

Détection d'opinions dans des tweets

Caroline Collet*, Alexandre Pauchet*, Laurent Vercoüter*, Khaled Khélif*

*LITIS-Avenue de l'Université - BP 8 76801 Saint-Étienne-du-Rouvray Cedex

Résumé. Twitter est à l'heure actuelle un des réseaux sociaux les plus utilisés au monde et analyser les opinions qui y sont contenues permet de fournir de précieuses informations notamment aux entreprises commerciales. Dans cet article, nous décrivons une méthode permettant de déterminer l'opinion d'un tweet en détectant dans un premier temps sa subjectivité, puis sa polarité.

1 Introduction

De nos jours, les institutions étatiques tout comme les entreprises, s'appuient souvent sur l'opinion publique pour orienter leurs décisions stratégiques. L'analyse automatique d'opinions a ainsi connu une véritable envolée depuis l'apparition des réseaux sociaux tels que Twitter. Selon Pak et Paroubek (2010), une opinion peut être soit positive, négative, ou neutre, ce qui revient à un problème de classification en 3 classes. Un second point de vue, que nous adopterons, consiste à considérer qu'une opinion ne peut pas être neutre, seulement objective. Ainsi, le problème peut être décomposé en une évaluation de l'objectivité dans un premier temps, suivie dans le cas d'un texte subjectif d'une seconde étape de détection de la polarité.

L'objectif de ces travaux, dans le cadre d'un projet européen de tourisme, est de présenter une méthode pour détecter la subjectivité puis la polarité d'un tweet anglais ou français. Dans la suite de cet article, nous présentons section 2 un état de l'art sur la détection d'opinions dans des textes. La section 3 décrit notre méthode de détection de subjectivité et son évaluation sur des tweets. La section 4 est dédiée à la détection de polarité. Enfin, la section 5 résume et analyse les résultats obtenus et propose des pistes pour nos travaux futurs.

2 La détection d'opinions dans des textes

La détection de subjectivité et de polarité sont des tâches similaires pouvant être résolues par le même type de méthodes. Il en existe 3 à l'heure actuelle : les méthodes symboliques, statistiques et hybrides.

Les méthodes symboliques créent un ensemble de règles de décision servant à classer un texte dans une catégorie. Ces règles peuvent être éditées de manière manuelle (très coûteux) ou semi-manuelle par un expert humain et sont directement appliquées aux textes à évaluer. Les règles définies peuvent être plus générales (moins coûteuses) en utilisant par exemple des formules telles que des calculs de valences Baccianella et Sebastiani (2010); Serban et Pécuchet (2012). Ces approches ne permettent néanmoins pas la détection d'opinions pour des structures linguistiques compliquées.