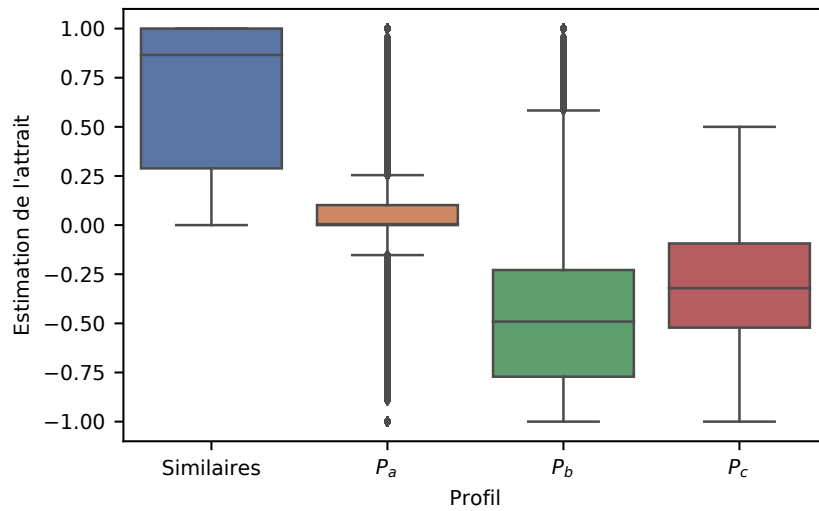
En prélude à l'étude des deux principaux changements proposés par notre système de recommandation, il est nécessaire d'observer la complémentarité des mesures de l'intérêt culturel et de la cadence de visite au travers de notre corpus. La figure 8.3 montre l'estimation de l'attrait des activités et des paires d'activités successives en fonction des différents profils en termes de préférences culturelles et d'organisation.

Concernant les activités considérées individuellement (figure 8.3a, reflétant la fonction u), on constate que les profils sur les préférences horaires (P_a à P_c) donnent des mesures évoluant dans une plage importante de valeurs. Cette mesure associée à ces profils offre donc un large choix de spectacles. A contrario, les profils culturels (C_1 à C_5) concentrent la plupart des mesures d'intérêt dans un espace de valeurs réduit ce qui indique que, pour les profils généralistes (dont la plage de valeurs est positive), il y aura de nombreux parcours équivalents possibles, tandis que les profils très spécifiques (C_5) n'entraîneront que très peu de diversités dans les choix secondaires.

Pour les paires d'activités (figure 8.3b, reflétant la fonction r), les mêmes profils sont utilisés pour la cadence de visite, car les actogrammes permettent d'évaluer l'utilité de ne rien faire entre deux spectacles, tandis que la mesure de l'intérêt est ici une mesure de la cohérence du parcours, favorisant les spectacles similaires.



(a) Attrait des activités prises individuellement



(b) Attrait des paires d'activités successives

FIGURE 8.3 – Attrait des activités en fonction du profil touristique

Avec cette métrique, la cadence de visite pénalise la grande majorité des successions possibles et ce pour tous les profils ce qui s'explique par la densité des spectacles, quelle que soit l'heure où une pièce finit, il y a toujours de très nombreuses autres qui commencent peu de temps après, alors que les actogrammes fournis dans PICOLOT (2016) sont construits sur une alternance entre les spectacles et les autres activités touristiques.

On observe l'effet inverse pour la mesure de similarité, celle-ci concentre les valeurs dans un intervalle réduit proche du maximum. Cette faible capacité à discriminer les paires d'activités est explicable par la fréquence des attributs associés aux pièces, une petite partie étant sur-représentée et le nombre peu élevé d'attributs différents par pièce rendant la mesure par similarité très peu efficace.

Ces résultats sur notre jeu de données confirment ceux du chapitre 6, l'intérêt culturel et la cadence de visite sont deux mesures complémentaires. Les nombreuses représentations sont bien discriminées en fonction des préférences sur l'organisation, mais la mauvaise distribution des genres théâtraux dans le corpus met en difficulté la mesure de l'intérêt. On voit ici la nécessité de bien choisir le paramètre α des fonctions u et r afin d'accorder une importance égale aux deux mesures.

8.5.2 Scénarios et paramètres

À partir des 30 profils touristiques, composés de goûts culturels et de préférences d'organisation, nous souhaitons mettre en évidence les effets induits par l'intégration de la cadence de visite et de l'effet de l'apogée-fin dans notre stratégie de recommandation.

Simulations et variations des paramètres

Le premier point à évaluer concerne l'utilisation conjointe de l'intérêt culturel et de la cadence de visite, grâce à l'utilisation des actogrammes. Ils doivent avoir une influence visible à l'échelle de leur définition, c'est-à-dire une variation des recommandations tout au long de la journée dans notre cas. Ainsi, nous effectuons une simulation par profil et pour chacune des 23 journées du festival, en faisant varier le paramètre α , qui contrôle l'importance des deux métriques, sur 5 degrés différents.

Le second point vise à étudier l'intégration de la loi de l'apogée-fin comme un critère déterminant pour la qualité du séjour. Comme précisé plus tôt, nous avons choisi de considérer des périodes d'une journée pour rendre compte de son effet (conformément aux travaux de KEMP et al. (2008)). Il est donc nécessaire de simuler des séjours de plusieurs jours de d'analyser la variation du nombre de spectacles proposés quotidiennement. Dans ce but, pour chacune des 3 semaines du festival, nous générons un parcours par profil en fonction de 4 différentes valeurs pour le paramètre λ , qui reflète l'importance de la loi de l'apogée-fin par rapport aux autres critères d'optimisation.

Ainsi, un total de 3 450 parcours journaliers et 2 760 parcours hebdomadaires ont été générés. Ils permettent d'observer les effets des deux changements dans la stratégie de recommandation tout en s'assurant que ceux-ci sont constants quelque soit le profil touristique ou le moment du séjour.

Systèmes et configurations

L'évaluation de notre système de recommandation se fait donc par l'analyse de parcours générés pour des profils touristiques issus d'enquêtes (présentées au chapitre 6). Nous souhaitons aussi démontrer l'efficacité des changements que nous proposons vis à vis d'autres systèmes de la littérature.

Aussi, afin de comparer notre modèle, nous réalisons les mêmes 6 210 simulations avec le système de recommandation présenté par WÖRNDL et al. (2017). Ce dernier a fait l'objet d'une étude sur le terrain montrant qu'il procure un bénéfice en termes d'expérience utilisateur lors de la proposition de visites culturelles sur une journée. WÖRNDL et al. (2017) assimilent aussi le problème à la recherche de chemin dans un graphe orienté, dont la construction est similaire à celle décrite dans la section 7.4. Une pondération est associée à chaque sommet et arc donnant respectivement l'utilité des activités et le coût du déplacement nécessaire entre deux d'entre elles. La génération du parcours quant à elle, se base sur l'algorithme de DIJKSTRA (1959) qui est modifié pour maximiser le ratio entre l'utilité d'ajouter une activité et la distance qu'il faut parcourir pour y accéder.

Le point d'utiliser cet autre système est de valider que l'utilisation des actogrammes en plus de l'intérêt culturel est indépendante d'autres considérations comme la modélisation du problème et la méthode de résolution.

La configuration de chaque système est détaillée ci-dessous :

Paramètres du PE-PVRP_c

Pour chacun des scénarios, l'instance du PE-PVRP_c associé est constituée avec les paramètres suivants :

- L'ensemble des activités A contient toutes les représentations contenues entre le début et la fin du séjour du profil.
- L'ensemble des périodes P est un découpage en journées, afin que toutes les activités d'une période donnée commencent le même jour.
- L'ensemble des contraintes C inclut deux types de contraintes :
 - On ne peut pas recommander deux représentations de la même pièce
 - Pour les profils ayant des horaires de début et fin quotidiens, les activités ne les respectant pas (en incluant les temps de trajet) sont exclues

- Le paramètre λ permettant d’influer sur l’importance de chaque composante de la fonction objectif varie afin de pouvoir étudier le comportement du système de recommandation.
- L’ensemble des paires d’activités \mathcal{D}_r est calculé en déterminant les couples d’activités pour lesquelles on peut aller de la première à la seconde (en considérant le temps de trajet et d’attente).

Fonctionnement du système de WÖRNDL et al. (2017)

Pour chaque instance du PE-PVRP_c, un problème similaire est résolu par le système de WÖRNDL et al. (2017). Les mêmes ensembles d’activités (A) et de contraintes (C) sont utilisés pour générer le graphe. Néanmoins, la méthode employée pour noter les points d’intérêt n’est pas utilisable avec notre jeu de données. Elle repose sur la popularité, une mesure qu’il nous est impossible d’appliquer puisqu’il y a de nombreux changements d’une édition à l’autre du « Festival OFF d’Avignon ». Par conséquent, leur algorithme est appliqué en utilisant comme fonction d’utilité des arcs les fonctions d’estimation de l’attrait des activités et des séquences, respectivement u et r en les pondérant par les poids associés dans λ . De plus, comme fonction de coût, le temps de trajet entre deux théâtres est utilisé et non la distance afin de ne pas défavoriser les spectacles dans d’autres villes.

8.5.3 Complexité en temps et en espace

Tous les tests ont été réalisés sur une même machine composée de 16 processeurs « Intel Xeon E5640 » et 31,4 GiB de mémoire vive.

La table 8.3 montre le temps de calcul des différents systèmes et sans surprise, l’algorithme de WÖRNDL et al. (2017) a des performances bien meilleures que notre système, tant sur la consommation mémoire que le temps d’exécution. Le problème étant NP-difficile, notre modélisation MILP a une complexité bien plus importante.

8.5.4 L’influence des actogrammes sur les recommandations quotidiennes

Nous nous concentrons dans un premier temps sur l’impact de l’ajout d’une mesure de la cadence de visite dans l’évaluation des activités par le système.

Système de recommandation	Période	Temps de calcul (moyenne)	Mémoire utilisée (moyenne)
PE-PVRP _c	Journée	5 min 37 s	3,45 GiB
	Semaine	30 min 1 s	12,0 GiB
WÖRNDL et al.	Journée	20,2 s	253,3 MiB
	Semaine	1 min 24 s	612,2 MiB

TABLE 8.3 – Comparaison des temps de calcul et de la mémoire utilisée des systèmes utilisés

La table 8.4 résume l'évaluation pour chaque système des propositions de visite en fonction des préférences culturelles et d'organisation du profil associé.

Système de recommandation	Valeur de α	Taille du parcours	Intérêt culturel	Cadence de visite
PE-PVRP _c	0,0	4,2	0,0256	0,73
	0,5	6,9	0,0444	0,65
	1,0	9,8	0,0629	0,57
WÖRNDL et al.	0,0	3,9	0,0256	0,62
	0,5	6,4	0,0439	0,51
	1,0	8,6	0,0627	0,48

 TABLE 8.4 – Mesures moyennes de la taille et qualité des parcours générés en fonction du paramètre α

On remarque tout d'abord que les valeurs de la mesure de l'intérêt culturel sont bien inférieures à celles de la cadence de visite. Cette différence est due à la méthode employée pour évaluer l'intérêt culturel du parcours. Étant donné la quantité de spectacles accessibles au visiteur en une journée, un parcours ne peut intégrer qu'une infime partie de l'offre culturelle.

L'augmentation proportionnelle de la mesure de l'intérêt culturel et de la taille des parcours par rapport au paramètre α est ici une manifestation d'une spécificité de notre corpus. L'offre pléthorique durant le festival, combinée à la distribution des attributs parmi les pièces rend les profils culturels peu discriminants comme nous l'avons remarqué plus tôt. Dans ce contexte, les systèmes trouvent toujours des spectacles avec de bons scores d'intérêt, quelque soit le moment de la journée.

La mesure de la cadence de visite, quant à elle, varie davantage, en fonction de l'importance de qu'accorde le profil touristique à cet aspect du parcours. Afin d'observer ces

variations, la figure 8.4 montre la densité de probabilité d'effectuer une activité à chaque instant de la journée en fonction du paramètre α utilisé dans les fonctions de mesure de l'attrait des activités et des séquences (les paramètres u et r du PE-PVRP_c) pour le système de recommandation que nous proposons.

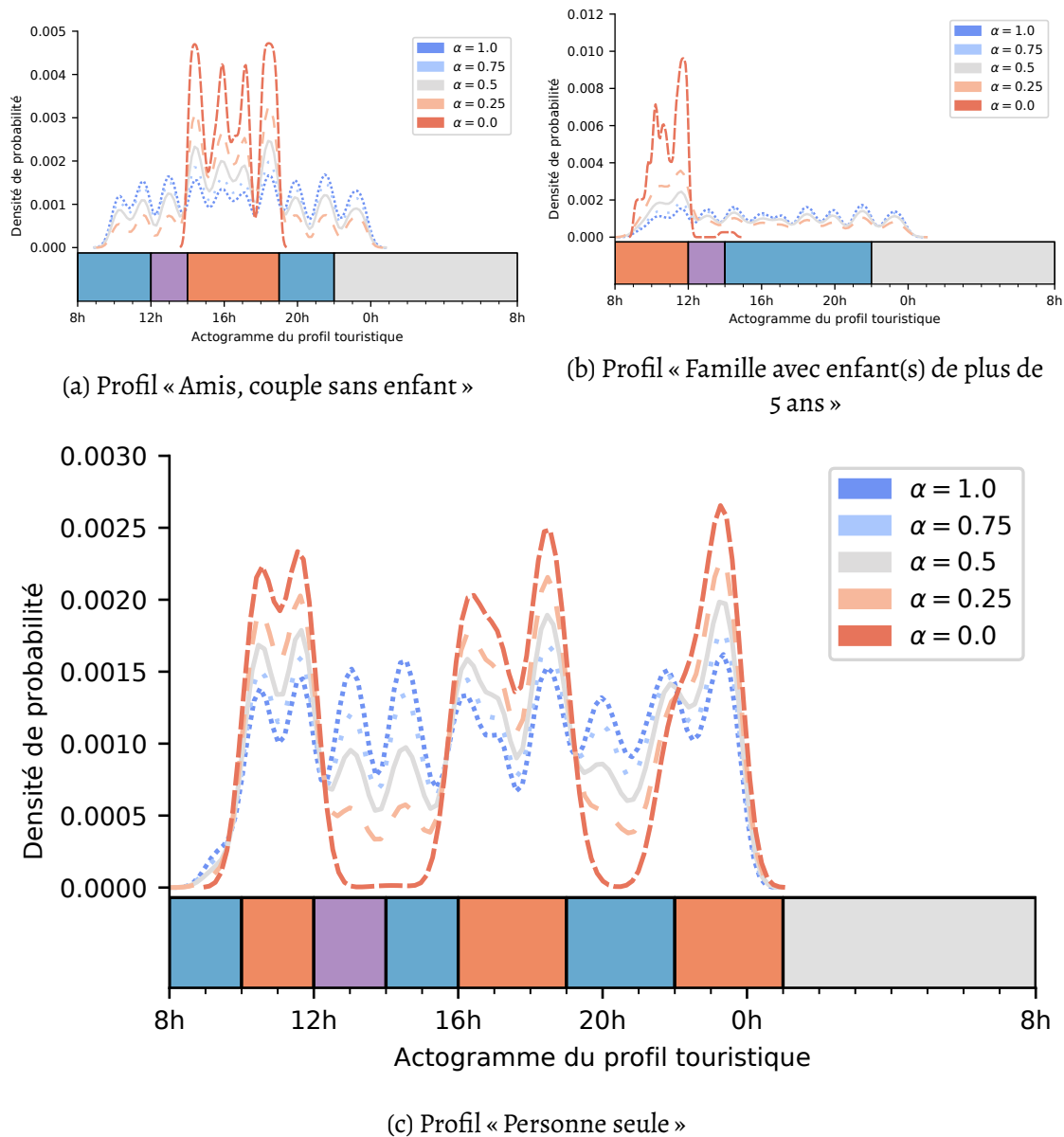


FIGURE 8.4 – Distribution des activités au long de la journée en fonction de l'importance relative de la cadence de visite (voir figure 6.3)

Nous constatons clairement un impact fort sur la composition du parcours, quel que soit le profil touristique. Lorsque l'accent est mis sur l'intérêt culturel (α proche de 1,0), la distribution des activités est équitable tout au long de la journée, quelle que soient les préférences d'organisation de l'utilisateur. Au contraire, lorsque seule la cadence de visite est considérée, le système sélectionne des représentations s'insérant au mieux dans l'emploi du temps du touriste.

