A Comparison of Deep-learning Algorithms for Classifying Strategic Items of Nuclear Export Control System

Mun-Young Ryu*, In-Chul Kim, Byung-Doo Lee, Sung-Ho Lee, Hyun-Jo Kim, Hyun-Sook Kim, Juang Jung, Seong-Mi Han, and Jin-Ha Choi
KAERI, 989-111 Daedeok-daero, Yuseong-gu, Daejeon, Republic of Korea
**Corresponding author: myryu@kaeri.re.kr*

1. Introduction

As a part of nuclear non-proliferation policy, goods and technologies related to weapon development (that is, strategic items) are banned from export, or may have restrictions from export. In general, strategic items is determined by experts with lots of relative experiences in nuclear industry. With the development of technology and growing international cooperation, the government has established more elaborated and specific export control policies, which can control the transfer of strategic items generated in various ways. However, the increase of international technology exchanges has led to problems such as burden of work, processing time delay and therefore it could be less efficient to execute and perform overall export licensing process.

Although deep-learning technology is rapidly applied in various ways, the development of support system in the nuclear field is at an insufficient level due to the characteristics of the nuclear industry, the use of specialized terminology and limited access to data. Therefore, this paper proposes a model that automatically learns documents related to strategic item classification and link data to each other helps to infer results.

2. Methods and Results

2.1 Data collecting and pre-processing

The data used in this analysis were collected from actual applications and manuals for strategic items' classification at KAERI. Data that are not suitable for language processing (e.g., design drawings, picture files, manuals in foreign languages, etc.) were excluded.

Then, the raw data were reprocessed through the steps in Fig.1. Pre-processing is a very important step in increasing the accuracy and reliability of deep learning. After converting the file into a batch, process the document according to the three steps: re-batch and labeling, tokenization, and corpus saving.

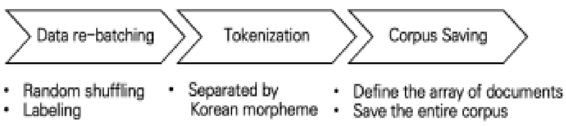


Fig. 1 Diagram of text data pre-processing

Data pre-processing prevents learning outcomes from being biased according to document order. After dividing, it into basic corpus units that could be analyzed, appropriate Python library was used to tokenize language. And then, the characters are digitized through word embedding. Due to the nature of the nuclear field, where terminology is commonly used, a generally available word models are not sufficient for analysis. This requires a domain-specific word model, so in this study, the new word model was trained on the entire corpus using Word2Vec.

2.2 Experimental Method

The overall experimental configuration, including pre-processing, is shown in the Fig.2.

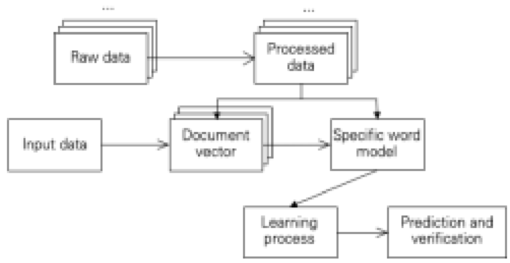


Fig. 2. Overall configuration of data learning model

For the data prepared in the previous chapter, three major algorithms were applied, which are considered having good performances for natural language processing recently: Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Long Short-Term Memory Units (LSTM).

