
Recommandation d'itinéraires

Recommandation et valorisation d'objets patrimoniaux

L. Rajaonarivo ¹, A. Fonteles ², C. Sallaberry ¹, P. Roose ¹,
MN. Bessagnet ¹, P. Etcheverry ¹, A. Lacayrelle ¹,
C. Marquesuzaà ¹, C. Cayere ¹, Q. Coudert ¹

1. Université de Pau et des Pays de l'Adour, LIUPPA, France

prenomcomplet.nom@univ-pau.fr

2. Indiana Wesleyan University, USA

prenomcomplet.nom@indwes.edu

RÉSUMÉ. Nous présentons un travail qui, dans un contexte de promotion du tourisme dans les Pyrénées, vise à valoriser les atouts méconnus relatifs au patrimoine culturel matériel (architecture et mobilier) et au patrimoine culturel immatériel (traditions, événements sociaux). L'objectif est double. Il s'agit tout d'abord de constituer une base de données patrimoniale fédérant des données géoréférencées hétérogènes puis, de proposer des services de diffusion de ces données dans le cadre de différents scénarios de valorisation. Ces services permettent notamment d'alimenter des applications touristiques Web et mobiles. Nous allons présenter des propositions visant la recommandation de points d'intérêts (POI) patrimoniaux dans des scénarios d'usage touristique.

ABSTRACT. In a context of valorization of the tourism in the « Pyrénées » area in France, we present a research work that aims at the use of unknown assets related to the tangible cultural heritage (architectural and furniture) as well as the intangible cultural heritage (traditions, social events). The goal is twofold. First of all, we have to design and to build a cultural heritage oriented database federating heterogeneous georeferenced data. Then, we have to propose data dissemination services according to various valorization scenarios. These services allow to supply Web and mobile tourism applications. The paper focuses on the recommendation of cultural heritage Points of Interest (POIs) within touristic scenarios.

MOTS-CLÉS : système de recommandation hybride, point d'intérêt (POI), contexte, génération d'itinéraires

KEYWORDS: hybrid recommender system, point of interest (POI), context, travel itinerary generation

1. Introduction

Le projet de recherche Européen FEDER TCVPYR (2017-2020) a pour objectif un inventaire du patrimoine bâti et du patrimoine culturel immatériel de la villégiature et du thermalisme dans le massif pyrénéen français. Il vise le développement de la connaissance et de la fréquentation des Pyrénées à travers ses atouts patrimoniaux. Ce riche patrimoine qui reste à inventorier, est autant architectural que paysager et culturel. Dans le cadre de ce projet pluridisciplinaire regroupant entre autres historiens, géographes, anthropologues, pour la partie informatique, l'objectif est double : (1) constituer une base de données fédérant des points d'intérêts (POI) patrimoniaux référencés par des experts, et (2) proposer des services de valorisation de ces POI dans le cadre de différents scénarios de valorisation. Ces services permettent (i) d'exporter tout ou partie des POI vers l'Open Data afin de les rendre accessibles à tous et/ou (ii) d'alimenter des applications touristiques. Nous avons identifié deux principaux challenges. Le premier réside dans la gestion de l'hétérogénéité des différentes sources de données que nous utilisons. Par exemple, les chargés d'inventaire (dépendants des Conseils Régionaux) utilisent actuellement deux logiciels différents (RenabLP2 et Gertrude¹), s'appuyant respectivement sur un modèle de données spécifique pour stocker les données du patrimoine. De manière similaire, pléthore de modèles sont utilisés dans le monde de l'Open Data (Foursquare², Datatourisme³, Wikivoyage⁴, WikiData⁵). Pour répondre à ce premier challenge, nous avons proposé un modèle unifié permettant d'intégrer toutes ces données (Bessagnet *et al.*, 2018) et (Fonteles *et al.*, 2018). Le second challenge sur lequel ce papier se focalise est lié à la valorisation des données du patrimoine stockées dans notre base de données unifiée. Nous proposons un ensemble d'algorithmes dédiés à la génération d'itinéraires « contextualisés » (i.e., des séquences de POI à visiter) adaptés aux utilisateurs d'une application mobile. Cette recommandation d'itinéraires prendra en compte le profil de l'utilisateur, sa position, les caractéristiques de son appareil ainsi que d'autres paramètres tels que le temps disponible pour effectuer la visite des POI ou le moyen de transport. Une recommandation pourra également être enrichie par des informations extraites de l'Open Data. Nous proposons une approche basée sur la notion de recommandation hybride. L'originalité de cette approche consiste à rapprocher des POI déjà parcourus par le touriste de ceux parcourus par d'autres touristes afin de mettre en exergue et d'évaluer des POI candidats selon le principe « un autre touriste qui a visité les mêmes POI que vous a aussi visité les suivants ». Ce principe est mis en œuvre dans notre approche grâce à l'intégration d'un algorithme d'optimisation de colonies de fourmis (*ACO: Ant Colony Optimization*) largement utilisés dans la recommandation

1. RenabLP2 et Gertrude sont des outils de création et de gestion de dossiers électroniques appliqués à la documentation de l'inventaire du patrimoine culturel. RenabLP2 est utilisé en région Occitanie et Gertrude l'est en région Nouvelle Aquitaine et Occitanie.

2. <https://fr.foursquare.com/>, visité le 04/02/2019

3. <http://www.datatourisme.fr/>, visité le 04/02/2019

4. <https://www.wikivoyage.org/>, visité le 04/02/2019

5. <https://www.wikidata.org/>, visité le 04/02/2019

Recommandation d'itinéraires

sur le commerce électronique (Lu, Guo, 2016)(Zhang, Pang, 2015)(Minjing *et al.*, 2017) mais peu courants pour la recommandation de POI.

Dans la section 2, nous présenterons les travaux connexes dans les domaines des systèmes de recommandation et de la génération d'itinéraires. La section 3 précisera les modèles et les algorithmes que nous avons élaborés dans le cadre d'une application touristique. Nous y détaillons notre démarche de recommandation de POI et de génération d'itinéraires ainsi que les premiers résultats de nos expérimentations.