8.5.5 L'effet apogée-fin sur des séjours d'une semaine

Le deuxième élément essentiel que nous voulons vérifier est la possibilité de créer un parcours tirant partie de l'effet apogée-fin, c'est-à-dire pour lequel le visiteur retirerait la meilleure expérience selon les principes de cette heuristique psychologique.

Nous présentons ici les résultats tirés de simulations mettant en scène un séjour d'une semaine pour chaque profil touristique. Le principal paramètre impliqué ici est le vecteur λ , donnant l'importance des métriques associées à l'attrait individuel, des paires et à la loi de l'apogée-fin. Nos tests montrent que déterminer ce paramètre est une tâche complexe. Nous nous contenterons ici d'exposer les trois configurations suivantes :

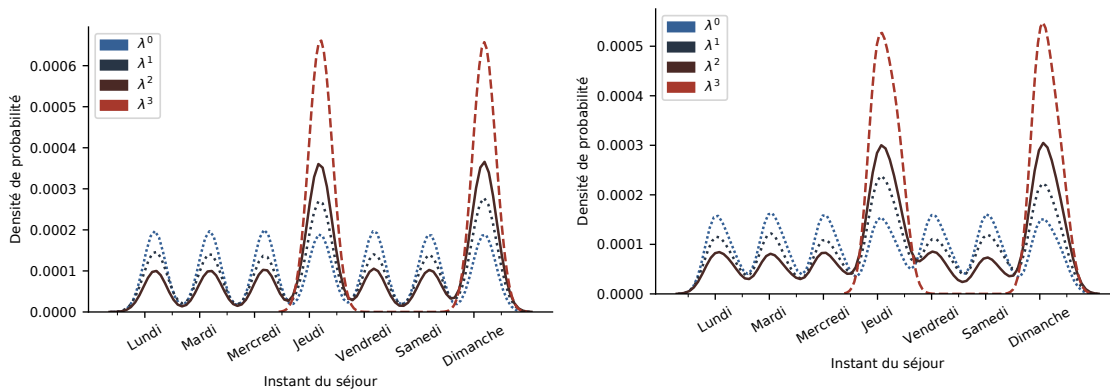
- λ^0 où l'effet apogée-fin est ignoré ;
- λ^1 où aucune métrique n'est mise en valeur par rapport à l'autre ;
- λ^2 favorisant l'effet apogée-fin par rapport aux deux autres ;
- λ^3 ne tenant compte que de la loi de l'apogée-fin.

La table 8.5 donne le nombre de spectacles par type de période (majeure si correspondant à l'apogée ou à la fin, mineure sinon) proposés par les deux systèmes. Avec l'algorithme de WÖRNDL et al. (2017), les activités du parcours sont équitablement réparties tout au long du séjour car il n'est pas prévu pour prendre en compte cette métrique. Pour notre système par contre, le paramètre λ montre qu'il est possible de faire varier les jours où les activités sont recommandées. Suivant l'importance donnée à l'effet apogée-fin, la différence entre les périodes majeures et mineures est plus ou moins marquée, ce qui est consistant.

Système de recommandation	Configuration	Nombre d'activités par périodes	
		Majeures	Mineures
PE-PVRP _c	λ^0	6,1	6,1
	λ^1	6,2	3,1
	λ^2	6,2	1,6
	λ^3		6,2
WÖRNDL et al.	–	6,0	

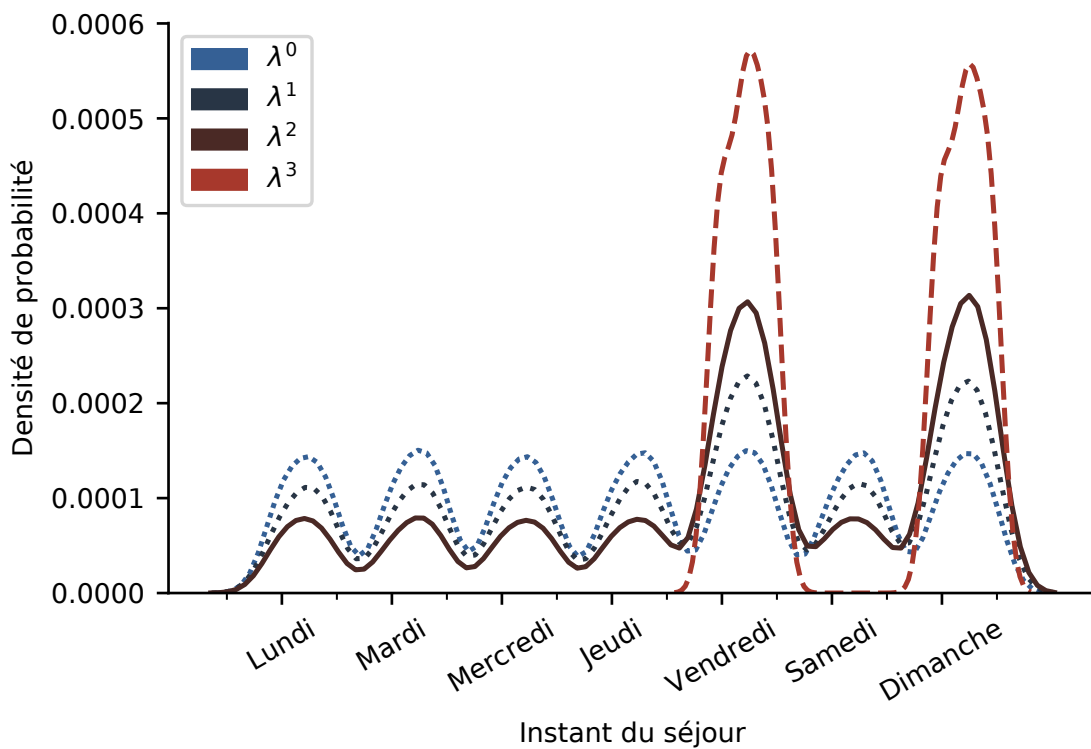
TABLE 8.5 – Nombre de recommandations par jour des systèmes évalués

La figure 8.5 expose la distribution des activités des parcours hebdomadaires générés par notre modèle en fonction des différentes configuration de λ . Nous mettons en évidence un impact non négligeable de ce dernier sur l'allure générale du parcours généré.



(a) Profil « Amis, couple sans enfant »

(b) Profil « Famille avec enfant(s) de plus de 5 ans »



(c) Profil « Personne seule »

FIGURE 8.5 – Distribution des activités tout au long du parcours en fonction du paramètre λ

Notre modèle permet de plus ou moins mettre l'accent sur deux jours durant la semaine. Dans sa dernière configuration où seul l'effet apogée-fin est considéré (λ^3), le système génère des parcours où toutes les activités se concentrent sur deux jours à cause de notre modélisation par différence de moyennes. Les configurations λ^1 et λ^2 quant à elles montrent qu'il est possible d'obtenir un compromis entre les différents critères, qui mène à la construction de parcours où chaque journée a son importance, tout en privilégiant deux pour marquer l'expérience du séjour. Déterminer la valeur optimale à utiliser

nécessite plus de tests, mais aussi une réflexion sur l'importance qu'il est souhaitable d'accorder à cet effet vis à vis de la globalité du parcours.

8.6 Conclusion

Les études sur l'organisation du séjour et la loi de l'apogée-fin dans le tourisme montrent qu'un changement de paradigme est nécessaire quant à la manière de penser les systèmes de recommandation de parcours touristiques. L'accumulation d'activités, sur le seul critère des goûts culturels du visiteur, n'est pas une stratégie qui convienne. Le PE-PVRP_c impose donc une nouvelle stratégie de recommandation conciliant plusieurs métriques, qui reflètent des effets sur des échelles différentes. Prendre en compte ces divers points de vue est nécessaire afin de construire des parcours riches, conformes aux souhaits du visiteur et dont il retirera une expérience plaisante.

L'utilisation des actogrammes pour la proposition de parcours est une méthode efficace pour pondérer les points d'intérêt et adapter le parcours aux préférences horaires du visiteur. Leur formulation au sein de la fonction d'utilité et non en tant que contrainte permet d'établir un compromis entre les différents types d'activités, tout en assurant des temps de repos au visiteur. De plus, leurs effets ne sont pas propre à notre modèle mais bien reproductibles avec d'autres algorithmes, ce qui montre les avantages et la flexibilité de ces outils.

Lorsqu'il s'agit d'un voyage de plusieurs jours, nous proposons de tenir compte de l'effet de la loi de l'apogée-fin, afin que le parcours soit plus marquant pour le visiteur. Pour y parvenir, c'est un équilibre subtil qu'il faut trouver afin de mettre en concurrence tous les critères de sélection des activités et proposer des parcours avec deux journées inoubliables, tout en ne délaissant pas le reste du séjour. Les simulations montrent que l'élaboration automatique de parcours reproduisant cet effet est possible, tout en contrôlant son importance sur la répartition des recommandations au long du parcours.

Le système de recommandation que nous proposons est une démonstration des atouts du PE-PVRP_c. Grâce à une évaluation de chaque activité, ainsi que des temps libres qui les séparent, nous pouvons proposer des séjours respectant les préférences culturelles et organisationnelles des touristes. Les actogrammes conduisent à des parcours comportant des moments intenses, entrecoupés de périodes calmes. Cette importance donnée aux temps libres, laisse au visiteur la possibilité de flâner et découvrir la ville, une caractéristique essentielle du tourisme urbain. L'intégration de la loi de l'apogée-fin en plus des autres critères, nécessite une nouvelle formulation mathématique mais permet de mettre en balance différents points de vue lors de l'élaboration du parcours. En proposant des

activités sur toute la durée du séjour, tout en mettant en avant deux périodes clés, les parcours proposés offrent une expérience satisfaisante au moment du voyage et mémorable par la suite.

Chapitre 9

Conclusion et perspectives

Sommaire

9.1	Conclusion	158
9.1.1	Visites au musée	158
9.1.2	Parcours en ville	159
9.2	Perspectives	161
9.2.1	Élargissement du cadre méthodologique	161
9.2.2	Visite cohérente au musée	162
9.2.3	Évaluation des critères en ville	162

9.1 Conclusion

L'élaboration d'un parcours culturel personnalisé est une tâche complexe pour laquelle l'assistance d'un agent expert peut être bénéfique. D'une part il permet de diminuer l'effort de collecte et d'analyse détaillée des informations mais surtout, en réduisant les possibilités à quelques propositions au visiteur, il lui permet de faire un choix plus facilement et duquel il en retirera une plus grande satisfaction.

Nos travaux reposent sur une démarche interdisciplinaire, en formalisant et intégrant des résultats et concepts issus de domaines scientifiques variés. Ainsi, les systèmes de recommandation proposés résultent de changements de paradigmes dans la manière de penser la recommandation culturelle, dans le bus d'améliorer l'expérience de visite.

Bien que des problématiques spécifiques soient soulevées dans chaque partie, nos travaux sur les parcours au musée et en ville se basent sur la MILP et une modélisation par recherche de chemin dans un graphe. Au musée, nous proposons d'optimiser le parcours en fonction du temps dans un graphe topologique, alors qu'en ville la solution inverse est adoptée, en optimisant un chemin suivant des déplacements au sein d'un graphe temporel.

9.1.1 Visites au musée

Dans une première partie, nous nous sommes intéressés à la visite au musée, afin de retarder la fatigue muséale. Bien que l'utilisation d'un système de recommandation soit une solution adéquate, ces outils souffrent particulièrement des problèmes de démarrage à froid et de sur-spécialisation dans ce contexte. Aussi, notre problématique était d'améliorer le profil utilisateur du visiteur, sans pour autant lui soumettre un questionnaire à compléter à son arrivée.

L'hypothèse en amont de nos travaux était que les textes officiels, issus de l'institution, permettent d'établir un point de vue du musée qui puisse être utilisé pour enrichir un parcours basé sur des préférences culturelles, ou proposer une visite composée des œuvres marquant l'identité de l'institution.

Le chapitre 4 montre qu'avec l'utilisation de méthodes de résumé automatique de texte, nous sommes capables d'attribuer un score à chaque expôt d'un musée, et ce de manière indépendante de la langue dans laquelle les notices sont rédigées. Le classement qui en découle plaçant en tête ceux considérés comme des chefs-d'œuvres par l'institution, cette nouvelle mesure du « prestige des œuvres » permet d'identifier un point de vue du musée sur sa propre collection. En regardant le musée comme un recueil de textes, il

apparaît que les cartels des expôts soient porteur d'une information allant au delà de la simple description de l'œuvre, le champs lexical et la sémantique utilisés pouvant indiquer les affinités du conservateur et des équipes de médiation.

