ADV MATH SCI JOURNAL

Advances in Mathematics: Scientific Journal **10** (2021), no.2, 1065–1069
ISSN: 1857-8365 (printed); 1857-8438 (electronic)
https://doi.org/10.37418/amsj.10.2.34

CLASSIFICATION OF COVID-19 SYMPTOM FOR CHATBOT USING BERT

Ho-yeon Park, Gun-doo Moon, and Kyoung-jae Kim[1]

ABSTRACT. Coronavirus disease (COVID-19) is a significant disaster worldwide from December 2019 to the present. Information on the COVID-19 is grasped through news media or social media, and researchers are conducting various research. This is because we are trying to shorten the time to be aware of the COVID-19 disaster situation. In this paper, we build a chatbot so that it can be used in emergencies using the COVID-19 data set and investigate how the analysis is changing the situation with deep learning.

1. INTRODUCTION

The World Health Organization (WHO) declared a "Pandemic" on the new coronavirus infection (COVID-19) on March 11, 2020, following Hong Kong flu in 1968 and swine flu in 2009. COVID-19 is an infectious disease with a high spreading power that, in December 2019, about 2428 million people in 215 countries are infected (as of August 28, 2020), and about 830,000 people are killed within eight months of the outbreak in Wuhan, China. This COVID-19 not only combines the disasters of environmental and social factors around the world, but the economic damage is rapidly increasing worldwide. Because COVID-19 is at risk for airborne droplets, it has become difficult to spend time

[1] *corresponding author*

2020 *Mathematics Subject Classification.* 68T07, 92B20.

Key words and phrases. Tweets, COVID, Disaster, Text Mining, Deep Learning, BERT.

Submitted: 12.02.2021; *Accepted:* 25.02.2021; *Published:* 04.03.2021.

outside the home unless people have a strong belief in the people they are meeting with. Such an environment has led to the state's border blockade. This situation has led the world's tourism economy to collapse in the travel industry, from international airlines to the smallest independent hotel and mart owners. According to the Organisation for Economic Co-operation and Development (OECD) policy report, the financial crisis, leading regional development, various types of jobs and businesses, and other factors supporting many local communities are being damaged by the collapse of the tourism industry. To prepare evidence for the damages of COVID-19, we intend to proceed with topic modeling by extracting economic articles from Korea and the world. This paper proposes a bidirectional encoder representations from transformers (BERT) chatbot model to utilize the problem for relation extraction through a disaster situation.

2. Related Works

The recent long-term spread of Corona 19 has negatively affected individuals and society, incomparable to the current epidemic outbreak. To overcome the COVID-19 situation that occurs face-to-face, we are trying to use many non-face-to-face cases. A chatbot is a deep learning-based technology that provides appropriate answers to questions and various related information through text conversations with people. The typical chatbot's structure is based on recurrent neural networks (RNN), and when one text sentence is input, it outputs one text sentence. Chatbot service has been researched representatively not only in shopping [1,2] and customer service [3] but also in other fields [4].

3. Methodology

Before developing deep learning embedding, text analysis was mainly used to identify trends in unstructured data. In the case of text analysis, the accuracy of the goal of the keyword to be searched can be studied in the deepest. However, research results are likely to involve the author's subjective judgment. Large amounts of data cannot be analyzed without the researcher's background knowledge and efforts on target keywords. In the early rule-based machine translation (RBMT) and statistical machine translation (SMT) period, the latent semantic analysis (LSA) algorithm that directly utilizes statistics has been proven effective in the research. In the era of neural machine translation (NMT), word embedding developed to sentence embeddings with deep learning. Representative examples of NMT are BERT embedding, and BERT has been popularly used in many studies in natural language processing (NLP) using pre-train and fine-tuning methods among various embedding techniques. Unlike a bidirectional long short term memory (Bi-LSTM) network, BERT has a fine-tuning network. Fine-tuning is an implementation that minimizes the cross-entropy loss of learning compared to the correct answer label. The cross-entropy loss of learning can be minimized because fine-tuning does not set each embedding of the dataset layer like Bi-LSTM. This is because the model from training data to word embedding and fine-tuning is updated at once.

4. Experiments and Results

To confirm the superiority of the methodology described above, we conduct several experiments. In this study, classification analysis was made using deep learning and BERT and then compared through a typical chatbot. The dataset was provided by CORD-19, a website designed for COVID-19 research. Wang et al. [5] developed CORD-19 website to contribute advances in natural language processing. The data used were 4,134 articles on COVID-19 and 100 interactive data sets for chatbots. The research paper's role-play data was used to understand the context to add many terms and intentions of the chatbot. Table 1 shows an example of the data set structure in this study.

