

Sous-bases k -faibles pour des règles d'association valides au sens de la confiance

Jean Diatta, Régis Girard

IREMIA, Université La Réunion
15, Avenue René Cassin- 97715-St Denis, FRANCE
{ jean.diatte, rgirard } @univ-reunion.fr

Résumé. Nous introduisons la notion de sous-base k -faible pour les règles d'association valides au sens de la confiance. Ces sous-bases k -faibles sont caractérisées en termes d'opérateurs de fermeture correspondant à des familles de Moore k -faiblement hiérarchiques.

1 Introduction

L'un des problèmes majeurs rencontrés dans la fouille des règles d'association valides au sens de la confiance est le nombre souvent très élevé de ces règles. Plusieurs solutions à ce problème ont été proposées ou considérées dans la littérature. Parmi ces solutions figurent les bases, c'est-à-dire, des familles génératrices minimales (Zaki et Ogihara, 1998; Pasquier et al., 1999). La plupart de ces bases se caractérisent en terme d'un opérateur de fermeture de Galois sur l'ensemble des motifs du contexte considéré. Or, cet opérateur de fermeture correspond à une famille de Moore m -faiblement hiérarchique, où $m \geq 2$ est un entier (Diatte, 2004). Plus précisément, les fermés de cet opérateur de fermeture coïncident avec les classes faibles associées à une certaine mesure de dissimilarité m -voies et forment donc, de ce fait, la hiérarchie m -faible associée à cette mesure de dissimilarité.

Dans cet article, nous considérons la caractérisation de ces bases pour les règles d'association, en remplaçant l'opérateur de fermeture de Galois par un opérateur de fermeture correspondant à la hiérarchie k -faible associée à une mesure de dissimilarité k -voies donnée, où $2 \leq k \leq m$. Pour chaque valeur de k , l'ensemble de règles ainsi caractérisé sera appelé sous-base k -faible. Ces sous-bases k -faibles offrent une approximation de l'ensemble des règles valides, relativement à des ensembles d'items (classes k -faibles) ayant un certain degré d'homogénéité exprimé par le biais d'un indice d'isolation. Par ailleurs, la possibilité d'associer une sous-base (k -) faible à une mesure de dissimilarité (k -voies) permet d'intégrer la sémantique de cette mesure de dissimilarité dans le choix des règles à générer.

2 Règles d'association

2.1 Définition générale

Étant donné un contexte binaire $\mathbb{K} = (\mathcal{E}, \mathcal{V})$, où \mathcal{E} désigne un ensemble fini d'entités et \mathcal{V} un ensemble fini de variables booléennes (ou attributs) définies sur \mathcal{E} . On appelle *motifs* les