



**HAL**  
open science

## **ANASTASIA : recommandation de séquences d'activités spatio-temporelles**

Diana Nurbakova, Léa Laporte, Sylvie Calabretto, Jérôme Gensel

► **To cite this version:**

Diana Nurbakova, Léa Laporte, Sylvie Calabretto, Jérôme Gensel. ANASTASIA : recommandation de séquences d'activités spatio-temporelles. Rencontres Jeunes Chercheurs en Recherche d'Information (RJCRI CORIA-CIFED), Mar 2016, Toulouse, France. pp.10. hal-01328062

**HAL Id: hal-01328062**

**<https://hal.science/hal-01328062>**

Submitted on 7 Jun 2016

**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.

---

# ANASTASIA : recommandation de séquences d'activités spatio-temporelles

Diana Nurbakova\* — Léa Laporte\* — Sylvie Calabretto\* — Jérôme Gensel\*\*

\* *Laboratoire d'InfoRmatique en Image et Système d'information (LIRIS)*

*UMR 5205 CNRS/INSA de Lyon*

*20 avenue Albert Einstein, 69621 Villeurbanne cedex*

\*\* *Laboratoire d'Informatique de Grenoble (LIG)*

*UMR 5217 CNRS/Université Grenoble Alpes*

*BP 53, 38041 Grenoble cedex 9*

---

*RÉSUMÉ. Avec l'augmentation du nombre et de la variété des activités accessibles par les utilisateurs, la recommandation personnalisée de séquences d'activités devient un enjeu important. Or, la plupart des systèmes de recommandation ne tiennent pas compte des contraintes temporelles liées aux activités, ce qui rend la recommandation difficile à suivre par un utilisateur. Dans cet article, nous décrivons une nouvelle approche pour la recommandation de séquences d'activités limitées dans le temps et concurrentes. Elle s'appuie sur l'historique des activités des utilisateurs pour extraire des motifs comportementaux et intègre différents éléments contextuels (popularité, informations démographiques et spatio-temporelles). Nous présentons un protocole d'évaluation et un jeu de données qui permettra l'évaluation de notre approche.*

*ABSTRACT. As amount of activities available for users and their variety have grown, personalised recommendation of activities sequences has become an important challenge. However, most of recommender systems do not consider temporal constraints of activities, making the recommendation hard for user to follow. In this article, we describe a novel approach for recommendation of competing activities limited in time. It makes use of historical records of users' activities in order to mine users' behavioral patterns, and combines different contextual elements (popularity, demographic and spatio-temporal information). We present an evaluation framework and a dataset that will allow us to evaluate our approach.*

*MOTS-CLÉS : système de recommandation, séquence d'activités, score personnalisé, contexte d'utilisateur.*

*KEYWORDS: recommender system, activities sequence, personalised score, user context.*

---

## 1. Introduction

Au cours des dernières années, les services et les réseaux sociaux basés sur la localisation (*LBSNs*) ont attiré des millions d'utilisateurs. Une des fonctionnalités qu'ils proposent est la recommandation de points d'intérêt (*POIs*). Un *POI* est une position physique correspondant à une localisation géographique d'une personne ou d'un événement et présentant un intérêt spécifique pour un utilisateur<sup>1</sup>. Les systèmes de recommandation de *POIs* proposent à l'utilisateur un ensemble de *POIs*, ordonnés selon le degré d'intérêt de l'utilisateur. Récemment, des travaux (Sang *et al.*, 2015 ; Yu *et al.*, 2015) se sont intéressés au problème de la recommandation de séquences de *POIs*. Un inconvénient de ces approches est qu'elles ne tiennent pas compte des contraintes temporelles associées aux *POIs* (*e.g.* horaires d'ouverture-fermeture).

Dans cet article, nous nous concentrons sur la recommandation de séquences d'activités spatio-temporelles, que nous nommons R-STAS et qui s'applique à plusieurs scénarios : événements distribués (conférences, festivals ou expositions) ; tourisme (recommandation d'animations pendant une croisière) ; crowdsourcing mobile. Nous définissons la notion d'*activité* comme un événement géo-localisé à durée limitée et associé à une ou plusieurs catégories.

