

Des motifs séquentiels généralisés aux contraintes de temps étendues

Céline Fiot, Anne Laurent,
Maguelonne Teisseire

Laboratoire d'Informatique de Robotique et de Microélectronique de Montpellier
161 rue Ada
34392 Montpellier Cedex 5
{fiot, laurent, teisseire}@lirmm.fr

Résumé. Dans de nombreux domaines, la recherche de connaissances temporelles est très appréciée. Des techniques ont été proposées aussi bien en fouille de données qu'en apprentissage, afin d'extraire et de gérer de telles connaissances, en les associant également à la spécification de contraintes temporelles (e.g.: fenêtre temporelle maximale), notamment dans le contexte de la recherche de motifs séquentiels. Cependant, ces contraintes sont souvent trop rigides ou nécessitent une bonne connaissance du domaine pour ne pas extraire des informations erronées. C'est pourquoi nous proposons une approche basée sur la construction de graphes de séquences afin de prendre en compte des contraintes de temps plus souples. Ces contraintes sont relâchées par rapport aux contraintes de temps précédemment proposées. Elles permettent donc d'extraire plus de motifs pertinents. Afin de guider l'analyse des motifs obtenus, nous proposons également un niveau de précision des contraintes temporelles pour les motifs extraits.

1 Introduction

Dans un certain nombre de domaines (détection de fraudes, de défaillances, analyse de comportements), la recherche de connaissances temporelles est non seulement utile mais nécessaire. Certaines techniques d'apprentissage permettent de gérer et de raisonner sur de telles connaissances, (Allen, 1990) a notamment défini des opérations sur des règles associées à des intervalles de temps. Des techniques d'extraction de connaissances cherchent quant à elles à extraire des épisodes récurrents à partir d'une longue séquence (Mannila et al., 1997), (Raissi et al., 2005) ou de bases de séquences (Agrawal et Srikant, 1995), (Massegli et al., 1998). La recherche de telles informations devient d'autant plus intéressante qu'elle permet de prendre en compte un certain nombre de contraintes entre les événements comme par exemple la durée minimale ou maximale séparant deux événements.

C'est dans ce cadre qu'a été introduite la recherche de motifs séquentiels généralisés dans (Srikant et Agrawal, 1996). Cette technique de fouille de données permet d'obtenir des séquences fréquentes respectant des contraintes spécifiées par l'utilisateur, à partir d'une base de données de séquences (par exemple les achats successifs de différents clients d'un supermarché). Différents algorithmes ont été proposés afin de gérer ces contraintes soit directement dans