

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.