Parmi les travaux sur les systèmes de recommandation, il y a très peu d'approches qui exploitent conjointement les centres d'intérêt des utilisateurs, l'ensemble des facteurs spatio-temporels et les habitudes de l'utilisateur afin de construire un itinéraire personnalisé des activités. La motivation du travail actuel peut être formulée selon les aspects suivants : (1) *Personnalisation et Contexte* : accorder les activités aux intérêts de l'utilisateur et son contexte ; (2) *Espace et Temps* : prendre en compte le caractère spatial et la disponibilité limitée des activités ; (3) *Séquence et Succession* : gérer de façon optimale les séquences d'activités concurrentes et interdépendantes.

Nous proposons une approche qui prend en compte les contraintes et dépendances temporelles entre activités au sein d'une séquence. En s'appuyant sur les historiques d'activités de l'utilisateur et sur ses centres d'intérêts, nous récupérons les habitudes, ou motifs de comportement, de l'utilisateur et calculons le score de pertinence des activités pour chaque utilisateur. Ensuite, nous utilisons les résultats obtenus et les horaires de disponibilité des activités pour construire une séquence d'activités. Afin d'évaluer l'approche proposée et en l'absence de collection adaptée, nous proposons un protocole d'évaluation et la construction d'un jeu de données.

L'article est organisé de la manière suivante. Nous discutons des travaux liés à la section 2. La section 3 décrit l'approche proposée. Dans la section 4 nous proposons une discussion sur le protocole d'évaluation. Enfin, la section 5 conclut l'article.

---

1. [http://www.w3.org/2010/POI/wiki/Main\\_Page](http://www.w3.org/2010/POI/wiki/Main_Page)

## 2. Travaux liés

Les travaux de l'état-de-l'art touchent deux domaines en recommandation : la recommandation de POIs et la recommandation de séquences de POIs. La recommandation de POIs considère des éléments indépendants restitués sous la forme d'une liste ordonnée selon la pertinence aux centres d'intérêt de l'utilisateur, tandis que la recommandation de séquences de POIs gère l'interdépendance géographique des éléments.

La plupart des algorithmes de recommandation de POIs supposent généralement l'indépendance entre les POIs et se basent sur le principe de l'ordonnement des POIs selon l'intérêt de l'utilisateur. Les centres d'intérêt sont indiqués explicitement dans le profil de l'utilisateur ou récupérés à partir des POIs visités, via leur position géographique, la fréquence de visite (*check-in*) ou leur popularité. Ils sont utilisés pour calculer un score de pertinence des POIs. Les POIs sont ensuite ordonnés selon le score calculé et les tops  $k$  résultats sont proposés à l'utilisateur (Zhang et Chow, 2015a ; Yang et Fang, 2014 ; Li *et al.*, 2015 ; Yang et Fang, 2014 ; Zhang et Chow, 2015b). Ce type d'algorithmes de recommandation exploite des combinaisons de différents facteurs (influence géographique, évolution des goûts de l'utilisateur, etc.), mais ne tient compte ni de l'aspect consécutif des visites, ni de la disponibilité temporelle limitée des POIs.

La recommandation de séquences de POIs permet la prise en compte de l'aspect consécutif des visites. Nous pouvons distinguer deux types d'approches. Dans la première approche, majoritairement présente dans la littérature, la recommandation de séquences de POIs est divisée en deux sous-problèmes : l'estimation des centres d'intérêt de l'utilisateur pour calculer un score de pertinence personnalisé pour chaque POI et la résolution d'un problème d'optimisation pour la construction d'un itinéraire optimal (par exemple, (Yu *et al.*, 2015 ; Zhang *et al.*, 2015 ; Schaller *et al.*, 2014b)). Le calcul de l'itinéraire est généralement effectué à l'aide d'algorithmes d'optimisation sous contraintes issus du domaine de la recherche opérationnelle (Vansteenwegen *et al.*, 2011 ; Gavalas *et al.*, 2014). Dans la deuxième approche, les algorithmes déterminent une séquence de POIs en calculant la probabilité de transition d'un POI à un autre (Sang *et al.*, 2015). Cependant, ces approches ne tiennent pas compte de la disponibilité temporelle limitée des POIs.

