E-learning Text Sentiment Classification Using Hierarchical Attention Network (HAN)

Abdessamad Chanaa (✉), Nour-eddine El Faddouli
Mohammed V University in Rabat, Rabat, Morocco
`abdessamad.chanaa@gmail.com`

Abstract—Massive Open Online Courses (MOOCs) have recently become a very motivating research field in education. Analyzing MOOCs discussion forums presents important issues since it can create challenges for understanding and appropriately identifying student sentiment behaviours. Using the high effectiveness of deep learning, this study aims to classify forum posts based on their sentiment polarity using two experiments. The first use the three known sentiment labels (positive/negative/neutral) and the second one employs sevens labels. The classification method implemented the Hierarchical Attention Network (HAN) algorithm; it combines the attention mechanism with a hierarchical network that simulates the same hierarchical structure of the document. The analysis of 29604 discussion posts from Stanford university affirms the effectiveness of our model. HAN achieved a classification accuracy of 70.3%, which surpassed the other prediction results using usual text classification models. These results are promising and have implications on the future development of automated sentiment analysis tool on e-learning discussion forum.

Keywords—Text sentiment classification, Hierarchical Attention Network (HAN), E-learning, Online discussion forum, Massive Open Online Courses (MOOCs)

1   Introduction

In recent years, the usage of discussion forums has increased exponentially; those platforms allow users to express thoughts and share opinions. E-learning discussion forums facilitate the interaction between learners, where each posted message or comment presents a learner personal sentiment/feeling towards learning objects.

The use of e-learning discussion forums platforms has exponentially increased especially in the Massive Open Online Courses (MOOCs). MOOCs discussion forums enhanced the interactions between the learning process actors, where learners can textually communicate and interact, generating a tremendous amount of textual data related to each learner during his learning process.

Text sentiment classification is one of the most decisive tasks in Natural Language Processing (NLP). It aims to automatically determine the sentiment orientation (also called sentiment polarity) of an analysed text, whether it is positive or negative, by

analyzing the subjective information in posts, such as point of view, attitude, perspective and mood. It could help to efficiently classify and organize messages and reviews, and hence, rapidly identify their orientations (positive or negative). In e-learning discussion forums, detecting a learner sentiment orientation across his learning process becomes a necessity to analyse his behaviour and yield. Text sentiment classification motivates researchers to collect and analyse opinions in wide range of applications such as predicting drop out rate [1], student satisfaction measurement [2] and learning objects' recommendation [3] [4].

Remarkable success has been achieved along with the wide application of artificial intelligence, more precisely machine learning methods on text sentiment classification [5]. However, those methods cannot fully exploit sentiment orientation/polarity relying on sentiment linguistic knowledge such as bag-of-words (the occurrence of words within a document) and sentiment lexicon, where the sentiment orientation of the text is determined based on the number of positive/negative words; those limits are more recurrent in particular fields application such as education. Deep learning is a machine learning sub-field that constitutes a recent, novel technique for data analysis, with encouraging results and considerable capabilities. It can automatically learn complex data representation (such as image and text) hierarchically through multiple levels of abstraction, by adding more depth (complexity) to the analysed model [6]. As deep learning has witnessed great success in the last few years, it has been successfully applied in many domains applications and fields. More precisely in text sentiment classification tasks [7], deep learning can learn text representations and semantic connections without the need for manual feature engineering.

Because of the lack of face-to-face interactions, MOOCs suffer from very low completion rates [8]. Most students who drop out are not able to engage consciously in the course materials or the discussion forums, as many learners are enduring to stay cognitively and emotionally engaged. Detecting the sentimental state of these students may be the first step to increase the success rate of these courses and decrease the dropout rate. To perform this, we need to better understand their experience through out the learning process. More commonly, learners tend to express their difficulties, opinions and thoughts toward course materials textually in discussion forums. Thus, in this paper, we study the sentiment factor expressed in those discussion forums using deep learning algorithms and text sentiment classification techniques, by creating a sentiment model dedicated to the educational field that could predict the sentiment orientation of each learner based on their posted messages in discussion forums.