Le système de recommandation exposé dans le chapitre 5 utilise ces résultats pour répondre au PVRP_m, et démontre qu'ils permettent de construire des parcours personnalisés. D'une part pour un visiteur désireux de découvrir le musée sans qu'il n'ait de préférence particulière, il est possible de lui proposer un parcours dans un temps limité. D'autre part, lorsque qu'utilisés conjointement avec des appétences culturelles, cette métrique indépendante permet de compléter un profil et ainsi inclure de la diversité dans le parcours en proposant des œuvres emblématiques. Enfin, une heuristique basique donnant des résultats similaires au modèle MILP, nous pouvons conjecturer que l'emploi des cartels décrivant les expôts améliore significativement les visites quelque soit l'algorithme de recommandation sous-jacent.

9.1.2 Parcours en ville

Dans la seconde partie de cette thèse, nous nous sommes tournés vers le tourisme en ville, dans le but de proposer des parcours où, en plus des goûts culturels, le rythme du séjour est mis au premier plan. En effet, les études menées en collaborations avec d'autres laboratoires mettent en évidence l'importance de l'organisation du séjour et la diversité des préférences d'emploi du temps des visiteurs. Ainsi, la problématique de nos travaux était d'étendre la stratégie de recommandation pour offrir une expérience de visite mêlant des temps forts et moments de repos durant le parcours.

L'hypothèse à l'origine de nos travaux repose sur l'élaboration d'un séjour à partir d'une évaluation à la fois multicritère et multiscalaire du parcours. Notre objectif était d'intégrer les actogrammes et de la loi de l'apogée-fin pour influencer sur l'organisation globale de la visite.

Une première étape, réalisée dans le chapitre 6, a été de mettre en lumière la présence de deux facteurs importants pour la qualité du séjour. En plus de l'adéquation avec les goûts culturels de ce dernier, nous avons proposé une nouvelle évaluation de la qualité d'un parcours touristique, basée sur une formalisation du concept d'actogramme. Cette mesure de la « cadence de visite » permet ainsi de considérer l'organisation du séjour vis à vis des comportements de mobilité du touriste. Les expériences, réalisées à partir des profils issus d'une enquête de terrain, montrent d'une part, que cette métrique peut être utilisée en complément d'une évaluation se basant sur les goûts culturels, mais aussi qu'elle produit des mesures variant significativement entre les profils.

Nous définissons dans le chapitre 7 le PE-PVRP_c. Ce dernier impose une stratégie de recommandation où le parcours est construit en suivant une vision multicritère et multiscalaire. Les deux métriques détaillées précédemment, que sont l'intérêt culturel et la cadence de visite, sont combinées et utilisées pour évaluer non seulement chaque activité du parcours mais aussi les séquences qu'elles forment. De plus, la loi de l'apogée-fin est au cœur du problème et amène à construire un parcours où deux périodes se démarquent des autres, favorisant la création d'un meilleur souvenir du voyage. Enfin, nous proposons une description du PE-PVRP_c sous la forme d'un objet mathématique qui puisse être exploité par un modèle MILP.

Cette stratégie consistant à agréger différents points de vue pour l'élaboration du parcours est incarnée par le système de recommandation que nous proposons dans le chapitre 8. Avec la considération des paires d'activités se succédant, le parcours n'est plus seulement construit comme un ensemble d'éléments mais en tant qu'une séquence, dont il est possible d'estimer les instants où le touriste assiste à une représentation, mais aussi ceux l'invitant à prendre son temps, pour flâner ou se reposer. Cette évaluation se fait en conciliant intérêt culturel et cadence de visite, sans que l'une de ces métriques ne soit une contrainte forte du système. Rien n'empêche alors le parcours de s'adapter, en dérivant des horaires choisis par le visiteur pour proposer une pièce particulière ou, au contraire, remplacer une représentation par un spectacle jugé moins pertinent car se déroulant à un moment propice. Finalement, nous avons proposé une première formulation de la loi de l'apogée-fin qui puisse être utilisé dans le contexte d'un modèle MILP. Celle-ci est intégré parmi un équilibre subtil, permettant de créer des parcours offrant des activités tout au long du séjour, tout en favorisant deux périodes clés afin de que le visiteur retire de son voyage, le meilleur souvenir.

Les expériences sur des profils types de visiteurs montrent deux résultats fondamentaux. D'une part, l'utilisation des actogrammes permet au visiteur de se voir proposer des parcours qui suivent le rythme qu'il préfère pour la journée. Ces résultats étant confirmé à partir de différents algorithmes, l'intégration de ces programmes d'activité nous apparaît comme une solution efficace pour prendre en compte les préférences horaires d'un touriste. D'autre part, lorsque le visiteur planifie un voyage de plusieurs jours, il est possible d'organiser le séjour de manière à favoriser deux journées, sans pour autant délaissé le reste du parcours. L'effet décrit par la loi de l'apogée-fin est ainsi mis en avant, tout en suivant à la fois les goûts culturels, et les préférences horaires de l'utilisateur.

Pour conclure, nous avons dans cette thèse présenté un cadre méthodologique pour la recommandation de parcours culturels profondément marqué par l'interdisciplinarité. À partir de celui-ci, nous avons mis en place différents systèmes de recommandation, au musée et en ville en nous appuyant sur des modèles de programmation linéaire mixte en

nombres entiers. Cependant, bien que ces derniers ait été évalués à partir d'enquêtes de terrain, leurs performances restent à être validées dans le cadre d'une utilisation réelle. Notamment, l'impact des parcours qu'ils élaborent sur la satisfaction du visiteur doit faire l'objet d'une attention particulière. Cela apporterait de plus amples arguments sur l'applicabilité des théories usées comme postulat dans le contexte des visites culturelles.

9.2 Perspectives

Nous avons mis en place dans cette thèse un cadre menant à une stratégie de recommandation basée sur une approche interdisciplinaire. Bien que cette collaboration permette d'explorer de nouvelles hypothèses et de proposer des méthodes innovantes, elle nécessite de travailler dans un cadre applicatif plus restreint. La question de la recommandation touristique et culturelle soulève encore de nombreuses interrogations, souvent à la croisée de plus d'un domaine de recherche.

9.2.1 Élargissement du cadre méthodologique

En observant le cadre méthodologique, nous pouvons mettre en lumière plusieurs aspects constituant des pistes d'amélioration des systèmes de recommandation proposés

Premièrement, bien que la modélisation au moyen de la programmation linéaire mixte en nombres entiers soit une pratique courante, une utilisation en conditions réelles nécessite des temps de réponse dans des délais raisonnables. Dans la littérature, de nombreuses heuristiques sont proposées pour donner de bonnes solutions au problème d'orientation, qui doivent être adaptées au contexte de la recommandation culturelle. Néanmoins, une modélisation de la loi de l'apogée-fin par une heuristique reste une problématique ouverte qui ouvrirait la voie à de nombreuses applications.

Deuxièmement, nous avons fait le choix de nous concentrer sur une recommandation à partir des envies et attentes exprimées explicitement par le visiteur, aujourd'hui, la plupart des systèmes de recommandation utilisent une stratégie hybride, incluant par exemple, une popularité des activités déduite à partir des retours par de précédent visiteurs.

Dernièrement, les formalismes proposés imposent au constituant d'un parcours d'avoir une durée fixe. Au musée, cette contrainte implique de déterminer pour chaque expôt, la durée que le visiteur serait amené à passer devant l'œuvre. Dans la littérature, nous pouvons retrouver quelques études en sociologie des publics sur les temps passés devant les œuvres, qui constituent un socle pour proposer des parcours plus réalistes.

Concernant nos travaux en ville, lever cette contrainte de durée fixe permettrait de proposer la visite de monuments ou lieux dont la durée dépend de l'intérêt que le touriste y accorde et nécessite là encore de prendre en compte les habitudes de consommation culturelle du voyageur.

9.2.2 Visite cohérente au musée

Au musée, l'utilisation du résumé automatique de texte a montré que les cartels des expôts sont empreints d'une marque trahissant un point de vue du musée sur sa propre collection. Néanmoins, cette analyse montre des faiblesses lorsqu'elle est confrontée à des corpus dont les textes sont très hétérogènes. Au delà d'une phase de pré-traitement uniformisant la longueur des cartels, l'utilisation de réseaux de neurones, qui ont montré leur efficacité dans le domaine du traitement automatique de la langue naturelle écrite, pourrait permettre d'affiner le classement des œuvres. De plus, ces derniers seraient aussi adéquats pour effectuer des rapprochements entre les expôts par *clustering*, amenant à considérer la pertinence de groupes d'expôts dans la visite, et ainsi voir celle-ci comme un récit cohérent au sein de la collection. Enfin, nos travaux se sont concentrés sur les musées d'art car la chronologie de la visite y a moins d'impact, aussi de nouveaux paradigmes et modèles seront nécessaires pour correctement traiter les parcours dans d'autres types d'institutions culturelles.

9.2.3 Évaluation des critères en ville

Lors de la proposition de séjours touristiques, une évaluation multicritère et à l'échelle du parcours dans son ensemble ouvre l'opportunité de considérer d'autres mesures. Nous avons évoqué plus tôt l'idée d'influer sur la cohérence culturelle ou l'alternance des styles durant le parcours, une autre piste à explorer est d'envisager la substitution de deux types d'activités lors de la construction du parcours, en fonction de la tolérance d'un visiteur à s'écarter d'un emploi du temps type pendant son séjour. Nous pouvons aussi noter les travaux de NAWIJN et BIRAN (2018) au sujet des émotions négatives, qui font partie intégrante de l'expérience de visite et peuvent avoir de multiples résultats positifs sur la satisfaction globale du touriste.

Nos travaux ont éludé la question des déplacements et de leur impact sur la qualité du parcours. D'une part, l'aspect multimodal a été exclu du fait de notre terrain d'étude mais doit être intégré pour faire face aux défis imposés par une métropole moderne, ou à ceux de la recommandation sur une aire urbaine plus étendue. D'autre part, nous pouvons aussi imaginer l'introduction de mesures utilisées par certains travaux de la littérature, telle

que la distance parcourue. Il se pourrait que cette dernière soit un facteur d'une métrique plus large qu'est celle de la convivialité de l'itinéraire emprunté. Comme nous l'avons soulevé dans le chapitre 3, la prédiction de l'itinéraire d'un pèlerin est une tâche complexe, et bien que des travaux aient identifiés de nombreux facteurs, une formalisation de ces derniers est aujourd'hui un verrou scientifique pour l'évaluation automatique du potentiel piétonnier.

Enfin, l'intégration de la loi de l'apogée-fin repose sur un découpage du séjour en période, pour lesquelles une valeur est attribuée en fonction des critères de qualité du parcours. Bien que nos choix aient été guidés par des études tirées de la littérature, la segmentation en périodes et leur pondération est un sujet ouvert.

Annexes

Annexe A

Le « Festival OFF d’Avignon », contexte et caractérisation du corpus

Sommaire

A.1	Présentation du terrain d’étude	168
A.1.1	Popularité lors des éditions passées et orientations des organisateurs	169
A.1.2	Données disponibles	170
A.2	Analyse des données textuelles	170
A.3	Analyse des données structurées	172
A.3.1	Informations spatio-temporelles	172
A.3.2	Attributs des spectacles et genre théâtral	175
A.4	Profils touristiques utilisés en simulation	176
A.4.1	Préférences culturelles et d’organisation du séjour	177
A.4.2	Itinéraire et mode de déplacement	178

A.1 Présentation du terrain d'étude

Cette étude s'articule autour de l'édition de 2015 du « Festival OFF d'Avignon », qui se présente comme l'un des plus grands festivals de spectacle vivant au monde, rassemblant des milliers de pièces sur trois semaines et regroupant des centaines de théâtres et compagnies indépendants.

À partir du programme officiel¹ dont un aperçu est donné par la figure A.1, nous avons collecté et organisé les informations concernant les pièces et les théâtres dans lesquelles elles sont jouées. Après un pré-traitement manuel retirant toutes les pièces pour lesquelles certaines informations sont manquantes, nous obtenons un catalogue de 1 348 pièces produites dans 127 théâtres sur 23 jours offrant un choix parmi 25 656 représentations.

11h	11h	12h30	12h40	12h40
durée 1h10 du 3 au 26 juillet	durée 1h du 3 au 26 juillet	durée 1h15 du 3 au 26 juillet	durée 1h15 du 4 au 26 juillet jours pairs	durée 1h15 du 5 au 25 juillet jours impairs relâche le 15 juillet
Théâtre [de 7 à 77 ans]	Chanson	Humour	Spectacle musical [à partir de 6 ans]	Spectacle musical
tarif : 17 € tarif adhérent public : 12 € tarif enfant [-16 ans] : 8 €	tarif : 15 € tarif adhérent public : 10,5 € tarif enfant [-12 ans] : 8 €	tarif : 16 € tarif adhérent public : 11 € tarif enfant [-14 ans] : 8 €	tarif : 17 € tarif adhérent public : 12 € tarif enfant [-6 ans] : 6 €	tarif : 17 € tarif adhérent public : 12 € tarif enfant [-6 ans] : 6 €
FAUTEUILS / GRADINS	FAUTEUILS / CHAISES	FAUTEUILS / CHAISES	FAUTEUILS / GRADINS	FAUTEUILS / GRADINS
L'odyssée de bric et de broc ... à peu près Homère Ulysse vainqueur de la guerre de Troie devra affronter une longue route semée d'embûches avant de retrouver sa femme Pénélope	A Capella Pour Les Mômes Vian, Dutronc, Souchon, Gotainer, Perret, Lapointe La démarche de ces artistes est singulière à chaque nouveau spectacle : Exercice de style. Baby Gagnon	Détention Dérisoire de Fanny Mermet, Jean-Jacques Devaux Le juge a dit "NON COUPABLE" après 30 ans de prison, il était temps ! Se retrouver dehors ne va pas être simple. Fanny en sera-t-	Le New Lyrique Boys Band : Chevauchée Lyrique ! de Fabrice Maitre Succès depuis 2012 en	Le New Lyrique Boys Band : La Do Ré etc de François Rollin Succès depuis 2012 le NLBB et François Rollin s'associent

FIGURE A.1 – Un aperçu du programme de l'édition 2015 du « Festival OFF d'Avignon »

Il est important de préciser que cet événement n'est pas à confondre avec le « Festival d'Avignon » (couramment appelé le « IN »), qui est lui aussi un festival de théâtre se déroulant en parallèle mais dont l'organisation est indépendante comme le détaille l'ouvrage de ETHIS (2002). Bien qu'en apparence très similaires, ces deux manifestations présentent des différences importantes rendant l'exploitation simultanée des deux collections de pièces complexe à évaluer. En effet, la programmation du « IN » est sélectionnée chaque

1. Disponible sur <https://www.avignonleoff.com> au format PDF dans la rubrique « archives ».

année en suivant une thématique par une équipe, tandis que l'association responsable du « OFF » n'assure pas le programme du festival mais expose les spectacles prenant place dans les théâtres participants. Ainsi, l'ampleur du catalogue proposé par le « IN » est bien moins importante (une soixantaine de pièces pour environ 200 représentations²) et les descriptifs des pièces sont travaillés par l'équipe en charge de la programmation en suivant la thématique choisie. À l'inverse, le « OFF » expose des milliers de pièces (pour des dizaines de milliers de représentations) qui d'une part ne s'axent pas autour d'une thématique unique et d'autre part dont les descriptions ne sont pas supervisées, mais enregistrées telles que formulées par chaque compagnie.

