

Énumération des occurrences d'une chronique

Thomas Guyet*, Philippe Besnard**, Ahmed Samet***,
Nasreddine Ben Salha***, Nicolas Lachiche****

*AGROCAMPUS-OUEST/IRISA UMR 6074 IRISA

**CNRS/IRIT

***INSA Strasbourg/Laboratoire ICube

****Université de Strasbourg/Laboratoire ICube

Résumé. Une chronique est une représentation du comportement d'un système dynamique formalisée comme un ensemble d'événements liés par des contraintes temporelles. Dans cet article, nous nous intéressons à l'énumération des occurrences exactes ou approchées d'une chronique en comparant différentes stratégies de reconnaissances. Nous proposons un algorithme de reconnaissance de chroniques avec différentes heuristiques et sa version approchée. L'efficacité de ces différentes approches est évaluée sur des données synthétiques.

1 Introduction

Pour un grand nombre de contextes applicatifs, des données de nature séquentielle sont collectées. Par exemple, il peut s'agir d'une séquence de soins dont un patient a bénéficié, d'une séquence d'achats qu'a effectué un client dans un supermarché ou encore de *log* de fonctionnement d'une machine industrielle. Il est intéressant d'explorer ces données pour extraire de l'information sur les situations qui se sont effectivement déroulées.

L'approche présentée dans cet article cherche à interroger une base de séquences avec un motif défini par l'utilisateur. Un motif correspond à une situation dont l'utilisateur cherche les occurrences. Généralement un motif tient compte de l'ordre des événements mais rarement de l'intervalle de temps entre ces événements. Pourtant, cette dimension temporelle peut contenir une information pertinente. Par exemple, les occurrences d'une situation où une hospitalisation survient après qu'une personne ait fait une chute est moins informatif qu'une situation correspondant à une hospitalisation survenue moins de 24h après la chute.

Dans la littérature, certaines approches retrouvent les occurrences d'une situation complexe dans des séquences d'événements en prenant en compte la dimension temporelle. Les méthodes de raisonnement temporel ou les logiques d'action répondent en particulier au problème de la reconnaissance de situations (Ulus et Maler, 2018; Artikis et al., 2012). Ces approches sont très expressives mais manquent d'efficacité. Les méthodes efficaces de reconnaissance font appel à des méthodes d'indexation (Kostakis et Papapetrou, 2017), mais sont peu expressives. Dans le domaine de la surveillance automatique, les approches basées sur les automates temporisés (Alur et Dill, 1994; Carle et al., 2011) permettent de spécifier des comportements avec des modèles formels. Ces représentations peuvent être difficiles à appréhender par des