

Détection de communautés dans les réseaux multiplexes : Application aux systèmes de recommandation

Issam Falih

LIPN - UMR CNRS 7030,
SPC, Université Paris 13
99 Av. J-B. Clément, 93430 Villetaneuse
issam.falih@lipn.univ-paris13.fr
<http://lipn.univ-paris13.fr/A3/>

Résumé. Dans cet article, nous proposons une approche de recommandation qui consiste à prédire la note qui sera donnée par un utilisateur à un produit. La détection de communautés dans les réseaux multiplexes est une étape primordiale dans cette approche. Celle-ci a été validée sur le jeu de données MovieLens et le résultat obtenu montre que l'utilisation de la détection de communautés pour la recommandation est une piste encourageante.

1 Introduction

Les systèmes de recommandation suggèrent à un utilisateur des items qui pourraient l'intéresser ou prévoient comment il préférerait chaque item. Ces systèmes aident les utilisateurs à choisir les items qui pourront les intéresser dans une collection. Le filtrage collaboratif est la technique la plus utilisée dans les systèmes de recommandation. Son principe général consiste à prévoir le niveau d'intérêt que porte un utilisateur à un certain item. Il est basé sur l'hypothèse que les utilisateurs ayant aimé des items identiques dans le passé ont un goût similaire et qu'ils apprécieront les mêmes items dans le futur. Généralement, les données sont sous la forme d'une matrice utilisateur-item où chaque utilisateur a donné son degré d'intérêt à certains items. Le tableau 1 illustre un exemple de cette matrice où les lignes représentent les utilisateurs et les colonnes représentent les items. Le problème de recommandation peut se formaliser comme un problème de complétion de cette matrice (notée \mathcal{R}) pour évaluer l'intérêt que porte un utilisateur u_k à un item i_j . Cet intérêt est exprimé ici sur une échelle de valeurs entières.

Malgré le fait que le filtrage collaboratif soit largement utilisé dans les systèmes de recommandation, il fait toujours face à certains défis. En effet, la matrice utilisateur-item est creuse. Un grand nombre d'items n'ont pas été évalués suffisamment par les utilisateurs et ceci rend la tâche de recommandation difficile ce qui affecte la qualité des recommandations par la suite.

	i_1	i_2	i_3	i_4	i_5	i_6
u_1	2	-	-	-	-	4
u_2	-	-	4	-	-	-
u_3	-	5	-	4	3	-
u_4	3	-	1	-	-	2

TAB. 1: Matrice (\mathcal{R}) utilisateur - item

Les techniques de détection de communautés peuvent servir à traiter le problème de recommandation. En effet, la détection de communautés peut être vue comme une généralisation du principe du filtrage collaboratif où le principe est de recommander à une personne les items les bien évalués par les membres de sa communauté. Les produits peuvent être aussi regroupés en communautés selon les motifs de leurs achats ce qui permet de recommander à un client des produits similaires à ce qu'il a aimé dans le passé. Des réseaux multiplexes [2, 3] sont aussi utilisés pour modéliser des interactions typées comme dans les sites de notation où les consommateurs peuvent évaluer sur une échelle de valeurs les produits proposés (hôtels, restaurants, films, etc.). Chaque niveau d'évaluation correspond à une couche dans le réseau multiplexe. Dans ce travail nous proposons une approche de recommandation. Nous appliquons différentes approches de détection de communautés pour retrouver les partitions des utilisateurs et des items.

Le reste de ce travail est organisé comme suit. Dans la section 2 nous présentons notre approche de recommandation. Puis, les expérimentations effectuées ainsi qu'une discussion des résultats dans la section 3. Enfin nous concluons dans la section 4.

2 Système de recommandation proposé

La matrice utilisateur-item peut être représentée comme une matrice d'adjacence d'un graphe biparti valué. En effet, le graphe biparti comprend l'ensemble des utilisateurs et l'ensemble des items. Ils sont liés par des notes qui représentent le niveau d'appréciation d'un item par un utilisateur. On suppose qu'il existe 5 niveaux d'appréciation. Afin d'appliquer les approches de détection de communautés, il est indispensable de transformer ce graphe biparti en graphe simple [5]. Pour ce faire, nous commençons par dissocier les différentes notes les unes des autres. On retrouve donc, cinq graphes bipartis dont chacun représente une note. Nous projetons par la suite les 5 graphes bipartis une fois sur les utilisateurs et une autre fois sur les items. A l'issue de cette étape, nous obtenons 5 graphes simples qui représentent la projection des utilisateurs (resp. items) sur les items (resp. utilisateurs). Chaque graphe correspond à un niveau d'appréciation. Ces différents graphes simples forment un graphe multiplexe. Nous obtenons deux graphes multiplexes. Un graphe multiplexe est un modèle de graphe facilitant la représentation des graphes multi-relationnels. Il comporte le même ensemble de nœuds mais ils sont reliés avec différents types de relations. Ici les relations sont les différents niveaux d'appréciation. Nous appliquons par la suite les approches de détection de communautés sur ces deux graphes multiplexes [2]. Nous obtenons deux partitions : une partition des items et une autre des utilisateurs. Cette étape représente la première partie de notre système de recommandation. La deuxième étape prend en entrée les deux partitions ainsi que la ma-

