

La prise en compte de la dimension temporelle dans la classification de données

Eloïse Loubier *, Bernard Dousset *

* I.R.I.T. (Institut de Recherche en Informatique de Toulouse),
118 route de Narbonne, 31062 TOULOUSE Cedex 9
{[loubier](mailto:loubier@irit.fr), [dousset](mailto:dousset@irit.fr)}@irit.fr;

Résumé. Dans un contexte d'ingénierie de la connaissance, l'analyse des données relationnelles évolutives est une question centrale. La représentation de ce type de données sous forme de graphe optimisé en facilite l'analyse et l'interprétation par l'utilisateur non expert. Cependant, ces graphes peuvent rapidement devenir trop complexes pour être étudiés dans leur globalité, il faut alors les décomposer de manière à en faciliter la lecture et l'analyse. Pour cela, une solution est de les simplifier, dans un premier temps, en un graphe réduit dont les sommets représentent chacun un groupe distinct de sommets : acteurs ou termes du domaine étudié. Dans un second temps, il faut les décomposer en instances (un graphe par période) afin de prendre en compte la dimension temporelle.

La plateforme de veille stratégique Tétralogie, développée dans notre laboratoire, permet de synthétiser les données relationnelles évolutives sous forme de matrices de cooccurrence 3D et VisuGraph, son module de visualisation, permet de les représenter sous forme de graphes évolutifs.

VisuGraph assimile les différentes périodes à des repères temporels et chaque sommet est placé en fonction de son degré d'appartenance aux différentes périodes. Ce prototype est aussi doté d'un module de la classification interactive de données relationnelles basé sur une technique de Markov Clustering, qui conduit à une visualisation sous forme de graphe réduit. Nous proposons ici de prendre en compte la dimension temporelle dans notre processus de classification des données. Ainsi, par la visualisation successive des différentes instances, il devient plus facile d'analyser l'évolution des classes au niveau intra mais aussi au niveau inter classes.

1 Introduction

L'étude de la migration des termes, en particulier de l'évolution des données relationnelles issues de la synthèse de grands corpus d'information est un aspect majeur dans l'ingénierie de la connaissance et en particulier dans le cadre de la veille. Dans ce contexte, le recours à la visualisation de données par des graphes apporte un réel confort aux utilisateurs, qui, de façon intuitive, peuvent s'approprier une forme de connaissance difficile à décrire autrement. Bien souvent, ces graphes sont trop complexes pour être étudiés dans leur globalité, il faut alors les décomposer de manière à faciliter la lecture et l'analyse des données. Une première simplification du graphe est réalisée par le biais de la classification en un graphe réduit dont les sommets représentent chacun un groupe distinct d'acteurs ou de