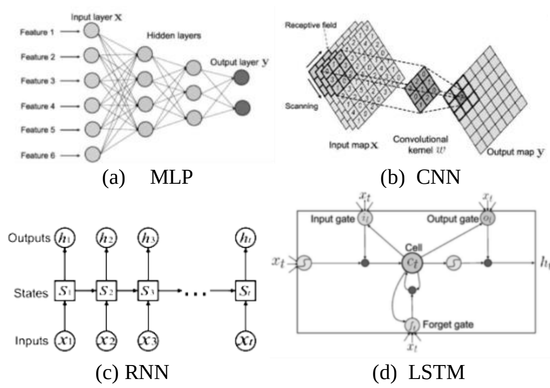


Fig. 3. NLP algorithms applied to this experiment

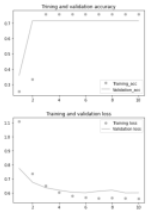
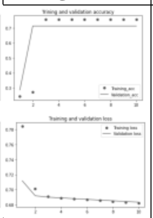
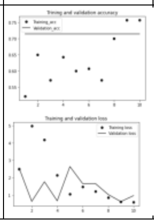
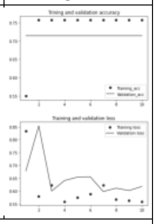
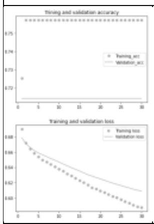
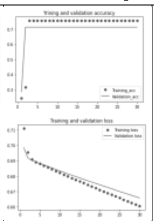
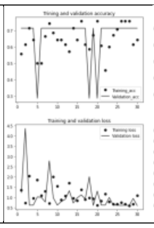
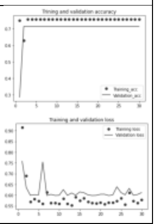
The environment of experiment was established with Python and the deep learning library provided by Keras was used. Based on this system, performance and features of each algorithm were reviewed compared with

Multi-Layer Perceptron (MLP). Fig.3 shows the algorithms and concept structures applied to this analysis. To ensure sufficient learning, the prepared data sets were classified into two groups, training and testing with the ratio of 8:2.

2.4 Results

The learning results are as shown in the Table I.

Table I: Accuracy and loss output for each algorithm

MLP	CNN	RNN	LSTM
			
epoch 10			
			
epoch 30			

As shown in the table, the number of iterations was 10 and 30 times. Depending on the algorithms, the accuracy and the convergence and speed of the graph have different forms. The feature is that while the accuracy of MLP and CNN increases constantly, RNN and LSTM take on a bit fluctuating form. This is the same as loss output. It is considered that the accuracy and stability of RNNs were lower than CNN and LSTM because of its characteristics affected by sequential.

Since the application documents for the strategic items' classification consist of summarized description of the items, the meaning is transmitted similarly even if the words are omitted, or the order has been changed back and forth. Also, applicants can write regardless of form because format of documents is not standardized. Therefore, CNN and LSTM algorithms, which are not affected by the sequential nature of the language, are considered appropriating for the model of classification support system.

3. Conclusions

Although many NLP studies have been conducted in various fields so far, in nuclear industry field, the transfer of knowledge is difficult and administrative burden is aggravated due to a system operated by a few skilled experts in the high security environment.

This paper proposes the appropriate algorithms that process related data and provide support for judgment whether the item is classified as strategic items. For this, NLP deep learning algorithms were applied to find the target information using the past data and natural language processing.

Therefore, it is expected both judges and applicants can easily access the results and take precautionary measures by comparing similarities between documents and predict the results of the classification, so those related tasks can be processed more quickly and efficiently. As a result, the study could be used as a basic design data for advanced intelligent systems and contribute to improving export control of nuclear fields. As a result, it is intended to establish an import-export control system that integrates related implementation tasks to increase the efficiency of the applicant's work and reduce administrative burden.

REFERENCES

[1] Tomas Mikolov, Wentau Yih, Geoffrey Zweig, Linguistic Regulations in Continuous Space Word Representations, 2013.
[2] Yoon Kim, Convolutional Neural Networks for Sentence Classification, EMNLP, 1746~1751, 2014.
[3] Kim U., H. Kim, M. Y. Yi, and D. Shin, Nuclear Export Control System Using Semi-Automatic Keyword Extraction, International Journal of Information and Electronics Engineering, Vol.4, No.4, 293~297, 2014.
[4] Hong, Woneui, Kim, U., Cho. S., Kim S., M. Y. Yi, and D. Shin, Export Control System Based on Case Based Reasoning Design and Evaluation, International Journal of Information and Electronics Engineering, Vol.20, No.3, 109~131, 2014.
[5] Kim D.S. and J.W.Kim, Research Trend Analysis Using Bibliographic Information and Citations of Cloud Computing Articles: Application of Social Network Analysis, Journal of the Korea Institute of Intelligent Transport System, 2014
[6] Zhiyong Cui, Ruimin Ke, and Yinhai Wang, Deep Stocked Bidirectional and Undirectional LSTM Recurrent Neural Network for Network-wide Traffic Speed Prediction, 2018