



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

Monsieur El Mahdi MERCHA

Soutiendra publiquement sa thèse de Doctorat en Informatique

Le Samedi 28 septembre 2024 à 10h00 au Grand Amphi à l'ENSIAS de Rabat

Intitulé de la thèse

**Heterogeneous Text Graph and Slightly Deep Graph
Convolutional Network for Comprehensive Multilingual
Sentiment Analysis**

Président :

Pr. Mohammed BENKHALIFA, PES, FSR, Université Mohammed V de Rabat

Directeur de thèse :

Pr. Houda BENBRAHIM, PES, ENSIAS, Université Mohammed V de Rabat

Rapporteurs :

Pr. Kamel SMAILI, Professeur des Universités, LORIA, Université Lorraine, France

Pr. Khalid CHOUGDALI, MCH, ENSA, Université Ibn Tofail, Kenitra

Pr. Yasser EL MADANI EL ALAMI, MCH, ENSIAS, Université Mohammed V de Rabat

Examineur :

Pr. Sanaa EL FKIHI, PES, ENSIAS, Université Mohammed V de Rabat

Invité :

Pr. Mohammed ERRADI, PES, ENSIAS, Université Mohammed V de Rabat



Résumé : L'apparition et la croissance rapide du Web, des réseaux sociaux et d'autres forums en ligne ont entraîné la génération continue et rapide de données textuelles exprimant des opinions. Plusieurs applications du monde réel se concentrent sur la détermination des sentiments exprimés dans ces données. En raison de la nature multilingue des données générées, il existe un besoin croissant de réaliser une analyse de sentiment multilingue pour prendre en charge les données dans différentes langues. L'analyse de sentiment multilingue désigne la tâche de comprendre les opinions, les sentiments et les émotions des personnes dans des textes écrits multilingues. Elle a été largement étudiée en raison de son importance dans la récupération d'informations pour la prise de décision dans divers domaines, tels que le marketing, la finance et la politique. Dans cette thèse, nous proposons une exploration approfondie des techniques utilisées pour mener une analyse de sentiment à travers les langues, en mettant particulièrement l'accent sur les méthodologies multilingues et interlangues, et nous introduisons une catégorisation systématique de ces méthodes.

De plus, nous proposons deux principales approches d'analyse de sentiment multilingue basées sur les réseaux de convolution graphique : la première est *Multilingual Sentiment Analysis Approach Based on Graph Convolution Network (MSA-GCN)*, où nous créons un seul graphe textuel hétérogène basé sur des informations séquentielles, sémantiques et statistiques pour modéliser l'ensemble des données d'un corpus multilingue, puis un réseau de convolution graphique légèrement profond est utilisé pour apprendre des représentations de documents prédictives. La deuxième approche est *"inductive multilingual sentiment analysis via slightly deep graph convolutional network and Bi-LSTM with attention (Ind-MSA)"* qui combine un réseau de convolution graphique légèrement profond et une mémoire à court et long terme bidirectionnelle avec un mécanisme d'attention dans une architecture unifiée pour capturer à la fois les sémantiques à courte et longue distance tout en prenant en charge l'ordre des mots. Des expériences complètes sur divers ensembles de données démontrent l'efficacité des deux approches. De plus, ces approches montrent des résultats remarquables dans diverses combinaisons de langues, révélant leur robustesse face à la variation linguistique.

Mots-clés : Analyse des sentiments multilingues Apprentissage automatique, apprentissage profond, Recherche d'information, Réseau de convolution de graphes, Traitement du langage naturel.

Abstract: The inception and rapid growth of the Web, social media, and other online forums have resulted in the continuous and rapid generation of opinionated textual data. Several real-world applications have been focusing on determining the sentiments expressed in these data. Owing to the multilinguistic nature of the generated data, there exists an increasing need to perform multilingual sentiment analysis (MSA) to support data in diverse languages. MSA refers to the task of understanding people's opinions, sentiments, and emotions in multilingual



written texts. It has been widely studied due to its significance in retrieving insights for decision-making in a variety of fields, such as marketing, finance, and politics. In this thesis, we offer a comprehensive exploration of techniques employed in conducting sentiment analysis across languages, with a special focus on multilingual and cross-lingual methodologies. We also introduce a systematic categorization of these methods. Furthermore, we propose two main multilingual sentiment analysis approaches based on graph convolution networks: the first one is a *“multilingual sentiment analysis approach based on graph convolution network (MSA-GCN),”* where we create a single heterogeneous text graph based on sequential, semantic, and statistical information to model the entire data of a multilingual corpus, and then a slightly deep graph convolution network is used to learn predictive document representations. The second approach is an *“inductive multilingual sentiment analysis via a slightly deep graph convolutional network and Bi-LSTM with attention (Ind-MSA)”* that blends a slightly deep graph convolutional network and bi-directional long short-term memory with attention mechanism into a unified architecture to capture both short- and long-distance semantics while supporting the word ordering. Comprehensive experiments on various datasets demonstrate the efficiency of both approaches. In addition, these approaches show prominent results in a variety of language combinations, revealing their robustness against language variation.

Keywords: Cross-lingual sentiment analysis, deep learning, graph convolutional network, information retrieval machine learning, multilingual sentiment analysis, natural language processing.