Keyword Extraction in Economics Literatures using Natural Language Processing

Soojeong Kim
School of Electrical Engineering
Korea University
Seoul, South Korea
sookelly@korea.ac.kr

Sunho Choi
School of Electrical Engineering
Korea University
Seoul, South Korea
schoi_@korea.ac.kr

Junhee Seok
School of Electrical Engineering
Korea University
Seoul, South Korea
jseok14@korea.ac.kr

***Abstract*—** Using Natural Language Process (NLP) as an efficient way to research paper is important when user feedback is sparse or unavailable. The task of text mining research paper is challenging, mainly due to the problem of unique characteristics such as jargon. Nowadays, there exist many language models that learn deep semantic representations by being trained on huge corpora. In this paper, we specify the NLP pre-processing process with Economics journal paper and apply it to a deep learning model to extract keywords. Here, we focus on the strength of NLP when applied to an unknown field. The analysis result shows the possibility and potential usefulness of the relationship research between keywords in research papers.

Keywords—Preprocessing, Natural Language Processing, BERT, Economics Journal Paper

I. Introduction

Natural Language Processing (NLP)[1] is necessary for data processing in various areas. In recent years, text mining in other fields, such as medicine has been rapidly evolving due to the development of natural language processing deep learning technology, for example, LSTM [2], CRF [3], NER [4], etc. Extracting knowledge from research papers using NLP has been successful in natural science and engineering. However, research is still insufficient in the humanities, society, economy, and management field compared to the amount of data present, as the definition and meaning of keywords are unclear and overlapped. In this study, we would like to solve this problem by extracting keywords from economics literature using a transfer learning model such as BERT. Economics research is the number one research area on "Categories by Rank" in "InCites Journal Citation Reports". We measured the performance by comparing the extracted keywords with the existing manualized keywords.

In order to figure out the feasibility, this study tries the preprocessing process in advance. Data processing research in economic journals has not yet been active. Moreover, it is difficult to directly incorporate existing NLP text mining technologies, such as Word2Vec [5], ELMo [6], and BERT [7]. This is because there is a difference between text expression used in daily life and jargon in journal papers.

It is too early to check the performance by applying those models. In this paper, we want to learn economics text mining to complement and apply an advanced version of natural language processing model.

This work was supported by the National Research Foundation of Korea grant (NRF-2019R1A2C1084778). Correspondence should be addressed to jseok14@korea.ac.kr

II. Related Works

A. Natural Language Processing

NLP [1] is an area where computers understand what human speaks and writes, manipulates it to new various research and application. While early research attempts in NLP were to enable computers interpret simplified language and communicate with people, many of recent NLP algorithms seek to extract meaningful content by statistical inference. The result is that each document in a corpus can be represented as a vector of features with associated meanings. These vectors can be used to evaluate the probability that a certain sentiment is expressed in a document, or that a particular issue is discussed.

B. Bidirectional Encoder Representation from Transformer (BERT)

i) Transfer-Learning

Transfer-learning allows BERT to do two tasks, pre-training and fine-tuning. For pre-training, BERT executes two unsupervised learning simultaneously. First is Masked Language Model (MLM) which is tokenizing masked random sentences. Second is Next Sentence Prediction (NSP) which is similar to binary classification. It discriminates whether sentence B is sequentially after sentence A. This helps BERT to understand the whole context of the language.

Fine-tuning is a process of modifying input and output layers. After the word vectors are generated and initialized with the pre-trained parameters, BERT activates softmax function with softmax layers with 3000 neurons. It (word vectors) is fully connected with vocabularies in word piece. BERT trains with cross entropy loss of the masked words by comparing those with one-hot encoding.

Fig. 1 is an example that shows the process of transfer learning by comparing it with traditional machine learning.

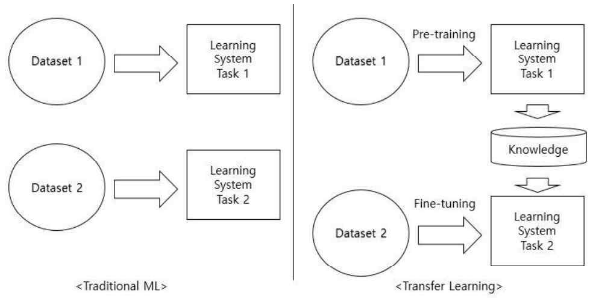

Fig. 1. Differences of Traditional ML and Transfer Learning

ii) Bidirectional

Former models of BERT, which is RNN [8] and LSTM [9] had time complexity problems and a lack of bidirectional features. To solve these, Transformer [11] entered the stage. It is much faster and simultaneously and deeply bidirectional. Transformer architecture is composed of an encoder and decoder. BERT uses the architecture of bidirectional Transformer, which is a part of the Transformer encoder, and it makes difference in pre-training model architecture. Figure 2 shows the architecture.