2. État de l'art

Dans ce qui suit nous considérons que modèle utilisateur/touriste et profil utilisateur/touriste sont synonymes. Ainsi, un profil comprend généralement les informations qui caractérisent l'utilisateur (genre, catégorie d'âge, activités socio-professionnelles), ses préférences (préférences thématiques et historiques) et/ou son groupe d'appartenance (enfant, adulte, personne âgée). Certaines de ces informations sont fournies directement par l'utilisateur (par exemple : genre, catégorie d'âge, préférences), d'autres peuvent être déduites de ses interactions (par exemple ses « J'aime », son groupe d'appartenance). Les préférences peuvent être déduites des notes que l'utilisateur attribue aux POI ou des commentaires qu'il fait. Selon les POI que l'utilisateur a visités, il est également possible de déduire à quel groupe d'utilisateurs il appartient. Dans (Kesorn *et al.*, 2017), par exemple, le profil utilisateur est caractérisé par son nom, son lieu de travail, son affiliation, son groupe d'appartenance (amis sur *Facebook*) et ses préférences sont déduites des POI que lui ou ses amis ont visités et partagés sur les réseaux sociaux (*Facebook*). Les approches proposées dans (Aliannejadi *et al.*, 2016) et (Logesh *et al.*, 2018), quant à elles, considèrent uniquement les préférences de l'utilisateur. Les travaux cités ici exploitent les commentaires partagés par des membres de réseaux sociaux (Yelp⁶, TripAdvisor⁷ et Foursquare⁸). Ils utilisent un système d'apprentissage pour déterminer, via ses commentaires, si un utilisateur a apprécié ou non une visite.

Un itinéraire est formé par une liste ordonnée de POI. L'itinéraire indique l'ordre de visite des POI et la durée estimative de visite de chaque POI et du trajet entre POI. Un système de génération d'itinéraires aide l'utilisateur à préparer son voyage. La majorité des systèmes de génération d'itinéraires se basent sur la résolution du *Orienteering Problem* (OP) qui est un système de *scoring* et trouve son origine dans la littérature de recherche opérationnelle. L'OP a été développé pour la première fois par Tsiligirides (Tsiligirides, 1984). Dans notre cas, l'OP consiste pour un touriste à collecter des points qui sont assignés à des POI. L'objectif est de maximiser le nombre de points collectés tout en respectant des contraintes (par exemple la durée de parcours de l'ensemble de l'itinéraire).

6. <https://www.yelp.com/writeareview/>

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

8. <https://fr.foursquare.com/>

Plusieurs travaux étendent l'*OP*, par exemple : TOP, TDOP et TDTOPTW où T signifie *Team* - équipe (prise en compte d'une équipe, chaque POI doit être visité par, au maximum, un membre de l'équipe), TD signifie *Time-dependent* (prise en compte du temps de déplacement entre les POI), et TW signifie *Time Window* (prise en compte des horaires d'ouverture et de fermeture des visites des POI). Le processus de génération d'itinéraires est souvent composé de deux phases : le *scoring* des POI et la construction d'itinéraires. Certaines approches intègrent une troisième phase qui est l'adaptation de l'itinéraire généré. Pour cela, l'utilisateur peut intervenir pour modifier l'itinéraire (ajouter ou enlever un POI, changer l'ordre de visite de POI) et le système doit être capable de s'adapter à cette modification en respectant le contexte utilisateur comme par exemple la durée de visite (Garcia *et al.*, 2010). Les approches de recommandation de POI peuvent être classées en trois catégories décrites ci-dessous : les approches basées sur le contenu, les approches basées sur le filtrage collaboratif et les approches hybrides.

Les approches basées sur le contenu utilisent des informations sur les POI et/ou sur l'utilisateur afin de faire correspondre les caractéristiques des POI aux préférences et au contexte de l'utilisateur. Plusieurs techniques peuvent être utilisées pour définir cette correspondance : les techniques d'activation-propagation (Bahramian *et al.*, 2017a), les systèmes à base d'agents (Costa *et al.*, 2012), les approches probabilistes (Costa *et al.*, 2012), ou encore les approches inspirées de la biologie (basées sur les réseaux de neurones (Bahramian *et al.*, 2017b) par exemple). De plus, (Kłopotek, 2009) considère que les systèmes de recommandation s'exposent à différents types de problèmes de démarrage à froid : les problèmes liés aux nouveaux utilisateurs (qui n'ont pas d'historique et/ou de profils), ceux liés aux nouveaux POI de la base de données (qui n'ont reçu aucune évaluation), enfin, les problèmes liés aux utilisateurs spécifiques (qui ont des préférences différentes des autres). Ainsi, les approches basées sur le contenu ne s'exposent pas aux problèmes de démarrage à froid dans le cas de l'ajout de nouveaux POI.

Les approches basées sur le filtrage collaboratif utilisent les informations concernant d'autres utilisateurs (âge, activités socio-professionnelles et/ou retours d'évaluation de visite) pour recommander des POI à un utilisateur donné. Plusieurs techniques sont utilisées par ces approches telles que les techniques de *clustering* permettant, par exemple, de grouper les utilisateurs selon leur profil (Wang *et al.*, 2013) ou selon leurs liens d'amitiés sur les réseaux sociaux, ou encore, les algorithmes de type « *colonies de fourmis* » dédiés au calcul de traces de phéromones laissées au passage des différents utilisateurs (Dennouni *et al.*, 2018). Dans ce dernier cas, la recommandation tient compte uniquement de la popularité d'un POI. L'avantage des approches basées sur le filtrage collaboratif est qu'elles permettent de recommander des POI sans avoir besoin d'informations précises sur ces derniers. Toutefois, le démarrage à froid est un véritable défi dans cette catégorie d'approches puisqu'il est nécessaire d'obtenir des informations relatives aux autres utilisateurs.