In this paper, we aim to exploit a deep learning algorithm named Hierarchical Attention Network (HAN) to classify learner's sentiment state in MOOCs discussion forums. By conducting a comparative study between HAN and different other text classification algorithms, we prove the efficiency of our predictive model over other different algorithms, by classifying messages into three classes: positive, negative and neutral. Then, we propose a more pointed text sentiment classification model that classify text into seven classes instead of three, where we predict not just the sentiment state of the text but also the degree of this state. This model could efficiently identify the sentiment orientation/polarity of any posted message during the learning

process. Therefore, having an overview of the global sentiment state of each learner on the platform.

The remainder of this paper is organised as follows. Section 2 gives an overview of text classification in e-learning and deep learning for text sentiment classification. Section 3 presents our structure, steps and method. We establish the experiment and the analysis of the results in Section 4. Finally, section 5 highlights the conclusion and discusses future work.

2 Related Works

The E-learning platform is an open platform where people can freely publish and share their materials. It leads to a large quantity of text material accumulating on the platform, including forum discussions, chat messages and reviews. Thus, organising such data will be an important issue. Researches in [9] proposed a full systematic overview of the current status of the educational text mining field, by analysing multiple text mining methods and their projection on different educational sources and resources. The study in [10] applied text mining and Feature-Sentiment-Pair (FSP) to identify user preferences in the MOOC. Similar studies conducted text data classification in the exam questions automatically into its learning levels using Bloom's taxonomy and Natural Language Processing (NLP) [11].

Online learning discussions forums can offer a lot of information about learners. In this direction, Hind Hayati et al. proposed to classify learners' messages in online discussions forums into four levels of cognitive presence, using Doc2Vec and machine learning algorithms [12]. Evermore in online discussion forums, more precisely in Twitter. Hamed AL-Rubaiee et al. [13], developed a framework to analyse tweets, based on arabic text sentiment classification using different machine learning algorithms such as Support Vector Machine (SVM) and Naive Bayes (NB). Another study focused on solving the educational and non-educational text classification problem based on deep Gaussian Processes (DGPs) and Word2vec [14]. Also, Abdessamad Chanaa et al. examined confusion state of learners by analysing MOOCs discussion forums using neural network classification algorithm and ontologies [15].

Remarkable success has been achieved along with the full application of deep learning methods on the e-learning field. However, the use of such models is still sparse comparing to other domain application fields. As an example, Hierarchical Attention Network (HAN) proves excellent performance in many domains [16] [17] [18]. In contrary to other classification algorithm and methods, that usually capture the sentimental state of the text as one unity without targeting the pertinence of specific words/sentences in the document. HAN hierarchically establish the message into numerical vector's representation by aggregating relevant words into the sentence vectors and next aggregating important sentences vectors to the final message vector. Those steps are done using double levels of attention mechanism. This method can capture significant words/sentences to the chosen context (the sentimental context in our case) and hence, classify the message into the corresponding class label. However, despite HAN effectiveness, as far as we know, it was never used in the educational

field except for [19], where it was used in computer vision and not NLP. In the next sections, we will describe our methodology based on HAN and then apply it to MOOCs forum posts.

3 Methodology

The overview of our proposed model is displayed in figure1. We divide our model into three phases, text pre-processing, word embedding based feature and the hierarchical attention network classification algorithm.
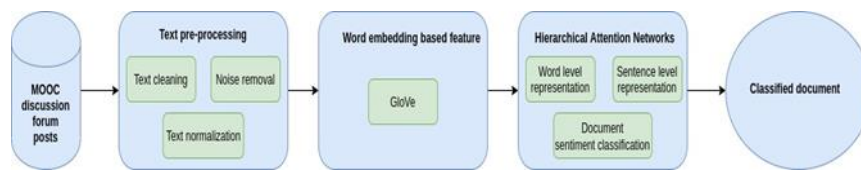

Fig. 1. The proposed model

3.1 Text pre-processing

Pre-processing is a crucial step for text classification applications. In this section, we introduced the necessary text transformations methods for data processing. Those methods used to remove noise, clean data and normalise text.

