
RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEURE ET DE LA RECHERCHE SCIENTIFIQUE
ECOLE SUPÉRIEURE EN INFORMATIQUE DE SIDI BEL ABBÈS (ESI-SBA)



MÉMOIRE DE FIN D'ÉTUDE

Pour l'obtention du diplôme d'ingénieur d'état

Spécialité : Système d'information et web (SIW)

Une approche Deep learning pour l'analyse des sentiments
multi-langages pour l'opérateur téléphonique Ooredoo

Réalisé par:

Laouni Hossem

Seddik Walid

Encadré par :

Pr BENSLIMANE Sidi Mohammed

Mme KLOUCHE Badia

Année Universitaire : 2019/2020

Résumé

L'analyse des sentiments, (aussi appelé l'opinion mining) est l'un des domaines récents de recherche émergent, en matière de classification de polarité des sentiments et de fouille de textes, en particulier avec le nombre considérable d'opinions disponibles dans les médias sociaux. Il concerne l'extraction automatique des opinions exprimées dans un texte donné. En raison de ses vastes applications, de nombreuses études ont été menées dans le domaine de l'Opinion Mining, en particulier dans les textes en arabe algérien dialectal, tandis que d'autres langues utilisées, telles que l'arabe, le français et l'anglais, ont fait l'objet d'une attention amoindrie. Les opérateurs sont exigeants et prêtent une grande importance à l'opinion de leur clientèle, qui utilise un langage arabe dialectal dans les différents réseaux sociaux, tout en introduisant les sentiments dans leur plate-forme respective et leur enjeu stratégique

Dans ce travail nous avons proposé une approche d'apprentissage profond **Deep Learning** pour l'analyse des sentiments **multi-langages** des commentaires des clients de l'opérateur téléphonique algérien **OOREDOO**. A cet effet, nous avons utilisé les réseaux de neurones convolutifs **CNN** ainsi que les réseaux Long Short-Term Memory **LSTM**, qui abordent l'analyse de l'Arabe Standard Moderne (ASM) et du dialecte algérien et des autres langues. Nous avons proposé **SentiOoredoo**, un outil qui permet d'analyser les sentiments des clients de l'opérateur téléphonique ooredoo représentés par des commentaires Facebook, des SMS, ainsi que des tweets en temps réelle avec l'api **Tweepy** de **Twitter**.

Dans notre travail, le jeu des données **SentiOoredoo dataset** est annoté manuellement. Nous avons aussi utilisé la dataset **Sentiment140** contenant plus de 1.6 millions de tweets considérée comme étant la plus grande dataset dans le domaine de l'analyse des sentiments .

Mots clés : Analyse de sentiment, Traitement automatique de la langue, Apprentissage Profond, Apprentissage automatique, Classification, TF-IDF, Word2Vec, CNN, LSTM .

Abstract

Sentiment analysis , a natural language processing task, has been a subject of interest in the past and in the present day for the research community. This problem has a lot of solutions offered with different machine learning classification techniques. One of the more recent techniques is deep learning which we find a lot of contributions have used as a solution to this type of task and also shows good results.

In this work we proposed a **deep learning** approach for the analysis of feelings **multi-languages** at **OOREDOO Algeria** , precisely for this work we used the convolutional neural networks **CNN** and also Long Short-Term Memory **LSTM** networks, which tackles analyzing at both Modern Standard Arabic (MSA) and Algerian Arabic and other languages.

Our system is called **SentiOoredoo** to analyze the feelings of the files of the ooredoo telephone operator which contains comments or SMS, and to analyze real-time feelings tweets are already posted on **Twitter** using API **Tweepy** .

We have proposed our manually annotated **SentiOoredoo dataset** and also we have used the largest dataset in this domain **Sentiment140** which contains 1.6 million tweets.

Key Words: Sentiment Analysis, Deep learning , Machine Learning, Text Classification, Opinion and Sentiment Extraction, , Convolutional neural networks CNN , Long Short-Term Memory LSTM .