

Identification des utilisateurs atypiques dans les systèmes de recommandation sociale

Benjamin Gras, Armelle Brun, Anne Boyer

Laboratoire Loria - Université de Lorraine
Campus Scientifique, 54506 Vandoeuvre-les-Nancy, France
{benjamin.gras, armelle.brun, anne.boyer}@loria.fr

Résumé. Malgré des performances très satisfaisantes, l'approche sociale de la recommandation ne fournit pas de bonnes recommandations à un sous-ensemble des utilisateurs. Nous supposons ici que certains de ces utilisateurs ont des préférences différentes de celles des autres, nous les qualifions d'atypiques. Nous nous intéressons à leur identification, en amont de la tâche de recommandation, et proposons plusieurs mesures représentant l'atypicité des préférences d'un utilisateur. L'évaluation de ces mesures sur un corpus de l'état de l'art montre qu'elles permettent d'identifier de façon fiable des utilisateurs recevant de mauvaises recommandations.

1 Introduction

Les systèmes de recommandation (SR), visent à recommander à des utilisateurs des ressources pertinentes pour eux. Le filtrage collaboratif (FC) (Resnick et al., 1994) est une des approches les plus populaires de la recommandation.

Bien que la qualité des recommandations fournies par le FC soit considérée comme satisfaisante en moyenne (Castagnos et al., 2013), certains utilisateurs ne reçoivent pas de recommandations de qualité. Le manque de données sur ces utilisateurs est une des raisons possibles expliquant cette mauvaise qualité. Ce problème est appelé démarrage à froid (Schein et al., 2001). Parmi les autres raisons évoquées dans l'état de l'art se trouve la trop grande différence des préférences de ces utilisateurs, par rapport à celles des autres (Haydar et al., 2012). C'est sur cette raison que nous nous focalisons dans cet article. En effet, le filtrage collaboratif suppose une cohérence entre les préférences des utilisateurs ; ces utilisateurs ne respectant pas ce critère, il semble normal qu'ils se voient proposer des recommandations de mauvaise qualité. Ces utilisateurs peuvent aussi être considérés comme des données aberrantes, ou *outliers*. Nous choisissons de les appeler des utilisateurs atypiques.

Notre objectif ici est d'identifier ces utilisateurs atypiques. Nous proposons, dans ce travail préliminaire, plusieurs mesures permettant de les identifier, en exploitant uniquement leurs préférences.

La section 2 se focalise sur les systèmes de recommandation et l'atypisme. Dans la section 3, nous proposons des mesures d'identification des utilisateurs atypiques. Ensuite, nous présentons les expérimentations menées pour valider ces mesures et nous concluons notre travail.

2 État de l'art

La recommandation sociale, ou filtrage collaboratif (FC), exploite les préférences d'utilisateurs (en général des notes sur des ressources) pour estimer des préférences inconnues. L'approche à base de mémoire, et notamment les k plus proches voisins (knn), exploite les similarités de préférences entre utilisateurs. Bien que simple à mettre en œuvre, intégrant dynamiquement les nouvelles préférences et fournissant des recommandations de qualité, cette approche ne passe pas à l'échelle. L'approche à base de modèle souffre moins du problème de passage à l'échelle. La technique de factorisation de matrices (Hu et al., 2008), la plus répandue, forme un sous-espace de caractéristiques latentes, dans lequel utilisateurs et ressources sont représentés, qui permet d'estimer les préférences inconnues.