Dans nos travaux, nous nous focaliserons sur la recommandation de séquences d'activités spatio-temporelles prenant en compte les contraintes et dépendances temporelles entre activités et les habitudes des utilisateurs.

## 3. Approche Proposée

Nous proposons une approche de recommandation de séquences d'activités, nommée ANASTASIA (« A Novel Approach for Spatio-Temporal Activities Sequences and Itineraries recommendAtion »). Nous adaptons des algorithmes de l'état de l'art pour tenir compte des centres d'intérêts de l'utilisateur et des contraintes et dépendances séquentielles, temporelles et spatiales. Notre approche est composée de 3 mo-

dules : un module d'extraction des motifs de séquences, un module de calcul des scores des activités et un module de construction d'itinéraires. Les données sont traitées pour extraire des motifs caractéristiques dans les historiques des utilisateurs et pour calculer des scores de pertinence pour chaque activité. Ces motifs et scores sont ensuite intégrés dans le calcul de l'itinéraire.

### 3.1. Module d'extraction des motifs de séquences

Le module d'extraction des motifs de séquences cherche à apprendre le comportement typique de l'utilisateur en récupérant des séquences d'activités consécutives. Deux activités sont considérées consécutives si l'intervalle temporel entre la fin de la première activité et le début de la deuxième est inférieur à un seuil fixé.

Nous proposons d'adapter les travaux de (Zhang *et al.*, 2014) sur l'extraction de motifs de séquences. (Zhang *et al.*, 2014) utilisent un graphe de transition localisation-localisation ( $L^2TG$ ) pour modéliser les séquences de POIs visités par l'utilisateur. Nous étendons ce principe en considérant deux graphes de transition : un graphe de transition activité-activité, noté  $A^2TG$ , et un graphe de transition catégorie-catégorie, noté  $C^2TG$ . Le premier graphe permet de modéliser les transitions entre activités uniques (e.g. concert avec date et lieu fixé). Le deuxième graphe permet une généralisation en modélisant les transitions entre catégories d'activités (e.g. spectacles).

Nous nous appuyons sur une liste d'activités antérieures de l'utilisateur et sur les horaires de ces activités pour construire le graphe  $A^2TG$ . Les nœuds du graphe, notés  $V = \{a_1, \dots, a_N\}$ , correspondent aux activités. Contrairement à (Zhang *et al.*, 2014) qui utilisent le nombre d'arcs sortants du nœud, nous proposons d'utiliser le nombre d'arcs entrants, noté  $InCount(a_i)$ . Cette modification est motivée par l'hypothèse que la satisfaction de l'utilisateur dépend des activités précédentes recommandées et réalisées. Par exemple, il n'est pas pertinent de recommander un restaurant à un utilisateur qui vient de manger dans une pizzeria. Les arcs du graphe reflètent les transitions d'une activité à une autre. Ils sont pondérés par le nombre de transitions d'une activité à une autre, que nous notons  $TransCount(a_i \rightarrow a_j)$ . La probabilité de transition d'une activité à une autre, notée  $P_T(a_i \rightarrow a_j)$ , est définie comme suit :

$$P_T(a_i \rightarrow a_j) = \begin{cases} \frac{TransCount(a_i \rightarrow a_j)}{InCount(a_j)}, & \text{si } InCount(a_j) \neq 0 \\ 0, & \text{si } InCount(a_j) = 0 \text{ et } a_i \neq a_j \\ 1, & \text{si } InCount(a_j) = 0 \text{ et } a_i = a_j \end{cases}$$

Notons que les activités que nous considérons sont uniques. Dans ce cas, la probabilité de transition  $P_T(a_i \rightarrow a_j)$  devient binaire :  $P_T(a_i \rightarrow a_j) = 1$  si la transition d'une activité à une autre existe, et  $P_T(a_i \rightarrow a_j) = 0$  sinon. Ainsi, nous proposons de passer au niveau hiérarchique supérieur, c'est-à-dire au niveau des catégories d'activités. Nous supposons que l'attribution des activités aux catégories est connue.

