A Study on Development of Deep-Learning Based Strategic Item Classification System for Nuclear Export Control

Mun-Young Ryu*, HJ Kim, IC Kim, and SM Han
Korea Atomic Energy Research Institute (KAERI), Republic of Korea
**Corresponding author: myryu@kaeri.re.kr*

Abstract

The international community has been operating several export control regimes to protect human lives and property from weapons of mass destruction. The Korean government also implements export control of strategic items and technologies by reflecting relevant guidelines in the Foreign Trade Act and Nuclear Safety Act. The first step in implementing export control is to determine whether the item is to be classified as a strategic item or not, which is conducted through a process of self-classification by exporters or a classification request to the government. According to relevant Korean laws and regulations, exporters must apply to the government for an item classification and proceed with export licensing if the result corresponds to strategic items. In the field of nuclear technology, it is difficult to secure open data due to the strict security standards imposed by the nature of the industry, and continuous experience and expertise in nuclear technology are required for classification judgment.

Therefore, this study analyzes the possibility of using deep-learning technology to determine whether items and technologies are strategic items and to utilize the history of the Korea Atomic Energy Research Institute (KAERI)'s strategic items export and import to find related information and support the prediction of strategic item classification. In addition, with KAERI's data, the simulation was conducted based on a specific word model and a deep-learning model suitable for the development of the strategic item classification system was explored. After preliminary tests, the results show that a few neural network models were found to be effective for the characteristic of documents, and the necessity of pre-processing of the document for accuracy improvement was confirmed. The results of the study will be used to develop a decision support system to determine whether a strategic item belongs to the nuclear trigger list, which will contribute to securing KAERI's strategic trade control system and improving practical efficiency.

1. Introduction

Generally, strategic items used to develop weapons of mass destruction are controlled strictly under nuclear non-proliferation regimes and experts with extensive experience in the nuclear industry classify those items. With the development of technology and growing international cooperation, the government has established more elaborate and specific export control policies, which can control both tangible and intangible transfers of strategic items generated in various ways. Due to the growing number of international technology exchanges, there have been problems such as delays of classification and permit process, burdens of work, and inefficiencies in the overall export licensing process. This means that executing and completing export licensing could be less efficient.

Despite the rapid application of deep-learning technology in various fields, the development of support systems in the nuclear field is insufficient due to the industry's characteristics, the use of specialized terminology and limited data access. Therefore, this paper proposes a system model that automatically learns documents related to strategic item

classification, connects data to each other to infer results, and help making decision about classification especially for exporters. Chapter 2 explains the process of selecting an algorithm suitable for predicting results, and Chapter 3 explains the system configuration.

2. Methods and Results

In order to select proper deep learning algorithms for item classifications, the experiment was organized in the following order: experimental data collection, data preprocessing, overall system configuration, algorithms application, and consideration of the results. (Fig.1)

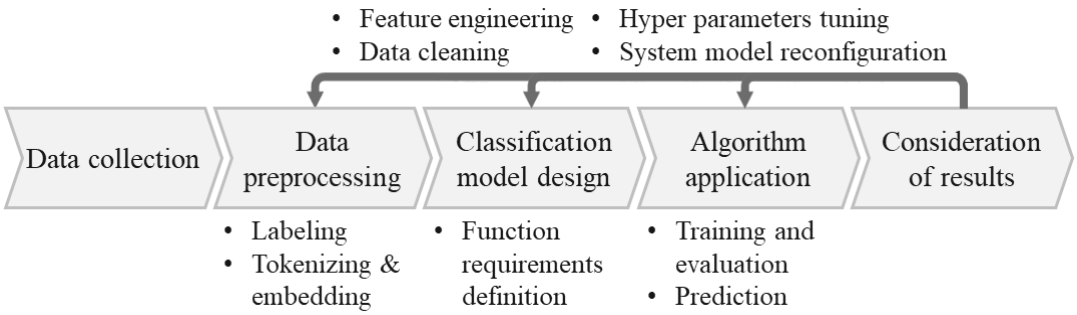

Fig. 1 Experiment Configuration

In general, deep learning methods are divided into supervised and unsupervised learning. Choosing the right method depends on the characteristics of learning data and problem-solving scenarios, supervised learning is appropriate in this study because it aim to determine whether the learning material is strategic or not.

