

Description interactive de l'intérêt de l'utilisateur via l'échantillonnage de motifs

Moditha Hewasinghage, Suela Isaj, Arnaud Giacometti, Arnaud Soulet

Université François-Rabelais de Tours, LI EA 6300
Campus de Blois, 41000 Blois
prenom.nom@univ-tours.fr

Résumé. La plupart des méthodes d'extraction de motifs requièrent que l'utilisateur formalise son intérêt avec une mesure d'intérêt et des seuils. L'utilisateur est souvent incapable d'explicitement son intérêt mais il saura juger si un motif donné est pertinent ou non. Dans cet article, nous proposons une nouvelle méthode de découverte de motifs interactive en supposant que seule une partie des données est intéressante pour l'utilisateur. En intégrant le retour utilisateur de motifs proposés un à un, notre méthode vise à échantillonner des motifs avec une probabilité proportionnelle à leur fréquence d'apparition au sein des transactions implicitement préférées par l'utilisateur. Nous démontrons que notre méthode identifie exactement les transactions implicitement préférées par l'utilisateur sous réserve de la consistance de ses retours. Des expérimentations montrent les bonnes performances de l'approche en terme de précision et rappel.

1 Introduction

La découverte de motifs est un outil puissant pour extraire des motifs intéressants au sein d'un jeu de données. Néanmoins, la plupart des approches reposent sur le paradigme de la recherche par requête qui peut être une contrainte à satisfaire ou une mesure à maximiser. Il est difficile à l'utilisateur final d'exprimer explicitement son intérêt sous la forme d'une telle requête. Pour cette raison, l'extraction interactive de motifs van Leeuwen (2014) vise à capturer cette requête en observant le retour de l'utilisateur au fur et à mesure que des motifs lui sont présentés. La plupart des méthodes itèrent un cycle en trois phases. Une phase d'extraction extrait des motifs pertinents. Les premières itérations passées, cette pertinence intègre les préférences apprises. En effet, une phase d'interaction glane les retours de l'utilisateur (e.g., notation des motifs) qui seront généralisés par une phase d'apprentissage en un modèle de préférence.

Dans ce contexte, le choix du modèle de préférence s'avère déterminant pour bien identifier l'intérêt de l'utilisateur. Par exemple, le modèle du produit pondéré associe un poids à chaque item et considère alors le score d'un motif comme le produit des poids de ses items Bhuiyan et al. (2012). Dans ce cas, aucune distinction sur les transactions ne sera faite. Les autres propositions de la littérature associent à chaque motif un vecteur de caractéristiques. Grâce aux retours de l'utilisateur, certains vecteurs seront jugés plus pertinents que d'autres et l'apprentissage de ce classement construira un modèle de préférence sur l'ensemble des motifs Rueping