

Extraction de chroniques discriminantes

Yann Dauxais*, David Gross-Amblard*, Thomas Guyet** et André Happe***

*Université Rennes-1/IRISA

prenom.nom@irisa.fr

**Agrocampus-Ouest/IRISA

*** CHRU Brest - équipe REPERE

Résumé. L'extraction de motifs séquentiels vise à extraire des comportements récurrents dans un ensemble de séquences. Lorsque ces séquences sont étiquetées, l'extraction de motifs discriminants engendre des motifs caractéristiques de chaque classe de séquences. Cet article s'intéresse à l'extraction des chroniques discriminantes où une chronique est un type de motif temporel représentant des durées inter-événements quantitatives. L'article présente l'algorithme *DCM* dont l'originalité réside dans l'utilisation de méthodes d'apprentissage automatique pour extraire les intervalles temporels. Les performances computationnelles et le pouvoir discriminant des chroniques extraites sont évalués sur des données synthétiques et réelles.

1 Introduction

La fouille de données temporelles, *e.g.* séries temporelles ou des séquences, est dédiée à l'analyse de données portant une information temporelle. De telles données sont largement rencontrées dans les domaines tels que la médecine, l'ingénierie ou la finance. Pour ces domaines, l'aspect temporel est primordial et il est crucial de proposer des approches capables d'analyser ces données en tenant compte de cette spécificité. Dans cet article, nous nous concentrons sur l'extraction de motifs dans des séquences d'évènements datés. Nous cherchons à extraire des motifs pouvant servir à discriminer précisément des comportements associés aux séquences Fradkin et Mörchen (2015). Il s'agit, par exemple, d'associer une suite d'évènements à l'état pathologique d'un patient.

L'extraction de motifs séquentiels, *i.e.* tenant compte de la séquentialité des évènements, mais pas de leurs dates, a été largement étudiée et fait l'objet de plusieurs états de l'art (Massegia et al., 2004; Mooney et Roddick, 2013). L'extraction de motifs séquentiels est efficace pour une contrepartie non-négligeable : l'information temporelle n'est pas prise en compte dans son intégralité. Les motifs décrivant une information temporelle riche tels que les motifs d'intervalles (Guyet et Quiniou, 2011) ou les chroniques (Dousson et Duong, 1999; Cram et al., 2012; Huang et al., 2012) captent une information plus riche. De ce fait, leur utilisation permet des prédictions plus précises.

D'autre part et bien que l'objectif d'extraction de motifs séquentiels soit la prédiction d'évènements, la plupart des approches de fouille se sont intéressées à extraire des motifs fréquents, *i.e.* qui apparaissent fréquemment dans la base d'exemples. L'extraction de motifs