Les approches hybrides combinent les approches précédentes. Différentes formes de combinaison ont été envisagées (Bartolini *et al.*, 2016) : (i) mise en œuvre séparée des approches basées sur le contenu et sur le filtrage collaboratif puis combinaison des résultats de recommandation, (ii) intégration de quelques traitements basés sur

Recommandation d'itinéraires

le contenu dans une approche collaborative, (iii) intégration de quelques traitements basés sur une approche collaborative dans une approche basée sur le contenu, (iv) proposition d'une approche générale qui intègre à la fois des traitements basés sur le contenu et sur une approche collaborative. Dans la première catégorie, nous pouvons citer les approches de (Kesorn *et al.*, 2017), (Aliannejadi *et al.*, 2016) et (De Pessemier *et al.*, 2015). Elles se différencient par les techniques de traitement de contenu et de prise en compte du filtrage collaboratif. Elles utilisent une fonction d'agrégation pour combiner les résultats de recommandation : (Kesorn *et al.*, 2017) et (Aliannejadi *et al.*, 2016) utilisent une fonction linéaire tandis que (De Pessemier *et al.*, 2015) utilisent une fonction de prédiction. L'approche que nous proposons se situe dans cette quatrième catégorie des approches hybrides. A chaque itération, nous recalculons le potentiel d'un POI candidat, c'est-à-dire son score de pertinence global, en tenant compte des POI déjà intégrés dans l'itinéraire en cours de construction.

Dans ce cadre, pour mettre en valeur les POI du patrimoine pyrénéen, tout en évitant les problèmes de démarrage à froid et en nous approchant le mieux possible des préférences des utilisateurs, nous envisageons de mettre en œuvre une approche hybride qui intègre à la fois des traitements basés sur le contenu et sur une approche collaborative. Cette approche nous permettra de composer une liste de POI pertinents tout en suggérant avec parcimonie un ou plusieurs POI originaux, en complément de ceux répondant au plus près aux préférences de l'utilisateur.

3. De la valorisation à la recommandation de données patrimoniales

Afin de collecter les informations patrimoniales, les chercheurs en SHS⁹ recueillent sur le terrain les informations qu'ils enregistrent par le biais de deux logiciels : RenabLP2 et Gertrude. Ces deux logiciels permettent d'exporter les données collectées au format XML. Cependant, les exports de RenabLP2 et de Gertrude sont très différents aussi bien au niveau de la structuration, de l'organisation que du nommage des informations. Pour exploiter ces données hétérogènes dans un cadre dédié au tourisme, nous avons conçu un modèle de données unifié et homogène (Bessagnet *et al.*, 2018) et (Fonteles *et al.*, 2018). Un premier prototype d'application fédèrent ces données est actuellement disponible sur <http://tcvpyr.iutbayonne.univ-pau.fr>.

Il convient à présent de valoriser ces informations patrimoniales. Nous proposons deux types de valorisation de POI patrimoniaux : via des ressources libres de type *Open Data* (voir détails dans (Bessagnet *et al.*, 2018)) et via une application mobile dont l'objectif sera de proposer des itinéraires de visite personnalisés, à destination d'un large public. Dans ce papier nous développerons davantage ce dernier point.

Une première particularité de notre approche réside dans l'utilisation de connaissances *a priori* : un thésaurus de classification de concepts patrimoniaux nous permet de catégoriser aussi bien les POI (point de vue des experts) que les préférences des utilisateurs (point de vue de l'utilisateur et/ou de l'application de recommandation). Nous disposons d'un thésaurus expert du domaine patrimonial qui comporte 4007 concepts

9. Sciences Humaines et Sociales

et 7946 liens. Ces connaissances permettent de déterminer des liens sémantiques entre POI, entre utilisateurs et entre POI et utilisateurs. Nous pouvons, par exemple, calculer la proximité sémantique entre un POI et des préférences utilisateur en mesurant leur similarité sur la base des connaissances modélisées sur ce thésaurus patrimonial. La seconde particularité de notre approche réside dans l'exploitation de traces d'interactions des utilisateurs (POI visités, notes, enchaînement de visites). Nous nous inspirons des approches de recommandation des produits commerciaux selon lesquelles « un client qui a consulté ce produit a aussi consulté ceux-ci ».

3.1. Modèles

Afin de préparer la valorisation du patrimoine via une application mobile, plusieurs modèles ont été définis.

3.1.1. Modèle Utilisateur

Le modèle utilisateur permet de décrire les éléments caractérisant les utilisateurs de notre application : genre, catégorie d'âge et préférences. Un utilisateur peut avoir plusieurs préférences thématiques (par exemple : parc, musée) et historiques (par exemple : XV^e siècle). La figure 1 illustre notre modèle utilisateur.

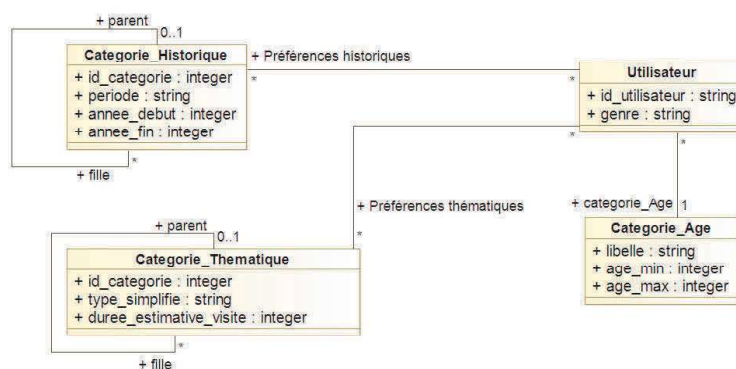


Figure 1. Modèle Utilisateur

3.1.2. Modèle de Contexte

Le contexte est composé de trois facettes : utilisateur, physique et ressource (voir figure 2). Le contexte utilisateur vient compléter le profil utilisateur avec des informations propres à une visite donnée. Ainsi, il permet de connaître le moyen de déplacement qui sera utilisé, la date, la durée et la zone géographique (ville qui doit être précisée par l'utilisateur s'il anticipe la préparation de sa visite sinon sa localisation courante). Le contexte physique décrit la localisation de l'utilisateur lors de sa visite. Cette facette change en temps réel en fonction du déplacement de l'utilisateur dans sa

Recommandation d'itinéraires

zone de visite. Le contexte ressource regroupe les informations liées notamment à son appareil mobile (niveau de batterie) et à son environnement (qualité de la connectivité et de la bande passante). Ces deux dernières facettes seront utilisées au moment du parcours de l'itinéraire pour adapter ce dernier si besoin.