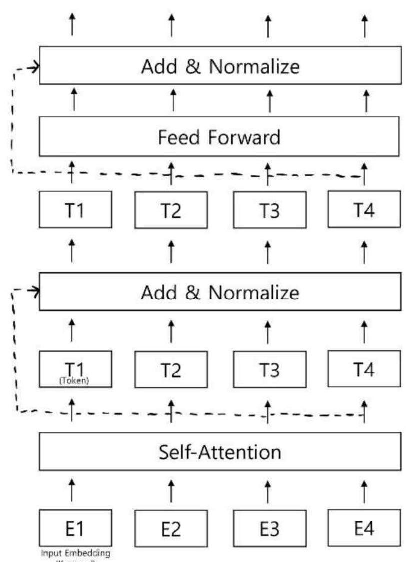

Fig. 2. Transformer Encoder of BERT

In the decoder, while OpenAI GPT [10] uses left-to-right architecture, the self-attention layer in Transformer works just like multi headed self-attention. As opposed to directional models, which read the text input sequentially such as word2vec, BERT's MLM objective can access both left and right context. It leads the Transformer model to pre-train in a deep bidirectional way. Figure 3 shows how the self-attention layer in Transformer processes.

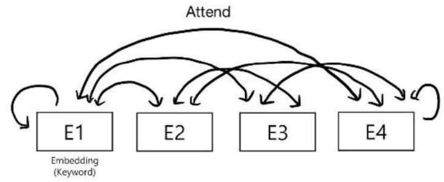

Fig. 3. Self-Attention

III. DATA AND METHODS

For keyword extraction, we apply the economics journal data records extracted from *Web of Science.* A total of 20 2019 SSCI edition journals ranked in the Impact Factor (IF) and Journal Citation Reports (JCR), which indices the influence of the paper, were extracted. This information is available on the site provided by Clarivate Analytics, a well-known developer of the paper's index management program, End Note. It can be obtained from "Incites Journal Citation Reports" [12].

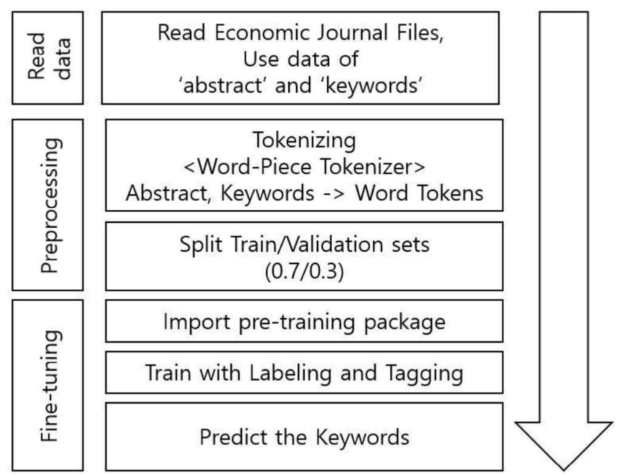

Fig. 4. Work flow of keyword extraction from Economics journal paper using BERT

The total number of the journal data records is 46,014. Specifically, we use only three data, which are 'document type', 'abstract', and 'author keywords'. If the data doesn't have at least one of the three elements, it is eliminated, and the number of the remains is 36,345 in total.

After loading the data files, Word-piece tokenizer does the tokenizing encode process, which is a sub-word tokenization algorithm used for BERT. It converts 'abstract' and 'author keywords' strings into a sequence of tokens. It split in words for word-based vocabulary or sub-words for sub-word-based vocabularies.

After preprocessing the data, we import the pre-trained module implemented with Torch package and Tensor Dataset. In order to train a deep bidirectional representation, we predict keywords by tagging tasks. The tokens split by the tokenizer are linked with tag of words. In fine-tuning, we add a classification layer to the last layer of the model in order to tag the keyword of tokens. We train the model with labeling and tagging with cross-entropy loss with batch size of 12.