Noise removal and text cleaning: This step comprises removing special characters, digits and pieces of text that interfere with the text analysis. Then we remove unnecessary information like links, tags and emoticons. Abbreviation and slang replacement are other techniques aim to remove noise by transforming text into its original formal language [20].

Normalization: Text normalisation is the process of transforming the text into a consistent form, in our text pre-processing we use two main techniques: Stemming and lemmatisation. Stemming includes gathering all words with the same semantic meaning into their single root form. For example, the words "studies", "studying" and "studied" will all be replaced by their root form "study". As for Lemmatisation, It relies on extracting morphological analysis by identifying the lemma of the word. For example, words like "better" and "worse" will be replaced respectively by "good" and "bad".

3.2 Word embedding based feature

Word embedding based feature is a technique in which each word from the vocabulary is mapped into an N dimension vector of real numbers. Several word embedding methods have been proposed to transform the text into a structured input vector for machine learning algorithms. The most common methods for features word embedding are Glove, FastText, Latent Semantic Analysis (LSA) and Word2Vec. Word2vec is a word embedding technique that predicts surrounding words of a given

one using a shallow neural network. Created by Tomas Mikolov [21], Word2vec have proved its efficiency by encoding every word in a text, hence the contextual relation between words. On the other hand, Glove algorithm used in our work, which is a word2vec extension, creates explicit word to word co-occurrence statistics across the text corpus. The result is a learning model that create better word embeddings. In our paper, we use a pre-trained Glove model [22].

3.3   Hierarchical Attention Network (HAN)

Introduced by Zichao Yang et al. [23], Hierarchical Attention Network (HAN) extracts the contextual information by aggregating important words into sentence vectors and important sentence vectors into document vectors. With its hierarchical structure that mirrors the hierarchical structure of documents; HAN algorithm has two levels of attention mechanisms that applied at the word level and sentence level to capture sentiment information based on the interactions of the words, not just their presence in isolation.

The idea behind HAN is that every document is made of important sentences and every sentence is made of important words. HAN applies the attention mechanism to every word of the sentence, we can then have more attention on the important words that represent sentiment since not all the words forming a sentence have the same significance in that sentence. Then, the same procedure is applied to the sentences forming the whole document, since not all sentences contribute to the same importance in a document. In other words, the attention mechanism will extract and accumulate words that assist more in building the meaning of the sentences (the sentiment meaning) in order to form a sentence vector, then the attention mechanism will focus on sentences that contribute more in document classification. The attention model consists of two parts: Bidirectional Gated Recurrent Units (GRU) [24] and Attention Neural Networks (ANN) [25]. Generally in machine/deep learning problems, the quality of dataset labelling/coding has a crucial significance. Since HAN algorithm focuses on the attention or the importance (the weight) of each word in the sentence and focuses on the importance (the weight) of each sentence in the posted message. It will accord efficiently the composing words/sentences of the message with the score given by coder, i.e., the model will coordinate this words/sentence hierarchical composition with the score given by coder for a better understanding of each post sentiment orientation. Moreover, in HAN, the relevance of words/sentences is decidedly context-dependent, this means the same word/sentence may be differentially relevant in a different context depending on its given score by coders. Therefore, with many trained messages in the dataset corpus, the predictive model will be able to automatically identify the sentiment polarity/orientation of any new given message.