2.1 Data pre-processing and preliminary experiments

Products or technologies classified as strategic items in the nuclear industry vary in form and type, ranging from R&D tasks, education programs, to facilities used in nuclear reactors and power plants. The exporters should submit documents indicating the performance, use, and technical characteristics of the goods, etc. to Korea Institute of Nuclear Nonproliferation and Control (KINAC) and the Korea Security Agency of Trade and Industry (KOSTI) in accordance with Article 14 of the Strategic Materials Import and Export Notice. In this study, item descriptions, manuals or specifications used in actual classification from 2010 to 2021 were collected and evaluated as raw data. Then, data are preprocessed and stored for the experiment, and the results are compared after each algorithm has been applied. At the same time, preprocessed data is labeled and randomly rearranged to calculate cosine similarity. Cosine similarity is mainly used to check the similarity between data, review search results, and check the significance of specific words and so on. This mechanism generally shows better performance because it includes direction in addition to simply calculating the distance from the coordinates. At last, the results were compared by applying deep learning algorithms Multi-layer perceptron (MLP), Constitutional neural network (CNN), Recurrent neural network (RNN), and Long short-term memory (LSTM) to data.

2.2 Results

As for the experimental environment, a Tensorflow-based Keras library was applied using Python. Fig.2 shows the results of experiments for each algorithm based on the code configured in the previous chapter. While MLP, CNN, and LSTM show stable learning results, RNN algorithms still produce unstable output after learning, making it difficult to expect reliable classification results from the data.

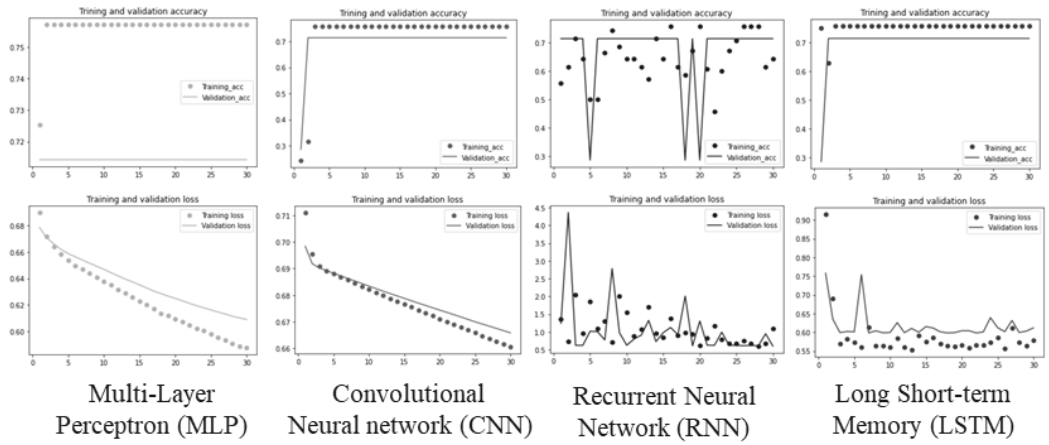

Fig. 2 Learning Results (Accuracy and loss) by Algorithms

This suggests that the data on the analysis of strategic items in this study are low in context and sequential interpretation dependence. On the other hand, CNN exhibits a steadily decreasing cost value in proportion to the increase in repetitions, and LSTM also provides improved learning results than RNN, and has the advantage of converging faster than CNN. Even though MLP algorithm is also similar to CNN, when developing actual code, the number of hidden layers and hyper-parameters must be adjusted directly. This makes it less convenient than other algorithms in system design. Therefore, LSTM is applied to the design of a classification support system model.

3. System Concept and General Requirements

This section presents a configuration plan for a strategic item classification support system based on experimental results. As described above, raw data is converted to learning data through preprocessing. By comparing and measuring similarity with other documents, most similar cases are displayed when new data is entered. Simultaneously, new data is added to the existing word model, and the classification result is displayed after deep learning.

