

# Intégration des Influences Géographique et Temporelle pour la Recommandation de Points d'Intérêt

Jean-Benoît Griesner\*, Talel Abdessalem\*,\*\*, Hubert Naacke\*\*\*

\*Télécom ParisTech, UMR CNRS LTCI, UMI CNRS IPAL Paris, France  
griesner@telecom-paristech.fr

\*\*School of Computing, National University of Singapore  
talel.abdessalem@telecom-paristech.fr

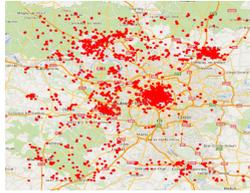
\*\*\*UPMC Université Paris 06, LIP6, Paris, France  
Hubert.Naacke@lip6.fr

**Résumé.** La recommandation de points d'intérêts (ou *POI*), est devenue un problème majeur avec l'émergence des réseaux sociaux (ou *LBSN*). À la différence des approches de recommandation traditionnelles, les données des *LBSN* présentent des caractéristiques géographique et temporelle importantes qui limitent les performances des algorithmes traditionnels existant. L'intégration de ces caractéristiques dans un unique modèle de factorisation pour augmenter la qualité de la recommandation n'a pas été un problème très étudié jusqu'à présent. Dans ce papier nous présentons GeoMF-TD, une extension d'un modèle de factorisation géographique avec des dépendances temporelles. Nos expérimentations sur un jeu de données réel montre jusqu'à 20% de gain sur la précision de la recommandation.

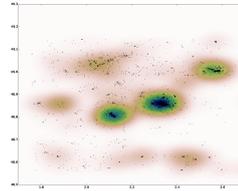
## 1 Introduction

La recommandation de POI dans les *LBSN* est l'activité de faire des suggestions personnalisées de POI à un utilisateur. Du fait des phénomènes d'explosion des données, cette tâche est aujourd'hui devenue une composante essentielle dans les *LBSN*. De nombreuses approches ont été proposées ces dernières années pour répondre à ce besoin (Hu et al. (2008); Lian et al. (2014); Ye et al. (2011)). Ces approches ont démontré leur efficacité et fourni de bons résultats expérimentaux. Cependant l'évolution de la mobilité des utilisateurs dans le temps n'est pas prise en compte. Or l'évolution temporelle des préférences joue un rôle essentiel dans la recommandation.

Ce papier vise à intégrer les caractéristiques temporelles des données, dans un modèle de factorisation. La contribution principale de ce travail est **GeoMF-TD** (Geographical Matrix Factorization with Time Dependencies). Nous montrons par nos expérimentations sur un jeu de données réel que notre modèle offre une meilleure qualité de classement. La suite de ce papier est organisé de la façon suivante. Dans la section 2 nous donnons un panorama succinct de quelques approches de recommandation de POI. La section 3 détaille un modèle de factorisation géographique. Nous présentons notre modèle GeoMF-TD dans la section 4. La section 5



(a) Check-ins de Flickr dans la ville de Paris.



(b) Phénomène de clustering spatial dans Paris par une estimation de densité 2D gaussienne.

FIG. 1: Représentation sur la carte des check-ins dans Paris et de l'estimation de densité.

expose nos résultats expérimentaux obtenus sur un dataset issu de Flickr. Enfin nous concluons dans la section 6.

## 2 Travaux connexes

De nombreuses méthodes ont été proposées pour résoudre le problème de la recommandation de POI. Un domaine de recherche important en recommandation sont les modèles de facteurs latents. Les techniques à base de facteurs latents ont démontré depuis le challenge Netflix de très bonnes performances et ont depuis été largement utilisées (Cheng et al., 2012; Shan et Banerjee, 2010). En marge des approches traditionnelles, il existe entre les POI une influence géographique liée à leur localisation, notamment du fait de la première loi de Tobler (Li et al. (2014)) selon laquelle la probabilité que quelqu'un visite un POI est inversement proportionnelle à la distance qui l'en sépare. De plus il a été observé par Zhang et Chow (2013) que dans un LBSN existe un *spatial clustering phenomenon*, que l'on peut voir sur les figures (1a) et (1b) des check-ins que nous avons collecté dans la région de Paris en France. Il y a peu de temps Ye et al. (2011) ont étudié l'influence géographique des POI dans un modèle traditionnel. De même Cheng et al. (2012) ont présenté récemment un modèle basé sur plusieurs centres gaussiens. Enfin Zhang et Chow (2013) ont proposé un framework de fusion personnalisée basé sur une estimation de densité des distributions de distances entre chaque paire de POI pour chaque utilisateur.