Pour construire le graphe  $C^2TG$  nous passons par une étape intermédiaire de substitution des activités par leurs catégories, puis nous appliquons le même principe que pour le graphe  $A^2TG$ . Les nœuds du graphe  $C^2TG$  représentent les catégories

correspondant aux activités effectuées par l'utilisateur et sont associés au nombre d'arcs entrants dans le nœud  $c_i$ , noté  $InCount(c_i)$ .  $InCount(c_i)$  est fonction des activités composant la catégorie  $c_i$ , tel que :  $InCount(c_i) = \sum_{a_j \in c_i} InCount(a_j)$ .

Les arcs correspondent aux transitions d'une catégorie à une autre et sont pondérés par le nombre de transitions,  $TransCount(c_i \rightarrow c_j)$ , défini comme la somme des valeurs de  $TransCount$  des activités de cette paire de catégories :  $TransCount(c_i \rightarrow c_j) = \sum_{a_k \in c_j, a_g \in c_i} TransCount(a_k \rightarrow a_g)$ . La probabilité de transition d'une catégorie à une autre est calculée de la façon suivante :

$$P_T(c_i \rightarrow c_j) = \begin{cases} \frac{TransCount(c_i \rightarrow c_j)}{InCount(c_j)}, & \text{si } InCount(c_j) \neq 0 \\ 0, & \text{si } InCount(c_j) = 0 \text{ et } c_i \neq c_j \\ 1, & \text{si } InCount(c_j) = 0 \text{ et } c_i = c_j \end{cases}$$

En s'appuyant sur la liste d'activités effectuées par l'utilisateur, nous proposons d'explorer l'importance d'une activité au sein d'une séquence, en calculant son poids, noté  $Weight(a_i)$ . L'objectif est de comprendre si une activité a été effectuée uniquement pour patienter entre deux activités ou si elle a présenté un intérêt pour l'utilisateur. Ce poids est basé sur la probabilité de choisir l'activité au hasard dans la collection des activités concurrentes. Elle est calculée de la manière suivante. Parmi les motifs de séquences d'utilisateur, nous choisissons les sous-séquences composées de 3 éléments, c'est-à-dire pour l'activité  $a_i$  nous cherchons une sous-séquence :  $a_{i-1} \rightarrow a_i \rightarrow a_{i+1}$ . Étant donnée la liste d'activités avec leurs fenêtres temporelles et leur durée, nous proposons de chercher toutes les activités,  $a_j \in Concurrents(a_i)$ , concurrentes à l'activité  $a_i$ . La probabilité de choix de l'activité  $a_i$  au hasard est calculée comme un ratio :  $PH(a_i) = \frac{1}{|Concurrents(a_i)|}$ , où  $|\cdot|$  est la cardinalité de l'ensemble. Le poids de l'activité est donnée par :  $Weight(a_i) = 1 - PH(a_i)$ .

### 3.2. Module de calcul des scores des activités

Le module de calcul des scores des activités vise à mesurer la pertinence d'une activité donnée par rapport aux centres d'intérêts d'un utilisateur. Nous proposons de calculer le score des activités en explorant différentes dimensions (spatiale, personnelle, sociale, financière) du contexte de l'utilisateur.

Chaque activité est associée à la localisation. Nous supposons que dans le cadre de la recommandation d'activités spatio-temporelles, la proximité géographique a moins d'influence que dans le cas de la recommandation de POIs et que les contraintes temporelles sont plus fortes et englobent une partie des contraintes géographiques. Nous intégrons le score spatial, noté  $Score_{spatial}(a_i)$ , dans le score global de pertinence d'une activité. Ce score explore l'influence de la distance entre les activités afin de récupérer la préférence de l'utilisateur à se déplacer ou à rester au même endroit.

