

Une plateforme d'analyse d'opinions en temps réel sur Twitter avec recommandation

Noureddine Azzouza*, Karima Akli-Astouati*
Samy Ait- Bachir*, Amira Oussalah*

*FEI- Département Informatique, Laboratoire RIIMA
Université de Sciences et de Technologie Houari Boumedién (USTHB)
Bab ezzouar, alger, Algérie
n_azzouza@esi.dz, kakli@usthb.dz, {samyab2, oussalah.amira}@gmail.com

1 Introduction

Le suivi des opinions propose beaucoup de défis dans le domaine du "Data Analytics ", qui est utilisée dans plusieurs domaines tel que le marketing, le "Sentiments Analysis" (Di Capua et al., 2015), ou même la détection des événements (Gaglio et al., 2016).

Dans notre projet, il est question de suivre la propagation des opinions en temps-réel sur le réseau social Twitter, et ce en utilisant l'outil Storm de la fondation Apache, puis d'utiliser des méthodes de visualisation de données afin d'obtenir une représentation dynamique des résultats obtenus.

2 Plateforme de suivi des opinions en Temps réel

Pour la conception de notre plateforme, nous avons opté pour l'utilisation d'une approche basée dictionnaire afin de déterminer la polarité et l'intensité d'une opinion dans un tweet et pour les pré-traitements sur les textes.

2.1 Module d'acquisition des tweets

Ce module est chargé de récupérer des tweets selon des mots-clés, en continu, puis de les transmettre au reste du système. Les tweets doivent être filtrés selon la langue.

2.2 Module de pré-traitement des tweets

Ce module est précédé par une étape de tokenisation qui va séparer les tokens des émojis et émoticônes. Ensuite ce module va effectuer trois étapes dans son processus : La correction lexicale, syntaxique et la normalisation des tokens. La Correction lexicale permet de faire une correction de l'argot, remplacer les acronymes et permettre aussi de reconnaître les expressions de langue. Dans ce module, les statuts sont étiquetés morpho-syntaxiquement. L'étiquetage permet de dissiper l'ambiguïté autour d'un terme pouvant avoir plusieurs étiquettes.

2.3 Module d'analyse des opinions

Ce module détermine la polarité et l'intensité des opinions présentées dans les tweets analysés. On commence par le calcul des valeurs des opinions véhiculées par les émojis/émoticônes qui sont considérés comme un moyen de transmettre une émotion, et donc laisser paraître des avis et des opinions. Les valeurs d'opinion de chacun des émoticônes/émojis sont récupérées à partir d'un dictionnaire. De la même façon, la valeur d'opinion des mots d'un tweet est calculée en faisant la somme des valeurs d'opinions de chacun des mots normalisés récupérés à partir d'un dictionnaire. Grâce aux étiquettes des mots, certaines ambiguïtés peuvent être levées. La valeur d'opinion finale du tweet est la moyenne des deux valeurs obtenues.

2.4 Modules de visualisation

La visualisation des données est faite via des représentations graphiques des scores calculés précédemment. Ces représentations se mettent à jour continuellement, pour refléter l'état des données en temps-réel. L'utilisateur visualise les données construites par le système grâce à des représentations graphiques dynamiques, s'adaptant avec le temps aux différentes opinions issues des tweets analysés. En effet, on peut retrouver le timeline des tweets avec polarités, une génération de listes dynamiques, un histogramme sur l'évolution du nombre de tweets et un graphe de suivi des opinions.

2.5 Modules de recommandation

Cette partie exploite les tweets collectés précédemment, et propose à l'utilisateur une courte liste de mots-clés recommandés. Ces mots-clés représentent les mots les plus employés par la twittosphère aux côtés des mots-clés de la recherche initiale. Nous avons réalisé des recommandations basant sur la fréquence des mots et sur le nombre de tweets dans lesquels ils apparaissent.

Références

- S. Gaglio, G. Lo Re et M. Morana (2016). A Framework for Real-time Twitter Data Analysis. *In Journal of Computer Communications*, January 2016.
- M. Di Capua, E. Di Nardo et A. Petrosino (2015). An Architecture for Sentiment Analysis in Twitter. *International Conference on E-learning*, Berlin, Germany. September 2015.

Summary

In this paper, our contribution is to achieve a platform for opinion tracking in real time about a subject or a person represented as a keyword. The results are presented in four dynamic graphic visualizations. Also, it is possible to issue positive or negative recommendations in the form of keyword list.