Table 1. Dataset example

Question	Answer	Label
I'm sick right now.	Just tell me in detail again.	0
I have a runny nose and fever.	Heat and cold. You have to rest for a while. It's an early symptom.	1
Neck, runny nose, and fever.	I think I got corona. It would be better to visit the hospital.	2
...	...	...

Each label value was divided into 0 for no symptom, 1 for cold and fever, and 2 for the virus. To create answers to the sentences, the dataset used the NLTK

(Natural Language Toolkit). NLTK is a tool that can be tokenized and is divided by measuring the length of words. In the NLTK process, text preprocessing for stopwords, numbers, punctuation marks, and POS tagging is performed. After NLTK preprocessing, sentences longer than the maximum length to be applied to the model are truncated. For short sentences, make sure that all words are padded to the full size. The decoder preprocessing puts a start-token as an input value and an end-token at the end of the sentence. In the experiment, the batch size was 2, the number of the epoch was 40, the embedding dimension was 256, and the neural network unit was 1,024. Table 2 shows the experimental results of this study.

TABLE 2. Dataset example

Model	F1Score	Sensitivity	Question	Answer
Base (RNN)	0.6288	0.6944	I have a little head pain. What do you think of coughing and snoring?	Just tell me in detail again.
CNN-LSTM	0.7527	0.7319		You better rest.
BERT	0.8113	0.8123		You better rest.

The F1 score was used to compare the performance between models. Compared with BERT, which showed the best performance, RNN performed the least, and CNN-LSTM was less than BERT.

5. CONCLUSIONS

In this paper, chatbots were constructed and compared using RNN, CNN-LSTM, and BERT. As a structure to generate output sentences through input sentences, the research paper dataset and interactive dataset were used as training data. As a result of the experiment, BERT was the best, followed by CNN-LSTM and RNN. However, unexpectedly, the chatbot in this paper lacked a data set of human common sense. Because of the lack of interactive data for chatbots, sentence delivery rates were low, and simple sentence delivery rates were excellent in many conversations. In future studies, a method of automatically generating complex interactive data centered around is needed. Besides, we look forward to the emergence of research using the model structure for transfer learning in deep learning using BERT and text similarity.

6. ACKNOWLEDGMENT

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2019S1A5A2A01050194).

REFERENCES

[1] M. CHUNG, E. KO, H. JOUNG, S. J. KIM: *Chatbot e-service, and customer satisfaction regarding luxury brands,* Journal of Business Research, **117** (2020), 587–595.

[2] L. CIECHANOWSKI, A. PRZEGALINSKA, M. MAGNUSKI, P. GLOOR: *In the shades of the uncanny valley: An experimental study of human–chatbot interaction,* Future Generation Computer Systems, **92** (2019), 539–548.

[3] B. SHEEHAN, H.S. JIN, U. GOTTLIEB: Customer service chatbots: Anthropomorphism and adoption. Journal of Business Research, **115** (2020), 14–24.

[4] C. S. KULKARNI, A. U. BHAVSAR, S. R. PINGALE, S. S. KUMBHAR: BANK CHAT BOT–An Intelligent Assistant System Using NLP and Machine Learning. IRJET (International Research Journal of Engineering and Technology, **4** (05) (2017).

[5] L. L. WANG, K. LO, Y. CHANDRASEKHAR, R. REAS, J. YANG, D. EIDE, K. FUNK, R. M. KINNEY, Z. LIU, W. MERRILL, P. MOONEY, D. MURDICK, D. RISHI, J. SHEEHAN, Z. SHEN, B. STILSON, A. WADE, K. WANG, C. WILHELM, B. XIE, D. RAYMOND, D. S. WELD, O. ETZIONI, S. KOHLMEIER: CORD-19: The COVID-19 Open Research Dataset. ArXiv. (2020)

DEPARTMENT OF MIS
DONGGUK UNIVERSITY-SEOUL
JANGCHUNG-DONG, JUNG-GU, SEOUL,
SOUTH KOREA.

DEPARTMENT OF MIS
DONGGUK UNIVERSITY-SEOUL
JANGCHUNG-DONG, JUNG-GU, SEOUL,
SOUTH KOREA.

DEPARTMENT OF MIS
DONGGUK UNIVERSITY-SEOUL
JANGCHUNG-DONG, JUNG-GU, SEOUL,
SOUTH KOREA.
Email address: kjkim@dongguk.edu