En nous appuyant sur les profils des utilisateurs, nous proposons de construire des vecteurs d'informations personnelles pour chaque utilisateur qui contiennent les attributs suivants : classes d'âges (<13, 14-17, 18-20, 21-30, 31-50, >50), genre (féminin,

masculin), nombre de participations aux éditions d'un événement ou à des événements similaires (0 fois, 1 fois, 2-5 fois, >5 fois). Les valeurs des attributs sont binaires. Ainsi, nous proposons de récupérer les fréquences de visites des activités par rapport aux attributs décrits et de calculer le score personnel  $Score_{pers}(a_i)$  de l'activité.

Nous proposons d'explorer l'influence sociale en nous basant sur la popularité des activités et les centres d'intérêt de l'utilisateur qui changent en fonction du groupe qui l'accompagne (seul, en couple, avec des amis, avec la famille, autre). Nous notons cette influence le score social,  $Score_{social}(a_i)$ , et l'intégrons dans le score global.

Certaines activités exigent des frais, d'autres sont gratuites. Les utilisateurs peuvent avoir des contraintes financières, identifiant le budget maximal à ne pas dépasser. Nous proposons d'explorer la préférence de l'utilisateur à exercer les activités qui exigent ou n'exigent pas des frais en se basant sur la fréquence de ces activités. Ainsi nous obtenons le score financier,  $Score_{finance}(a_i)$ .

Le score de pertinence de l'activité,  $Score(a_i)$ , sera ainsi fonction des différentes dimensions du contexte de l'utilisateur, c'est-à-dire :  $Score(a_i) = \alpha \cdot Score_{spatial}(a_i) + \beta \cdot Score_{pers}(a_i) + \gamma \cdot Score_{social}(a_i) + \delta \cdot Score_{finance}(a_i)$ . À notre connaissance, il n'existe pas, d'approche qui exploite toutes les dimensions envisagées. (Sang *et al.*, 2015) utilisent les facteurs personnels de l'utilisateur, les relations temporelles et géographiques entre les POIs, tandis que (Zhang et Chow, 2015b) proposent d'utiliser l'influence spatio-temporelle et la popularité des POIs.

### 3.3. Module de construction d'itinéraire

Le module de construction d'itinéraire cherche un parcours optimal, c'est-à-dire une séquence d'activités qui maximise la satisfaction de l'utilisateur par rapport aux activités pratiquées. Le problème de construction d'itinéraire peut être formulé comme un problème d'optimisation et plus précisément comme le problème nommé *Team Orienteering Problem with Time Windows* (TOPTW) (Vansteenwegen *et al.*, 2009). (Vansteenwegen *et al.*, 2009) formule ce problème de la façon suivante : soient un ensemble de POIs  $\{a_i\}, i = 1, \dots, n$ , des scores  $S_i$  associés à chaque point  $a_i$ , le temps  $T_i$  de visite de chaque POI  $a_i$ , des horaires  $[O_i, C_i]$  de la disponibilité de chaque POI  $a_i$ , le point de départ (point 1) et le point de destination (point  $n$ ), le temps  $t_{ij}$  de déplacement entre les points  $a_i$  et  $a_j$  et le budget temporel  $T_{max}$ , de trouver  $m$  parcours limités par  $T_{max}$  qui maximisent le score total collecté. Le parcours est constitué de visites de sous-ensemble de POIs, chacun ne pouvant être visité qu'une seule fois et pendant ses horaires de disponibilité. Il est permis d'arriver à POI avant son ouverture et y attendre. Dans notre cas, nous considérons des activités au lieu de POIs.

Pour résoudre ce problème, nous proposons d'adapter l'algorithme ILS (*Iterative Local Search*) introduit par (Vansteenwegen *et al.*, 2009) en le modifiant pour prendre en compte le caractère séquentiel des activités. ILS est un algorithme itératif heuristique qui cherche à maximiser le score en ajoutant une activité (nœud) au parcours existant à chaque itération. Pour qu'un nœud puisse être ajouté, il faut vérifier