trice utilisateur-item. L’algorithme 1 montre les grandes lignes de cette deuxième étape. Elle commence par identifier les utilisateurs appartenant à la même communauté que l’utilisateur en question. De la même manière, on identifie la communauté dont l’item appartient. La note prédite sera une fonction d’agrégation des liens entre la communauté d’utilisateurs et la communauté des items. Différentes fonctions d’agrégation peuvent être appliquées notamment la valeur la plus fréquente, la moyenne ou la médiane.

3 Résultats

Critères d’évaluation : Pour évaluer notre approche, on considère les différents critères d’évaluation de la recommandation à savoir la précision, le rappel, F1-mesure, l’erreur moyenne (MAE) et l’erreur type (RMSE).

Résultats : Nous avons appliqué différentes approches de détection de communautés dans les réseaux multiplexes sur le jeu de données Movielens. Ce jeu de données contient 943 utilisateurs (\mathcal{U}), 1682 films (\mathcal{M}) et 100 000 votes (\mathcal{Q}). Les votes sont sur une échelle de valeurs entières comprises entre 1 et 5. Une grande proportion des votes sont des notes de 3 et 4 (votes satisfaisants). Il y a globalement peu de votes d’utilisateurs insatisfaits (1 ou 2). Nous appliquons une procédure de 5-validation croisée. Pour les approches de détection de communautés, nous avons testé les deux familles d’approches notamment celles basées sur les schémas d’agrégation de couches et celles basées sur l’exploration simultanée des différentes couches du multiplexe. Pour l’agrégation des couches, nous avons utilisé les algorithmes Walktrap [4] et Louvain [1]. Pour la deuxième qui explore simultanément les différentes couches du réseau multiplexe nous avons utilisé l’algorithme muxLicod [2]. Le tableau ci-dessous synthétise les résultats obtenus. LA représente l’utilisation de l’approche d’agrégation de couches, et PA représente l’utilisation de l’approche d’agrégation des partitions.

	MAE	RMSE	Precision	Recall	F1-mesure
muxlicod	0.9635	1.2773	0.2274	0.2134	0.2202
LA louvain	0.8352	1.1509	0.3113	0.2521	0.2779
LA walktrap	0.8216	1.1155	0.2642	0.2233	0.2420
PA louvain	0.8713	1.1917	0.2532	0.2032	0.2245
PA walktrap	0.8801	1.2023	0.2705	0.2011	0.2283

TAB. 2: Résultat de recommandation sur la base Movielens en utilisant différentes méthodes de clustering et en considérant la médiane comme une fonction d’agrégation.

Nous observons que les approches basées sur les schémas d’agrégations des couches ou d’agrégations des partitions donnent une qualité de recommandation meilleure que l’approche qui explore simultanément les différentes couches du multiplexe. En comparant les différentes partitions retrouvées, on remarque qu’ils ont un partitionnement divergent. On pourra utiliser les méthodes d’ensemble clustering pour combiner les partitions retrouvées. Ces premiers résultats sont encourageant mais nécessitent bien sûr confirmation sur d’autres jeux de données. Ce travail est actuellement en cours.

4 Conclusion

Dans ce papier, nous avons présenté une nouvelle approche de recommandation. Les premiers résultats ont montré que l'utilisation des algorithmes de détection de communautés pour la recommandation est une piste encourageante. Une confirmation sur d'autres jeux de données reste à faire. Une extension pour l'intégration des attributs des items et des utilisateurs dans les réseaux multiplexes est aussi à l'étude.

Remerciements

Ce travail est financé dans le cadre du projet FUI HERMES.

Références

- [1] Blondel, V. D., J.-l. Guillaume, et E. Lefebvre (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics : Theory and Experiment 2008*, P10008.
- [2] Falih, I., M. Hmimida, et R. Kanawati (2015). Une approche centrée graine pour la détection de communautés dans les réseaux multiplexes. *Revue des Nouvelles Technologies de l'Information Extraction et Gestion des Connaissances, RNTI-E-28*, 377–382.
- [3] Kanawati, R. (2014). Détection de communautés dans les réseaux multiplexes : état de l'art. *RNTI RNTI-A-7*, 20.
- [4] Pons, P. et M. Latapy (2006). Computing communities in large networks using random walks. *J. Graph Algorithms Appl. 10(2)*, 191–218.
- [5] Yeung, A., C. Man, N. Gibbins, et N. Shadbolt (2007). Understanding the semantics of ambiguous tags in folksonomies.