A.1.1 Popularité lors des éditions passées et orientations des organisateurs

Dans l'état de l'art (voir chapitre 3), nous avons souligné que certains travaux utilisent la popularité comme une mesure de l'attrait des points d'intérêt. Cela peut être fait en utilisant une mesure de la fréquentation globale (à partir des réseaux sociaux comme dans CENAMOR et al. (2017)) ou une note de satisfaction pondérée par la similarité entre les précédents visiteurs et le profil actuel (parfois basée sur du texte comme les travaux de BUCUR (2015)).

Or, bien que le « Festival OFF d'Avignon » existe depuis 1966, chaque édition est unique car liée aux nombreux théâtres sélectionnant indépendamment leur programmation en fonction des compagnies souhaitant présenter leur spectacle. Ainsi, bien que certaines pièces se produisent plusieurs années de suite, une grande majorité de l'offre est renouvelée.

Par ailleurs, les organisateurs de l'évènements sont attachés au principe de flânerie et à la découverte de pépites, des pièces provenant de compagnies méconnues vers lesquelles le visiteur ne se serait pas dirigé de prime abord. Dans ce contexte, utiliser la popularité des spectacles ou de leur compagnie dans les éditions passées créerait un déséquilibre entre les habitués de la manifestation et les productions épisodiques.

Pour ces raisons, ainsi que la complexité de la tâche d'extraction et d'analyse des données, nous n'utilisons pas de mesure de popularité, mais nous nous basons principalement sur les envies exprimées par le visiteur.

2. Chiffres donnés par la FAQ du site <https://www.festival-avignon.com>.

A.1.2 Données disponibles

Le corpus de données est donc exclusivement issu du programme officiel et se compose de la liste des théâtres et pièces proposées aux visiteurs durant la manifestation.

Pour chaque pièce nous disposons des éléments suivants :

- Le titre de la pièce ;
- une description ;
- la durée d'une représentation ;
- les horaires de chacune des représentations (dates et heures) ;
- le style de pièce et les publics visés ;
- le nom et une courte description de la compagnie la jouant ;
- les intervenants (auteurs, interprètes, metteur en scène...) ;
- les tarifs disponibles ;
- le théâtre (et la plupart du temps la salle) dans lequel elle est jouée.

Concernant les théâtres, nous avons recueilli :

- Le nom du théâtre ;
- une description ;
- son adresse postale ;
- ses coordonnées géographiques ;
- sa programmation ;
- ses différentes salles ;
- des attributs pour chaque salle (accès handicapé, climatisation...).

Sur l'ensemble de ces données, nous pouvons distinguer les informations textuelles pour lesquelles il n'y a aucune contrainte (hormis celle du nombre de caractères imposée par la base de données) des données structurées dont les formats sont pré-établis par l'éditeur du site.

A.2 Analyse des données textuelles

Dans un premier temps, nous avons souhaité appliquer les mêmes traitements aux pièces de théâtre qu'aux expôts de musées, c'est-à-dire les différencier à partir des descriptions fournies par les organismes officiels. Ainsi, à chaque pièce est associé un document textuel composé de son titre et de sa description auquel un pré-traitement standard³ (lemmatisation et suppression des mots outils) est appliqué.

3. Toutes les opérations ont été effectuées en utilisant la bibliothèque NLTK en version 3.4.4 introduite par BIRD et al. (2009).

Premièrement, la figure A.2 montre que le nombre de termes par document est sensiblement équivalent à celui des termes uniques, c'est à dire que peu d'entre eux apparaissent plusieurs fois dans un même texte. En effet, en moyenne un terme a 1,12 occurrences dans un document.

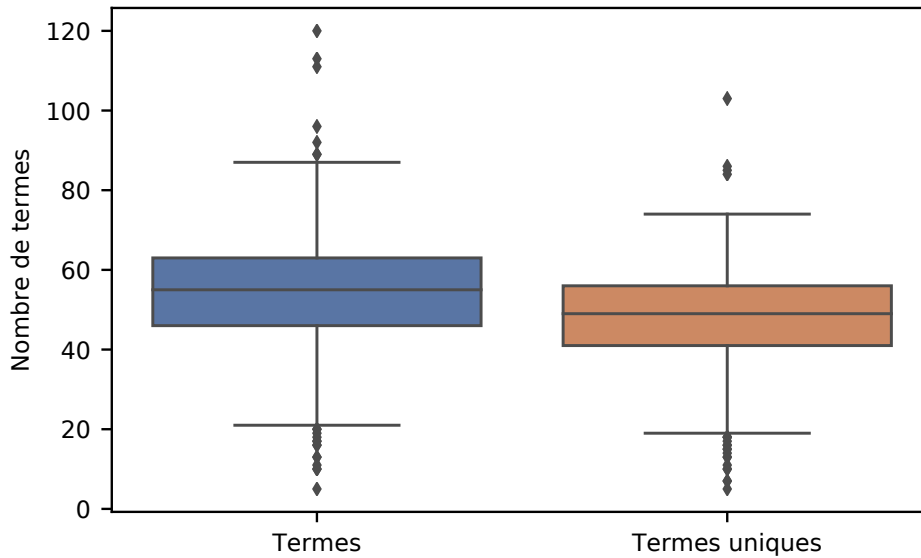


FIGURE A.2 – Nombre de termes par document représentant les pièces de théâtre

La figure A.3 quant à elle montre que la plupart des termes employés dans une description ne sont utilisés nulle part ailleurs. En effet, 29 % des documents ne partagent pas plus de 2 termes en commun.

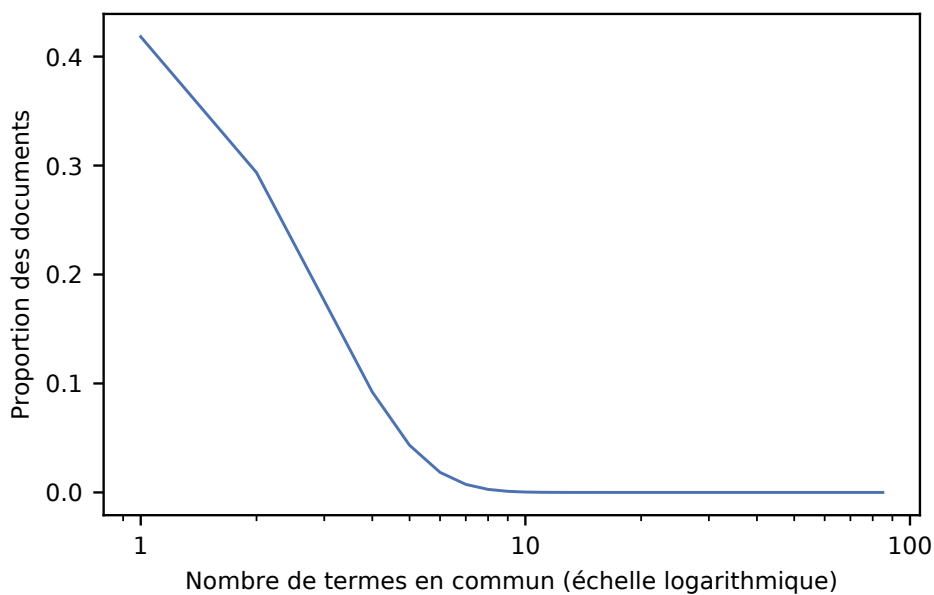


FIGURE A.3 – Proportion des documents partageant des termes en commun

Pour ces raisons, les méthodes classiques de pondération utilisées en traitement au-

tomatique de la langue naturelle écrite comme le *TF-IDF* (SPARCK JONES 1972 ; TORRES-MORENO 2014) ne sont pas capables de bien différencier les spectacles les uns par rapport aux autres en utilisant une requête de l'utilisateur. Nous pouvons conjecturer que cela est lié à la méthode employée par l'organisme responsable de l'évènement pour construire son offre culturelle. En effet, alors que les cartouches des œuvres sont tous écrits par un petit groupe d'experts, chaque compagnie écrit la description des spectacles qu'elle propose.

Conséquemment, il est évidemment impossible d'utiliser une méthode dérivée de l'analyse de sentiments afin de discriminer les pièces entre elles, puisqu'il n'existe qu'une seule source par spectacle et qu'elle provient de la compagnie qui le produit.

En conclusion, nous avons choisi de ne pas utiliser les ressources textuelles à disposition, car les techniques classiques d'analyse du langage naturel écrit ne produisent pas de résultats pertinents.

A.3 Analyse des données structurées

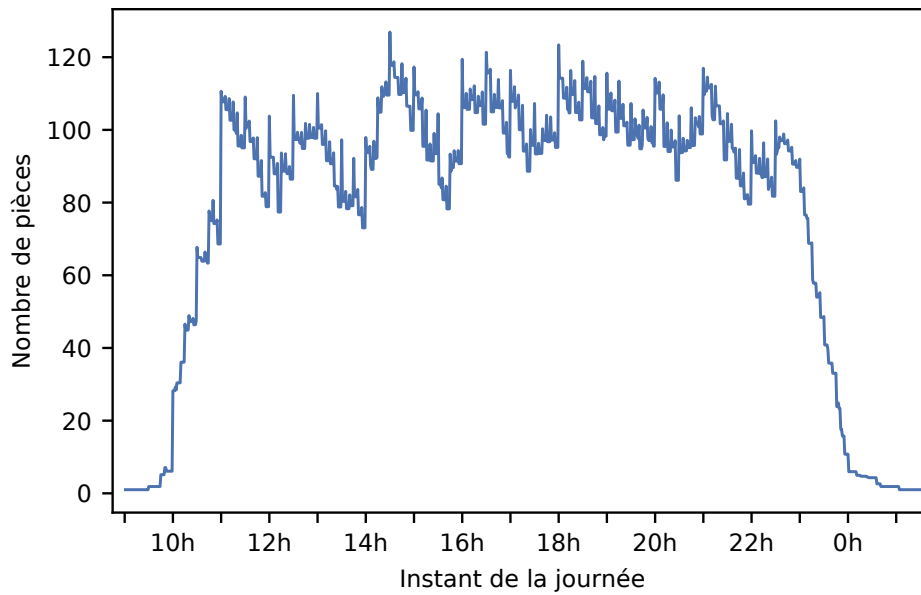
Le reste des données disponibles sont des données structurées, parmi lesquelles nous retrouvons les informations spatio-temporelles qui permettent à un système de recommandation de construire un parcours réalisable, c'est-à-dire en tenant compte des temps de trajet et n'incluant pas deux propositions en parallèle. Enfin, des éléments caractérisant le genre des spectacles sont utilisés, afin de les différencier et créer des profils touristiques.

A.3.1 Informations spatio-temporelles

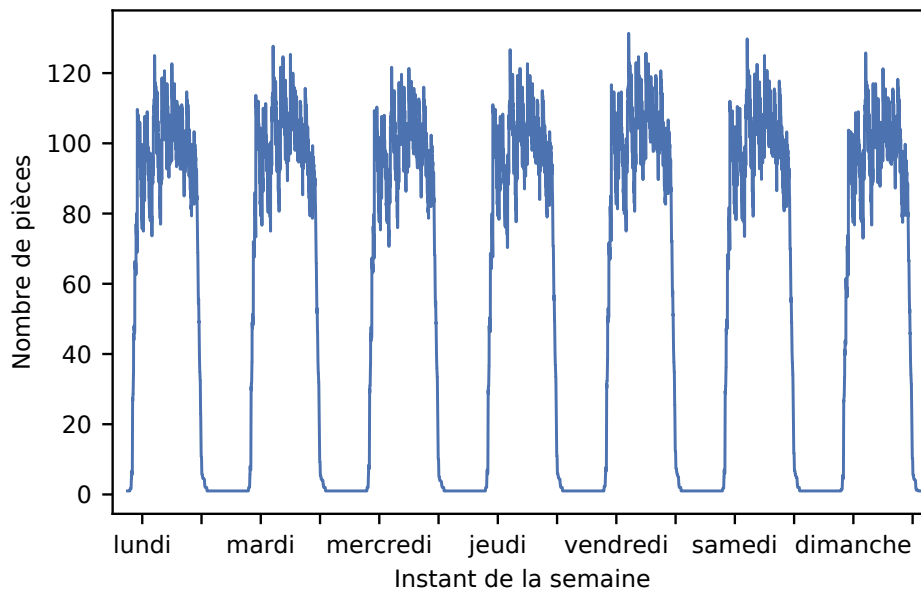
L'offre culturelle durant le « Festival OFF d'Avignon » est très dense, la figure A.4 montre que sur une journée, à chaque instant se déroulent simultanément entre 72 et 103 représentations (1^{er} et 3^e quartiles) et que cette abondance de choix est constante tout au long d'une semaine, offrant en moyenne 1 115 spectacles par jour.

Toutefois, l'univers des possibles est réduit par plusieurs facteurs. D'une part, nous avons fait le choix de ne recommander qu'une seule fois une pièce. Or, ces dernières sont jouées entre 17 et 23 fois tout au long du festival pour une représentation par jour pour 88 % des pièces. Sélectionner une représentation élimine donc toutes les autres représentations du même spectacle. D'autre part, le fait d'assister à une représentation occupe le temps disponible du visiteur entre 1 h et 1 h15 min mais peut atteindre 3 h comme le montre la figure A.5⁴.

4. Nous représentons les données par une densité de probabilité afin de lisser la courbe et ainsi améliorer la lisibilité.