Dans la littérature, les utilisateurs que nous appelons atypiques sont nommés déviants, anormaux, etc. (Del Prete et Capra, 2010) et la définition qui en est faite varie légèrement. Les travaux dédiés à leur identification sont peu nombreux. La mesure d'anormalité (Del Prete et Capra, 2010; Haydar et al., 2012) aussi appelée coefficient de déviance, déviance, etc., est la plus utilisée pour les identifier. Elle représente la propension qu'a un utilisateur à noter différemment des autres. Elle exploite l'écart entre les notes d'un utilisateur sur des ressources et la note moyenne sur ces ressources (équation (1)).

$$Anormalite(u) = \frac{\sum_{r \in R_u} |n_{u,r} - \bar{n}_r|}{\|R_u\|} \quad (1)$$

où $n_{u,r}$ est la note que l'utilisateur u a donné à la ressource r , \bar{n}_r la note moyenne sur r , R_u l'ensemble des ressources notées par u et $\|R_u\|$ leur nombre. Les utilisateurs dont l'anormalité est très élevée sont considérés comme atypiques. Bien que peu complexe, cette mesure ne tient pas compte du comportement propre à chaque utilisateur et les ressources sur lesquelles les utilisateurs ne sont pas unanimes vont injustement augmenter l'anormalité.

(Bellogín et al., 2011) définit un indicateur de clarté qui identifie les utilisateurs ambigus (instables) dans leur notation, basé sur la mesure de l'entropie. Il a l'inconvénient d'identifier comme instables des utilisateurs dont les préférences évoluent ou dont les préférences diffèrent en fonction des domaines des ressources. Cette mesure ne nous paraît pas adéquate car l'approche sociale peut leur proposer des recommandations de bonne qualité.

(Bellogín et al., 2011; Haydar et al., 2012; Griffith et al., 2012) identifient un lien entre l'erreur commise sur chaque utilisateur et ses caractéristiques (nombre de notes, de voisins, etc.). (Haydar et al., 2012), forme des clusters d'utilisateurs et identifie un cluster d'atypiques : des utilisateurs avec une forte erreur et un fort indice d'anormalité. Cependant, nous pensons que les atypiques ne sont pas toujours similaires entre eux (ce qui en ferait d'ailleurs des utilisateurs non atypiques au sens de la recommandation sociale), le clustering échouera probablement à former des clusters d'atypiques. C'est dans ce sens que va le travail présenté dans (Ghazanfar et Prugel-Bennett, 2011), qui clusterise des utilisateurs et propose de considérer comme atypiques les utilisateurs qui ne sont proches du centre d'aucun des clusters formés.

L'identification d'utilisateurs atypiques peut être associée à l'identification de données aberrantes (outliers) : un outlier est une donnée qui dévie tellement des autres données que cela laisse penser qu'elle a été générée par un mécanisme différent. Les méthodes statistiques et le clustering sont également très utilisées dans le domaine de la détection d'outliers (Aggarwal, 2013).

3 Nouvelles mesures d'identification d'utilisateurs atypiques

Partant des travaux de la littérature, nous proposons de nouvelles mesures permettant l'identification d'utilisateurs atypiques.

CorrKMax - Nous pensons que l'approche *knn* échoue sur les utilisateurs n'ayant pas suffisamment d'utilisateurs similaires. *CorrKMax* représente la similarité moyenne qu'a un utilisateur u avec ses k utilisateurs les plus similaires (équation (2)).

$$CorrKMax(u) = \frac{\sum_{v \in V(u)} CorrPearson(u, v)}{\|V(u)\|} \quad (2)$$

où $CorrPearson(u, v)$ est la corrélation de Pearson entre u et v . $V(u)$ représente les k utilisateurs les plus similaires à u . Nous pensons que les utilisateurs associés à une faible valeur de $CorrKMax(u)$ recevront des recommandations de mauvaise qualité.

Anormalité_{CR} (Anormalité avec Controverse sur les Ressources) - Cette mesure repose sur la mesure d'anormalité de l'état de l'art. Elle suppose que l'écart sur une ressource controversée n'a pas le même sens qu'un même écart sur une ressource consensuelle, ce qui n'est pas considéré par la mesure d'anormalité de l'état de l'art. Nous proposons de pondérer les notes par le degré de controverse de la ressource, fonction de l'écart-type de ses notes.

