

Extraction de motifs temporels à partir de séquences d'événements avec intervalles temporels

Thomas Guyet*, René Quiniou**

*AGROCAMPUS-OUEST

65 rue de Saint-Brieuc, CS 84215, 35 042 Rennes Cedex
thomas.guyet@agrocampus-ouest.fr,

**INRIA, Centre de Rennes - Bretagne Atlantique
Campus de Beaulieu, F - 35 042 Rennes Cedex
rene.quiniou@irisa.fr

Résumé. La fouille de base de données séquentielles a pour objet l'extraction de motifs séquentiels représentatifs. La plupart des méthodes concernent des motifs composés d'événements liés par des relations temporelles basées sur la précédence des instants. Pourtant, dans de nombreuses situations réelles une information quantitative sur la durée des événements ou le délai inter-événements est nécessaire pour discriminer les phénomènes. Nous proposons deux algorithmes, QTIAPriori et QTIPrefixSpan, pour extraire des motifs temporels composés d'événements associés à des intervalles décrivant leur position dans le temps et leur durée. Chacun d'eux ajoute aux algorithmes GSP et PrefixSpan une étape de catégorisation d'intervalles multi-dimensionnels pour extraire les intervalles temporels représentatifs. Les expérimentations sur des données simulées montrent la capacité des algorithmes à extraire des motifs précis en présence de bruit et montrent l'amélioration des performances en temps de calcul.

1 Introduction

Dans de nombreux domaines d'application, médecine (Quiniou et al. (2003)), agronomie (Le Ber et al. (2006)), e-commerce (Massegli et al. (2009)), industrie (Giannotti et al. (2006)), IHM (Cram et al. (2009)), les méthodes de fouille de données sont utilisées pour extraire des motifs séquentiels à partir de séquences. La fouille de séquences s'est principalement intéressée à l'extraction de motifs dont la dimension temporelle ou spatiale repose sur un ordonnancement simple des items (occurrences dans le temps ou dans l'espace des objets composant les motifs). Cependant, dans de nombreux domaines d'application un tel ordonnancement n'est pas suffisant : une caractérisation quantitative des durées des événements ou des délais entre les occurrences de ces événements est nécessaire pour affiner les motifs extraits et produire les connaissances utilisables pour différencier certains comportements. Par exemple, deux achats effectués à un jour, un mois ou un an d'intervalle ne sont pas corrélés de la même façon.

Notre objectif est de définir une méthode de fouille de séquences temporelles composées d'occurrences d'événements pouvant avoir une durée. Le but est d'extraire les motifs séquen-