4   Experiment

4.1   Dataset

In our experiment, we used the Stanford MOOCPosts dataset that contains 29604 learner forum posts from eleven Stanford university public online classes [26]. It is an open discussion forum where students, as well as tutors, are free to create threads, ask questions, give answers, express opinions, reveal difficulties and share knowledge. Learners are also free to textually examine learning objects and communicate their thoughts and struggles toward learning materials and course structures. Those courses were chosen equally from three different domains: medicine, humanities/sciences and education. Three different independent coders coded each post. Each post in the MOOCPosts dataset was scored across six different dimensions (confusion, urgency, opinion, question, answer and sentiment).

In the sentiment dimension, coders ranked the sentiment of the post on a rational numbers scale of 1 to 7. A score of "1" means the post is very negative, while "7" means it is extremely positive.

4.2   Experiment setup

In our analysis, we conducted two experiments on the dataset. In the first one, our goal is to assess whether the post is positive, negative or neutral. We considered positive posts to be the ones scoring 4.5 or above, negative ones scoring 3.5 or below and the neutral posts are the ones scoring 4. As for the second experiment, we aim to classify posts into seven classes (strongly positive, positive, weakly positive, neutral, weakly negative, negative or strongly negative). Table 1 describes the coders' score of the dataset posts and their respective chosen classes of our two experiment.

Table 1. Coders scores and their respective sentiment classes

Coders score	First experiment	Second experiment
6 < score <= 7		Strongly positive
5 < score <= 6	Positive	Positive
4 < score <= 5		Weakly positive
score = 4	Neutral	Neutral
3 <= score < 4		Weakly negative
2 <= score < 3	Negative	Negative
1 <= score < 2		Strongly negative

The experiment involves three phases. First, in the process of text pre-processing, we used Natural Language Toolkit (NLTK) library, as for the second step, we used the GloVe pre-trained models [22], this pre-trained model contains 42 billion tokens, 1.9 million uncased vocabularies and 300-dimensional vectors. As for the classification phase, we used the PyTorch open-source library alongside with Python programming language.

The core data set contains 29604 posts, the training set and the test set are divided randomly according to 92:8, that is 27 204 posts for the training set and 2400 posts in the test set. The overall distribution of the labels in test sets is balanced in the first experiment. In our implementation, we set the word/sentence dimension as 50. For training, we used a batch size of 16. We adopt stochastic gradient descent with a momentum of 0.9. We also pick a learning rate of 0.1 over 10 epochs (data overfits if we use too many epochs).

We compared the Hierarchical attention network performance to other four classification algorithms: Naive Bayes (NB), Logistic regression (LR), Support Vector Machine (SVM) and Multi-layer perception (MLP) using Scikit-learn library [27]. Based on the confusion matrix, we used four evaluation metrics: accuracy (Acc), precision (Prec), recall (Rec), and F1 score (F1).

4.3 Experimental results and discussion

In this section, we present the comparing results of different algorithms against HAN. Table 2 is divided into two categories where we present results of the first experiment (3 labels) and the second experiment (7 labels).

Table 2. Results of the two experiments

	First experiment				Second experiment			
	Acc	*F1*	*Rec*	*Prec*	*Acc*	*F1*	*Rec*	*Prec*
SVM	0.347	0.199	0.347	0.719	0.556	0.408	0.553	0.648
NB	0.412	0.406	0.412	0.410	0.336	0.361	0.336	0.418
MLP	0.332	0.177	0.332	0.319	0.549	0.389	0.549	0.413
LR	0.334	0.174	0.334	0.344	0.544	0.389	0.544	0.362
HAN	0.703	0.701	0.703	0.721	0.666	0.642	0.297	0.641

Table.2 shows that our experimental results display a very significant performance overall other baseline methods. A comparison with other machine learning methods verifies the effectiveness of hierarchical attention network that reflect the hierarchical structure of the MOOC reviews. HAN encourages the model to focus on the important words/sentences in the text during the learning process, which outstandingly enhance the discriminative ability for classification.

