

Combiner analyse syntaxique de surface et apprentissage supervisé pour la fouille d'opinion ciblée : expérimentations sur des données d'opinion concernant les livres

Jeanne Villaneau*, Stefania Pecore*
Farida Saïd^{1, **}, Pierre-François Marteau*

* IRISA, Campus de Tohannic, Université de Bretagne-Sud, 56000 Vannes
jeanne.villaneau, stefania.pecore, farida.said, pierre-francois.marteau@univ-ubs.fr

** LMBA, Campus de Tohannic, Université de Bretagne-Sud, 56000 Vannes

Résumé. La fouille d'opinion ciblée est une tâche complexe, susceptible de bénéficier de l'apport d'approches variées. Nos expérimentations testent des combinaisons de méthodes sur un corpus d'avis d'internautes concernant les livres. Sur ces données et pour ce qui concerne la polarité de l'opinion, des résultats prometteurs ont été obtenus par une approche basée sur une analyse linguistique de surface et un lexique enrichi par les informations discriminantes basées sur des méthodes de classification statistiques supervisées.

1 Introduction

La fouille d'opinion est devenue un champ de recherche important du Traitement Automatique des Langues (TAL) et sa maturité est attestée par les nombreux états de l'art dont elle a fait l'objet [Chapate et al. (2015); Feldman (2013); Liu (2012), etc.]. Son objectif est d'analyser l'ensemble des opinions des internautes sur un objet donné. La tâche se veut plus précise avec la fouille d'opinion ciblée (ABSA; Aspect Based Sentiment Analysis) [Liu (2012)] : les cibles correspondent aux différentes caractéristiques des objets concernés et les avis des internautes sont analysés en fonction de ces cibles. L'ABSA a fait l'objet de différents challenges, parmi lesquels SemEval-2014, 2015 et 2016 [Pontiki (2016)].

Ciblée ou non, la fouille d'opinion est une tâche extrêmement complexe pour de multiples raisons : forme indirecte de l'expression, humour, subjectivité, nécessité d'un traitement fin de la négation, etc. Les approches utilisées sont nombreuses et variées, généralement divisées entre méthodes statistiques, généralement supervisées [Pang et Lee (2008)] et méthodes basées sur un lexique d'opinion¹, à base de règles [Neviarouskaya et al. (2010)] ou statistiques [Wilson et al. (2005)], une tendance actuelle étant de combiner différentes approches.

La fouille d'opinion ciblée peut être partagée en trois sous-tâches : (i) détection de phrases ou portion de phrases porteuses d'opinion, (ii) détection de la cible relative à l'opinion émise, (iii) détection de la polarité et éventuellement, de l'intensité de l'opinion en question. Nos expérimentations concernent les deux dernières. Par ailleurs, les travaux présentés analysent des

1. Pour un état de l'art des méthodes basées sur l'utilisation d'un lexique, cf. Taboada et al. (2011).