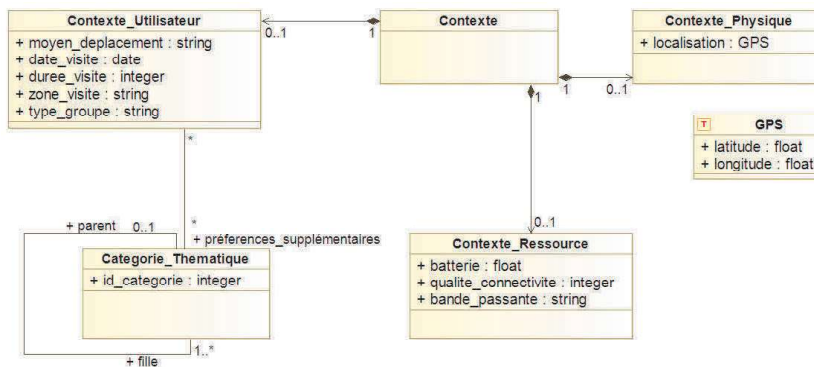


Figure 2. Modèle de Contexte

3.1.3. Modèle d'Itinéraire

Un itinéraire est une succession d'étapes dans lesquelles figurent les POI à visiter. Il est défini par sa date de génération, le moyen de déplacement utilisé et l'ensemble des étapes à enchaîner pendant la visite. Chaque étape désigne un ou plusieurs POI correspondant à l'étape. De même, une étape pointe l'étape précédente et/ou la suivante. La figure 3 illustre ce modèle. Les informations sur la durée estimative de chaque déplacement et de l'ensemble de la visite sont calculées avec l'itinéraire. Tous ces modèles sont utilisés dans notre démarche de génération d'itinéraires.

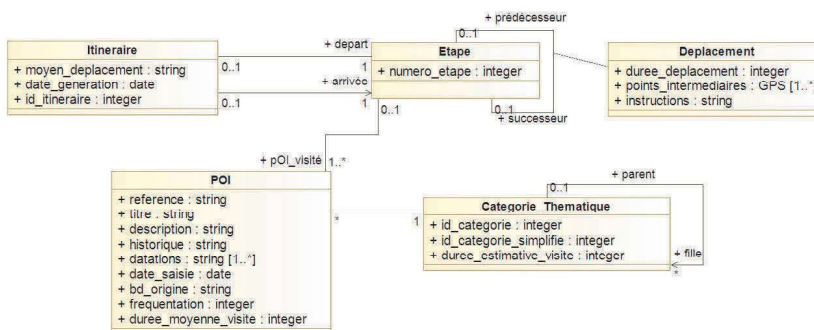


Figure 3. Modèle d'Itinéraire

3.2. Recommandation de POI pour la génération d'itinéraires

Notre approche de génération d'itinéraires fait partie des approches TDOP (voir section 2) car nous considérons le temps de déplacements entre les POI.

Algorithme 1 : Génération d'itinéraires - pseudo-code

Input : Ensemble des POI

Output : L'itinéraire composé d'une liste de POI

- 1 Filtre des POI de la zone (Ensemble des POI) : **Ensemble des POI candidats**
 - 2 *Scoring* de POI (Ensemble des POI candidats) : **Ensemble des POI candidats avec leur score**
 - 3 Sélection du POI de départ (Ensemble des POI candidats) : **POI de départ**
 - 4 Ajout d'un POI dans l'itinéraire (POI de départ) : **Itinéraire en cours de construction**
 - 5 **while** (*Il reste au moins un POI candidat ET Il reste du temps*) **do**
 - 6 *Scoring* de POI (Ensemble de POI candidats restants) : **Ensemble des POI candidats avec leur score**
 - 7 Sélection du POI suivant (Ensemble de POI candidats) : **POI suivant**
 - 8 Ajout d'un POI dans l'itinéraire (POI suivant) : **Itinéraire en cours de construction**
-

La première étape de notre approche (ligne 1, algorithme 1) permet de filtrer les POI figurant dans la zone de visite de l'utilisateur. La deuxième étape supporte le *scoring* de l'ensemble des POI candidats obtenus lors du filtre (cette fonction de *scoring* est décrite dans la formule 1 de la section 3.2.1). Elle nécessite les informations relatives aux POI et les informations utilisateur (profil et contexte). La troisième étape, quant à elle, vise la sélection du POI de départ parmi cet ensemble : par défaut, le POI de score le plus élevé. Le POI de départ est ensuite ajouté dans l'itinéraire en cours de construction. Les itérations suivantes (ligne 5, algorithme 1) visent le *scoring* des POI restants suivi de l'ajout du POI de score le plus élevé dans l'itinéraire. Nous relançons le processus de *scoring* de POI candidats à chaque nouvelle itération afin d'intégrer le paramètre de temps de trajet et celui, plus original, de popularité par rapport au dernier POI de l'itinéraire en cours de construction. Cette forme de popularité prendra notamment en compte le fait que plusieurs utilisateurs qui ont visité le dernier POI ont aussi visité le POI candidat. L'itération s'arrête quand il n'y a plus de POI candidat ou bien quand il ne reste plus de temps suffisant pour poursuivre la visite. A la fin de ces étapes, nous obtenons un itinéraire qui est défini par une liste ordonnée de POI avec les durées estimatives de visite et de trajet entre les POI.

3.2.1. Entrées/sorties de données pour le *scoring* de POI

La génération automatique d'un itinéraire pour un utilisateur particulier nécessite le *scoring* des POI. C'est au niveau de cette fonction de *scoring* que se distinguent les différentes approches de génération d'itinéraires. Notre fonction de *scoring* (voir formule 1) correspond à la somme pondérée de trois valeurs : la pertinence du POI par rapport aux préférences de l'utilisateur ($P_{préférence}$), la pertinence du POI en

Recommandation d'itinéraires

terme de trajet et de durée de visite ($P_{temporelle}$) et la pertinence du POI selon le comportement des autres utilisateurs ($P_{sociale}$). Cette fonction est appelée aux étapes 2 et 6 de l'algorithme 1.

$$P(O_e, t) = \alpha * P_{preference}(O_e) + \beta * P_{temporelle}(O_e, t) + \gamma * P_{sociale}(O_e, t) \quad (1)$$

avec $\alpha + \beta + \gamma = 1$, où O_e est le POI évalué et t est le moment de calcul de *scoring*. Les valeurs des coefficients α , β et γ sont également réparties à l'initialisation et, ensuite, peuvent évoluer selon le poids que l'utilisateur souhaite donner à ses préférences, son temps de parcours ou encore à la popularité des POI.