que son insertion ne rend pas le parcours infaisable, c'est-à-dire, toutes les activités planifiées après ce nœud doivent satisfaire les contraintes de leurs fenêtres temporelles. Deux mesures sont ainsi introduites pour chaque nœud inclus au parcours : (1) le temps d'attente au nœud  $a_i$  dans le cas d'arrivée  $arrive_i$  avant l'ouverture :  $Wait_i = \max [0, O_i - arrive_i]$ ; et (2) le délai maximal de début de l'activité au nœud  $a_i$ , noté  $s_i$ , c'est-à-dire le délai maximum pendant lequel le début de l'activité au nœud  $a_i$  peut être retardé sans rendre infaisable l'activité suivante. Étant donnée la durée de l'activité  $T_i$  au point  $a_i$ , le moment de fermeture,  $C_i$ , et l'activité suivante  $a_{i+1}$ , ce délai est :  $MaxShift_i = \min [C_i - s_i - T_i, Wait_{i+1} + MaxShift_{i+1}]$ .

Pour chaque nœud  $a_k$  non inclus dans le parcours, l'algorithme détermine sa meilleure position d'insertion entre les nœuds  $a_i$  et  $a_j$ , c'est-à-dire celle qui a le plus bas coût temporel d'insertion, déterminé par :  $Shift_k = t_{ik} + Wait_k + T_k + t_{kj} - t_{ij}$ , avec  $Shift_k \leq Wait_j + MaxShift_j$ , condition de faisabilité. Au lieu de maximiser directement le score, l'algorithme maximise le ratio du score par le coût temporel :  $Ratio_k = \frac{S_k^2}{Shift_k}$ .

Après avoir inséré le nœud  $a_k$  les paramètres des nœuds doivent être mis à jour selon les formules présentées dans le Tableau 1. Trois cas sont considérés : les paramètres du nœud  $a_k$ , des nœuds suivants et des nœuds précédents dans le parcours.

	Nœud $a_k$	Nœuds suivants	Nœuds précédents
Temps d'arrivée	$arrive_k = s_i + T_i + t_{ik}$	$arrive_{j^*} = arrive_j + Shift_{j-1}$	
Temps de début de visite	$s_k = arrive_k + Wait_k$	$s_{j^*} = s_j + Shift_j$	
Temps d'attente	$Wait_k = \max [0, O_k - arrive_k]$	$Wait_{j^*} = \max [0, Wait_j - Shift_{j-1}]$	
Délais maximal		$MaxShift_{j^*} = \max [0, MaxShift_j - Shift_j]$	$MaxShift_i = \min [C_i - s_i - T_i, Wait_{i+1} + MaxShift_{i+1}]$
Coût temporel		$Shift_j = \max [0, Shift_{j-1} - Wait_j]$	

**Tableau 1.** Les formules de mise à jour des paramètres des nœuds après l'insertion d'un nouveau nœud  $a_k$  dans le parcours.

Pour prendre en compte l'influence du caractère consécutif des activités, nous proposons de pondérer le score d'activité  $a_k$  par la probabilité de transition entre l'activité courante  $a_k$  et l'activité précédente  $a_{k-1}$  :  $Ratio_k = \frac{(S_k * P_T(a_{k-1} \rightarrow a_k))^2}{Shift_k}$ , où  $S_k$  est le score de l'activité  $a_k$  calculé par le module de calcul des scores des activités,  $P_T(a_{k-1} \rightarrow a_k)$  est une probabilité de transition entre l'activité  $a_{k-1}$  et l'activité  $a_k$ ; et  $Shift_k$  représente le coût temporel d'insertion de l'activité  $a_k$  dans le parcours.



#### 4. Discussion autour de l'évaluation

Bien qu'il existe des protocoles et jeux de test pour l'évaluation de la recommandation, il n'y a pas, à notre connaissance, de jeux de données et de protocole adaptés pour l'évaluation de séquences de POIs.

Différents types de jeux de données et de collections peuvent être utilisés pour l'évaluation des algorithmes de recommandation de POIs : issus de campagnes d'évaluation internationales comme TREC, issus des traces d'utilisation de LSBNs, issus d'applications développées spécialement pour cette tâche ou issus de la recherche opérationnelle. Ces jeux de données présentent des limites qui empêchent de les utiliser pour l'évaluation de séquences d'activités spatio-temporelles.