L'*Anormalité_{CR}* d'un utilisateur u est présentée dans l'équation (3).

$$Anormalite_{CR}(u) = \frac{\sum_{r \in R_u} ((n_{u,r} - \bar{n}_r) * contr(r))^2}{\|R_u\|} \quad (3)$$

où $contr(r)$ est la controverse associée à une ressource r . Cet indice est basé sur l'écart-type normalisé des notes sur r . Où $contr(r) = 1 - \frac{\sigma_r - \sigma_{min}}{\sigma_{max} - \sigma_{min}}$, avec σ_r est l'écart-type des notes de r . σ_{min} et σ_{max} sont respectivement le plus petit et le plus grand écart-type de notes possibles parmi les ressources.

Le calcul d'*Anormalité_{CR}* est d'une complexité comparable à celle de l'anormalité de l'état de l'art. Elle peut donc être calculée fréquemment et ainsi prendre en compte les nouvelles préférences des utilisateurs.

Anormalité_{CRU} (Anormalité avec Controverse sur les Ressources et profil Utilisateur) - Ni *Anormalité* ni *Anormalité_{CR}* ne tiennent compte du comportement général de l'utilisateur auquel elles s'appliquent. Un utilisateur sévère peut être identifié comme atypique alors qu'il n'est atypique que dans sa manière de noter, et non dans ses préférences et qu'il recevra probablement des recommandations de qualité. Pour éviter ce biais, nous proposons de centrer les notes de chaque utilisateur par rapport à sa moyenne de notes. L'*Anormalité* d'un utilisateur u , notée *Anormalité_{CRU}*(u) est calculée selon l'équation (4) :

$$Anormalite_{CRU}(u) = \frac{\sum_{r \in R_u} [(|n_{u,r} - \bar{n}_u - \bar{n}_{C_r}|) * contr_C(r)]^2}{\|R_u\|} \quad (4)$$

où \bar{n}_{C_r} représente la moyenne des notes centrées des utilisateurs sur r , $contr_C(r)$ est l'indice de controverse associé à r , calculé à partir de l'écart-type des notes sur la ressource, centrées par rapport aux utilisateurs.

Le calcul de *Anormalité_{CRU}*(u) est certes plus coûteux en temps que *Anormalité_{CR}*(u), mais devrait permettre une identification plus précise des utilisateurs atypiques. Notons que ces deux dernières mesures sont indépendantes de l'approche de recommandation utilisées (*knn* ou factorisation de matrices), à l'opposé de la mesure *CorrKMax*.

4 Expérimentations

Dans cette section, nous évaluons les mesures d'identification des utilisateurs atypiques que nous proposons, en comparaison avec celles de l'état de l'art.

Nous utilisons le corpus de données de l'état de l'art MovieLens, composé de 100 000 notes (de 1 à 5) de 943 utilisateurs sur 1 682 films (ressources). Une division du corpus en 2 sous-corpus de 80% (pour l'apprentissage) et 20% (pour le test) est effectuée.

L'état de l'art souligne que les utilisateurs pour lesquels le système manque de données (démarrage à froid) reçoivent de mauvaises recommandations. De façon à ne pas biaiser notre évaluation, nous écartons du corpus les utilisateurs associés à du démarrage à froid : ceux avec moins de 20 notes dans le corpus d'apprentissage (Schickel-Zuber et Faltings, 2006). Le corpus est alors réduit à 821 utilisateurs.

4.1 Corrélations entre les mesures et l'erreur de recommandation

Nous évaluons ici la corrélation entre l'erreur commise (RMSE) par une approche *knn* et 4 mesures d'identification des utilisateurs atypiques : la mesure d'*Anormalité* de l'état de l'art, *Anormalité_{CR}*, *Anormalité_{CRU}* et *CorrKMax*. Les corrélations observées sont présentées dans le Tableau 1.

	<i>Anormalité</i>	<i>Anormalité_{CR}</i>	<i>Anormalité_{CRU}</i>	<i>CorrKMax</i>
RMSE	0,453	0,504	0,546	-0,22

TAB. 1 – *Corrélations entre les mesures d'identification et la RMSE (approche knn)*

La mesure *Anormalité* de l'état de l'art présente une corrélation de 0,453 avec la RMSE. Cette corrélation significative confirme le lien existant entre l'anormalité d'un utilisateur et l'erreur commise par une approche *knn* : plus un individu est anormal (atypique), plus l'erreur commise sera élevée. A l'opposé, moins il est anormal, moins l'erreur sera élevée.

