

Extension de l'algorithme Apriori et des règles d'association au cas des données symboliques diagrammes et sélection des meilleures règles par la régression linéaire symbolique

Filipe Afonso

Lamsade et Ceremade-Université Paris 9 Dauphine/ Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, France
afonso@ceremade.dauphine.fr

Résumé. Cet article présente l'extension de l'algorithme Apriori et des règles d'association au cas des données symboliques diagrammes. La méthode proposée nous permet de découvrir des règles au niveau des concepts. Notamment, 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. Enfin, nous proposons une méthode de sélection des meilleures règles d'association selon la régression linéaire symbolique.

1 Introduction

L'algorithme Apriori défini par [Agrawal et Srikant, 1994] a pour but de découvrir des règles d'association à partir de données nominales issues du panier de la ménagère. Les règles extraites sont du type lait \rightarrow beurre traduisant le fait que si du lait est présent dans le panier de la ménagère alors il y a aussi du beurre. A partir de là, des travaux ont voulu exploiter 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 présent dans le panier de la ménagère 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]. [Kuok et al., 1998] font de même en exploitant les ensembles flous. Ainsi, cet article s'inscrit dans le prolongement de ces travaux. Nous traitons la découverte de règles d'association et l'extension de l'algorithme Apriori au cas des variables symboliques diagrammes afin d'extraire des règles non plus au niveau des individus mais au niveau des concepts. Ainsi, après avoir rappelé certaines définitions à propos des données symboliques, nous présentons l'algorithme Apriori et son extension aux cas des données diagrammes, c'est-à-dire lorsque chaque case de notre matrice de données contient plusieurs modalités pondérées telles que la somme des poids soit égale à un. Par la suite, nous étendons les règles d'association au cas de ces données et nous utilisons la régression linéaire afin d'étudier la qualité des règles d'association symboliques. Finalement, nous terminons par un exemple d'application où nous étudions les comportements d'achat dans des magasins non plus au niveau