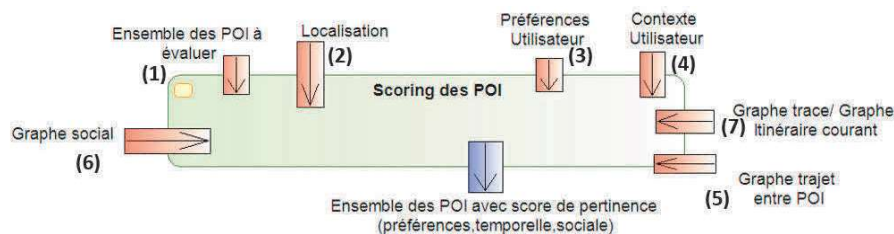


Figure 4. Données en entrée/sortie pour le scoring des POI

La fonction de *scoring* prend en entrée (voir figure 4) :

1. l'ensemble des POI à évaluer ;
2. la localisation correspondant à la position du dernier POI de l'itinéraire en cours de construction ;
3. les préférences utilisateurs composées des préférences thématiques notées $PREF_T$ et des préférences historiques notées $PREF_H$;
4. le contexte utilisateur (mode de déplacement, date, durée, zone de visite et type de groupe (avec des enfants, entre adultes ou avec des personnes âgées)) ;
5. le graphe G_{trajet} contenant les informations sur le temps de trajet entre les POI à disposition ;
6. le graphe G_{social} contenant les informations sur les évaluations de POI données par les utilisateurs ainsi que les informations relatives à leur trace de passage ;
7. la liste des POI dans l'itinéraire courant.

Le graphe G_{trajet} est défini par $G_{trajet} = \langle V, E \rangle$ où V est un ensemble de POI et E est un ensemble de trajets permettant de lier deux POI. Dans notre approche, nous considérons que tous les POI sont liés deux à deux entre eux. La définition des valeurs de ce graphe est faite une seule fois sauf s'il y a des mises à jour telles qu'un ajout de POI ou des changements de durée de trajet entre certains POI. La figure 5(a) illustre le graphe de trajets entre POI. Chaque arête a un triplet de valeurs : d_{ij}^1 (durée de trajet à pied entre POI_I et POI_J), d_{ij}^2 (durée de trajet en vélo entre POI_I et POI_J) et d_{ij}^3 (durée de trajet en voiture entre POI_I et POI_J).

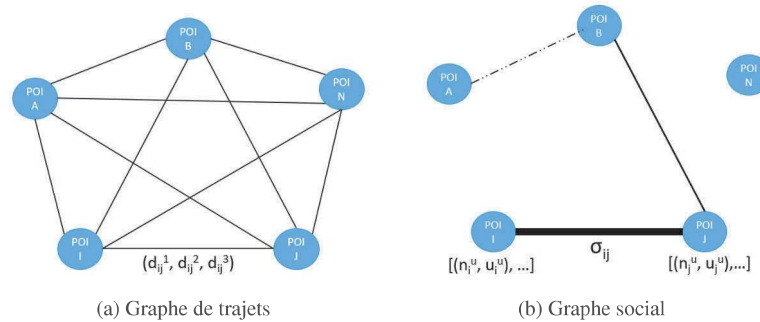


Figure 5. Graphes de données

Le graphe G_{social} est défini par $G_{social} = \langle V, E \rangle$ où V est l'ensemble des POI et E est un ensemble de trajets réalisés entre deux POI. La figure 5(b) illustre les informations sur les traces de l'ensemble des utilisateurs du système. Chaque nœud possède une liste d'évaluations dont chaque élément est une information donnée par un utilisateur. Chaque élément est formé d'un couple de valeurs : n_i^u (note donnée par l'utilisateur u au POI_I) et u_i^u (identité de l'utilisateur u qui a visité le POI_I). Chaque arête possède une valeur qui représente l'intensité du lien entre deux POI selon les interactions des autres utilisateurs. Nous adoptons l'hypothèse que si deux POI sont fréquemment visités successivement, ils présentent un lien fort entre eux selon les utilisateurs. Pour mettre en œuvre cette hypothèse, nous nous inspirons des approches basées sur la colonie de fourmis pour lesquelles un utilisateur est représenté par une fourmi. A chaque fois qu'une fourmi passe d'un POI à un autre, elle dépose une quantité de phéromones sur son trajet. Lorsque la quantité de phéromones entre deux POI est élevée, les deux POI présentent un lien fort entre eux. Sur la figure 5(b), les phéromones sont représentées par des traits. Plus l'épaisseur de l'arête est grande, plus le taux de phéromones sur cette arête est élevé. L'avantage des approches de type « algorithmes de fourmis » est la prise en compte du nombre de passages mais également de l'éventuelle absence de passage pendant une longue durée.

Le flux de sortie obtenu à l'issue de l'exécution de la fonction de *scoring* est composé du même ensemble de POI initialement proposé en entrée. A chaque POI est désormais associé un score de pertinence.

Nous rappelons que la fonction de *scoring* est composée des trois sous-fonctions $P_{preference}$, $P_{temporelle}$ et $P_{sociale}$ (formule 1) détaillées dans la section 3.2.2.

3.2.2. Fonctions dédiées au scoring de POI

1. Pertinence par rapport aux préférences ($P_{preference}$)

Nous adoptons l'hypothèse que si l'utilisateur s'intéresse à une catégorie thématique ou à une catégorie de période historique alors il peut s'intéresser aux catégories sémantiquement proches de ses catégories. La fonction de pertinence en terme de pré-

Recommandation d'itinéraires

férences mesure la similarité sémantique entre les catégories thématiques et/ou historiques auxquelles appartient le POI à évaluer et les catégories thématiques et/ou historiques définies comme préférences de l'utilisateur.

$$P_{preference}(O_e) = \mu * Sim_{thema}(O_e, PREF_T) + \nu * Sim_{histo}(O_e, PREF_H) \quad (2)$$

avec $\mu + \nu = 1$

Les Sim_{thema} et Sim_{histo} mesurent la similarité sémantique entre un POI et les préférences de l'utilisateur conformément à leurs caractéristiques thématique et historique respectives. Nous utilisons la mesure de similarité proposée par Wu et Palmer (Wu, Palmer, 1994) incluant la distance entre deux concepts dans un thésaurus de référence et la profondeur de leur concept parent en commun le plus proche (*LCS : Least Common Subsumers*).

