

Extraction de motifs séquentiels contextuels

Julien Rabatel^{*,**}, Sandra Bringay^{*,***}

^{*}LIRMM, Université Montpellier 2, CNRS

161 rue Ada, 34392 Montpellier Cedex 5, France

^{**}TecNALIA, Cap Omega, Rond-point Benjamin Franklin - CS 39521

34960 Montpellier, France

^{***}Dpt MIAP, Université Montpellier 3, Route de Mende

34199 Montpellier Cedex 5, France

{rabatel,bringay}@lirmm.fr

Résumé. Les motifs séquentiels traditionnels ne tiennent généralement pas compte des informations contextuelles fréquemment associées aux données séquentielles. Dans le cas des séquences d'achats de clients dans un magasin, l'extraction classique de motifs se focalise sur les achats des clients sans considérer leur catégorie socio-professionnelle, leur sexe, leur âge. Or, en considérant le fait qu'un motif séquentiel est spécifique à un contexte donné, un expert pourra adapter sa stratégie au type du client et prendre les décisions adéquates. Dans cet article, nous proposons d'extraire des motifs de la forme « *l'achat des produits A et B suivi de l'achat du produit C est spécifique aux jeunes clients* ». En mettant en valeur les propriétés formelles de tels contextes, nous développons un algorithme efficace d'extraction de motifs séquentiels contextuels. Les expérimentations effectuées sur un jeu de données réelles montrent les apports et l'efficacité de l'approche proposée.

1 Introduction

La découverte de motifs séquentiels présente un éventail important d'applications, telles que l'étude de comportement des utilisateurs, de données issues de capteurs, de puces ADN, etc. Les motifs séquentiels visent à extraire des ensembles d'items fréquemment associés au cours du temps. Par exemple, dans le but d'analyser les séquences d'achats de clients dans un magasin, un motif séquentiel peut être « *fréquemment, les clients achètent les produits A et B ensemble, puis achètent le produit C* ». Un tel motif apporte une information générale sur le comportement des clients. Cependant, les données traitées contiennent très souvent des informations additionnelles telles que l'âge ou le sexe des clients.

Les motifs séquentiels traditionnels ne permettent pas de tenir compte de ces informations additionnelles. Or, l'intérêt pour le décideur est immédiat car l'association entre un motif et un contexte lui permet d'adapter sa stratégie et ainsi de mieux coller à la réalité. Par exemple, l'expert pourra obtenir des réponses aux questions « *Quels sont les comportements spécifiques aux clients âgés ?* », « *Existe-t-il des comportements spécifiques aux jeunes hommes ?* » ou encore « *Quels sont les comportements généraux, qui ne dépendent pas du contexte ?* ». Dans cet