In the second experiment, even with a higher number of labels HAN also outperforms methods without hierarchical structures except with the recall where we get a lower result of 0,297 comparing to the other methods. Although, since the precision is high 0,641, this means that the algorithm was very finicky in classifying and it misses a lot of true positive results. However, since the F1-score metric is high (F1-score is used to combine precision and recall into one coherent metric), the results are still significant. Multi-layer Perceptron (MLP), which is also a neural network algorithm, does not perform well compared to the other machine learning models. Although the power of neural network algorithms, Comparing MLP with the Hierarchical Attention network, HAN shows significant high performance; this demonstrates the effectiveness of the hierarchical structure of HAN against other algorithms.

To give more insight into the quality of the results, we realised a JavaScript web-based application where we integrated our HAN pre-trained model into our university learning management system. As presented in Figure 2 for the first experiment and Figure 3 for the second one, we extracted the message from the database related to his owner. Thus, showing the polarity results (probability percentages) of each message according to his belonging class will give a chance for the tutor and the system to give more judgement of each message. Therefore, having an overview of the global learners' sentiments state during the learning process.
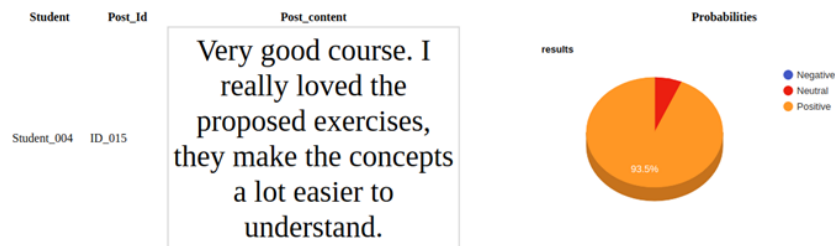

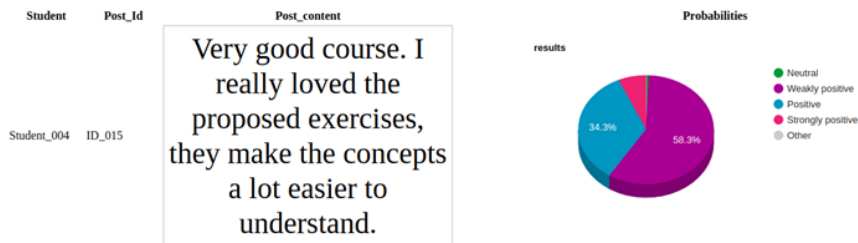**Fig. 2.** Results of a message according to the first experiment (3 labels)

Fig. 3. Results of a message according to the second experiment (7 labels)

As an important component of the MOOC learning process, the discussion forums have a big impact on the student experience. Discussion forums are a mean for learners to express clearly their thoughs, opinions and struggles toward learning contents. The students textually interactions enhance the sentiment and cognitive engagement toward the course. Our predictive model based on attention mechanism of deep learning and the hierarchical simulation of the document (the posted message) predicts with high efficiency the orientation of each message on discussion forums. The analysis of sentiment in discussion forums will shed light on the most active and less active students in the analysed learning process. This will give the opportunity to tutor and the system to track closely the sentiment orientation of each subscribed learner in the course. One possible use of this model is to gauge the learner sentimental satisfaction through the number of positive/negative posted messages on the platform. This predictive model will help the system to distinguish learner with negative sentimental orientation and early predict their drop out. This model can also be integrated on MOOCs sentiment-based recommender systems for better decision-making toward less engaged learners.

This model trained over 29604 messages related directly to the educational field. It can be integrated into any LMS hosting MOOCs materials to easily measure learner's sentimental satisfaction at any given time via his posted messages. Therefore, the model can determine the learner's global satisfaction by combining the sentimental state along with other social and cognitive factors.