(a) Au cours de la journée



(b) Au cours de la semaine

FIGURE A.4 – Nombre de pièces jouées pendant le « Festival OFF d'Avignon »

Enfin, 115 des 127 théâtres participants se situent en centre-ville (figure A.6), dans un espace peu étendu, ce qui mène à une discrimination des rares pièces situées dans les villes alentours à cause des temps de trajet beaucoup plus importants pour relier deux pièces situées dans des lieux éloignés.

Pour conclure, bien qu'il existe quelques facteurs limitant le champs des possibles, il subsiste en moyenne pas moins de 255 mille milliards de parcours réalisables chaque jour.

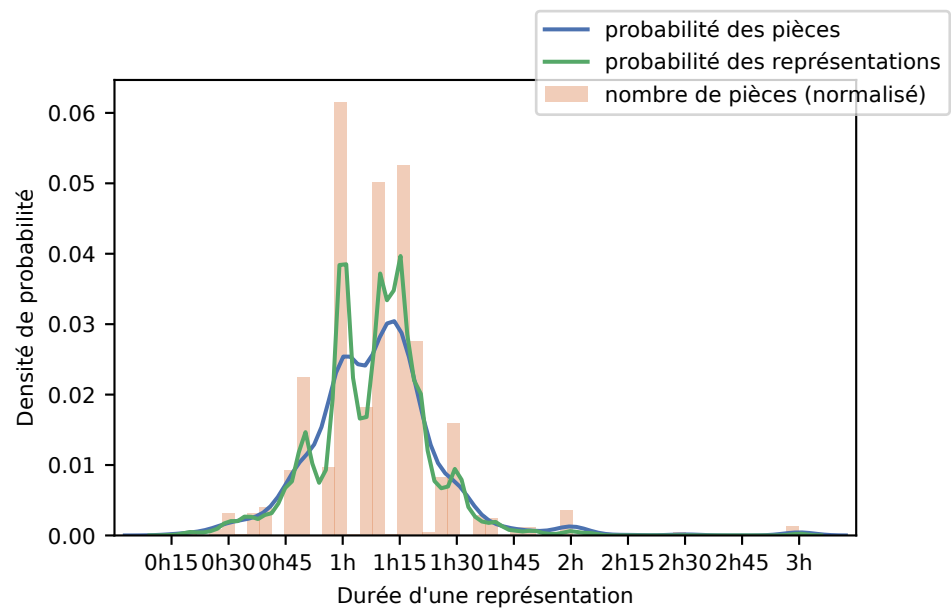


FIGURE A.5 – Durée des pièces de théâtre



FIGURE A.6 – Localisation des théâtres du centre-ville d'Avignon pendant le « Festival OFF d'Avignon » (par PICOLOT 2016)

A.3.2 Attributs des spectacles et genre théâtral

Afin de créer des profils touristiques, nous avons vu qu'il était impossible d'utiliser les données textuelles, trop disparates. Nous nous tournons donc vers l'utilisation des caractéristiques des spectacles qui déterminent les genres théâtraux de chaque pièce. Ces informations elles aussi sont fournies par chaque compagnie mais les valeurs sont limitées à un ensemble de genres déterminés, qui serviront de mots-clés pour l'utilisateur.

La figure A.7 recense tous les attributs possibles ainsi que leur importance dans le corpus. De plus, la figure A.8 illustre les liens qu'entretiennent ces attributs les uns envers les autres dans le corpus. Chaque arc représente une caractéristique et sa taille reflète le nombre de pièces associée à cette dernière. Un lien entre deux arcs est présent lorsqu'il existe des pièces possédant les caractéristiques de ces derniers. La taille de ce lien indique la proportion des pièces possédant la paire d'attributs correspondants.

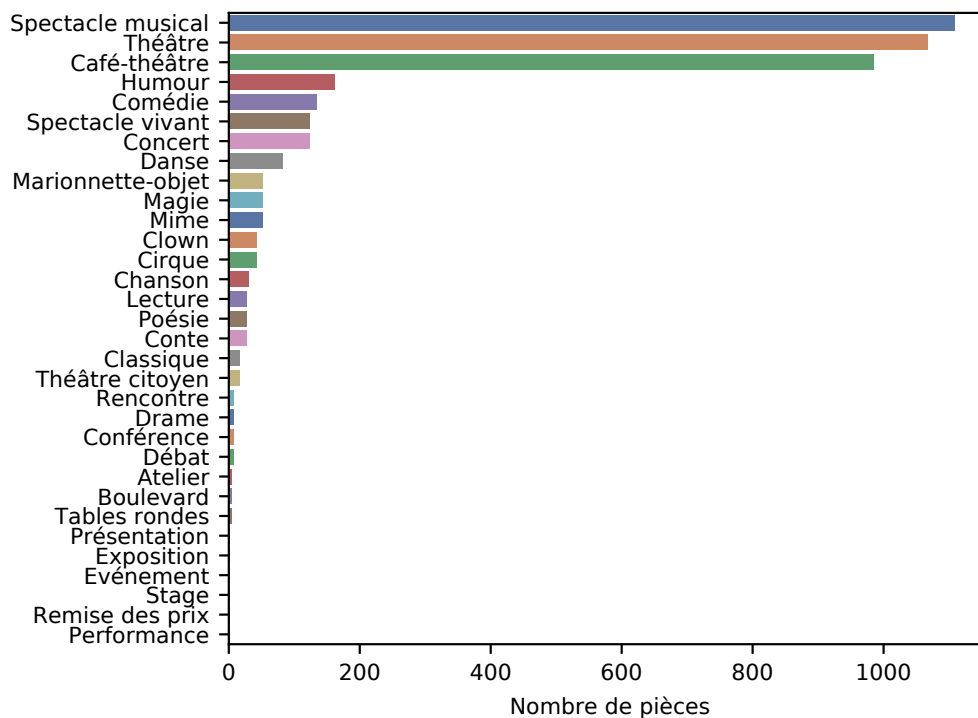


FIGURE A.7 – Importance des attributs caractérisant les pièces

On remarque que 3 des 32 genres théâtraux sont surreprésentés par rapport aux autres. De plus, 54,3 % des spectacles ayant un des ces attributs majeurs, n'ont aucun attribut mineur (autre que les 3 plus importants). Cet état de fait les rend indispensables dans les profils généralistes, puisqu'ils sont les seuls à couvrir une part importante du corpus, mais aussi dans les profils spécifiques, puisqu'ils permettent d'éliminer cette même portion du corpus.

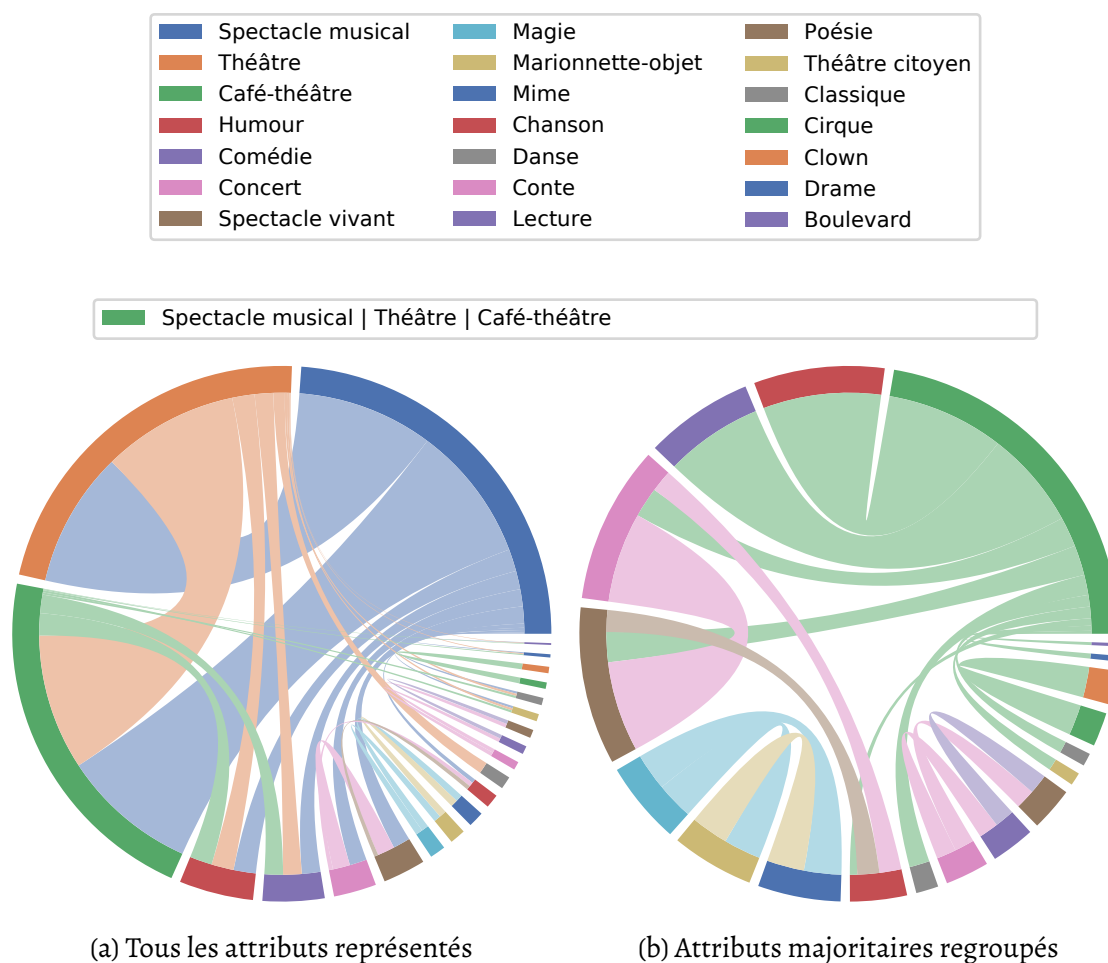


FIGURE A.8 – Relations entre les attributs

Sur la figure A.8b, on peut identifier des groupes d'attributs entretenant des liens forts entre eux et très peu, voire aucun, avec les autres (comme le triplet « Magie », « Marionnette-objet », « Mime »). Ces groupes constituent de bons candidats pour des profils marginaux très spécifiques avec une forte discrimination.

L'analyse de ce corpus permet de construire des profils touristiques, en termes de préférences culturelles, qui ciblent différentes catégories de spectacles et de représenter des goûts très variables.

A.4 Profils touristiques utilisés en simulation

Dans le chapitre 8, nous proposons un système de recommandation ayant pour but de générer des parcours personnalisés en s'adaptant aux goûts du touriste. Afin d'évaluer le

comportement de notre méthode, nous établissons un ensemble de profils touristiques réalistes utilisables dans les simulations. Dans un premier temps ces profils sont décrits par des goûts culturels et préférences horaires puis, par la suite, nous détaillons des considérations sur leur mobilité durant le séjour, avant de conclure par une analyse de ces profils vis-à-vis de l'offre culturelle.

A.4.1 Préférences culturelles et d'organisation du séjour

Le premier aspect de ces profils touristiques est bien évidemment lié aux attentes du visiteur. On distingue les préférences culturelles concernant le type de contenu culturel des préférences d'organisation qui se basent sur l'organisation spatiale et temporelle du parcours.

Contenu culturel

Nous avons choisi d'exprimer les préférences culturelles par un ensemble de mots-clés issus du corpus de données décrit dans la section précédente. La table A.1 expose les cinq profils culturels retenus en donnant pour chacun les genres théâtraux qui l'attirent ou au contraire qui sont à éviter.

Les variantes C_1 , C_2 et C_3 sont des profils généralistes ciblant le théâtre, mais sans adéquation négative pour aucun genre théâtral. Les deux derniers sont, quant à eux, des profils antagonistes très discriminants, car en plus de cibler un domaine très spécifique, ils excluent une partie de l'offre de l'évènement.

Organisation du séjour

Concernant les préférences sur l'organisation du parcours, nous utilisons des actogrammes, qui sont des représentations d'une journée type d'un individu (voir le chapitre 3). Dans le chapitre 6 nous détaillons les travaux de PICOLOT (2016) et sélectionnons puis comparons des actogrammes associés à différents profils touristiques.

De nouveau, les profils touristiques et leur actogramme employés sont les suivants :

P_a « Famille avec enfant(s) de plus de 5 ans »

P_b « Amis, couple sans enfant »

P_c « Personne seule »

Annexe A. Le « Festival OFF d'Avignon », contexte et caractérisation du corpus

Profil culturel	Goûts exprimés sur les genres théâtraux	
C_1	Apprécie : – Spectacle musical – Danse	– Chanson – Théâtre
C_2	Apprécie : – Poésie – Conte	– Lecture – Théâtre
C_3	Apprécie : – Boulevard – Comédie – Humour	– Drame – Théâtre
C_4	Apprécie : – Comédie – Humour N'apprécie pas : – Clown – Cirque	– Classique – Théâtre – Marionnette-objet – Magie
C_5	Apprécie : – Clown – Cirque N'apprécie pas : – Théâtre	– Marionnette-objet – Magie – Classique

TABLE A.1 – Profils culturels utilisés lors des simulations

Ces actogrammes, tirés d'une enquête de terrain, couvrent une palette variée d'emplois du temps comme le montre notre étude de cas en section 6.4 et sont donc conservés pour décrire les différents profils touristique en matière d'organisation du séjour.

A.4.2 Itinéraire et mode de déplacement

Une des spécificités du tourisme en ville vis à vis des visites de musées est la place prépondérante de la temporalité dans le parcours. Choisir d'effectuer une pièce détermine celles que l'on peut effectuer avant et après celle-ci en raison de ses horaires et sa durée mais aussi de l'emplacement du théâtre dans laquelle elle est jouée à cause des temps de parcours non négligeables. Or, intégrer le temps de trajet d'un théâtre à un autre implique de calculer l'itinéraire qu'empruntera un touriste, ce qui, comme nous l'avons vu dans le chapitre 3 est un problème complexe à modéliser et constitue un domaine de recherche en soi.

Afin d'estimer ce temps de trajet, nous utilisons les données cartographiques d'OpenS-

treetMap⁵ ainsi que le moteur de routage OSRM⁶ afin de calculer les plus courts chemins entre deux localisations en utilisant les paramètres suivants :

- Lors d'un trajet dans une même ville : déplacement à pied à $2,5 \text{ km h}^{-1}$. Ceci correspond à la vitesse délimitant la flânerie et la marche lente, d'après BAUDER (2015).
- Lors d'un trajet entre deux villes différentes : en voiture en respectant les limitations de vitesse. Le « Parking des Italiens » est utilisé comme point intermédiaire entre la marche et le déplacement véhiculé pour les trajets impliquant un théâtre du centre ville d'Avignon et un autre situé en dehors.