Au-delà de la dimension géographique, la dimension temporelle est un autre facteur important. Récemment Gao et al. (2013) ont proposé un framework de recommandation de destinations avec des effets temporels. Ils ont notamment montré comment modéliser deux des principales propriétés des données (c'est-à-dire la non-uniformité, et la consécutive) avec la factorisation de matrices. Leur idée est de diviser les données en session et d'imposer une régularisation temporelle entre chaque session Une autre approche récente a été proposée par Yuan et al. (2013). Les auteurs incorporent deux scores : un score de recommandation basé sur l'influence des POI, et un autre score basé sur la popularité temporelle des POI. Puis ils normalisent ces deux scores.

### 3 Recommandation de POI dans les LBSN

Soient  $\mathbf{u} = \{u_1, u_2, \dots, u_m\} \subset U^m$  un ensemble d'utilisateurs et  $\mathbf{p} = \{p_1, p_2, \dots, p_n\} \subset P^n$  un ensemble de POI. Soit  $\mathbf{C} \in \mathbb{R}^{m \times n}$  la matrice utilisateur-POI contenant  $m$  utilisateurs et  $n$  POI. La valeur  $c_{u,j}$  dans  $\mathbf{C}$  correspond à la fréquence de visite de l'utilisateur  $u$  au POI  $i$ .

**Factorisation Pondérée de Matrices.** Le but de la factorisation de matrices est d'approximer une matrice  $\mathbf{C}$  par le produit de deux matrices  $\mathbf{P} \in \mathbb{R}^{m \times k}$  et  $\mathbf{Q} \in \mathbb{R}^{n \times k}$  de facteurs latents de dimension  $k \ll \min(m, n)$ . Récemment Hu et al. (2008) ont prouvé que la factorisation de matrices pondérée (WMF) donnait de bons résultats avec les données des LBSN. Cette approche propose de résoudre le problème d'optimisation suivant,

$$\min_{\mathbf{P}, \mathbf{Q}} \left\| \mathbf{W} \odot (\mathbf{R} - \mathbf{P}\mathbf{Q}^T) \right\|^2 + \gamma (\|\mathbf{P}\|^2 + \|\mathbf{Q}\|^2) \quad (1)$$

où  $\odot$  est le produit élément par élément de deux matrices (i.e. le produit d'Hadamard). L'idée principale de WMF est de fixer un seuil de confiance minimum pour chaque POI, visité ou non, encodé dans la matrice  $\mathbf{W}$ .

**Modélisation de l'Influence Géographique.** La modélisation de l'influence géographique pour la recommandation de POI a été largement étudiée dans les travaux précédents Ye et al. (2011); Cheng et al. (2012); Zhang et Chow (2013); Liu et Xiong (2013). Récemment Lian et al. (2014) ont proposé un modèle efficace de factorisation géographique (appelé GeoMF) pour intégrer directement cette influence dans le modèle de WMF. GeoMF divise l'espace en  $L$  zones carré de même taille  $\mathbb{L} = \{g_1, g_2, \dots, g_L\}$  et calcule pour chaque POI son aire d'influence sur chaque zone basée sur une distribution normale des distances. Ainsi GeoMF augmente les matrices traditionnelles  $\mathbf{P}$  et  $\mathbf{Q}$  avec deux matrices de facteurs géographiques  $\mathbf{X} \in \mathbb{R}^{m \times L}$  et  $\mathbf{Y} \in \mathbb{R}^{n \times L}$ . Ainsi nous calculons pour chaque POI  $i$  et pour chaque zone  $g_l$  l'influence géographique gaussienne que  $i$  a sur  $g_l$  :  $y_i^l = \frac{1}{\sigma} K\left(\frac{d(i,l)}{\sigma}\right)$  où  $K(\cdot)$  est la distribution normale standard et  $\sigma$  la déviation standard. Avec ce modèle géographique augmenté, nous obtenons le score de recommandation de l'utilisateur  $u$  pour le POI  $i$  de la façon suivante,  $\widehat{c}_{u,i} = \mathbf{P}_u \mathbf{Q}_i^T + \mathbf{X}_u \mathbf{Y}_i^T$

### 4 GeoMF avec Dépendances Temporelles

Le modèle GeoMF repose sur l'hypothèse que l'espace est homogène et isotropique. Cependant les aires d'influence de deux POI distincts peuvent être très différentes en réalité en considérant différents paramètres autres que la distance. En particulier les effets temporels des fréquences de visite des POI peut jouer un rôle significatif comme l'explique Gao et al. (2013). Dans cette section nous introduisons une approche qui tient compte de cet impact temporel.