2. Pertinence temporelle ($P_{temporelle}$)

La pertinence temporelle est définie à partir du temps de trajet ($d_{(O_d, O_e)}^i$) entre le dernier POI (O_d) de l'itinéraire en cours de construction et le POI candidat (O_e). Elle dépend également du temps restant pour faire la visite ($duree_restante$). La formule 3 définit la pertinence temporelle d'un POI.

$$P_{temporelle}(O_e, t) = \frac{duree_restante(t) - d_{(O_d, O_e)}^i}{duree_restante(t)} \quad (3)$$

où i est le moyen de déplacement de l'utilisateur et $d_{(O_d, O_e)}^i$ est la durée du trajet entre O_d et O_e à l'instant t . Ainsi, le score de pertinence temporelle d'un POI est d'autant plus élevé (proche de 1) que la durée du trajet estimée pour rejoindre le POI candidat est faible.

3. Pertinence sociale ($P_{sociale}$)

La pertinence sociale est définie à partir de l'historique, à l'instant t , du nombre de passages relevés entre le POI candidat et le dernier POI dans l'itinéraire en cours de construction. On parle d'intensité de lien entre POI. La popularité du POI candidat est également considérée dans ce calcul. La fonction d'évaluation de pertinence sociale est définie comme suit :

$$P_{sociale}(O_e, t) = \delta * popularite(O_e, t) + \phi * intensiteLien(O_d, O_e, t) \quad (4)$$

avec $\delta + \phi = 1$

où t correspond à l'horodatage du *scoring*, O_d est le dernier POI dans l'itinéraire en cours de construction, *popularite* est la popularité de O_e et *intensiteLien* représente le nombre de passages relevés entre O_d et O_e .

La popularité d'un POI (formule 5) est définie par la fréquentation et les retours d'évaluation (les notes) des utilisateurs. Son évaluation est faite par rapport aux popularités des POI voisins.

$$popularite(O_e, t) = \lambda * \frac{(|V| + 1) - rang_{notes}(O_e, V, t)}{(|V| + 1)} + \theta * \frac{(|V| + 1) - rang_{freq}(O_e, V, t)}{(|V| + 1)} \quad (5)$$

avec $\lambda + \theta = 1$ et où t correspond à l'horodatage du *scoring*, V est un ensemble de POI qui est formé par des POI voisins de O_e (pour des raisons de place, nous avons simplifié la formulation de V qui, en réalité doit être $V(O_e)$). Ainsi, $rang_{notes}(O_e, V, t)$ est une fonction permettant d'utiliser la moyenne des notes attribuées au POI O_e afin de déterminer, à l'instant t , son rang dans la liste des POI de V ordonnés par moyenne décroissante des notes qui leur sont associées respectivement. De la même façon, la fonction $rang_{freq}(O_e, V, t)$ permet d'exploiter la fréquentation du POI O_e et celle des POI de V afin de déterminer le rang de O_e dans la liste des POI de V ordonnés par niveau de fréquentation décroissant. Ainsi, le score de popularité d'un POI est d'autant plus élevé (proche de 1) que ses $rang_{notes}$ et $rang_{freq}$ par rapport aux voisins sont proches de 1.

L'intensité de lien entre deux POI O_i et O_j est définie comme suit :

$$intensiteLien(O_i, O_j, t) = \frac{\tau_{ij}(t)}{Q_{max}} \quad (6)$$

où t correspond à l'horodatage du *scoring*.

τ_{ij} et Q_{max} sont expliqués en détail ci-dessous.

En se référant à l'approche « colonie de fourmis », un utilisateur est désormais représenté par une fourmi. Pour cela, lorsqu'un utilisateur visite un POI O_i , puis un POI O_j , il dépose des phéromones sur le trajet entre O_i et O_j (voir section 3.2.1, figure 5(b)). A chaque passage sur le trajet $O_i - O_j$ une quantité Q (prédéfinie) de phéromones est cumulée sur l'arête correspondante du graphe G_{social} . $\tau_{ij}(t)$ correspond à la quantité de phéromones cumulées entre O_i et O_j à l'instant t . Enfin, Q_{max} correspond à la valeur de τ maximale dans le graphe G_{social} ($Q_{max} = Max(\tau_{ij})$) où chaque τ est la quantité de phéromones cumulées sur les chemins entre les POI d'une zone géographique donnée). Q_{max} est dynamique et évolue dans le temps. Ainsi, lorsque la quantité de phéromones sur un trajet $O_i - O_j$ est élevée (proche de Q_{max}), à cause de nombreux passages, un visiteur pourrait être invité à visiter le POI O_j lorsqu'il visite le POI O_i ou inversement.

Il est à noter que $\tau_{ij}(t)$ est stocké sur le graphe G_{social} . Cette quantité de phéromones est mise à jour à chaque nouveau passage. Ainsi, le principe de renforcement de phéromones entre O_i et O_j est défini par :

$$\tau_{ij}(t') = \tau_{ij}(t) + Q \quad (7)$$

tel que $t' > t$.

Q vaut 1 à chaque passage. De plus, la quantité de phéromones sur un trajet (notée

Recommandation d'itinéraires

τ_{ij}) diminue progressivement si personne ne l'emprunte pendant un certain temps. Le principe d'évaporation de phéromones est défini par l'équation suivante :

$$\tau_{ij}(t') = (1 - \rho) * \tau_{ij}(t) \quad (8)$$

où ρ est le taux d'évaporation.