5 Conclusion and Future Work

Due to the lack of face-to-face interactions in MOOCs, our method attends to explore the sentimental textual interactions in a MOOC discussion forum. Based on Deep learning and automated text analysis, we conducted a comparative study between different classification algorithm to build a predictive model able to predict the sentimental orientation of every posted message in the discussion forum. Our method is divided into three steps: text pre-processing, modelling language into numerical vectors (Word embedding based feature) and deep learning-based automated text classification. Our method based on Hierarchical Attention Networks (HAN) algorithm outperforms other known classification algorithms. Therefore, this predictive model will present an efficient tool for tutor and system to gauge learner sentiment orientation via the posted messages throughout the learning process. Consequently, this will enhance the supervision over learner's sentimental and cognitive engagement toward content.

In future work, we will build a vector presenting different sentimental behaviour of learners during his learning process. This vector will be based mainly on the number of negative, positive and neutral messages produced by each learner. Besides, we also intend to combine sentiment analysis with a questionnaire survey to explore the psychological and social motivations of learners in MOOCs.

6 References

[1] S. Halawa, D. Greene, and J. Mitchell, "Dropout Prediction in MOOCs using," eLearning Papers, vol. 37, no. March, pp. 1–10, 2014.

[2] K. F. Hew, X. Hu, C. Qiao, and Y. Tang, "What predicts student satisfaction with moocs: A gradient boosting trees supervised machine learning and sentiment analysis approach," Computers & Education, vol. 145, p. 103724, 2020. https://doi.org/10.1016/j.compedu.2019.103724

[3] A. Chanaa and N.-e. El Faddouli, "Context-aware factorization machine for recommendation in Massive Open Online Courses (MOOCs)," 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems, WITS 2019, pp. 1–6, 2019. https://doi.org/10.1109/wits.2019.8723670

[4] A. Chanaa and N.-e. El Faddouli, "Predicting learners need for recommendation using dynamic graph-based knowledge tracing," in Artificial Intelligence in Education, I.I.Bittencourt, M. Cukurova, K. Muldner, R. Luckin, and E. Mill ́an, Eds. Cham: Springer International Publishing, pp. 49–53, 2020. https://doi.org/10.1007/978-3-030-52240-7_9

[5] A. Shah, K. Kothari, U. Thakkar, and S. Khara, "User review classification and star rating prediction by sentimental analysis and machine learning classifiers," in Information and

Communication Technology for Sustainable Development. Springer, 2020, pp. 279–288. https://doi.org/10.1007/978-981-13-7166-0_27

[6] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015. https://doi.org/10.1038/nature14539

[7] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing [Review Article]," IEEE Computational Intelligence Magazine, vol. 13, no. 3, pp. 55–75, 2018. https://doi.org/10.1109/mci.2018.2840738

[8] C. H. Aydin, "Moocs as change agents," The 2018 OpenupEd trend report on MOOCs, p. 18, 2018.

[9] R. Ferreira-Mello, M. Andr´e, A. Pinheiro, E. Costa, and C. Romero, "Text mining in education," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, no. March, 2019. https://doi.org/10.1002/widm.1332

[10] N. Dina, R. Yunardi, and A. Firdaus, "Utilizing text mining and feature-sentiment-pairsto support data-driven design automation massive open online course," InternationalJournal of Emerging Technologies in Learning (iJET), vol. 16, no. 1, pp. 134–151, 2021. https://doi.org/10.3991/ijet.v16i01.17095

[11] K. Jayakodi, M. Bandara, I. Perera, and D. Meedeniya, "Wordnet and cosine similaritybased classifier of exam questions using bloom's taxonomy."International Journal ofEmerging Technologies in Learning, vol. 11, no. 4, 2016. https://doi.org/10.3991/ijet.v11i04.5654

[12] H. Hayati, A. Chanaa, M. Khalidi Idrissi, and S. Bennani, "Doc2Vec &Na¨ıve Bayes: Learners' Cognitive Presence Assessment through Asynchronous Online Discussion TQ Transcripts," International Journal of Emerging Technologies in Learning (iJET), vol. 14, no. 08, p. 70, apr 2019. [Online]. Available: https://online-journals.org/index.php/ijet/article/view/9964 https://doi.org/10.3991/ijet.v14i08.9964

[13] H. Al-Rubaiee, R. Qiu, K. Alomar, and D. Li, "Sentiment Analysis of Arabic Tweets in e-Learning," Journal of Computer Science, vol. 12, no. 11, pp. 553–563, 2016.

