

Extension de l'algorithme Apriori et des règles d'association aux cas des données symboliques diagrammes et intervalles

Filipe Afonso^{*,**}, Edwin Diday^{*}

^{*}Ceremade-Université Paris Dauphine, Place du Maréchal de Lattre de Tassigny,
75775 Paris Cedex 16, France

afonso@ceremade.dauphine.fr, diday@ceremade.dauphine.fr

^{**}Lamsade-Université Paris Dauphine

Résumé. Nous traitons l'extension de l'algorithme Apriori et des règles d'association aux cas des données symboliques diagrammes et intervalles. La méthode proposée nous permet de découvrir des règles d'association au niveau des concepts. Cette extension implique notamment de nouvelles définitions pour le support et la confiance afin d'exploiter la structure symbolique des données. Au fil de l'article, l'exemple classique du panier de la ménagère est développé. Ainsi, plutôt que d'extraire des règles entre différents articles appartenant à des mêmes transactions enregistrées dans un magasin comme dans le cas classique, nous extrayons des règles d'association au niveau des clients afin d'étudier leurs comportements d'achat.

1 Introduction

L'algorithme Apriori défini dans (Agrawal et Srikant 1994) a pour but d'extraire des règles d'association à partir de données classiques issues du panier de la ménagère. Les règles sont du type lait→beurre traduisant le fait que "si du lait est présent dans le panier de la ménagère alors il y a aussi du beurre". Dès lors, des travaux ont exploité la complexité des données afin d'accélérer l'exécution de l'algorithme Apriori ou d'enrichir les règles d'association. Ainsi, (Wang et al. 2000) et (Cai et al. 1998) découvrent des règles pondérées par l'importance d'un même article dans le panier alors que (Srikant et al. 1997) et (Han et Fu 1995) exploitent les relations de taxonomie dans les données. De plus, les règles d'association sont étendues aux données quantitatives et intervalles, notamment (Srikant et Agrawal 1996) et (Miller et Yang 1997) alors que dans (Kuok et al. 1998), les auteurs s'intéressent aux ensembles flous. Ainsi, cet article s'inscrit dans le prolongement de ces travaux. Nous étendons l'algorithme Apriori aux cas des variables symboliques à valeurs diagrammes et intervalles afin d'extraire des règles d'association non plus au niveau des individus mais au niveau des concepts. Cette extension implique notamment de nouvelles définitions pour le support et la confiance afin de tirer parti de la structure symbolique des données. Finalement, nous terminons par une application où nous étudions les comportements d'achat dans des magasins non plus au niveau des transactions comme dans le cas classique mais au niveau des clients. Pour chaque client, nous agrégeons les articles achetés grâce à un diagramme construit avec la proportion de chaque article par rapport aux achats totaux du client et nous enrichissons nos règles par l'ajout d'une variable intervalle sur les montants dépensés par le client.