Conformément aux préconisations de (Anagnostopoulos *et al.*, 2017) et de (Sriphaew, Sombatsricharoen, 2015), nous faisons le choix d'une évaporation lente. Comme (Anagnostopoulos *et al.*, 2017), nous fixons la valeur de ρ à 0.1 afin de garder longtemps la trace de précédentes visites tout en évitant la saturation des arcs en phéromones. Ainsi, la sérendipité est prise en compte dans le cadre de cette approche. En effet, cette gestion de l'intensité des liens consiste à proposer un nouveau POI visité par des touristes qui ont eux aussi visité le POI précédent, puis le nouveau POI, quand bien même ils n'auraient pas du tout le même profil que le touriste pour lequel on est en train de construire l'itinéraire. Ce nouveau POI ne respecte pas forcément toutes les préférences d'un touriste donné mais représente un intérêt important pour la communauté et, ainsi, permet d'étendre les centres d'intérêt du dit touriste.

3.3. Expérimentation

Nous avons mené une première série d'expérimentations selon différents scénarios et hypothèses de travail sur un jeu de données composé d'un échantillon de 120 POIs localisés à Bagnères-de-Luchon. Nous avons pour objectif d'expérimenter (i) l'algorithme de génération d'itinéraires et (ii) les pondérations α et β . Nous commentons un scénario qui n'intègre pas de contraintes sociales telles que la popularité des POIs ou l'intensité de liens entre eux (en effet, $\gamma = 0$). Dans ce scénario, nous avons défini le paramétrage suivant :

Point de départ : "place de la mairie"

Critères temporels : Temps de visite = "60 minutes"; Mode de transport = "à pied"

Préférences : $PREF_T = \{\text{"musée/casino", "sculpture/peinture"}\}$; $PREF_H = \{\text{"époque contemporaine"}\}$

Nous avons étudié deux hypothèses :

Hypothèse 1 (H1) : $\alpha = 0.5$; $\beta = 0.5$; $\gamma = 0$; $\mu = 0.5$; $\nu = 0.5$

Hypothèse 2 (H2) : $\alpha = 0.6$; $\beta = 0.4$; $\gamma = 0$; $\mu = 0.5$; $\nu = 0.5$

Nous obtenons 2 itinéraires décrits sur la figure 6. H1 (figure 6(a)) donne un itinéraire et des POIs majoritairement différents de H2 (figure 6(b)). Le choix des coefficients α et β dans le cadre de H2 resserre la contrainte de préférence - en effet les POIs proposés respectent les contraintes $P_{preference}$ avec un score moyen de 0.87 (contre 0.67 dans H1). Ce choix de coefficients relâche la contrainte de temps de déplacement court - malgré un circuit plus long, les POIs proposés respectent toujours les contraintes $P_{temporelle}$ avec un score moyen de 0.94 (contre 0.81 dans H1). Ici, nous constatons un effet inattendu avec une augmentation du nombre de POIs à visiter dû à certaines distances plus longues compensées par des temps de visite de POIs plus courts pour nombre d'entre eux. Le temps de parcours de l'itinéraire additionné au temps de visite de chaque POI respecte toujours la contrainte de temps correspondant aux critères temporels.

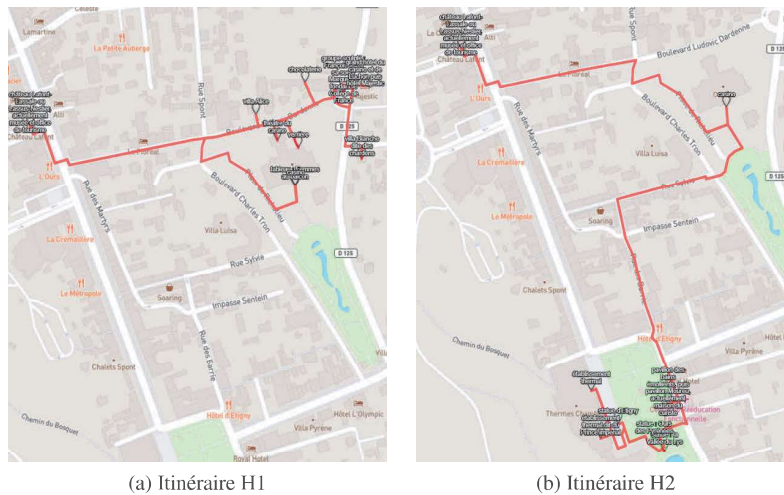


Figure 6. Exemple de génération d'itinéraire selon 2 hypothèses

L'algorithme propose des itinéraires composés de POIs pertinents. Ainsi, nous constatons que les POIs proposés respectent l'ensemble des contraintes : la moyenne des scores de pertinence des POIs de l'itinéraire H2 est de 0.90 tandis que celui de l'itinéraire H1 est de 0.74. H2 donne également de meilleurs résultats dans d'autres expérimentations mais demande encore à être confirmé par des tests de plus grande envergure. Les coefficients de pondération α et β jouent parfaitement leur rôle de régulation des contraintes. Après de nouvelles expérimentations, nous pourrions proposer des valeurs d'initialisation par défaut de ces paramètres.

4. Conclusion et perspectives

Nous avons présenté une architecture générale pour la fédération de données patrimoniales hétérogènes puis la valorisation de ces données. Nous avons focalisé cette présentation sur la partie recommandation d'itinéraires composés de POI patrimoniaux. Il s'agit de croiser les profils (position géographique, mode de transport, fenêtre de disponibilité, centres d'intérêts, etc.) des utilisateurs et la base de données patrimoniales pour mettre en place un système de recommandation d'itinéraires de visite comprenant des POI pertinents. Ces itinéraires sont composés de POI conformes au profil utilisateur, mais aussi d'un nombre raisonnable de POI considérés originaux du point de vue de l'utilisateur tout en méritant le détour selon la communauté. Un thésaurus de classification de concepts patrimoniaux nous permet de catégoriser aussi bien les POI que les préférences des utilisateurs. La proximité sémantique entre un POI et des préférences utilisateur peut ainsi être mesurée sur la base de ce thésaurus patrimonial. Enfin, nous intégrons l'usage d'un algorithme dit de « colonies de fourmis » afin de proposer des POI appréciés par l'ensemble des utilisateurs et qui pourront paraître originaux car ne correspondant pas exactement aux préférences de l'utilisa-

Recommandation d'itinéraires

teur courant. Ce dosage subtil de POI originaux peut-être adapté via une modification dynamique des valeurs des paramètres α , β et γ de la formule 1.