Anormalité_{CR} augmente la corrélation de 11% (0,504). La controverse associée aux ressources permet donc d'améliorer l'estimation de la qualité des recommandations fournies aux utilisateurs. *Anormalité_{CRU}* prend également en compte le profil de l'utilisateur. Une corrélation de 0,546 est obtenue (amélioration supplémentaire de 8%, et donc de 20% par rapport à l'état de l'art). La prise en compte des particularités de notation des utilisateurs permet donc d'améliorer l'estimation de la qualité des prédictions fournies aux utilisateurs.

La faible corrélation de la mesure *CorrKMax* (-0,22) indique que, contrairement à notre intuition, la qualité des voisins d'un utilisateur n'est pas corrélée à la qualité des recommandations dans le cadre d'une approche de recommandation *knn*.

4.2 Erreur en prédiction sur les utilisateurs atypiques

Une corrélation représente le lien entre deux variables sur un ensemble d'observations. Cependant, il est possible qu'un lien existe sur une seule partie des observations, ce qui ne sera pas reflété par la corrélation. Ici, nous nous intéressons uniquement aux utilisateurs qualifiés d'atypiques, c'est-à-dire à ceux ayant des valeurs d'anormalité les plus extrêmes. Par conséquent, dans la suite des expérimentations, nous nous intéressons uniquement à la répartition

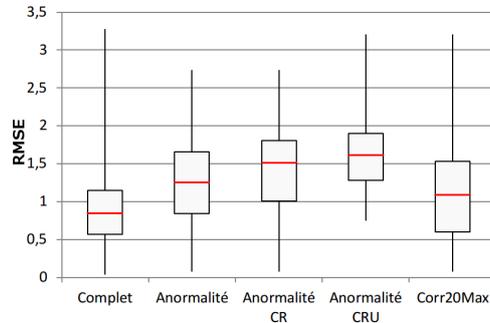


FIG. 1 – Répartition de la RMSE des utilisateurs atypiques avec l'approche *knn*

des erreurs observées sur les utilisateurs considérés comme atypiques. Pour représenter la répartition de ces erreurs, nous exploitons les quartiles et la médiane de ces erreurs, sur les 4 mesures d'anormalité. Plus les erreurs sont élevées, plus nous pouvons considérer que la mesure est de qualité. Nous comparons ces répartitions à celle associée à l'ensemble total des utilisateurs, dénommé Complet (Figure 1).

Nous utilisons un pourcentage d'utilisateurs atypiques fixe, car les mesures n'ont pas des valeurs d'anormalité comparables. Nous avons fixé expérimentalement ce seuil à 6% des utilisateurs, cela correspond à environ 50 utilisateurs parmi les 821 utilisateurs.

L'erreur médiane sur l'ensemble des utilisateurs (Complet) est de 0,82. Celle de *Anormalité* est de 1,26, ce qui correspond à une augmentation de l'erreur de plus de 50%, elle est d'ailleurs équivalente au troisième quartile de l'ensemble complet. Cependant, 25% des utilisateurs qualifiés d'atypiques ont une RMSE plus faible que la RMSE médiane sur l'ensemble des utilisateurs. Par conséquent, *Anormalité* semble sélectionner un grand nombre d'utilisateurs dont les recommandations sont de bonne qualité. *Anormalité_{CR}* et *Anormalité_{CRU}* présentent de meilleurs résultats qu'*Anormalité*. *Anormalité_{CRU}* est la plus performante : plus de 75% des utilisateurs sélectionnés ont une RMSE supérieure à 1,25 : 75% des utilisateurs *Anormaux_{CRU}* font partie des 25% de l'ensemble complet des utilisateurs à recevoir les moins bonnes recommandations. Enfin, environ 50% des utilisateurs sélectionnés avec *CorrKMax* reçoivent de bonnes recommandations, ce qui confirme les premières conclusions obtenues avec la corrélation.