Les collections issues de campagnes d'évaluation, comme par exemple la tâche de suggestion contextuelle de TREC (TREC CS) (Dean-Hall *et al.*, 2014), fournissent un ensemble de profils d'utilisateurs, de contextes et de POIs, mais ne tiennent pas compte ni des contraintes temporelles entre POIs, ni des aspects séquentiels. Les jeux de données issus de LSBNs comme Yelp<sup>2</sup>, Foursquare<sup>3</sup>, Twitter<sup>4</sup>, Flickr<sup>5</sup> fournissent un ensemble de POIs et de *check-ins* effectués par les utilisateurs. Néanmoins, ces jeux de données sont souvent agrégés ce qui rend impossible la récupération des préférences personnelles. Les jeux de données issus d'applications développées spécialement pour cette tâche (comme (Schaller *et al.*, 2014a)) sont pertinents pour le problème considéré. Néanmoins, ils ne sont pas disponibles. Enfin, les jeux de données issus de la recherche opérationnelle sont disponibles en ligne<sup>6</sup>. La limitation principale de ces jeux de données est qu'ils ne fournissent aucune information sur les POIs ou les utilisateurs : tous les scores sont pré-calculés.

Dans ce contexte, nous proposons de construire un nouveau jeu de données, composé d'une collection d'activités spatio-temporelles et de profils d'utilisateurs. La collection annotée devra contenir également une hiérarchie des catégories d'activités utilisée pour attribuer une catégorie à chaque activité. La collection d'activités doit former un événement distribué et être composée d'activités concurrentes caractérisées par des horaires de disponibilité et une localisation. Nous considérons les données de 994 animations organisées lors d'une croisière de 7 nuits, disponibles en ligne<sup>7</sup>. Nous souhaitons annoter les données des activités par des utilisateurs afin de créer des profils et d'extraire les centres d'intérêt des utilisateurs, en lançant un projet de crowdsourcing sur Amazon Mechanical Turk<sup>8</sup>. La tâche consistera dans un premier temps, à compléter un questionnaire portant sur les caractéristiques personnelles, l'expérience

---

2. [www.yelp.com](http://www.yelp.com)

3. <https://foursquare.com>

4. <https://twitter.com/>

5. <https://www.flickr.com/>

6. <http://www.mech.kuleuven.be/en/cib/op>

7. <http://disneycruiselineblog.com/2015/07/personal-navigators-7-night-eastern-caribbean-cruise-on-disney-fantasy-itinerary-a-june-20-2015/>

8. <https://www.mturk.com>

et les goûts d'utilisateur, et dans un second temps, à annoter des activités disponibles. Le jeu de données composé sera librement distribué.

Les métriques utilisées pour évaluer les systèmes de recommandation (Shani et Gunawardana, 2011) ne sont pas adaptées au cas de la recommandation de séquences et ne tiennent pas compte de l'aspect séquentiel et de la disponibilité limitée des éléments recommandés, ce qui rend l'évaluation difficile. Nous proposons le protocole d'évaluation suivant. Dans un premier temps, nous évaluerons séparément le calcul des scores d'intérêt via des mesures connues en recommandation (comme celles utilisées par TREC CS (Précision au rang 5 (P@5), Mean Reciprocal Rank (MRR) (Dean-Hall *et al.*, 2014) ou Erreur Absolue Moyenne (MAE) (Shani et Gunawardana, 2011)) et la construction de l'itinéraire par rapport à une solution optimale. Dans un second temps, nous utiliserons le crowdsourcing pour évaluer des résultats obtenus.

## 5. Conclusion et perspectives

Dans cet article, nous avons proposé une nouvelle approche de recommandation de séquences d'activités spatio-temporelles. Elle construit des séquences d'activités personnalisées, en tenant compte de motifs de comportement typique d'utilisateur et de plusieurs facteurs qui sont utilisés pour calculer le score personnalisé des activités.

Nous pouvons distinguer les perspectives à court terme et à long terme. À court terme, nous allons implémenter l'approche proposée, construire le jeu de données décrit et tester l'approche sur ce jeu de données. À long terme, nous allons proposer une approche probabiliste qui tiendra compte du temps de déplacement stochastique et de toute la chaîne des déplacements antérieurs de l'utilisateur.

