

# Extraction de sentiments et d'opinions basée sur des règles

Sigrid Maurel\*, Paolo Curtoni\*, Luca Dini\*

\* CELI France, SAS  
12-14, rue Claude Genin  
38000 Grenoble  
{maurel, curtoni, dini}@celi-france.com  
<http://www.celi-france.com>

**Résumé.** Nous présentons ici trois méthodes différentes pour effectuer une classification automatique de textes d'opinion. La première méthode est symbolique, la seconde statistique et la dernière, hybride, est une combinaison des deux premières. Nous montrons comment la combinaison des méthodes symbolique et statistique permet de tirer parti des avantages des deux méthodes, à savoir la robustesse de l'apprentissage automatique statistique et la possibilité de configuration manuelle offerte par la méthode symbolique, permettant une utilisation dans des applications réelles. Les textes classés par ces méthodes viennent de sources informationnelles non structurées de type forum sur Internet.

## 1 Introduction

### 1.1 Motivation

Cet article s'intéresse à la classification de textes d'opinion en langue française. Dans ce cas précis, la classification a pour objectif l'analyse de sentiments exprimés dans différents types de textes comme par exemple dans des forums de discussion sur Internet où les internautes échangent des avis et s'entraident. Les textes issus de forums sur Internet constituent des sources d'informations spontanées et récentes, incontournables pour acquérir, au jour le jour, des connaissances sur les consommateurs, pour anticiper leurs besoins et leurs attentes afin de tenter d'améliorer la relation client/fournisseur. En analysant ces textes d'opinion le fournisseur d'un produit ou d'un service peut mieux réagir aux desiderata de ses clients, le client peut de son côté s'inspirer des sentiments et opinions d'autres clients sur le produit auquel il s'intéresse et profiter ainsi d'une aide à la décision (acquérir ou non le produit, choisir plutôt le produit A ou le produit B, etc.).

Comme le montrent de nombreux travaux de socio- et psycho-linguistique (Sproull et Kiesler, 1991), la communication médiée par ordinateur favorise l'expression des émotions, sentiments et opinions souvent contrôlés ou réprimés dans des cadres de communication plus traditionnels visant à étudier le point de vue des consommateurs (interviews face à face, enquêtes fermées, enquêtes ouvertes, etc.). De là, naît l'intérêt des analystes pour ces sources d'informations.