Il est important de noter que les entretiens avec les professionnels et organisateurs nous indiquent qu'il faut considérer un temps d'attente entre le moment où le touriste arrive à un théâtre et le moment où celui-ci peut prendre place dans la salle. Il lui faut, entre autre, valider, voire acheter, son billet. De plus, les compagnies conseillent d'arriver au moins 15 min en avance, temps que nous ajoutons à celui nécessaire au déplacement entre deux activités.

De plus, pour créer une plus grande diversité, deux variantes sont utilisées pour chacun des profils :

- Le visiteur est à portée de toutes les activités du début à la fin ;
- Le visiteur doit quotidiennement commencer depuis le « Parking des Italiens » à partir de 9:15:00 et terminer au même lieu au plus tard à 21:45:00.

Elles traduisent respectivement le cas d'un visiteur logeant en centre ville et celui résidant aux alentours, devant se déplacer en voiture afin d'atteindre le centre ville.

Au total, nous avons donc un ensemble de 30 profils touristiques différents à utiliser pour analyser le comportement du système de recommandation que nous proposons. Les résultats présentés dans le chapitre 8 montrent que nous arrivons à produire des parcours s'adaptant à cette diversité de situations.

5. <https://www.openstreetmap.org>

6. <http://project-osrm.org>

Glossaire

Actogramme Méthode de représentation des comportements d'individus issue de l'étude des populations qui permet de rendre compte de l'enchaînement des activités et des déplacements. ii, 7, 10, 38, 45, 88, 94, 95, 100–103, 106–108, 112, 120, 129–131, 145–149, 154, 159, 160, 177, 178, 184

Loi de l'apogée-fin « Heuristique psychologique dans laquelle les gens jugent une expérience en grande partie en fonction de ce qu'ils ont ressenti à son apogée (c'est-à-dire à son point le plus intense) et à sa fin, plutôt qu'en fonction de la somme totale ou moyenne de chaque moment de l'expérience. » (KAHNEMAN – 2000) ii, 7, 10, 11, 39–41, 45, 110, 111, 113, 114, 116, 117, 124, 126, 127, 133, 135–145, 147, 152–154, 159–161, 163, 185

MILP programmation linéaire mixte en nombres entiers (*Mixed Integer Linear Programming*) 8, 10, 11, 62, 68, 71, 75, 79–82, 127, 133, 136, 149, 158–161

OP problème d'orientation (*Orienteering Problem*) 43, 44, 71, 124, 161

PCTSP problème de voyageur de commerce collectant des prix (*Prize-Collecting TSP*) 43, 44, 124

PE-PVRP_c problème de recommandation de visite personnalisée en ville considérant l'effet de l'apogée-fin (*Peak-end rule driven Personalized Visit Recommendation Problem in city*) 10, 111, 114, 116, 118, 120, 121, 126, 128, 132–134, 137, 138, 144, 148, 149, 154, 160, 185

PTP problème du parcours rentable (*Profitable Tour Problem*) 43, 44, 124

PVRP_m problème de recommandation de visite personnalisée au musée (*Personalized Visit Recommendation Problem in museum*) 10, 63, 66–73, 159, 185

Résumé automatique de texte « Un résumé automatique est un texte généré par un logiciel, qui est cohérent et contient une quantité importante d'informations pertinentes provenant du texte source. Son taux de compression τ est inférieur à un tiers de la longueur du document original. » (TORRES-MORENO – 2014) i, 5, 51, 54–57, 59, 60, 64, 77, 80, 84, 158, 162

Système de recommandation « Tout système produisant des recommandations individualisées ou permettant de guider l'utilisateur de manière personnalisée vers des solutions utiles ou intéressantes au sein d'un vaste espace de possibilités. » (BURKE – 2002) ii, 3–10, 16–19, 21, 23–28, 30, 31, 33, 34, 38, 39, 41–45, 50–54, 60, 62, 63, 65, 68, 72, 75, 76, 78, 83, 84, 88, 91, 95–98, 108, 110–115, 118, 119, 124, 126, 133, 145, 148, 149, 151, 154, 158–161, 172, 176, 179

TSP problème de voyageur de commerce (*Traveling Salesman Problem*) 43–45

TTDP problème de conception de voyage touristique (*Tourist Trip Design Problem*) 43, 44, 63, 89, 111, 120, 124

Liste des figures

3.1	Représentation 3D dans l'espace et le temps des activités d'une journée (par KRAAK 1988)	37
3.2	Représentation d'un actogramme correspondant à un profil de visiteur du « Festival OFF d'Avignon » (voir le chapitre 6)	38
3.3	Illustration du processus de sélection des expériences selon la loi de l'apogée-fin	40
4.1	Utilisation conjointe des préférences personnelles et d'analyse du texte pour estimer l'intérêt des expôts	51
4.2	Processus de calcul du prestige des expôts par adaptation du résumé automatique de texte	55
4.3	Prestige indépendant des expôts calculé en adaptant l'algorithme ENERTEX	59
5.1	Architecture et composants principaux de notre système pour la recommandation de visites personnalisées au musée	64
5.2	Les différents niveaux d'information géographique à modéliser pour un musée et une visite	65
5.3	Plans et graphes associés des différents musées	77
5.4	Efficacité en fonction de l'intérêt des expôts dans le « Musée de l'Orangerie »	80
5.5	Efficacité en fonction de l'intérêt des expôts dans la « National Gallery » .	81
5.6	Efficacité en fonction du nombre de chefs-d'œuvre dans le parcours pour le « Musée de l'Orangerie »	82
5.7	Efficacité en fonction du nombre de chefs-d'œuvre dans le parcours pour la « National Gallery »	82
5.8	Parcours conseillés pour deux profils différents dans la « National Gallery »	83
6.1	Exemples de parcours touristiques	91
6.2	Relations entre les intérêts déclarés pour les types de lieux culturels . . .	93

Liste des figures

6.3	Actogrammes des profils touristiques de PICOLOT (2016) pour le « Festival OFF d'Avignon »	95
6.4	Différences entre les actogrammes de profils touristiques	102
6.5	Parcours proposés par PICOLOT	104
6.6	Parcours « Variation personne seule », adapté à l'actogramme correspondant	106
7.1	Les différentes relations entre deux intervalles temporels (selon ALLEN 1984)	115
7.2	Graphe associé à une infime partie du « Festival OFF d'Avignon »	122
8.1	Architecture et composants principaux de notre système pour la recommandation de parcours touristiques en ville	127
8.2	Utilisation des actogrammes comme outil de mesure du confort	130
8.3	Attrait des activités en fonction du profil touristique	146
8.4	Distribution des activités au long de la journée en fonction de l'importance relative de la cadence de visite (voir Figure 6.3)	151
8.5	Distribution des activités tout au long du parcours en fonction du paramètre λ	153
A.1	Un aperçu du programme de l'édition 2015 du « Festival OFF d'Avignon »	168
A.2	Nombre de termes par document représentant les pièces de théâtre	171
A.3	Proportion des documents partageant des termes en commun	171
A.4	Nombre de pièces jouées pendant le « Festival OFF d'Avignon »	173
A.5	Durée des pièces de théâtre	174
A.6	Localisation des théâtres du centre-ville d'Avignon pendant le « Festival OFF d'Avignon » (par PICOLOT 2016)	174
A.7	Importance des attributs caractérisant les pièces	175
A.8	Relations entre les attributs	176

Liste des tableaux

2.1	Stratégie et données requises des différentes approches de recommandation (selon AGGARWAL 2016a)	20
4.1	Principaux éléments caractérisant les musées étudiés	57
5.1	Variables de décisions utilisées pour la résolution du PVRP _m	69
5.2	Comparaison des temps de calcul des approches proposées	79
6.1	Répartition des répondants de l'enquête de BÉGUIN	92
6.2	Organisation du séjour des visiteurs	94
6.3	Durée du séjour des visiteurs du « Festival OFF d'Avignon »	96
6.4	Matrice des pénalités appliquées pour l'évaluation de la cadence de visite	105
6.5	Évaluation croisée de parcours touristiques	107
8.1	Principales variables de décision associées au PE-PVRP _c	134
8.2	Variables de décision utilisées pour rendre compte de l'effet apogée-fin .	137
8.3	Comparaison des temps de calcul et de la mémoire utilisée des systèmes utilisés	150
8.4	Mesures moyennes de la taille et qualité des parcours générés en fonction du paramètre α	150
8.5	Nombre de recommandations par jour des systèmes évalués	152
A.1	Profils culturels utilisés lors des simulations	178

Bibliographie

Références

- AGGARWAL, Charu C. (2016a). « An Introduction to Recommender Systems ». In : *Recommender Systems : The Textbook*. Sous la dir. de Charu C. AGGARWAL. Cham, Switzerland : Springer International Publishing, p. 1-28. ISBN : 978-3-319-29659-3. DOI : 10.1007/978-3-319-29659-3_1.
- AGGARWAL, Charu C. (2016b). « Knowledge-Based Recommender Systems ». In : *Recommender Systems : The Textbook*. Sous la dir. de Charu C. AGGARWAL. Cham, Switzerland : Springer International Publishing, p. 167-197. ISBN : 978-3-319-29659-3. DOI : 10.1007/978-3-319-29659-3_5.
- AGGARWAL, Charu C., éd. (2016c). *Recommender Systems : The Textbook*. Cham, Switzerland : Springer International Publishing. 498 p. ISBN : 978-3-319-29659-3. DOI : 10.1007/978-3-319-29659-3.
- AHUJA, Ravindra K., Thomas L. MAGNANTI et James B. ORLIN (1993). *Network Flows : Theory, Algorithms, and Applications*. Upper Saddle River, NJ, USA : Prentice-Hall, Inc. ISBN : 978-0-13-617549-0.
- ALLEN, James F. (1^{er} juil. 1984). « Towards a general theory of action and time ». In : *Artificial Intelligence* 23.2. ISSN : 0004-3702. DOI : 10.1016/0004-3702(84)90008-0. URL : <http://www.sciencedirect.com/science/article/pii/0004370284900080>.
- AMAR, Georges et Véronique MICHAUD (17 nov. 2009). *La marche au coeur des mobilités : Etat des connaissances*. Lyon : CERTU. 64 p. ISBN : 978-2-11-098225-4.
- AMINI, Massih-Reza et Gaussier ERIC (avr. 2013). *Recherche d'Information - applications, modèles et algorithmes*. Sous la dir. d'EYROLLES. Algorithmes. Eyrolles. 1-233. URL : <https://hal.archives-ouvertes.fr/hal-00881257>.
- ANDREACOLA, Florence (2 déc. 2015). « Nouvelles modalités d'appropriation et de partage au musée : les pratiques multiples de l'expérience de visite à l'époque de la culture numérique ». Thèse de doct.
- ARDISSONO, Liliana, Tsvi KUFLIK et Daniela PETRELLI (18 oct. 2011). « Personalization in Cultural Heritage : The Road Travelled and the One Ahead ». In : *User Modeling and User-Adapted Interaction* 22.1-2, p. 73-99. ISSN : 0924-1868. DOI : 10.1007/s11257-011-9104-x.
- BALAS, Egon (2007). « The Prize Collecting Traveling Salesman Problem and its Applications ». In : *The Traveling Salesman Problem and Its Variations*. Sous la dir. de Gregory GUTIN et Abraham P. PUNNEN. Combinatorial Optimization. Boston, MA : Springer US, p. 663-695. ISBN : 978-0-306-48213-7. DOI : 10.1007/0-306-48213-4_14.

- BARBEHENN, Michael (1998). « A Note on the Complexity of Dijkstra's Algorithm for Graphs with Weighted Vertices ». In : *IEEE Transactions on Computers* 47.2, p. 263. ISSN : 0018-9340. DOI : <http://doi.ieeecomputersociety.org/10.1109/12.663776>.
- BAUDER, Michael (2015). « Using GPS Supported Speed Analysis to Determine Spatial Visitor Behaviour ». In : *International Journal of Tourism Research* 17.4, p. 337-346. DOI : 10.1002/jtr.1991. eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jtr.1991>. URL : <https://onlinelibrary.wiley.com/doi/abs/10.1002/jtr.1991>.
- BÉGUIN, Camille (2015). « P'ART COURT ». Mém. de mast. Université d'Avignon et des Pays de Vaucluse.
- BELTING, Hans, Arthur DANTO, Jean GALARD, Martina HANSMANN, Neil MACGREGOR, Werner SPIES et Matthias WASCHEK (22 nov. 2000). *Qu'est-ce qu'un chef-d'œuvre?* Gallimard. 224 p. ISBN : 2070759636.
- BEN TICHA, Sonia (11 nov. 2015). « Recommandation personnalisée hybride ». thesis. Université de Lorraine. URL : <http://www.theses.fr/2015LORR0168>.
- BENOUARET, Idir (25 jan. 2017). « Un système de recommandation contextuel et composite pour la visite personnalisée de sites culturels ». thesis. Compiègne. URL : <http://www.theses.fr/2017COMP2332>.
- BERRE, Daniel Le, Pierre MARQUIS et Stéphanie ROUSSEL (2013). « Planning Personalised Museum Visits ». In : *Proceedings of the Twenty-Third International Conference on Automated Planning and Scheduling, ICAPS 2013, Rome, Italy, June 10-14, 2013*. Sous la dir. de Daniel BORRAJO, Subbarao KAMBHAMPATI, Angelo ODDI et Simone FRATINI. AAAI. ISBN : 978-1-57735-609-7. URL : <http://www.aaai.org/ocs/index.php/ICAPS/ICAPS13/paper/view/6025>.
- BIRD, Steven, Ewan KLEIN et Edward LOPER (2009). *Natural Language Processing with Python*. 1st. O'Reilly Media, Inc. ISBN : 978-0-596-51649-9.
- BORRÀS, Joan, Antonio MORENO et Aida VALLS (15 nov. 2014). « Intelligent tourism recommender systems : A survey ». In : *Expert Systems with Applications* 41.16, p. 7370-7389. ISSN : 0957-4174. DOI : 10.1016/j.eswa.2014.06.007. URL : <http://www.sciencedirect.com/science/article/pii/S0957417414003431>.
- BUCUR, Cristian (1^{er} jan. 2015). « Using Opinion Mining Techniques in Tourism ». In : *Procedia Economics and Finance*. 2nd GLOBAL CONFERENCE on BUSINESS, ECONOMICS, MANAGEMENT and TOURISM 23, p. 1666-1673. ISSN : 2212-5671. DOI : 10.1016/S2212-5671(15)00471-2. URL : <http://www.sciencedirect.com/science/article/pii/S2212567115004712>.
- BURKE, Robin (1^{er} nov. 2002). « Hybrid Recommender Systems : Survey and Experiments ». In : *User Modeling and User-Adapted Interaction* 12.4, p. 331-370. ISSN : 1573-1391. DOI : 10.1023/A:1021240730564.
- CENAMOR, Isabel, Tomás de la ROSA, Sergio NÚÑEZ et Daniel BORRAJO (1^{er} mar. 2017). « Planning for tourism routes using social networks ». In : *Expert Systems with Applications* 69, p. 1-9. ISSN : 0957-4174. DOI : 10.1016/j.eswa.2016.10.030. URL : <http://www.sciencedirect.com/science/article/pii/S0957417416305693>.
- CHARDONNEL, Sonia et Thomas THÉVENIN (2012). « Les apports de la Time-Geography dans les représentations spatio-temporelles ». In : *École thématique MODyS du CNRS sur modélisation et visualisation des dynamiques spatiales*. Fréjus, France.