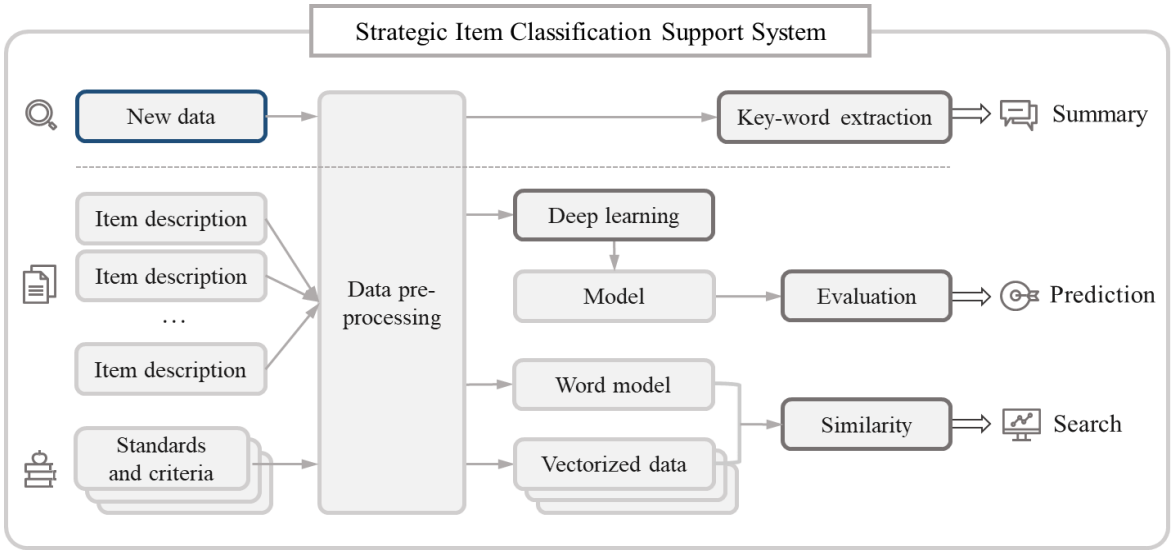

Fig. 3 System Configuration

Overall process leading to data input, learning, and output of the strategic items

classification support system configured from the experiment is as shown in Fig.3. The system predicts possible categories and results based on existing classification histories and the strategic items control list. All data should be processed in both Korean and English, and the major functional requirements of the system are as follows:

- ✓ Summary of relevant documents : Applied text summary technology
- ✓ Searching similar cases and control categories : Showing similar cases & classification results and strategic item categories & regulation information
- ✓ Prediction of classification results : Predicting results based on learned data

Particularly, KAERI has developed and operated its own import and export control system since 2020 to systematically control strategic items and nuclear material and improve researchers' convenience. It is expected to be effectively operated by linking the two systems proposed in this study.

4. Conclusions

Until now, many studies on natural language processing (NLP) using artificial intelligence have been conducted in various fields, but there have not been enough cases applied to the concept of strategic items in the nuclear field. In addition, compared to other technologies, the current system, which has low information openness and relies on a few skilled experts, is challenging to transfer knowledge. So artificial intelligence technology can be a way to effectively achieve nuclear nonproliferation and security with less manpower and cost. In particular, this system can support exporters to actively use their own information related to strategic items and the review results can be predicted to help prepare for follow-up tasks.

The preliminary experiment was conducted to predict the results of classifying items using major deep learning models (MLP, CNN, RNN, LSTM) that perform well in natural language processing, and to display past most similar cases by applying the TF-IDF embedding method. As a result, the LSTM algorithm showing stable performance results was selected as the classification support system model. It also proposed a system concept based on artificial intelligence for document summaries, similarity measurements, and prediction.

Since the learning data is based on documents written by the applicant himself, it is inevitable to rely on the details recorded by the author. The terms vary depending on the writer, and the frequency of use of various loanwords and abbreviations is high. Therefore, it is necessary to consider additional data processing to increase system performance in the future. Then, it is expected to efficiently increase accuracy by automating the process of reducing semantic redundancy, especially through the standardization process of frequently used terms. The results of this study will be used to develop a classification support system to determine whether the item is belonging to a strategic item list or not, which will contribute to securing KAERI's strategic item export control system and improving practical efficiency. Therefore, it is expected to reduce the administrative and economic burden on the control practice and further secure domestic and foreign technological competitiveness in strategic material management, contributing to overall strategic material export control.

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] Zhiyong Cui, Ruimin Ke, and Yinhai Wang, Deep Stocked Bidirectional and Undirectional LSTM Recurrent Neural Network for Network-wide Traffic Speed Prediction, 2018
[6] Foreign Trade Act, No.18885, 2022
[7] Strategic Materials Import and Export Notice, No. 2022-224, 2023