## Remerciements

Ce projet est soutenu par la Région Rhône Alpes.

## 6. Bibliographie

- Dean-Hall A., Clarke C. L. A., Kamps J., Thomas P., Voorhes E., « Overview of the TREC 2014 Contextual Suggestion Track », *NIST Special Publication 500-308 : The Twenty-Third Text REtrieval Conference Proceedings (TREC 2014)*, 2014.
- Gavalas D., Konstantopoulos C., Mastakas K., Pantziou G., « A survey on algorithmic approaches for solving tourist trip design problems », *Journal of Heuristics*, vol. 20, n° 3, p. 291-328, 2014.
- Li X., Cong G., Li X.-L., Pham T.-A. N., Krishnaswamy S., « Rank-GeoFM : A Ranking Based Geographical Factorization Method for Point of Interest Recommendation », *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '15*, ACM, New York, NY, USA, p. 433-442, 2015.

- Sang J., Mei T., Xu C., « Activity Sensor : Check-In Usage Mining for Local Recommendation », *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 6, n° 3, p. 41 :1-41 :24, April, 2015.
- Schaller R., Harvey M., Elswailer D., « Detecting Event Visits in Urban Areas via Smartphone GPS Data », in M. de Rijke, T. Kenter, A. P. de Vries, C. Zhai, F. de Jong, K. Radinsky, K. Hofmann (eds), *Advances in Information Retrieval*, vol. 8416 of *Lecture Notes in Computer Science*, Springer International Publishing, p. 681-686, 2014a.
- Schaller R., Harvey M., Elswailer D., « Relating User Interaction to Experience During Festivals », *Proceedings of the 5th Information Interaction in Context Symposium, IiX '14*, ACM, New York, NY, USA, p. 38-47, 2014b.
- Shani G., Gunawardana A., « Evaluating Recommendation Systems », in F. Ricci, L. Rokach, B. Shapira, P. B. Kantor (eds), *Recommender Systems Handbook*, Springer US, p. 257-297, 2011.
- Vansteenwegen P., Souffriau W., Oudheusden D. V., « The orienteering problem : A survey », *European Journal of Operational Research*, vol. 209, n° 1, p. 1-10, 2011.
- Vansteenwegen P., Souffriau W., Vanden Berghe G., Van Oudheusden D., « Iterated Local Search for the Team Orienteering Problem with Time Windows », *Computers and Operations Research*, vol. 36, n° 12, p. 3281-3290, December, 2009.
- Yang P., Fang H., « Exploration of Opinion-aware Approach to Contextual Suggestion », in E. M. Voorhees, A. Ellis (eds), *NIST Special Publication 500-308 : The Twenty-Third Text REtrieval Conference Proceedings (TREC 2014)*, 2014.
- Yu Z., Xu H., Yang Z., Guo B., « Personalized Travel Package With Multi-Point-of-Interest Recommendation Based on Crowdsourced User Footprints », *Human-Machine Systems, IEEE Transactions on*, vol. PP, n° 99, p. 1-8, 2015.
- Zhang C., Liang H., Wang K., Sun J., « Personalized Trip Recommendation with POI Availability and Uncertain Traveling Time », *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, CIKM '15*, ACM, New York, NY, USA, p. 911-920, 2015.
- Zhang J.-D., Chow C.-Y., « GeoSoCa : Exploiting Geographical, Social and Categorical Correlations for Point-of-Interest Recommendations », *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '15*, ACM, New York, NY, USA, p. 443-452, 2015a.
- Zhang J.-D., Chow C.-Y., « Spatiotemporal Sequential Influence Modeling for Location Recommendations : A Gravity-based Approach », *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 7, n° 1, p. 11 :1-11 :25, October, 2015b.
- Zhang J.-D., Chow C.-Y., Li Y., « LORE : Exploiting Sequential Influence for Location Recommendations », *Proceedings of the 22Nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL '14*, ACM, New York, NY, USA, p. 103-112, 2014.