

Fouille de motifs temporels négatifs

Katerina Tsesmeli*, Manel Boumghar***, Thomas Guyet**
René Quiniou*, Laurent Pierre***

*Univ Rennes, Inria, CNRS, IRISA
prenom.nom@inria.fr

**Agrocampus Ouest, IRISA - UMR 6074

***EDF R&D Saclay

Résumé. Dans cet article nous étudions le problème de l'extraction de motifs fréquents contenant des événements positifs, des événements négatifs spécifiant l'absence d'événement ainsi que des informations temporelles sur le délai entre ces événements. Nous définissons la sémantique de tels motifs et proposons la méthode NTGSP basée sur des approches de l'état de l'art. Les performances de la méthode sont évaluées sur des données commerciales fournies par EDF (Électricité de France).

1 Introduction

Dans de nombreux domaines d'application tel que le diagnostic, la santé ou le marketing, les praticiens s'intéressent aux événements qui sont corrélés à des événements indésirables ou qui les déclenchent. Souvent, l'occurrence de la situation indésirable s'explique par la présence d'une action spécifique, mais aussi par l'absence de certains événements (Cao et al., 2016). Par exemple, dans le cadre du marketing, si un client de supermarché n'a pas reçu de promotions depuis longtemps, il a une très forte probabilité de choisir la concurrence, alors que dans le cas contraire il resterait fidèle à son enseigne. Pour réagir au mieux, il est important de découvrir les événements ainsi que leurs contextes d'occurrence ou d'absence afin de déterminer la meilleure action à exécuter pour éviter la situation indésirable, comme l'attrition en marketing. De plus, il est important de connaître les caractéristiques temporelles des situations indésirables, *i.e.* de quelle manière il faut anticiper l'occurrence de ces situations. Par exemple, ne pas envoyer d'offre promotionnelle dans les trois jours précédant le jour habituel de courses d'un client peut conduire ce client à acheter moins de produits, mais lui envoyer une telle offre trop tôt peut n'avoir aucun effet. Une information temporelle, comme les délais acceptables, peut améliorer la précision des prédictions et de la recommandation d'action auprès des décisionnaires.

Notre but est d'extraire d'une base de séquence d'événements datés des motifs temporels indiquant l'absence de certains événements appelés *événements négatifs* (Cao et al., 2016). Ainsi, le motif négatif $p = \langle a \neg c b \rangle$ indique que a est suivi fréquemment de b sans la présence de c entre eux. De plus, nous souhaitons traiter la dimension temporelle, *i.e.* savoir quel laps de temps s'écoule entre une occurrence de a et une occurrence de b en l'absence de c .

Peu de travaux se sont intéressés à la fouille de motifs séquentiels négatifs et, à notre connaissance, le présent travail est le premier qui concerne la fouille de motifs temporels né-