Nous pouvons conclure que *Anormalité_{CRU}* identifie de façon fiable les utilisateurs atypiques : ceux recevant des recommandations de mauvaise qualité avec une approche *knn*.

5 Conclusion et perspectives

Notre objectif dans ce premier travail était d'identifier, en recommandation sociale, les utilisateurs qui reçoivent des recommandations de mauvaise qualité, avant que des recommandations ne leur soient proposées. Nous avons fait l'hypothèse que ces utilisateurs avaient des préférences différentes des autres utilisateurs : des utilisateurs atypiques. Nous avons proposé plusieurs mesures exploitant la similarité de préférence avec les autres utilisateurs, l'écart des notes par rapport aux autres, le consensus de notation sur les ressources et le profil de nota-

tion des utilisateurs. Nous avons montré que, sur un corpus de l'état de l'art, ces informations permettaient de prédire fiablement la mauvaise qualité des recommandations faites à un utilisateur. Nous pouvons donc conclure que les utilisateurs présentant des préférences différentes des autres utilisateurs reçoivent effectivement des recommandations de mauvaise qualité. Ces mesures peuvent donc être utilisées pour anticiper une mauvaise recommandation et la suite de ce travail portera naturellement sur la prise en compte des préférences atypiques des utilisateurs pour leur fournir des recommandations de qualité.

Références

- Aggarwal, C. (2013). An introduction to outlier analysis. In *Outlier Analysis*, pp. 1–40. Springer.
- Bellogín, A., P. Castells, et I. Cantador (2011). Predicting the performance of recommender systems : An information theoretic approach. In *Proc. of the ICTIR'11 conf.*, pp. 27–39.
- Castagnos, S., A. Brun, et A. Boyer (2013). Utilité et perception de la diversité dans les systèmes de recommandation. In *CORIA 2013*, Neuchâtel, Suisse.
- Del Prete, L. et L. Capra (2010). differs : A mobile recommender service. In *Proc. of the MDM '10 conf.*, pp. 21–26.
- Ghazanfar, M. et A. Prugel-Bennett (2011). "fulfilling the needs of gray-sheep users in recommender systems, a clustering solution". In *ISCI conference*.
- Griffith, J., C. O'Riordan, et H. Sorensen (2012). Investigations into user rating information and predictive accuracy in a collaborative filtering domain. In *Proceedings of the SAC Conf.*, pp. 937–942.
- Haydar, C., A. Roussanaly, et A. Boyer (2012). Clustering users to explain recommender systems' performance fluctuation. In *Foundations of Intelligent Systems*, Volume 7661 of *Lecture Notes in Computer Science*, pp. 357–366. Springer.
- Hu, Y., Y. Koren, et C. Volinsky (2008). Collaborative filtering for implicit feedback datasets. In *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, ICDM '08*, Washington, DC, USA, pp. 263–272. IEEE Computer Society.
- Resnick, P., N. Iacovou, M. Suchak, P. Bergstrom, et J. Riedl (1994). Grouplens : An open architecture for collaborative filtering of netnews. In *Proc. of CSCW'94 conf.*, pp. 175–186.
- Schein, A. I., A. Popescul, L. H. Ungar, et D. Pennock (2001). Generative models for cold-start recommendations. In *Proceedings of the 2001 SIGIR workshop on recommender systems*.
- Schickel-Zuber, V. et B. Faltings (2006). Overcoming incomplete user models in recommendation systems via an ontology. In *Proc. of the 7th WebKDD conf.*, Berlin, pp. 39–57. Springer.

Summary

The social approach in recommender systems relies on the hypothesis that users' preferences are coherent between users. Although it provides high quality recommendations, some users never get any high quality recommendations. We think these users have preferences different from others and we propose measures to identify them. The experiments conducted show that they improve the identification of these users, in comparison with the literature.