[14] H. Wang, J. Z. B, Z. Tang, and S. S. B, "Educational and Non-educational Text," pp. 415–423, 2017. https://doi.org/10.3844/jcssp.2016.553.563

[15] A. Chanaa and N.-E. El Faddouli, "Bert and prerequisite based ontology for predicting learner's confusion in moocs discussion forums," in Artificial Intelligence in Education,I.I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, and E. Mill´an, Eds.Cham:Springer International Publishing, pp. 54–58, 2020. https://doi.org/10.1007/978-3-030-52240-7_10

[16] S. Gao, M. T. Young, J. X. Qiu, H. J. Yoon, J. B. Christian, P. A. Fearn, G. D. Tourassi, and A. Ramanthan, "Hierarchical attention networks for information extraction from cancer pathology reports," Journal of the American Medical Informatics Association, vol. 25, no. 3, pp. 321–330, 2018. https://doi.org/10.1093/jamia/ocx131

[17] H. Ying, F. Zhuang, F. Zhang, Y. Liu, G. Xu, X. Xie, H. Xiong, and J. Wu, "Sequential recommender system based on hierarchical attention network," IJCAI International Joint Conference on Artificial Intelligence, vol. 2018-July, pp. 3926–3932, 2018. https://doi.org/10.24963/ijcai.2018/546

[18] S. Singhania, N. Fernandez, and S. Rao, "3han: A deep neural network for fake news detection," in International Conference on Neural Information Processing. Springer, 2017, pp. 572–581. https://doi.org/10.1007/978-3-319-70096-0_59

[19] A. Kaur, "Attention Network for Engagement Prediction in the Wild," pp. 516–519, 2018.

[20] S. Dhuliawala, D. Kanojia, and P. Bhattacharyya, "SlangNet: A WordNet like resource for English slang," Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016, no. May, pp. 4329–4332, 2016.

[21] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representa tions ofwords and phrases and their compositionality," Advances in Neural Information Processing Systems, pp. 1–9, 2013.

[22] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1532–1543. [Online]. Available: https://www.aclweb.org/anthology/D14-1162 https://doi.org/10.3115/v1/d14-1162

[23] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention net works for document classification," 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016 - Proceedings of the Conference, pp. 1480–1489, 2016. https://doi.org/10.18653/v1/n16-1174

[24] K. Cho, B. Van Merri¨enboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, pp. 1724–1734, 2014. https://doi.org/10.3115/v1/d14-1179

[25] Y. Kim, C. Denton, L. Hoang, and A. M. Rush, "Structured Attention Networks," pp.1–21, 2017. [Online]. Available: http://arxiv.org/abs/1702.00887

[26] A. Agrawal, J. Venkattraman, S. Leonard, and A. Paepcke, "YouEDU: Addressing Confusion in MOOC Discussion Forums by Recommending Instructional Video Clips," Proceedings of the 8th International Conference on Educational Data Mining, pp.297–304, 2015.

[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg et al., "Scikit-learn: Machine learning in python," Journal of machine learning research, vol. 12, no. Oct, pp. 2825–2830, 2011. https://doi.org/10.3389/fninf.2014.00014

7 Authors

Abdessamad Chanaa is a Ph.D. candidate at Mohammadia School of Engineers (EMI), Mohammed V University in Rabat, Morocco. Email: abdessamad.chanaa@gmail.com.

Nour-eddine El Faddouli is a Full Professor at Mohammadia School of Engineers (EMI), Mohammed V University in Rabat, Morocco. Email: nfaddouli@gmail.com.