




BERT and Prerequisite Based Ontology for Predicting Learner's Confusion in MOOCs Discussion Forums

Abdessamad Chanaa^(✉)  and Nour-Eddine El Faddouli 

RIME Team, MASI Laboratory, E3S Research Center, Mohammadia School
of Engineers (EMI), Mohammed V University (UM5), Rabat, Morocco
abdessamad.chanaa@gmail.com, nfaddouli@gmail.com

Abstract. The use of Massive Open Online Courses (MOOCs) is rapidly increasing due to the convenience and ease that provide to learners. However, MOOCs suffer from high drop out rate owing mostly to the confusion and frustration going with the learning process. Based on MOOCs discussion forums, this paper aims to explore different levels of confusion in specific concept using prerequisite based ontology for extracting relevant posts, and Bidirectional Encoder Representations from Transformers (BERT) classification algorithm to describe the degree of confusion for each post. The analysis of discussion posts from Stanford University dataset affirms the effectiveness of our model. BERT achieve good classification accuracy; this will help in early drop out detection and also facilitate future support for learners in confusion state.

Keywords: Massive Open Online Courses (MOOCs) · Confusion · Text classification · Prerequisite based ontology · BERT classification algorithm

1 Introduction

Over the few past years, Massive Open Online Courses (MOOCs) have witnessed a significant evolution in the academic and industrial community. MOOCs give more flexibility and convenience in taking the course through many helpful learning experiences to students, such as video lectures, assignments, exams. It also provides the opportunity to connect and collaborate with others through discussion forums. Despite this great success, MOOCs still suffer from a high drop out rate [7]. Although many causes exist for this problem, students' confusion and frustration are one of the main reason behind it. Confusion can be defined as a blockage or dilemmas where the learner is uncertain how to proceed with the learning process. In MOOCs, there are several ways for learners to express confusion through retaking assessment or rewatching/slowing down videos. However, in most cases, learners tend more to reveal their confusion via online discussion forums through questions and posts, where each learner can express clearly his struggles in more details [2, 11].

Due to the absence of physical access to tutors, it is harder to early detect learners confusion about a particular concept or learning materials. On the other hand, Deep learning [5] and natural language processing (NLP) [9] are two artificial intelligence subfields used widely in e-learning. They aim to analyse learners posted messages and predict their different behaviours. Bidirectional Encoder Representations from Transformers (BERT) algorithm, is a new technique that reveals a very high performance over previous NLP techniques [1]. Published by Jacob Devlin in 2018 [6], BERT is based on attention mechanism that learns contextual relations between words in the text [10].

Our work aims to explore different levels of confusion in MOOCs discussion forums based on predetermined knowledge concept. We used ontologies to extract related terms to the chosen concept, then classify the selected messages using BERT classification algorithm. This method will help to identify the overall confusion level at each step of the learning process. It can also help distinguish learners with learning difficulties, then prepare them for a future process to increase their learning engagement and prevent drop out.

2 Proposed Approach

The overview of our proposed model is displayed in Fig. 1. Our approach is mainly composed of two principal subsystems: a prerequisite based ontology and a text classification using the BERT algorithm. First, at every end of the coursework, an intelligent tutor introduces the knowledge concept that should be acquired at this level of the learning process. The concepts generally do not exist alone; some concepts are the prerequisites of other concepts; Thus, for a student to master a chosen concept, he should usually master its prerequisites. Therefore, Based on OWL prerequisites ontology [3], we extract all the prerequisites of the introduced concept. The use of ontologies provides a useful tool for the representation of concepts, performance and relationships more adequately. After extracting all related prerequisites concepts, we filter by those concepts/terms all the messages posted in the discussion forums of MOOCs in this period. In this way, we only get posts related to the concept in question, alongside posts with the prerequisite concepts. This method will help to classify the confusion level based on only the chosen concept.

Before performing text classification, text pre-processing is a crucial step. Pre-processing transforms text into a more straightforward form for better performance for classification algorithms. We first perform noise removal and text cleaning (removing special characters, digits, lowercasing, ...) then we proceed with normalization which includes transforming the text into a consistent form through two main techniques stemming and lemmatization [4].

The final step consists of BERT classification algorithm; BERT is a deep learning algorithm given state-of-the-art results on multiple natural language processing tasks. BERT is based on multi-layer bidirectional Transformer encoder, and multi-head attention network. It is published by Google, and it is trained based on the corpus of 3.3 billion texts. The model is able to learn the context of a word based on all its neighbourhood.

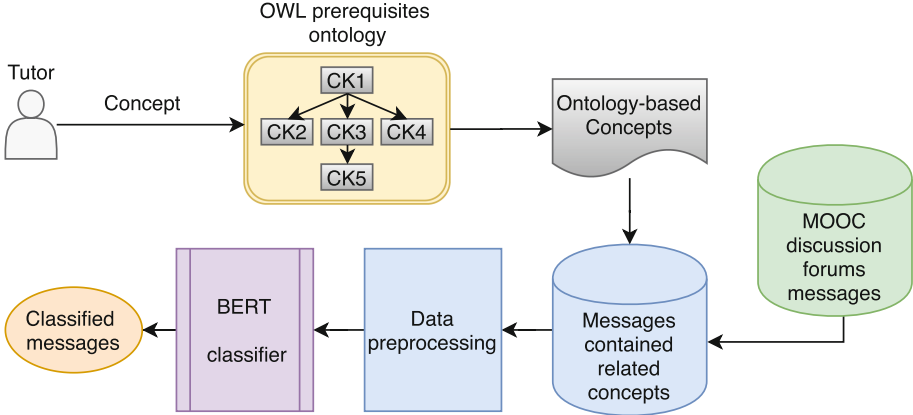


Fig. 1. Proposed approach workflow

BERT’s attention model between encoder and decoder is crucial, and it is a function that maps the input (a query q and a key-value pairs k and v) to output as presented in Eq. 1:

$$Atten(q, K, V) = \sum_i \frac{e^{q \cdot k_i}}{\sum_j e^{q \cdot k_j}} v_i \quad (1)$$

Based on contextual features within sentences and sequential features within the surrounding ones, we use the BERT classification algorithm to classify the selected messages into three different levels: confused, unconfused and neutral.