IV. EXPERIMENT RESULT

We adjusted the epoch for the BERT experiment to confirm that the overfitting point is epoch 9. Figure 5 describes the train/validation loss of the following procedure. The train loss continuously and rapidly reduced while the validation loss increased slightly after the 4th epoch.

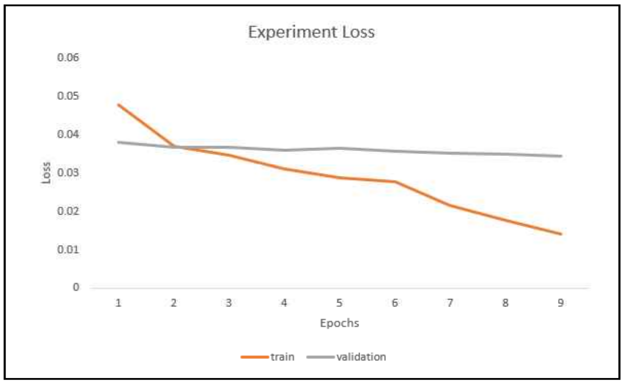

Fig. 5. A train loss graph according to the epochs

We extracted predicted keywords from the input keywords, true keywords, and true keywords in abstract as experiment result. If there were at least one well predicted keyword in true keywords in abstract, then ‘At least one keyword’ value was counted as 1. Of the total 3,635 input keywords set, 2,673 prediction keywords set had at least one successfully predicted keyword. Thus, approximately 73.53% of the ‘At last one keyword’ value was derived.

The fine-tuning was carried out for 25,441 papers, which are the training dataset, until reaching at least 96% of keyword tagging accuracy. The evaluation performance with the validation dataset of 10,903 papers showed the mean values of Precision and Recall being 0.307 and 0.244, respectively. Table I shows the fine-tuning and evaluation result of the experiment using BERT.

TABLE I. THE ACCURACY RESULT OF EXPERIMENT

	At least one keyword prediction	*Precision*	*Recall*
BERT	0.735 (8 epochs)	0.307	0.244

V. CONCLUSION

In this paper, we proposed a keyword extraction and prediction method with economic journal article datasets using NLP. Since the accuracy results are still low in numbers, we plan to collect more than 120,000 articles in economics journals to further develop our research using high-level fine-tuning algorithms and techniques such as Transfer Learning and Few-Shot Learning. In addition, various recently developed artificial intelligence technologies will be further used to enhance accuracy and continue to improve relationship research between keywords. We believe this experiment can also present a strong argument for the potential economists to help make computational linguistics – already a field with many applications – into a much more powerful toolset for understanding behavior and predicting outcomes.

REFERENCES

[1] Gobinda G. Chowdhury, “Natural Language Processing,” Annual Review of Information Science and Technology, 2003.

[2] T. Mikolov, M. Karafiat, L. Burget, J. Cernocky, and S. Khudanpur, “Recurrent neural network network based language model,” INTERSPEECH, pp.1045-104, 2010.

[3] Zhiheng Huang, Wei Xu, and Kai Yu, “Bidirectional LSTM-CRF models for sequence tagging”, CoRR, abs/1508.01991, 2015.

[4] Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer., “Neural architectures for named entity recognition”, NAACL-HLT, 2016.

[5] Yoav Goldberg and Omer Levy, “word2vec explained: deriving Mikolov et al.’s negative-sampling wordembedding method”, arXiv preprint arXiv:1402.3722, 2014.

[6] Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer, “Deep contextualized word representations”, NAACL, 2018.

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding” arXiv preprint arXiv:1810.04805, 2018.

[8] E. Choi, A. Schuetz, W. F. Stewart, and J. Sun, “Using recurrent neural network models for early detection of heart failure onset,” J. Amer. Med. Informat. Assoc., 2016.

[9] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, pp. 1735–1780, 1997

[10] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever, “Improving language understanding with unsupervised learning”, Technical report, OpenAI, 2018.

[11] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, “Attention is all you need.”, Advances in Neural Information Processing Systems, pages 6000–6010, 2017.

[12] Incites Journal Citation Reports, https://jcr.clarivate.com/JCRJournalHomeAction.action?pg=JRNLHOME&categoryName=ECONOMICS&year=2019&edition=SSCI&categories=GY