

Analyse de sentiments à base d'apprentissage profond dans les forums de discussions MOOC

Ahlem Mejri*, Dr. Mariem Mahfoudh*,**, Dr. Nihel Kooli***

*Université de Kairouan, Tunis, mejriahlem25@gmail.com

**Laboratoire MIRACL, Université de Sfax, Tunis, mariem.mahfoudh@gmail.com

*** Direction générale de l'armement de France, France, Nihel.Kooli@gmail.com

1 Introduction

Selon (Liu, 2012), l'analyse des sentiments (AS) est "le domaine d'étude qui analyse les opinions, les sentiments, les évaluations, les attitudes et les émotions des gens à partir du langage écrit". Elle vise à analyser les avis pour suivre l'humeur du public à propos d'un produit ou d'un sujet particulier. Notre travail s'intéresse au domaine d'apprentissage en ligne, plus particulièrement à l'AS des avis publiés sur les cours en lignes ouvertes et massives (MOOC). Le MOOC est un type de formation en ligne qui offre aux étudiants/apprenants la possibilité de suivre une formation à distance. Au cours de la formation, les apprenants ont tendance à discuter et publier fréquemment leurs avis sur les forums de discussions, rendant ainsi le travail des instructeurs très difficile en matière d'analyse des avis. Des travaux de la littérature ont cherché à analyser automatiquement les polarités des avis (Nikolić et al., 2020). Cependant, ils se sont concentré sur l'identification des sentiments selon trois classes : négatifs, neutres et positifs.

Nous utilisons des techniques de TAL avec une approche basée sur l'apprentissage profond (CNN) pour déterminer leurs polarités émotionnelles. Les avis sont classés en fortement positif, positif, fortement négatif, négatif, ou neutre.

2 Méthode proposée

Notre approche d'AS sur les MOOC comprend cinq étapes principales :

1. Collection des données : consiste à collecter les commentaires des apprenants de site MOOC Coursera. Ils sont 140320 avis.
2. Pré-traitement des données : consiste à appliquer certaines techniques de TAL, comme la tokenisation, le stemming, la suppression des mots vides sur les avis afin de bien les préparer à la phase de l'analyse.
3. Extraction des caractéristiques : utilise le TF/IDF et le word embedding pour transformer les avis en représentations numériques.

4. Classification des sentiments : adapter les représentations numériques des revues aux réseaux de neurones profonds CNN. Nous avons choisi cet algorithme suite à une comparaison. A titre d'exemple, l'algorithme Random nous a donné 88% comme Accuracy, SVM 90% alors que CNN(Modelseq3) 93%.
5. évaluation : la réparation des avis à cinq classes a donné : 22460 avis positive, 106516 fortement positive, 5923 neutre, 2554 négative et 2867 avis fortement négative). Ex.
– 1. it's not really attractive (négative); – 2. BOring (Très négative); – 3. Great course - I recommend it for all (Très positive); – 4. nice teacher (positive); – 5. It's ok (neutre).
Le tableau 1 donne une comparaison avec quelques autres approches.

TAB. 1 – *Tableau comparatif de performance dans le dataset Coursera*

Auteurs	Technique	Précision
Rossi et Gnawali (2014)	SVM + noyaux linéaires	89%
Chaplot et al. (2015)	SentiWordNet + Réseau neuronal	77%
Notre Travail	TAL + CNN	93%

3 Conclusion

Nous avons proposé une approche d'analyse de sentiments sur le corpus MOOC Coursera. Elle se base sur le modèle CNN et permet d'identifier cinq classes (fortement négatif, négatif, neutre, positif, fortement positif) avec une précision de 93%. Cette classification nous aidera à mieux comprendre les besoins des apprenants pour une meilleure recommandation des cours.

Références

- Chaplot, D. S., E. Rhim, et J. Kim (2015). Predicting student attrition in moocs using sentiment analysis and neural networks. In *AIED Workshops*, Volume 53, pp. 54–57.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies* 5(1), 1–167.
- Nikolić, N., O. Grljević, et A. Kovačević (2020). Aspect-based sentiment analysis of reviews in the domain of higher education. *The Electronic Library*.
- Rossi, L. A. et O. Gnawali (2014). Language independent analysis and classification of discussion threads in coursera mooc forums. In *15th International Conference on Information Reuse and Integration*, pp. 654–661. IEEE.