3 Experiment Settings

In our experiment, we used the Stanford MOOCPosts dataset that contains 29 604 learner forum posts from eleven Stanford university public online classes [2]. Those courses were chosen equally from three different domains: medicine, humanities/sciences and education. Each post was coded by three different independent coders. Each post in the MOOC Posts dataset was scored across six different dimensions, including the confusion. In the confusion dimension, coders ranked the confusion of the post on a scale of 1 to 7. A score of 1 means the post writer is not confused, while 7 means he is perplexed. We re-score the posts to limit confusion in three degrees, in the manner that posts with label inferior to 3.5 have the new label of 0 (unconfused). Posts with a label of 4 get the new label 1 (neutral), and the messages with score superior to 4.5 will be assigned the label 2 (confused).

As for the prerequisite based Ontology, we opt for the OWL (Ontology Web Language) which is a layer used below RDF (Resource Description Framework) to express logical constraints governing RDF triplets. We use this language to

build our ontology, which is mainly used to build our vocabulary. By vocabulary, we mean the set of prerequisites concepts, their relationships that are formally expressed. As for the querying of data, it is done through a specific query language called SPARQL. We Use Python programming language for filtering posts by the resulting prerequisite concepts and Natural Language Toolkit (NLTK) library for messages pre-processing. Therefore, the final resulting corpus is input to Bert algorithm.

We use the BERT-Base pre-trained model, which has 12 layers, 768 hidden states, 12 heads and 110M parameters. The batch size is set to 30. We use Adam-optimizer [8] as a learning rate optimization algorithm with hyper-parameters set to $\beta_1 = \beta_2 = 0,9$.

4 Results and Evaluation

In our experiment, we took the example of «**Statistic in medicine**» as a chosen concept. The OWL ontology generates **14** different prerequisites concepts/terms: «probability», «median», «frequency», «mean», «function», «standard deviation»... From the 29 604 posts, only **7203** contain the chosen concept and its prerequisites. After pre-processing, then BERT classification algorithm, we achieve **68,16%** accuracy score, which is very high for this small corpus. The obtained result show significant performance, the combination of OWL prerequisite based ontology with text classification help the system to build a model that can predict the overall confusion score around a given concept on each course session. Also, this model helps us distinguish individual learners with high confusion rates. This will aid taking precaution to early preventing losing learners motivation and engagement; Since learners with much higher confusion are more likely to drop out the learning process.

5 Conclusion

In this paper, we explore the combination of OWL ontology with BERT classification algorithm on Mooc forum posts to analyse learners' confusion level through posted messages. This method shows high efficiency. It will provide practical guidance for improving student engagement and early preventing their drop out.

In future work, we will build a vector presenting different confused behaviour of learners during his learning process. This vector will be based mainly on the number of confused, unconfused and neutral messages produced by each learner. Also, in order to better evaluate our approach, We aim to process answers from directed questions such as interviews and questionnaires. Those methods of data collection might also be applicable to classify confusion.

References

1. Adhikari, A., Ram, A., Tang, R., Lin, J.: DocBERT: BERT for document classification. arXiv preprint [arXiv:1904.08398](https://arxiv.org/abs/1904.08398) (2019)

2. Agrawal, A., Venkatraman, J., Leonard, S., Paepcke, A.: YouEDU: addressing confusion in MOOC discussion forums by recommending instructional video clips (2015)
3. Antoniou, G., Van Harmelen, F.: Web ontology language: OWL. In: Staab, S., Studer, R. (eds.) *Handbook on Ontologies*, pp. 67–92. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24750-0_4
4. Balakrishnan, V., Lloyd-Yemoh, E.: Stemming and lemmatization: a comparison of retrieval performances (2014)
5. Chanaa, A., El Faddouli, N.E.: Deep learning for a smart e-learning system. In: *Proceedings of the 2nd International Conference on Smart Digital Environment*, pp. 197–202 (2018)
6. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. arXiv preprint [arXiv:1810.04805](https://arxiv.org/abs/1810.04805) (2018)
7. Jordan, K.: MOOC completion rates: the data (2015). <http://www.katyjordan.com/MOOCproject.html>. Accessed 05 May 2020
8. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
9. Robinson, C., Yeomans, M., Reich, J., Hulleman, C., Gehlbach, H.: Forecasting student achievement in MOOCs with natural language processing. In: *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge*, pp. 383–387 (2016)
10. Vaswani, A., et al.: Attention is all you need. In: *Advances in Neural Information Processing Systems*, pp. 5998–6008 (2017)
11. Yang, D., Wen, M., Howley, I., Kraut, R., Rose, C.: Exploring the effect of confusion in discussion forums of Massive Open Online Courses. In: *Proceedings of the Second (2015) ACM Conference on Learning @ Scale*, pp. 121–130 (2015)