**GeoMF-TD.** Dans le modèle GeoMF présenté à la section précédente, les check-ins sont considérés comme étant statiques. En particulier l'aire d'influence de chaque POI suit une loi normale invariable et fixée à priori, basée uniquement sur les distances. Cependant la mobilité des utilisateurs change généralement avec l'évolution du temps, et ces changements peuvent

être exploités pour modéliser le fait qu'un POI  $j$  peut être dans l'aire d'influence d'un autre POI  $i$  sans pour autant être vraiment négatif. Conformément à cette intuition nous proposons d'affecter à chaque paire utilisateur-zone  $(u_i \times g_l) \in \mathbb{L} \times U$  un coefficient temporel  $\psi_i^l \in [0, 1]$  exprimant un "decay" temporel entre le dernier check-in fait par l'utilisateur dans la zone  $g_l$  et l'instant cible  $t_q$ . Ce decay est une fonction monotone décroissante, de façon à donner la plus forte contribution aux check-ins les plus récents.

**Impact temporel et densité spatiale.** Nous proposons de modéliser la proximité de  $t_q$  avec l'instant  $t_{last}^{i,l}$  du dernier check-in fait par l'utilisateur  $i$  dans la zone  $g_l$ . Ainsi nous définissons nos coefficients  $\psi_i^l$  :

$$\psi_i^l = 1 - f\left(\left|t_q \cdot t_{last}^{i,l}\right|\right) \quad (2)$$

où  $\left|t_q \cdot t_{last}^{i,l}\right|$  représente une mesure de l'écart temporel entre  $t_{last}^{i,l}$  et  $t_q$ , et  $f(\cdot)$  correspond à la fonction de decay suivante :  $f(t) = e^{-\lambda \times t}$  avec  $\lambda = \frac{1}{H}$ ,  $H \in \mathbb{R}$  le taux de décroissance de la fonction de decay. Pour modifier l'aire d'influence de chaque POI  $i$  dans la grille  $g_{l \in \mathbb{N}^L}$  nous proposons de fusionner le modèle GeoMF avec nos coefficients temporels. Nous définissons pour l'utilisateur  $u_i$  et le POI  $p_j$  nos coefficients d'influence  $\mathbf{Z}_{i,j} \in \mathbb{R}^L$  :  $\mathbf{Z}_{i,j} = \psi_i^l \odot \mathbf{Y}_j^T$ . Avec ces nouveaux coefficients, Eq 3 devient :  $\widetilde{c_{u,i}} = \mathbf{P}_u \mathbf{Q}_i^T + \mathbf{X}_u \mathbf{Z}_{u,i}^T$ . L'idée principale de ces coefficients spatio-temporels  $\mathbf{Z}_{u,i}$  est de diminuer la "négativité" de POI potentiellement négatifs. Remarquons que la complexité du calcul des  $\psi_i^l$  dépend du nombre de zones et du nombre d'utilisateurs, ce qui est bien moins que la complexité de notre précédente solution (Griesner et al., 2015).

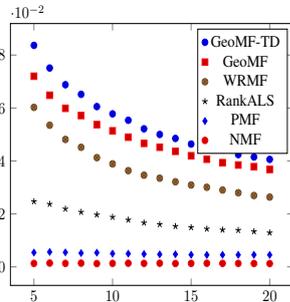
## 5 Évaluation expérimentale

Dans cette section nous présentons une validation expérimentale du modèle présenté dans la section précédente, en utilisant un jeu de données réel issu du LBSN Flickr<sup>1</sup> que nous avons collectées sur 5 ans, enrichi par des données d'Open Street Map<sup>2</sup>.

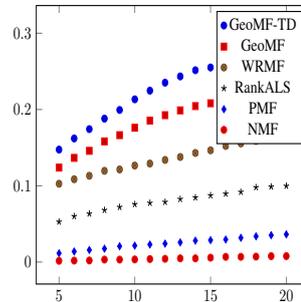
**Environnement expérimental et Métriques d'évaluation.** Nous sélectionnons les utilisateurs ayant au moins 5 check-ins dans des POI distincts, situés dans la région de Paris en France. La figure (1a) présente la distribution des check-ins sélectionnés. Les données utilisées pour les tests correspondent à 57743 check-ins, 3114 utilisateurs et 4544 POI. Pour nos expérimentations nous fixons la taille de chaque zone à 10kms<sup>2</sup>, et le paramètre  $H$  contrôlant le decay à  $4.0 \cdot 10^{-2}$ . L'ordinateur utilisé pour exécuté nos expérimentations possède un processeur Intel Core i7-2600 64 bit, et 8GB de RAM. Nous divisons notre jeu de données en un ensemble d'apprentissage et un ensemble de test comme suit. Pour chaque utilisateur  $u_i \in U$  nous incluons 20% des POI visités les plus récents dans l'ensemble de test, et les 80% restant constituent l'ensemble d'apprentissage. Ensuite nous calculons la *precision@N* and le *recall@N*, correspondant respectivement à la proportion de POI pertinents sur le nombre total  $K$  de POI recommandés, et la proportion de POI pertinents sur le nombre de POI visités précédemment.

1. Flickr : <https://www.flickr.com/>

2. Open Street Map : <https://www.openstreetmap.org>



(a) Comparaison des précisions Top@k.