- CHAUDHARI, Kinjal et Ankit THAKKAR (9 oct. 2019). « A Comprehensive Survey on Travel Recommender Systems ». In : *Archives of Computational Methods in Engineering*. ISSN : 1886-1784. DOI : 10.1007/s11831-019-09363-7.
- CHIEN, Hung-Yu, Shyr-Kuen CHEN, Ching-Yang LIN, Jia-Ling YAN, Wei-Chen LIAO, Huan-Yi CHU, Kuan-Ju CHEN, Bo-Fan LAI et Yi-Ting CHEN (1^{er} déc. 2013). « Design and Implementation of ZigBee-Ontology-Based Exhibit Guidance and Recommendation System ». In : *International Journal of Distributed Sensor Networks* 9.12, p. 248535. ISSN : 1550-1477. DOI : 10.1155/2013/248535.
- CLEVELAND, Donald et Ana CLEVELAND (1983). *Introduction to Indexing and Abstracting*. Libraries Unlimited. 384 p. ISBN : 978-1-59884-976-9.
- COJUHARENCO, Irina et Dmitry RYVKIN (2008). « Peak-End rule versus average utility : How utility aggregation affects evaluations of experiences ». In : *Journal of Mathematical Psychology* 52.5, p. 326-335. ISSN : 0022-2496. DOI : 10.1016/j.jmp.2008.05.004.
- CONTRIBUTEURS D'OPENSTREETMAP (2019). *Données cartographiques d'OpenStreetMap*. URL : <https://www.openstreetmap.org>.
- CONTRIBUTEURS DU PROJET OSRM (2019). *Projet OSRM*. URL : <http://project-osrm.org>.
- DAVEY, Gareth (2005). « What is museum fatigue ». In : *Visitor Studies Today* 8.3, p. 17-21.
- DESROSIERS, Christian et George KARYPIS (2011). « A Comprehensive Survey of Neighborhood-based Recommendation Methods ». In : *Recommender Systems Handbook*. Sous la dir. de Francesco RICCI, Lior ROKACH, Bracha SHAPIRA et Paul B. KANTOR. Boston, USA : Springer US, p. 107-144. ISBN : 978-0-387-85820-3. DOI : 10.1007/978-0-387-85820-3_4.
- DIJKSTRA, E. W. (1^{er} déc. 1959). « A note on two problems in connexion with graphs ». In : *Numerische Mathematik* 1.1, p. 269-271. ISSN : 0945-3245. DOI : 10.1007/BF01386390. URL : <https://doi.org/10.1007/BF01386390>.
- EHRGOTT, Matthias (2005). *Multicriteria optimization*. T. 491. Springer Science & Business Media.
- ELLEGÅRD, Kajsa (17 sept. 2018). *Thinking Time Geography : Concepts, Methods and Applications*. Abingdon, Oxon ; New York : Routledge. 172 p. ISBN : 978-1-138-57379-6.
- ETHIS, Emmanuel (2002). *Avignon, le public réinventé*. URL : <https://halshs.archives-ouvertes.fr/halshs-00007306>.
- EXPÓSITO, Airam, Simona MANCINI, Julio BRITO et José A. MORENO (2019). « Solving a Fuzzy Tourist Trip Design Problem with Clustered Points of Interest ». In : *Uncertainty Management with Fuzzy and Rough Sets : Recent Advances and Applications*. Sous la dir. de Rafael BELLO, Rafael FALCON et José Luis VERDEGAY. Studies in Fuzziness and Soft Computing. Cham : Springer International Publishing, p. 31-47. ISBN : 978-3-030-10463-4. DOI : 10.1007/978-3-030-10463-4_2. URL : https://doi.org/10.1007/978-3-030-10463-4_2.
- FALK, John H. (2009). *IDENTITY AND THE MUSEUM VISITOR EXPERIENCE*. Left Coast Press.
- FALK, John H., John J. KORAN, Lynn D. DIERKING et Lewis DREBLOW (1985). « Predicting Visitor Behavior ». In : *Curator : The Museum Journal* 28.4, p. 249-258. ISSN : 2151-6952. DOI : 10.1111/j.2151-6952.1985.tb01753.x.
- FEILLET, Dominique, Pierre DEJAX et Michel GENDREAU (1^{er} mai 2005). « Traveling Salesman Problems with Profits ». In : *Transportation Science* 39.2, p. 188-205. ISSN : 0041-1655. DOI : 10.1287/trsc.1030.0079.

- FERNÁNDEZ, Silvia, Eric SANJUAN et Juan-Manuel TORRES-MORENO (2007). « Textual Energy of Associative Memories : Performant Applications of Enertex Algorithm in Text Summarization and Topic Segmentation ». In : *MICAI*, p. 861-871. DOI : 10.1007/978-3-540-76631-5_82.
- FILEP, Sebastian et Margaret DEERY (1^{er} déc. 2010). « Towards a Picture of Tourists' Happiness ». In : *Tourism Analysis* 15.4, p. 399-410. ISSN : 1083-5423. DOI : 10.3727/108354210X12864727453061.
- GAVALAS, Damianos, Charalampos KONSTANTOPOULOS, Konstantinos MASTAKAS et Grammati PANTZIOU (1^{er} juin 2014). « A survey on algorithmic approaches for solving tourist trip design problems ». In : *Journal of Heuristics* 20.3, p. 291-328. ISSN : 1572-9397. DOI : 10.1007/s10732-014-9242-5.
- GILMAN, Benjamin Ives (1916). « Museum Fatigue ». In : *The Scientific Monthly* 2.1, p. 62-74. ISSN : 0096-3771.
- GORRINI, Andrea et Viola BERTINI (1^{er} jan. 2018). « Walkability assessment and tourism cities : the case of Venice ». In : *International Journal of Tourism Cities* 4.3, p. 355-368. ISSN : 2056-5607. DOI : 10.1108/IJTC-11-2017-0072.
- GUNN, Clare A. et Turgut VAR (2002). *Tourism Planning : Basics, Concepts, Cases*. Psychology Press. 466 p. ISBN : 978-0-415-93268-4.
- HÄGERSTRAND, Torsten (1^{er} jan. 1970). « WHAT ABOUT PEOPLE IN REGIONAL SCIENCE? » In : *Papers in Regional Science* 24.1, p. 7-24. ISSN : 1056-8190. DOI : 10.1111/j.1435-5597.1970.tb01464.x.
- HEIDORN, P. Bryan et Qin WEI (10 sept. 2008). « Automatic Metadata Extraction From Museum Specimen Labels ». In : *International Conference on Dublin Core and Metadata Applications* 0.0, p. 57-68. ISSN : 1939-1366. URL : <http://dcpapers.dublincore.org/pubs/article/view/919>.
- HOPFIELD, J J (1982). « Neural networks and physical systems with emergent collective computational abilities ». In : *Proceedings of the National Academy of Sciences* 79.8, p. 2554-2558. ISSN : 0027-8424. DOI : 10.1073/pnas.79.8.2554.
- INTERNATIONAL BUSINESS MACHINES CORP. (2017). *IBM ILOG CPLEX Optimization Studio. CPLEX User's Manual*. Version 12.8. URL : <https://cplex.com>.
- KAHNEMAN, Daniel (1973). *Attention and effort*. T. 1063. Prentice-Hall series in experimental psychology. Prentice-Hall. 246 p. ISBN : 9780130505187.
- KAHNEMAN, Daniel (25 sept. 2000). « Evaluation by moments : Past and future ». In : *Choices, Values, and Frames*. Sous la dir. de Daniel KAHNEMAN et Amos TVERSKY. Cambridge University Press, p. 693-708. ISBN : 978-1-107-65106-7.
- KAHNEMAN, Daniel (25 oct. 2011). *Thinking, Fast and Slow*. New York : Farrar, Straus and Giroux. 499 p. ISBN : 978-0-374-27563-1.
- KEMP, Simon, Christopher D. B. BURT et Laura FURNEAUX (jan. 2008). « A test of the peak-end rule with extended autobiographical events ». In : *Memory & Cognition* 36.1, p. 132-138. ISSN : 1532-5946. DOI : 10.3758/MC.36.1.132.
- KNUTH, Donald E. (4 juil. 1997). *The Art of Computer Programming : Volume 3 : Sorting and Searching*. Addison-Wesley Professional. 1037 p. ISBN : 078-5-342-89685-5.
- KOSMOPOULOS, Dimitrios et Georgios STYLARAS (1^{er} juil. 2018). « A survey on developing personalized content services in museums ». In : *Pervasive and Mobile Computing* 47, p. 54-77. ISSN : 1574-1192. DOI : 10.1016/j.pmcj.2018.05.002.

- KRAAK, Menno-Jan (1988). « The space-time cube revisited from a geovisualization perspective ». In : *Proceedings of the 21st International Cartographic Conference (ICC) 'Cartographic Renaissance', Durban, 10-16 August*.
- KRACAUER, Siegfried (1995). *Rues de Berlin et d'ailleurs*. Gallimard. 183 p. ISBN : 978-2-07-072922-7.
- KURATA, Yohei et Tatsunori HARA (2013). « CT-Planner4 : Toward a More User-Friendly Interactive Day-Tour Planner ». In : *Information and Communication Technologies in Tourism 2014*. Sous la dir. de Zheng XIANG et Iis TUSSYADIAH. Springer International Publishing, p. 73-86. ISBN : 978-3-319-03973-2. DOI : 10.1007/978-3-319-03973-2_6.
- LAMSFUS, Carlos, Aurkene ALZUA-SORZABAL, David MARTIN, Zigor SALVADOR et Alex USANDIZAGA (2009). « Human-centric Ontology-based Context Modelling in Tourism. » In : *Mediterranean Conference on Information Systems*. 64, p. 677-690.
- LEQUIME, Julie (2010). « Le tourisme urbain et patrimonial, nouvel enjeu pour les villes. L'exemple de Toulouse ». Mém. de mast. Université de Toulouse.
- LIKERT, R. (1932). « A technique for the measurement of attitudes. » In : *Archives of Psychology* 22 140, p. 55-55.
- LIM, Kwan Hui (2016). « Personalized Recommendation of Travel Itineraries based on Tourist Interests and Preferences ». In : *Late-breaking Results, Posters, Demos, Doctoral Consortium and Workshops Proceedings of the 24th ACM Conference on User Modeling, Adaptation and Personalisation (UMAP 2016), Halifax, Canada, July 13-16, 2016*. Sous la dir. de Federica CENA, Michel DESMARAIS et Darina DICHEVA. T. 1618. CEUR Workshop Proceedings. CEUR-WS.org. URL : http://ceur-ws.org/Vol-1618/DC%5C_6.pdf.
- LOPS, Pasquale, Marco de GEMMIS et Giovanni SEMERARO (2011). « Content-based Recommender Systems : State of the Art and Trends ». In : *Recommender Systems Handbook*. Sous la dir. de Francesco RICCI, Lior ROKACH, Bracha SHAPIRA et Paul B. KANTOR. Boston, USA : Springer US, p. 73-105. ISBN : 978-0-387-85820-3. DOI : 10.1007/978-0-387-85820-3_3.
- LU, Chun, Philippe LAUBLET et Milan STANKOVIC (2016). « Travel Attractions Recommendation with Knowledge Graphs ». In : *Lecture Notes in Computer Science*. Springer International Publishing, p. 416-431. DOI : 10.1007/978-3-319-49004-5_27.
- LYKOURANTZOU, Ioanna, Xavier CLAUDE, Yannick NAUDET, Eric TOBIAS, Angeliki ANTONIOU, George LEPOURAS et Costas VASSILAKIS (2013). « Improving museum visitors' Quality of Experience through intelligent recommendations : A visiting style-based approach ». In : *Workshop Proceedings of the 9th Int. Conf. on Intelligent Environments*. Sous la dir. de Juan A. BOTÍA et Dimitris CHARITOS. T. 17. Ambient Intelligence and Smart Environments. IOS Press, p. 507-518. ISBN : 978-1-61499-285-1. DOI : 10.3233/978-1-61499-286-8-507.
- MANNING, Christopher D. et Hinrich SCHÜTZE (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA, USA : MIT Press. ISBN : 978-0-262-13360-9.
- MEEHAN, K., T. LUNNEY, K. CURRAN et A. McCAUGHEY (mar. 2013). « Context-aware intelligent recommendation system for tourism ». In : *2013 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, p. 328-331. DOI : 10.1109/PerComW.2013.6529508.
- MELTON, Arthur W. (1935). *Problems of installation in museums of art*. Washington DC : American Association of Museums. 269 p. DOI : 10.1037/11526-000.