L'application fédératrice est un premier prototype de visualisation des POI référencés dans la base ont été rendus publics¹⁰. Les travaux que nous menons offrent un accès aux données par les Systèmes d'Information Touristiques (SIT) via Wikipedia permettant ainsi aux offices de tourisme de bénéficier de ces données patrimoniales (Bessagnet *et al.*, 2018).

Nous envisageons d'expérimenter cette première version de notre application de recommandation sur un panel d'utilisateurs varié et plus important. Cette expérimentation permettra notamment de définir des valeurs d'initialisation des coefficients des différentes formules de mesure de pertinence. Nous allons également comparer les résultats de notre système avec ceux issus d'une génération d'itinéraire uniquement basée sur des critères géographiques et temporels. De plus, nous travaillons sur une nouvelle version de recommandation qui intègre des éléments de contexte *physique* et *ressource* de l'utilisateur qui auront pour conséquence d'adapter dynamiquement un itinéraire. Enfin, nous envisageons d'enrichir les recommandations de notre système avec des données de l'Open Data (Datatourisme, par exemple) afin de proposer aux utilisateurs des événements culturels proches de leurs localisation et préférences.

5. Remerciements

Ce projet est réalisé dans le cadre du programme de recherche Européen TCV-PYR (2017-2020), financé par l'Union Européenne (FEDER) en partenariat avec les régions Occitanie et Nouvelle-Aquitaine.

Bibliographie

- Aliannejadi M., Mele I., Crestani F. (2016). User model enrichment for venue recommendation. In *Asia information retrieval symposium*, p. 212–223.
- Anagnostopoulos A., Atassi R., Becchetti L., Fazzino A., Silvestri F. (2017). Tour recommendation for groups. *Data Mining and Knowledge Discovery*, vol. 31, n° 5, p. 1157–1188.
- Bahramian Z., Abbaspour R., Claramunt C. (2017b, 09). A cold start context-aware recommender system for tour planning using artificial neural network and case based reasoning. *Mobile Information Systems*, vol. 2017b, p. 1–18.
- Bahramian Z., Ali Abbaspour R., Claramunt C. (2017a). A context-aware tourism recommender system based on a spreading activation method. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-4/W4, p. 333–339.
- Bartolini I., Moscato V., Pensa R. G., Penta A., Picariello A., Sansone C. *et al.* (2016). Recommending multimedia visiting paths in cultural heritage applications. *Multimedia Tools and Applications*, vol. 75, n° 7, p. 3813–3842.

10. Code source disponible sur : <https://git.univ-pau.fr/liuppa/TCVPYR/importateur>

- Bessagnet M.-N., Etcheverry P., Lacayrelle A., Marquesuzaà C., Rajaonarivo L., Roose P. *et al.* (2018, octobre). Leveraging heterogeneous Cultural Heritage data to promote tourism. In *Open Source Geospatial Research and Education Symposiums (OGRS)*. Lugano, Switzerland. Consulté sur <https://hal.archives-ouvertes.fr/hal-01976405>
- Costa H., Furtado B., Pires D., Macedo L., Cardoso A. (2012). Context and intention-awareness in pois recommender systems. In *6th acm conf. on recommender systems, 4th workshop on context-aware recommender systems, recsys*, vol. 12, p. 5.
- Dennouni N., Peter Y., Lancieri L., Slama Z. (2018, 10). Towards an incremental recommendation of pois for mobile tourists without profiles. *International Journal of Intelligent Systems and Applications*, vol. 10, p. 42-52.
- De Pessemier T., Dhondt J., Vanhecke K., Martens L. (2015). Travelwithfriends: a hybrid group recommender system for travel destinations. In *Workshop on tourism recommender systems (tours15), in conjunction with the 9th acm conference on recommender systems (recsys 2015)*, p. 51–60.
- Fonteles A. S., Bessagnet M.-N., Lacayrelle A., Sallaberry C. (2018, November). Un environnement pour la valorisation de données patrimoniales hétérogènes. In *Spatial Analysis and GEomatics (SAGEO)*. Montpellier, France. (To be published)
- Garcia A., Arbelaitz O., Linaza M. T., Vansteenwegen P., Souffriau W. (2010). Personalized tourist route generation. In *International conference on web engineering*, p. 486–497.
- Kesorn K., Juraphanthong W., Salaiwarakul A. (2017). Personalized attraction recommendation system for tourists through check-in data. *IEEE Access*, vol. 5, p. 26703–26721.
- Kłopotek M. A. (2009). Approaches to “cold-start” in recommender systems. *Studia Informatica : systems and information technology*, vol. Vol. 1(12), p. 47–54.
- Logesh R., Subramaniaswamy V., Vijayakumar V., Li X. (2018). Efficient user profiling based intelligent travel recommender system for individual and group of users. *Mobile Networks and Applications*, p. 1–16.
- Lu Q., Guo F. (2016). A novel e-commerce customer continuous purchase recommendation model research based on colony clustering. *International Journal of Wireless and Mobile Computing*, vol. 11, n° 4, p. 309–317.
- Minjing P., Xinglin L., Ximing L., Mingliang Z., Xianyong Z., Xiangming D. *et al.* (2017). Recognizing intentions of e-commerce consumers based on ant colony optimization simulation. *Journal of Intelligent & Fuzzy Systems*, vol. 33, n° 5, p. 2687–2697.
- Sriphaew K., Sombatsrichaoren K. (2015). Food tour recommendation using modified ant colony algorithm. In *5th international conference on computing and informatics, icoci 2015*.
- Tsiligirides T. (1984). Heuristic methods applied to orienteering. *Journal of the Operational Research Society*, vol. 35, n° 9, p. 797–809.
- Wang J., Lin K., Li J. (2013). A collaborative filtering recommendation algorithm based on user clustering and slope one scheme. In *Computer science & education (iccse), 2013 8th international conference on*, p. 1473–1476.
- Wu Z., Palmer M. (1994). Verbs semantics and lexical selection. In, p. 133–138.
- Zhang X., Pang X. (2015). Analysis on the mobile electronic commerce recommendation model based on the ant colony algorithm. In *2015 international conference on industrial technology and management science*. Atlantis Press.