HAL
open science

STAS : Recommendation of spatio-temporal activity sequences

Diana Nurbakova, Sylvie Calabretto, Léa Laporte, Jérôme Gensel

To cite this version:

Diana Nurbakova, Sylvie Calabretto, Léa Laporte, Jérôme Gensel. STAS : Recommendation of spatio-temporal activity sequences. Journée Scientifique de l'Arc 6, Nov 2016, Lyon, France. 2016. hal-01575728

HAL Id: hal-01575728

https://hal.science/hal-01575728v1

Submitted on 21 Aug 2017

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.

COMMUNAUTÉS DE RECHERCHE ACADÉMIQUE RhôneAlpes

T.I.C. ET USAGES INFORMATIQUES INNOVANTS

ARC 6 : Technologies de l'Information et de la Communication et Usages Informatiques Innovants

Axe 3 : Mondes numériques pour l'humain et la société : conception, comportements et usages

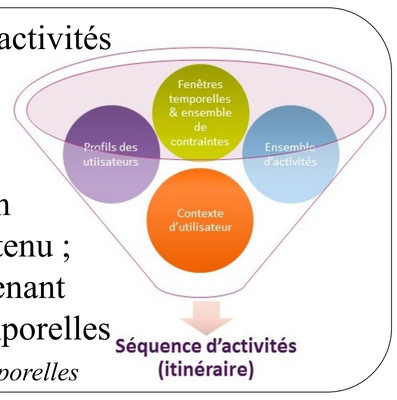
STAS : Recommandation de séquences d'activités spatio-temporelles

La recommandation de séquences d'activités spatio-temporelles est de plus en plus utile et demandée avec la pénétration des systèmes de localisation et des réseaux géo-sociaux dans la vie quotidienne. Nous proposons une approche personnalisée de recommandation de séquences d'activités en contexte mobile et dynamique.

POINTS CLES

Recommandation personnalisée de séquences d'activités tenant compte des préférences de l'utilisateur :

- *Extraction des motifs caractéristiques de comportement de l'utilisateur* ;
- *Calcul des scores personnalisés des activités* en tenant compte de leurs catégories et de leur contenu ;
- *Construction de l'itinéraire personnalisée* en tenant compte de l'ensemble des contraintes spatio-temporelles



Mots-clés : recommandation de séquences, activités spatio-temporelles

PROBLEMATIQUE

Question clé : Étant donné l'historique des activités effectuées par l'utilisateur, comment lui recommander la séquence d'activités qui lui convient le mieux ?

Q1 : Comment extraire *des motifs comportementaux caractéristiques de l'utilisateur* à partir de son historique ?

Q2 : Comment calculer *un score personnalisé* pour chaque activité ?

Q3 : Comment *organiser les activités dans une séquence* en tenant compte de *contraintes spatio-temporelles* ?

Problème STAS (recommendation of **S**patio-**T**emporal **A**ctivities **S**equences)

Positionnement :

	POI Rec	Trip Rec	STAS
Unité de Recommandation	POI, liste de POIs	Séquence de POIs	Séquence d'activités
Granularité	POI (POI = Localisation)	POI (POI = Localisation)	Activité
Granularité : Localisation	∞ : 1 (1:1)	∞ : 1 (1 : 1)	∞ : ∞ (∞ : 1)
Accessibilité Limitée	X	✓	✓
Temps de Déplacement	X	✓	✓
Unité Unique	X	X	✓
Visite Unique	✓	✓	✓
Orientation au futur	X	X	✓

** POI Rec – Recommandation des POIs, Trip Rec – Recommandation des parcours touristiques, STAS – Recommandation des séquences d'activités spatio-temporelles*

Applications :

- Grands événements distribués : conférences et festivals ;
- Tourisme (ex. itinéraire lors d'une croisière, parcours thématique)

Références :

Nurbakova, D., Laporte, L., Calabretto, S., Gensel, J. (2016). ANASTASIA : recommandation de séquences d'activités spatiotemporelles. RJCRI CORIA-CIFED 2016. pp. 325-334

Sang, J., Mei, T., & Xu, C. (2015). Activity Sensor: Check-In Usage Mining for Local Recommendation. *ACM Trans. Intel. Syst.. 6*, pp. 41:1--41:24. ACM.

Vansteenwegen, P., Souffriau, W., Vanden Berghe, G., & Van Oudheusden, D. (2009). Iterated local search for the team orienteering problem with time windows. *Comput. Oper. Res., 36*(12), pp. 3281-3290.

Zhang, J.-D., & Chow, C.-Y. (2015). Spatiotemporal Sequential Influence Modeling for Location Recommendations: A Gravity-based Approach . *ACM Trans. Intel. Syst.* 7 (1), pp. 11:1--11:25. ACM.

TRAVAUX ENGAGES / RESULTATS

Approche Proposée

I. *Extraction des motifs caractéristiques*

Extraction des séquences

Prise en compte des activités successives issues de l'historique de l'utilisateur (LORE (Zhang, 2015); HCS (Sang *et al.*, 2015))

Construction du Graphe de Transition Activité-Activité

- **nœud** – activité a_i caractérisée par le nombre d'**arcs entrants,** $InCount$
- **arc** – transition d'une activité à l'autre, pondérée par la fréquence de transition, $TransCount$. Activités uniques impliquent $TransCount = 1$

* *Adaptation de L2TG (LORE (Zhang, 2015))*

Construction du Graphe de Transition Catégorie-Catégorie

- **nœud** – catégorie c_i des activités

$$InCount(c_i) = \sum_{a_j \in c_i} InCount(a_i)$$

- **arc** - transition d'une catégorie à l'autre

$$TransCount(c_i \to c_j) = \sum_{a_k \in c_j, a_g \in c_i} TransCount(a_g \to a_k)$$

