



جامعة محمد الخامس بالرباط
Université Mohammed V de Rabat

École Nationale Supérieure d'Informatique et d'Analyse des Systèmes
Centre d'Études Doctorales en Sciences des Technologies de l'Information et de l'Ingénieur

AVIS DE SOUTENANCE DE THÈSE DE DOCTORAT

Madame Naaima BOUDAD

Soutiendra publiquement sa thèse de Doctorat en Informatique

Le Vendredi 17 Mars 2023 à 16H00 au Grand amphi à l'ENSIAS

Intitulé de la thèse

**SENTIMENT ANALYSIS IN ARABIC: FROM CONVENTIONAL MACHINE
LEARNING TO TRANSFER LEARNING**

Devant le Jury composé de :

Président :

Pr. Mohammed RZIZA, PES, Faculté des Sciences, Université Mohammed V, Rabat

Directeur de thèse :

Pr. Rachid OULAD HAJ THAMI, PES, ENSIAS, Université Mohammed V de Rabat

Co-Directeur de thèse :

Pr. Rdouan FAIZI, PES, ENSIAS, Université Mohammed V de Rabat

Rapporteurs :

Pr. Karim BAINA, PES, ENSIAS, Université Mohammed V de Rabat

Pr. Abdelalim SADIQ, PES, Faculté des Sciences, Université Ibn Tofail, Kenitra

Pr. Mohammed RAMDANI, PES, FST Mohammedia, Université Hassan II, Casablanca

Examineur :

Pr. Mahmoud El HAMLAOUI, PH, ENSIAS, Université Mohammed V de Rabat



Résumé: Opinion mining ou analyse des sentiments est un récent domaine de recherche qui consiste à créer des outils automatiques capables d'extraire et identifier les émotions humaines exprimées dans les textes. Avec la croissance rapide des quantités de données textuelles créées par les utilisateurs des réseaux sociaux, et vu l'intérêt des opinions recueillies pour les domaines commerciaux, économiques et politiques, un grand intérêt a été porté par la recherche au domaine des analyses des sentiments. Ainsi, plusieurs travaux de recherche ont été réalisés dans ce domaine. Cependant, la majorité de ces recherches ont été menées surtout sur des langues européennes (notamment l'Anglais), et que peu d'études ont traité les langues riches morphologiquement comme l'Arabe. Dans cette thèse, nous examinons différentes méthodes d'apprentissage automatique pour mener une classification supervisée des sentiments exprimés au niveau des textes Arabe. Nous menons d'abord un ensemble d'expériences de classification en combinant les algorithmes d'apprentissage classiques avec multiples représentations du texte basées sur des variables caractéristiques de type syntactiques, sémantiques et statistique. Les résultats de ces expériences démontrent l'importance et l'impact du choix des variables caractéristiques adéquates pour représenter le texte. Ensuite, nous abordons l'utilisation des méthodes d'apprentissage profond pour permettre l'extraction et la sélection automatique des variables caractéristiques. Nous constatons que l'utilisation d'un réseau de neurones à convolution avec des vecteurs d'incorporation de mots pré-entraînés en arabe surpasse les méthodes basées sur un large ensemble de caractéristiques spécifiques à la langue Arabe. Finalement, nous présentons une approche d'apprentissage par transfert basée sur des modèles de type BERT pour la classification des sentiments des textes écrits en dialecte marocain. Dans ce cadre, nous présentons notre propre modèle basé sur l'architecture du BERT et pré-entraîné avec des textes en dialecte marocain. Nous évaluons donc les performances de l'approche d'apprentissage par transfert pour l'analyse des sentiments en moyennant notre modèle marocain et d'autres modèles similaires arabes et multilingues. En conclusion, nous montrons que cette approche peut conduire à des meilleurs résultats.

Mots-clés : Analyse de Sentiment, Apprentissage Automatique, Apprentissage profond, Classement de texte.

Abstract: Sentiment analysis (SA), also termed opinion mining, is an active sub-research area of Natural Language Processing. This discipline aims to define the sentiment polarity of a piece of writing. With the increasing popularity of social media platforms, new opportunities and challenges arise as sentiment analysis can be employed to have insights about individuals' opinions. The application of the SA on several different domains has attracted interest from researchers, especially for English language that has already accomplished important improvement and success. In contrast, research on Arabic SA is still in its early stages and more research is needed to achieve the similar degree of success as English. Actually, most of the existing studies are limited to Modern Standard Arabic



جامعة محمد الخامس بالرباط
Université Mohammed V de Rabat

(MSA) language and a few have dealt with Arabic dialects and noisy text such as social media messages. In this thesis, we investigate different machine learning approaches to carry out sentiment classification of multidialectal Arabic texts. Specifically, we first report experiments evaluating handcrafted features-based approach for SA by employing a combination of a set of syntactic, semantic and statistic text features with classical supervised classifiers. Results of these experiments show that selecting appropriate features is a crucial task to improve performance of classification. As using handcrafted features is limited and makes the task more complicated, we explore the effectiveness of using deep learning approaches to automatically learn features and overcome the need of adapting features to each dialect. We find that using a convolution neural network with Arabic pre-trained word embedding vectors outperforms methods based on large set of multidialectal Arabic specific features. Finally, we propose a Bert-based transfer learning approach for Moroccan SA, to tackle the low resource issue in this Arabic variant. In this context, we introduce our Moroccan Bert-based model and perform extensive experiments to investigate the performance of transferring contextual embedding from Moroccan dialect, Arabic language, and multilingual modelling to Moroccan SA downstream task. We conclude that the TL approach leads to improved results even when using a Moroccan model trained on a small corpus.

Keywords: Sentiment Analysis, Text classification, Deep Learning, Transfer learning, BERT-based models.

