

Approche Fouille de Texte pour la détection précoce de tendances économiques

Marilyne Latour*,** Antoine Sigwalt**,***

*Université Grenoble-Alpes, GRESEC, F-38040 Grenoble
marilyne.latour@ac-grenoble.fr,
<http://gresec.u-grenoble3.fr/>

**ReportLinker, 4 Rue Montrochet, 69002 Lyon
marilyne.latour@reportlinker.com
antoine.sigwalt@reportlinker.com
<http://www.reportlinker.com>

***Ecole Centrale Lyon
antoine.sigwalt@ecl2014.ec-lyon.fr
<http://www.ec-lyon.fr>

Résumé. Cet article présente un retour d'expérience sur de la fouille de données complexes dans un processus d'extraction des connaissances dans un contexte industriel. À partir de données volumineuses non structurées issues de dépêches d'actualités économiques et selon certains traitements linguistiques et économétriques, notre objectif est de prédire des tendances économiques dans des séquences d'événements d'actualités. Pour cela, trois étapes sont primordiales : (i) l'extraction d'indicateurs économiques par des techniques linguistiques (comme les indices boursiers, les taux de change, les noms des monnaies ou encore les cours des matières premières...), (ii) l'annotation, par le recours à des terminologies externes, de ces indicateurs économiques : les données extraites portent alors des étiquettes permettant de les identifier, (iii) leur superposition à des modèles statistiques. À la suite de ce traitement, nous pouvons vérifier si il existe une corrélation entre des indicateurs économiques relevés par l'étude linguistique pour un secteur d'activité donné et sur un territoire donné (la production d'un élément *A* sur le prix d'un élément *B* par exemple). L'intérêt de cette méthode est d'apporter des outils linguistiques en complément des méthodes statistiques utilisées habituellement pour faire émerger des données cointégrées. L'article décrit ensuite les expérimentations effectuées et tire les premières conclusions sur divers aspects de cette méthode.

1 Introduction

Dans les ensembles volumineux et non structurés des données, l'accès facilité et enrichi à l'information est devenu un véritable marché économique : la compréhension d'un secteur d'activité - notamment ses prévisions de marché - est un angle de recherche particulièrement