

# Extension des bases de données inductives pour la découverte de chroniques

Alexandre Vautier\*, Marie-Odile Cordier\*, René Quiniou\*

\* Irisa - Projet DREAM Campus de Beaulieu 35042 RENNES Cedex, France  
{Alexandre.Vautier,Marie-Odile.Cordier,Rene.Quiniou}@irisa.fr

**Résumé.** Les bases de données inductives intègrent le processus de fouille de données dans une base de données qui contient à la fois les données et les connaissances induites. Nous nous proposons d'étendre les données traitées afin de permettre l'extraction de motifs temporels fréquents et non fréquents à partir d'un ensemble de séquences d'événements. Les motifs temporels visés sont des chroniques qui permettent d'exprimer des contraintes numériques sur les délais entre les occurrences d'événements.

## 1 Introduction

L'enjeu de la *fouille de données* est d'extraire des connaissances cachées à partir d'une grande quantité de données disponibles. En ne disposant, a priori, d'aucune information particulière sur les données, on souhaite découvrir des connaissances cachées sous forme de *motifs*. Dans ce cadre, on appelle motif un concept (une loi, une règle, une classe,...) synthétisant un ensemble de données.

Les bases de données inductives (BDIs), introduites par Imielinski et Mannila (Imielinski et Mannila 1996), sont nées de la nécessité de formaliser la fouille de données et d'établir des liens clairs avec les concepts et les algorithmes utilisés dans le domaine des bases de données. L'un des buts des BDIs, est de récupérer l'efficacité de cette gestion au profit de la fouille de données. Dans une BDI, les données et leurs motifs sont gérés de la même façon. Ainsi, le processus de fouille de données devient un processus interactif dans lequel l'utilisateur soumet des requêtes à la BDI qui lui fournit en résultat des motifs qui sont, soit entreposés directement dans la BDI, soit « induits » par un processus basé sur la recherche de motifs intéressants, par exemple fréquents.

Dans les BDIs actuelles (Lee et De Raedt 2003, Dzeroski 2002, Imielinski et Mannila 1996, De Raedt 2002) les données sont des séquences de symboles et les motifs sont des séquences complexes de symboles. Nous proposons d'adjoindre une dimension temporelle numérique aux BDIs. En effet, de nombreux problèmes de découverte de connaissances nécessitent d'intégrer d'autres phénomènes qu'une simple succession de symboles. Ainsi, dans une BDI étendue au temps, une donnée est une *séquence d'événements* et un motif est une *chronique* (Dousson et Ghallab 1994) constituée d'un ensemble de variables temporelles typées et soumis à des contraintes numériques.

La partie 2 expose les techniques d'extraction de motifs temporels proches des chroniques. La partie 3 détaille l'extension des BDIs : une relation de généralité entre chroniques permettant de structurer l'espace de recherche et un critère de reconnaissance utilisé pour définir la fréquence d'une chronique dans une séquence d'événements sont introduits. La partie 4 expose comment traiter une requête à la base de donnée. L'algorithme de l'espace des versions (Mitchell 1997) est modifié pour utiliser les chroniques minimales et fréquentes d'une séquence d'événements calculées par une adaptation d'un outil de fouille de données FACE (Dousson et Duong 1999). Enfin une illustration et des expérimentations montrent l'intérêt de cette approche.