

Extraction d'opinions appliquée à des critères

Benjamin Duthil*, François Troussel*
Gérard Dray*, Pascal Poncelet**, Jacky Montmain*

*EMA-LGI2P, Parc Scientifique Georges Besse, 30035 Nîmes Cedex, France
prénom.nom@mines-ales.fr

**LIRMM UM2 - CNRS 5506, 161 Rue Ada, 34095 Montpellier Cedex 5, France
Pascal.Poncelet@lirmm.fr

Résumé. Les technologies de l'information et le succès des services associés (e.g., blogs, forums,...) ont ouvert la voie à un mode d'expression massive d'opinions sur les sujets les plus variés. Récemment, de nouvelles techniques de détection automatique d'opinions (*opinion mining*) ont fait leur apparition et via des analyses statistiques des avis exprimés, tendent à dégager une tendance globale des opinions exprimées par les internautes. Néanmoins une analyse plus fine de celle-ci montre que les arguments avancés par les internautes relèvent de critères de jugement distincts. Ici, un film sera décrit pour un scénario décousu, là il sera encensé pour une bande son époustouflante. Dans cet article, nous proposons, après avoir caractérisé automatiquement des critères dans un document, d'en extraire l'opinion relative. A partir d'un ensemble restreint de mots clés d'opinions, notre approche construit automatiquement une base d'apprentissage de documents issus du web et en déduit un lexique de mots ou d'expressions d'opinions spécifiques au domaine d'application. Des expériences menées sur des jeux de données réelles illustrent l'efficacité de l'approche.

1 Introduction

Avec le développement du Web, de plus en plus de documents textuels sont disponibles et de plus en plus d'outils permettent de rechercher de l'information pertinente. Connaître l'opinion des personnes sur un produit, rechercher et classer des documents, indexer de manière automatique des documents sont des problématiques d'actualité (e.g. Xu et al. (2011); Bai (2011); Morinaga et al. (2002)). Par exemple, dans le cas de l'opinion de cinéphiles, de nombreux outils sont disponibles pour connaître l'avis général des spectateurs d'un film. Traditionnellement, pour extraire ces opinions, deux grandes approches existent : (i) celles basées sur un corpus d'apprentissage qui s'appuient généralement sur une analyse syntaxique et de co-occurrence des mots ; (ii) et celles utilisant un "dictionnaire" spécifique (e.g. *SenticNet* Cambria et al. (2010)) pour obtenir des orientations sémantiques d'un mot (Xu et al. (2011); Kamps et al. (2004)). Même si ces approches sont très efficaces, elles souffrent pour les premières de la nécessité de devoir constituer un corpus d'apprentissage. Dans des travaux précédents (Harb et al. (2008)), nous avons mis en avant le fait que l'expression d'opinions