(b) Comparaison des Recall Top@k.

FIG. 2: Comparaison des performances de GeoMF-TD avec d'autres approches.

**Résultats.** Nous avons implémenté l'approche GeoMF-TD avec LibRec<sup>3</sup>. et nous nous comparons à 5 autres méthodes, à savoir : NMF (Lee et Seung (2000)), PMF, RankALS (Takács et Tikk (2012)), WRMF (Hu et al. (2008)) et GeoMF (Lian et al. (2014)). Les figures 2a et 2b présentent les résultats comparatifs des valeurs de  $precision@N$  et de  $recall@N$ . Comme expliqué dans les sections précédentes, les coefficients temporels introduits permettent d'augmenter la précision globale. Les figures 2a et 2b montrent un gain moyen de 9 % pour la précision.

## 6 Conclusions et Travaux futurs

Dans ce papier nous proposons une analyse des algorithmes de factorisation de matrices basés sur l'influence géographique et temporelle pour la recommandation de POI. Notre contribution est GeoMF-TD, un modèle de factorisation qui considère les caractéristiques temporelles des données. Nous présentons des comparaisons entre 5 différentes autres approches. Notre évaluation expérimentale montre que GeoMF-TD fournit de très bonnes performances. Un travail futur sera de prendre en compte l'intégration en-ligne des changements de préférences des utilisateurs de façon personnalisée. Ce problème fait référence au défi de la dynamique de la recommandation, largement étudié dans de récentes études Gueye et al. (2013, 2015).

## 7 Remerciements

Ce travail a été en partie financé par la Chaire de recherche de Télécom ParisTech Big Data and Market Insights.

## Références

Cheng, C., H. Yang, I. King, et M. R. Lyu (2012). Fused matrix factorization with geographical and social influence in location-based social networks. AAI, 2012.

3. La librairie Java LibRec est accessible ici : <http://www.librec.net/>

- Gao, H., J. Tang, X. Hu, et H. Liu (2013). Exploring temporal effects for location recommendation on location-based social networks. *RecSys '13*.
- Griesner, J., T. Abdesslem, et H. Naacke (2015). POI recommendation : Towards fused matrix factorization with geographical and temporal influences. In *RecSys 2015*.
- Gueye, M., T. Abdesslem, et H. Naacke (2013). Technique de factorisation multi-biais pour des recommandations dynamiques. *EGC'13*.
- Gueye, M., T. Abdesslem, et H. Naacke (2015). Dynamic recommender system : using cluster-based biases to improve the accuracy of the predictions. *Advances in Knowledge Discovery and Management*.
- Hu, Y., Y. Koren, et C. Volinsky (2008). Collaborative filtering for implicit feedback datasets. *ICDM '08*.
- Lee, D. D. et H. S. Seung (2000). Algorithms for non-negative matrix factorization. In *In NIPS*, pp. 556–562. MIT Press.
- Li, T. J.-J., S. Sen, et B. Hecht (2014). Leveraging advances in natural language processing to better understand tober's first law of geography. *SIGSPATIAL '14*. ACM.
- Lian, D., C. Zhao, X. Xie, G. Sun, E. Chen, et Y. Rui (2014). Geomf : Joint geographical modeling and matrix factorization for point-of-interest recommendation. *KDD '14*.
- Liu, B. et H. Xiong (2013). Point-of-interest recommendation in location based social networks with topic and location awareness. *ICDM'13*.
- Shan, H. et A. Banerjee (2010). Generalized probabilistic matrix factorizations for collaborative filtering. *ICDM '10*.
- Takács, G. et D. Tikk (2012). Alternating least squares for personalized ranking. *RecSys '12*, New York, NY, USA, pp. 83–90. ACM.
- Ye, M., P. Yin, W.-C. Lee, et D.-L. Lee (2011). Exploiting geographical influence for collaborative point-of-interest recommendation. *SIGIR '11*.
- Yuan, Q., G. Cong, Z. Ma, A. Sun, et N. M. Thalmann (2013). Time-aware point-of-interest recommendation. *SIGIR '13*.
- Zhang, J.-D. et C.-Y. Chow (2013). igslr : Personalized geo-social location recommendation : A kernel density estimation approach. *SIGSPATIAL'13*.

## Summary

Providing personalized point-of-interest (POI) recommendation has become a major issue with the rapid emergence of location-based social networks (LBSNs). Unlike traditional recommendation approaches, the LBSNs application domain comes with significant geographical and temporal dimensions, which limit performances of most of traditional recommendation algorithms. Fusing geographical and temporal influences into a single factorization model has not been much investigated yet, as far as we know. In this paper, we present GeoMF-TD, an extension of geographical matrix factorization with temporal dependencies. Our experiments on a real dataset shows up to 20% benefit on recommendation precision.