Estimation de la Probabilité de Transition (Niveau des Catégories)

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

II. *Calcul des scores des activités*

Score Catégoriel

proximité aux catégories des activités effectuées par l'utilisateur :

$$Score_{cat}(a_i, u) = \frac{|A_{u,c(a)}|}{|A_u|}$$

- $A_{u,c(a)} = \{a_j\}: a_j \in c(a_i)$, $c(a_i)$ est une catégorie de l'activité a_i
- A_u est l'ensemble des activités effectuées par l'utilisateur u

Score textuel

similarité de contenu textuel entre l'activité et les précédentes activités effectuées par l'utilisateur, basée sur les descriptions des activités :

$$Score_{text}(a_i, u) = \alpha \cdot \cos(U_{pos}, \vec{e}) - (1 - \alpha) \cdot \cos(U_{neg}, \vec{e})$$

- U_{pos} est un profil positif de l'utilisateur composé des vecteurs TD-IDF des activités effectuées par l'utilisateur
- U_{neg} est un profil négatif composé des activités non effectuées par l'utilisateur
- $\vec{e}$ est un vecteur TD-IDF de l'activité a_i
- $\cos(\cdot,\cdot)$ est la similarité cosinus

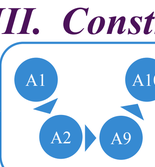
Score hybride

$$Score(a_i, u) = \beta \cdot Score_{cat}(a_i, u) + (1 - \beta) \cdot Score_{text}(a_i, u)$$

- $Score_{cat}$ est un score catégoriel
- $Score_{text}$ est un score textuel

III. *Construction d'itinéraire*

Séquence en cours



Id : Activité : Localisation : Début – Fin : Durée : Score : Wait : MaxShift

A1 : Character Meet & Greet Ticket Distribution : Port Adventures Desk : 11h30 – 15h00 : 10 min : 5 : 0 : 1h20
A2 : Walking Ship Tour : Preludes : 13h00 – 13h30 : 30 min : 5 : 1h20 : 0
A9 : Mandatory Life Boat Drill : Assembly Station : 16h00 – 16h30 : 30 min : 5 : 2h30 : 0
A10 : Sailing Away : Deck Stage : 16h30 – 17h15 : 45 min : 5 : 0 : 0

Calcul du Ratio et Insertion du nœud

Id : Activité : Shift : Ratio : $Ration_{NEW} = \frac{(S*P_T)^2}{Shift}$

A3 : Poolside Jams with Cruise Staff DJ : 0h45 : 0,089 : $= \frac{\left(\frac{2*2}{3}\right)^2}{45} = \mathbf{1,027}$ **max!**

A4 : Spa Open House : 1h30 : 0,044 : $= \frac{(2*0)^2}{90} = \mathbf{0}$

A6 : Acupuncture Demonstration : 2h00 : 0,033 : $= \frac{(2*0)^2}{120} = \mathbf{0}$

A7 : Disney Vacation Club: Members Celebration :1h45 : 0,01 : $= \frac{\left(\frac{1*2}{3}\right)^2}{105} = \mathbf{0,004}$

A8 : Spa Raffle : 2h00 : 0,033 : $= \frac{(2*0)^2}{120} = \mathbf{0}$

Recherche du meilleur nœud à ajouter parmis des candidats

Id : Activité : Localisation : Début – Fin : Durée : Score
A3 : Poolside Jams with Cruise Staff DJ : Deck 11 Stage : 13h45 – 15h15 : 30 min : 2
A4 : Spa Open House : Senses Spa & Salon : 13h45 – 15h00 : 75 min : 2
A6 : Acupuncture Demonstration : Senses Spa & Salon : 14h00 – 15h00 : 60 min : 2
A7 : Disney Vacation Club: Members Celebration : D Lounge : 14h30 – 15h15 : 45 min : 1
A8 : Spa Raffle : Senses Spa & Salon **:** 15h00 – 15h30 : 30 min : 2

Adaptation de l'algorithme *Iterated Local Search* (ILS) (Vansteenwegen et al., 2009)

Evaluation de l'approche

I. Jeu de données : Construction d'un jeu de données (en absence de jeu de données de référence)

II. Expérimentations :

- *Précision de l'estimation des intérêts d'utilisateur :*

Modèle	Binaire	Intérêt gradué
$Score_{cat}$	0.31742	0.35617
$Score_{text}$	0.29457	0.36634
Score(a, u)	**0.32711**	**0.36710**

Evaluation des séquences d'activités (similarité avec la vérité terrain) :

Modèle	ILS_TP	ILS	Gain
$Score_{cat}$	0.28278	0.27947	*1.19%*
$Score_{text}$	0.26472	0.18941	*39.76%*
Score(a, u)	**0.29150**	0.26659	*9.34%*

LIRIS DRIM – Distribution, Recherche d'Information et Mobilité – STeamer – LIG – Spatial and Temporal Information Systems

LIRIS (Laboratoire d'InfoRmatique en Image et Système d'information),
équipe **DRIM** (Distribution, Recherche d'Information, Mobilité)
LIG (Laboratoire d'Informatique de Grenoble),
équipe **STeamer** (Spatial and Temporal Information Systems)

ED 512 INFOMATHS Ecole doctorale — INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON

La Région Auvergne-Rhône-Alpes

Diana Nurbakova[1], Sylvie Calabretto[1], Léa Laporte[1], Jérôme Gensel[2]

[1] LIRIS, INSA de Lyon ; [2] LIG, Université Grenoble Alpes

[1]{diana.nurbakova, sylvie.calabretto, lea.laporte}@insa-lyon.fr ; [2] Jerome.Gensel@imag.fr