Effective Combination of Contextual Embedding based Tensor Space Model and Transformer for Text Classification

SoonKwan Kwon, Han-joon Kim *

Department of Electrical and Computer Engineering, University of Seoul, Korea

ghghfhddl@naver.com, khj@uos.ac.kr

Abstract. In this paper, we propose a deep learning architecture that effectively combines our previous tensor space representation model, contextual word embedding and Transformer model for more accurate text classifier. This is a type of multimodal architecture that not only considers the contextual information of words in the tensor space model, but also can learn the sequential information of words. Through experiments using 2 English news datasets, we proved that our method is superiror to conventional ones.

Keywords: ELMo, word embedding, tensor space model, TextCNN, transformer, text classification

1 Introduction

In this paper, we propose a deep learning architecture that effectively combines our previous tensor space representation model [1], contextual word embedding and Transformer model [2] for more accurate text classifier. Our tensor space model represents a single document as a *term-by-concept* matrix, and however it does not consider polysemy problem and word sequence information. To achieve more reliable text classification, we devised a multimodal deep learning architecture that not only includes the contextual information of words in the tensor space model, but also can learn the sequential information of words.

2 Proposed Method

To learn the sequential information with a tensor space model, we have utilized the Transformer model widely used in NLP task. As a result, our proposed architecture has a multimodal (or 2-channel) deep learning as shown in Fig. 1. Each of documents is represented as a two-dimensional matrix with an ELMo [3] embedding vector containing contextual and semantic information, and simultaneously, the sequential information of words within the document forms a sequence channel; the former is

* Corresponding Author: Han-joon Kim (khj@uos.ac.kr)

**This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2022R1A2C1011937).

input to the TextCNN layer [4], and the latter is input to the Transformer layer. The embedding channel input is converted into a feature map with context and semantic information learned through the internal CNN layer, and the sequence channel input is converted into a feature map with sequential information information learned through the Transformer encoder. As a result, the two feature maps trained contain weight vectors focused on the main dimensions that can contribute to classification as they pass through the Self-Attention layers. Lastly, their results are input into a DNN layer that outputs the class. Table 1 shows that for two English news datasets, the proposed model (denoted as bold) outperforms our previous 1-channel tensor space model as well as conventional classification models.

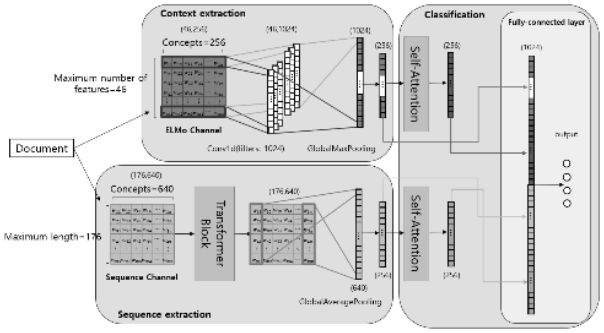

Fig. 1. Multimodal deep learning architecture for simultaneously learning contextual, semantic, and sequential information within classified documents

Table 1. Performance comparison of classification models

Dataset	Model	Accuracy (%)
News Category	1-channel tensor space model	74.9%
	2-channel tensor space model	**83.7%**
	Transformer classifier	65.5%
	SVM	71.0%
	Naïve Bayes	71.7%
	Logistic Regression	69.2%
AG news	1-channel tensor space model	91.8%
	2-channel tensor space model	**95.8%**
	Transformer classifier	89.1%
	SVM	91.3%
	Naïve Bayes	92.0%
	Logistic Regression	91.3%

References

1. Kim, H. J., and Chang, J. Y., A Semantic Text Model with Wikipedia-based Concept Space, Journal of Society for e-Business Studies, Vol. 16, No. 3, pp. 107-123, 2014.
2. Shaheen, Z., Wohlgenannt, G., and Filtz, E., Large scale legal text classification using transformer models, arXiv preprint arXiv:2010.12871, 2020.
3. Peters, M., Neumann, M. et al, Deep contextualized word representations, Proceedings of the international Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1, pp.2227-2237, 2018.
4. Yoon, K.: Convolutional neural networks for sentence classification, Proceedings of the International Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1746-1751,2014.