- MINNEN, David, Tracy WESTEYN, Thad STARNER, J WARD et Paul LUKOWICZ (2006). « Performance metrics and evaluation issues for continuous activity recognition ». In : *Performance Metrics for Intelligent Systems* 4, p. 141-148.
- MITTAL, Rohin et Vaibhav SINHA (jan. 2017). « A personalized time-bound activity Recommendation System ». In : *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)*. Institute of Electrical et Electronics Engineers (IEEE). DOI : 10.1109/ccwc.2017.7868485.
- MUSÉE DE L'ORANGERIE (2015). *Musée de l'Orangerie*. URL : <https://www.musee-orangerie.fr>.
- NATIONAL GALLERY (2015). *National Gallery*. URL : <https://www.nationalgallery.org.uk>.
- NAUDET, Yannick, Ioanna LYKOURTZOU, Eric TOBIAS, Angeliki ANTONIOU, Jenny ROMPA et George LEPOURAS (2013). « Gaming and Cognitive Profiles for Recommendations in Museums ». In : *8th International Workshop on Semantic and Social Media Adaptation and Personalization, SMAP 2013, Bayonne, France, December 12-13, 2013*. IEEE, p. 67-72. DOI : 10.1109/SMAP.2013.16. URL : <https://ieeexplore.ieee.org/xpl/conhome/6734196/proceeding>.
- NAWIJN, Jeroen et Avital BIRAN (20 mar. 2018). « Negative emotions in tourism : a meaningful analysis ». In : *Current Issues in Tourism* 22.19, p. 2386-2398. ISSN : 1368-3500, 1747-7603. DOI : 10.1080/13683500.2018.1451495.
- NAWIJN, Jeroen, Ondrej MITAS, Yeqiang LIN et Deborah KERSTETTER (2013). « How do we feel on vacation? A closer look at how emotions change over the course of a trip ». In : *Journal of Travel Research* 52.2, p. 265-274.
- NIELSEN, JAKOB (1^{er} jan. 1993). « Chapter 5 - Usability Heuristics ». In : *Usability Engineering*. San Diego : Morgan Kaufmann, p. 115-163. ISBN : 978-0-12-518406-9. DOI : 10.1016/B978-0-08-052029-2.50008-5. URL : <https://doi.org/10.1016/B978-0-08-052029-2.50008-5>.
- OBERLANDER, Jon, Mick O'DONNELL, Chris MELLISH et Alistair KNOTT (1^{er} jan. 1998). « Conversation in the museum : experiments in dynamic hypermedia with the intelligent labelling explorer ». In : *New Review of Hypermedia and Multimedia* 4.1, p. 11-32. ISSN : 1361-4568. DOI : 10.1080/13614569808914693.
- OCDE, UNION EUROPÉENNE, NATIONS UNIES et ORGANISATION MONDIALE DU TOURISME (6 avr. 2017). *Compte satellite du tourisme : recommandations concernant le cadre conceptuel 2008*. OECD. 129 p. ISBN : 978-92-64-27417-4. DOI : 10.1787/9789264274174-fr.
- PARK, Jeong-Yeol et SooCheong (Shawn) JANG (2013). « Confused by too many choices? Choice overload in tourism ». In : *Tourism Management* 3, p. 1-12. ISSN : 0261-5177. DOI : 10.1016/j.tourman.2012.05.004.
- PAZZANI, Michael J. et Daniel BILLSUS (2007). « Content-Based Recommendation Systems ». In : *The Adaptive Web : Methods and Strategies of Web Personalization*. Sous la dir. de Peter BRUSILOVSKY, Alfred KOBSA et Wolfgang NEJDL. Berlin, Heidelberg : Springer Berlin Heidelberg, p. 325-341. ISBN : 978-3-540-72079-9. DOI : 10.1007/978-3-540-72079-9_10.
- PICOLOT, Delphine (2016). « Recherche et construction de parcours culturels au sein du festival OFF d'Avignon ». Mém. de mast. Université d'Avignon et des Pays de Vaucluse.
- POLI, Marie-Sylvie (15 déc. 2014). « La notion de valeur appliquée à l'évaluation muséale ». In : DAIGNAULT, Lucie et Bernard SCHIELE. *Les musées et leurs publics : Savoirs et enjeux*. Presses de l'Université du Québec, p. 105-121. ISBN : 978-2-7605-4148-1.

- PORTER, M.F. (1^{er} jan. 1980). « An algorithm for suffix stripping ». In : *Program* 14.3, p. 130-137. ISSN : 0033-0337. DOI : 10.1108/eb046814.
- POWERS, David M.W. (2011). « Evaluation : From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation ». In : *Journal of Machine Learning Technologies* 2(1), p. 37-63.
- RICCI, Francesco, Lior ROKACH et Bracha SHAPIRA (2011). « Introduction to Recommender Systems Handbook ». In : *Recommender Systems Handbook*. Sous la dir. de Francesco RICCI, Lior ROKACH, Bracha SHAPIRA et Paul B. KANTOR. Boston, USA : Springer US, p. 1-35. ISBN : 978-0-387-85820-3. DOI : 10.1007/978-0-387-85820-3_1.
- RICCI, Francesco, Lior ROKACH, Bracha SHAPIRA et Paul B. KANTOR, éd. (2011). *Recommender Systems Handbook*. Boston, USA : Springer US. 842 p. ISBN : 978-0-387-85820-3. DOI : 10.1007/978-0-387-85820-3.
- RIJSBERGEN, C. J. Van (1979). *Information Retrieval*. 2nd. Newton, MA, USA : Butterworth-Heinemann. ISBN : 0408709294.
- RODRÍGUEZ, Beatriz, Julián MOLINA, Fátima PÉREZ et Rafael CABALLERO (2012). « Interactive design of personalised tourism routes ». In : *Tourism Management* 33.4, p. 926-940. ISSN : 0261-5177. DOI : 10.1016/j.tourman.2011.09.014.
- ROES, Ivo, Natalia STASH, Yiwen WANG et Lora AROYO (2009). « A Personalized Walk Through the Museum : The CHIP Interactive Tour Guide ». In : *CHI'09 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '09. New York, NY, USA : ACM, p. 3317-3322. ISBN : 978-1-60558-247-4. DOI : 10.1145/1520340.1520479.
- ROUILLARD, José (juil. 2008). « Contextual QR Codes ». In : *2008 The Third International Multi-Conference on Computing in the Global Information Technology (iccg 2008)*. 2008 The Third International Multi-Conference on Computing in the Global Information Technology (iccg 2008), p. 50-55. DOI : 10.1109/ICCGI.2008.25.
- ROUNDS, Jay (2004). « Strategies for the curiosity-driven museum visitor ». In : *Curator* 47.4, p. 389-412.
- RUOTSALO, Tuukka, Krister HAAV, Antony STOYANOV, Sylvain ROCHE, Elena FANI, Romina DELIAI, Eetu MÄKELÄ, Tomi KAUPPINEN et Eero HYVÖNEN (1^{er} mai 2013). « SMART-MUSEUM : A mobile recommender system for the Web of Data ». In : *Journal of Web Semantics* 20, p. 50-67. ISSN : 1570-8268. DOI : 10.1016/j.websem.2013.03.001.
- SALTON, G., A. WONG et C. S. YANG (nov. 1975). « A Vector Space Model for Automatic Indexing ». In : *Communication of the ACM* 18.11, p. 613-620. ISSN : 0001-0782. DOI : 10.1145/361219.361220.
- SANTOS, Filipe, Ana ALMEIDA, Constantino MARTINS, Ramiro GONÇALVES et José MARTINS (31 oct. 2017). « Using POI functionality and accessibility levels for delivering personalized tourism recommendations ». In : *Computers, Environment and Urban Systems* 77. ISSN : 0198-9715. DOI : 10.1016/j.compenvurbsys.2017.08.007.
- SAUSSURE, Ferdinand de (1916). *Cours de linguistique générale*. Sous la dir. de Charles BALLY et Albert SECHEHAYE. Paris : Payot.
- SCHAFER, J. Ben, Dan FRANKOWSKI, Jon HERLOCKER et Shilad SEN (2007). « Collaborative Filtering Recommender Systems ». In : *The Adaptive Web : Methods and Strategies of Web Personalization*. Sous la dir. de Peter BRUSILOVSKY, Alfred KOBZA et Wolfgang NEJDL. Lecture Notes in Computer Science. Berlin, Heidelberg : Springer, p. 291-324. ISBN : 978-3-540-72079-9. DOI : 10.1007/978-3-540-72079-9_9.

- SCHOUTEN, Frans (1^{er} sept. 1987). « Psychology and exhibit design : A note ». In : *International Journal of Museum Management and Curatorship* 6.3, p. 259-262. ISSN : 0260-4779. DOI : 10.1080/09647778709515075.
- SCHOUTEN, Frans (1^{er} juin 1995). « Improving visitor care in heritage attractions ». In : *Tourism Management* 16.4, p. 259-261. ISSN : 0261-5177. DOI : 10.1016/0261-5177(95)00014-F.
- SCHRIJVER, Alexander (1998). *Theory of Linear and Integer Programming*. Wiley Series in Discrete Mathematics & Optimization. John Wiley & Sons. ISBN : 978-0-471-98232-6.
- SCHWARTZ, Barry (18 jan. 2005). *The Paradox of Choice : Why More Is Less, Revised Edition*. Harper Collins. 304 p. ISBN : 978-0-06-000569-6.
- SERRELL, Beverly (1997). « Paying Attention : The Duration and Allocation of Visitors' Time in Museum Exhibitions ». In : *Curator : The Museum Journal* 40.2, p. 108-125. ISSN : 2151-6952. DOI : 10.1111/j.2151-6952.1997.tb01292.x. URL : <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2151-6952.1997.tb01292.x>.
- SIMON, Gwendal (29 nov. 2010). « Pratiques touristiques dans la métropole parisienne : une analyse des mouvements intra-urbains ». Thèse de doct. Université Paris-Est.
- SPARCK JONES, Karen (1^{er} jan. 1972). « A statistical interpretation of term importance in automatic indexing ». In : *Journal of Documentation* 28.1, p. 11-21. ISSN : 0022-0418. DOI : 10.1108/eb026526.
- SPECK, Jeff (2013). *Walkable city : How downtown can save America, one step at a time*. macmillan.
- STOCK, Oliviero (1993). « Intelligent Multimedia Interfaces ». In : sous la dir. de Mark T. MAYBURY. Menlo Park, CA, USA : American Association for Artificial Intelligence. Chap. ALFRESCO : Enjoying the Combination of Natural Language Processing and Hypermedia for Information Exploration, p. 197-224. ISBN : 0-262-63150-4.
- SU, Xing, Hanghang TONG et Ping JI (juin 2014). « Activity recognition with smartphone sensors ». In : *Tsinghua Science and Technology* 19.3, p. 235-249. DOI : 10.1109/tst.2014.6838194.
- THÉVENIN, Thomas, Sonia CHARDONNEL et Élodie COCHEY (2007). « Explorer les temporalités urbaines de l'agglomération de Dijon : Une analyse de l'Enquête-Ménage-Déplacement par les programmes d'activités ». In : *Espace populations sociétés. Space populations societies* 2007/2-3, p. 179-190.
- TORRES-MORENO, Juan-Manuel (2014). *Automatic Text Summarization*. Wiley-ISTE. ISBN : 978-1-84821-668-6.
- VANSTEENWEGEN, Pieter et Dirk VAN OUDHEUSDEN (1^{er} juil. 2007). « The Mobile Tourist Guide : An OR Opportunity ». In : *OR Insight* 20.3, p. 21-27. ISSN : 1759-0477. DOI : 10.1057/ori.2007.17. URL : <https://doi.org/10.1057/ori.2007.17>.
- VÉRON, Eliséo et Martine LEVASSEUR (1983). *Ethnographie de l'exposition : l'espace, le corps et le sens*. Paris : Centre Georges Pompidou. ISBN : 978-2-902706-01-3.
- WANG, Kuo-Ching, Po-Chen JAO, Hsi-Chen CHAN et Chia-Hsun CHUNG (jan. 2010). « GROUP PACKAGE TOUR LEADER'S INTRINSIC RISKS ». In : *Annals of Tourism Research* 37.1, p. 154-179. DOI : 10.1016/j.annals.2009.08.004.
- WASTIAUX, Michel (mar. 2009). « La qualité d'usage des espaces publics en France ». In : *Colloque Tourisme urbain patrimoine et qualité urbaine en Europe*. Rennes, France, p. 40-43.
- WÖRNDL, Wolfgang, Alexander HEFELE et Daniel HERZOG (1^{er} mar. 2017). « Recommending a sequence of interesting places for tourist trips ». In : *Information Technology & Tourism*

17.1, p. 31-54. ISSN : 1943-4294. DOI : 10 . 1007 / s40558 - 017 - 0076 - 5. URL : <https://doi.org/10.1007/s40558-017-0076-5>.

ZANCANARO, Massimo, Tsvi KUFLIK, Zvi BOGER, Dina GOREN-BAR et Dan GOLDWASSER (2007). « Analyzing Museum Visitors' Behavior Patterns ». In : *User Modeling 2007*. Sous la dir. de Cristina CONATI, Kathleen MCCOY et Georgios PALIOURAS. Berlin, Heidelberg : Springer Berlin Heidelberg, p. 238-246. ISBN : 978-3-540-73078-1.

Publications personnelles

MATHIAS, Mayeul, Assema MOUSSA, Juan-Manuel TORRES-MORENO, Fen ZHOU, Marie-Sylvie POLI, Didier JOSSELIN, Marc EL-BÈZE, Andréa CARNEIRO LINHARES et Françoise RIGAT (sept. 2014). « Optimisation using Natural Language Processing : Personalized Tour Recommendation for Museums ». In : *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems*. T. 2. Annals of Computer Science and Information Systems. IEEE, p. 439-446. DOI : 10 . 15439 / 2014F336.

MATHIAS, Mayeul, Fen ZHOU, Juan-Manuel TORRES-MORENO, Didier JOSSELIN, Marie-Sylvie POLI et Andréa Carneiro LINHARES (25 sept. 2016). « Personalized sightseeing tours : a model for visits in art museums ». In : *International Journal of Geographical Information Science* 31.3, p. 591-616. ISSN : 1365-8816. DOI : 10 . 1080 / 13658816 . 2016 . 1233332.

MATHIAS, Mayeul, Camille BÉGUIN, Juan-Manuel TORRES-MORENO, Didier JOSSELIN, Delphine PICOLOT, Fen ZHOU et Marie-Sylvie POLI (14 juil. 2017). « Quality of Experience for Personalized Sightseeing Tours : Studies and Proposition for an Evaluation Method ». In : *Computational Science and Its Applications – ICCSA 2017*. Lecture Notes in Computer Science. Springer International Publishing, p. 380-395. ISBN : 978-3-319-62398-6. DOI : 10 . 1007 / 978 - 3 - 319 - 62398 - 6 _ 27.

Recommandation de parcours culturels personnalisés :

Étude interdisciplinaire de la proposition automatique de visites

Par

Mayeul MATHIAS

Équipe d'encadrement

- M. **Juan-Manuel TORRES-MORENO** Maître de Conférences HDR, LIA, Avignon Université
- M. **Didier JOSSELIN** Directeur de recherches, CNRS, UMR ESPACE
- M. **Fen ZHOU** Maître de Conférences HDR, CERI-SN, IMT Lille Douai