International Journal of Advanced Culture Technology Vol.9 No.4 268-273 (2021)
DOI https://doi.org/10.17703/IJACT.2021.9.4.268

IJACT 21-12-31

A Study on Korean Sentiment Analysis Rate Using Neural Network and Ensemble Combination

[1]YuJeong Sim, [2]Seok-Jae Moon, [3]Jong-Youg Lee

[1]Graduate School of Smart Convergence Kwangwoon University, Korea
[2]Professor, Institute of Information Technology, KwangWoon University, Korea
[3]Professor, Ingenium College of Liberal Arts, KwangWoon University, Korea
{tladbwjd, msj8086, jyonglee}@kw.ac.kr

Abstract

In this paper, we propose a sentiment analysis model that improves performance on small-scale data. A sentiment analysis model for small-scale data is proposed and verified through experiments. To this end, we propose Bagging-Bi-GRU, which combines Bi-GRU, which learns GRU, which is a variant of LSTM (Long Short-Term Memory) with excellent performance on sequential data, in both directions and the bagging technique, which is one of the ensembles learning methods. In order to verify the performance of the proposed model, it is applied to small-scale data and large-scale data. And by comparing and analyzing it with the existing machine learning algorithm, Bi-GRU, it shows that the performance of the proposed model is improved not only for small data but also for large data.

Keywords: *Neural network, Ensemble combination, Bi-GRU, LSTM, Korean Sentiment*

1. INTRODUCTION

In this paper, we try to improve the existing algorithm in this paper. In order to analyze textual user review data at higher performance and faster speed than conventional methods, the GRU (Gated Recurrent Units) [1] model, a variant of LSTM [2] that excels in sequential data such as language modeling and speech recognition, is forwarded. In addition, Bi-GRU that uses the result in the reverse direction is used. GRU does not have a memory to store and memorize the cell state, so if the amount of data increases, the result may be worse than that of LSTM *(Long Short-Term Memory)*. However, when a small amount of data is used as in this paper, since the amount of computation is smaller than that of LSTM, good results can be expected in terms of computation time and inference accuracy. In addition, Bi-GRU shows better performance because it preserves future information and uses two states in combination, unlike GRU that only preserves past information. With this Bi-GRU model, training and testing are processed in parallel to improve overall performance. In order to reduce individual differences between models trained in parallel and to improve generalization errors, bagging, one of the ensemble techniques, was combined. That is, the proposed model is defined as Bagging-Bi-GRU. In order to verify the model Bagging-Bi-GRU proposed in this paper, word inference was performed using the Word2vec technique. And compared with Bi-GRU, which is an existing machine learning algorithm, it was shown that the proposed Bagging-Bi-GRU improved accuracy by 4.3% compared to Bi-GRU. The structure

Manuscript received: August 11, 2021 / revised: November 2, 2021 / accepted: December 2, 2021
Corresponding Author: msj8086@kw.ac.kr
Tel: + 82-01-7753-8086, Fax: +82-2-916-4751
Department of Information Security Engineering, Institute of Information Technology

of this paper is as follows. Chapter 2 describes the proposed model and experimental environment, and Chapter 3 describes analyzes the experiment result of the proposed model. Finally, Section 4 describes the conclusion.

2. PROPOSAL MODEL

2.1 Proposed Model Block Diagram

The block diagram of the model proposed in this paper is shown in <Figure 1>. The first task is to create 9 bootstrap sample [3, 4] data through restoration random extraction from the collected original data set. As the second task, polarity detection is performed to determine positive/negative using Word2vec [5] and Bi-GRU for each bootstrap sample data. Finally, we conduct an experiment in which the outcome variables of each model are aggregated by a majority vote method. The experimental design in this paper is shown in <Figure 1> and consists of three steps.

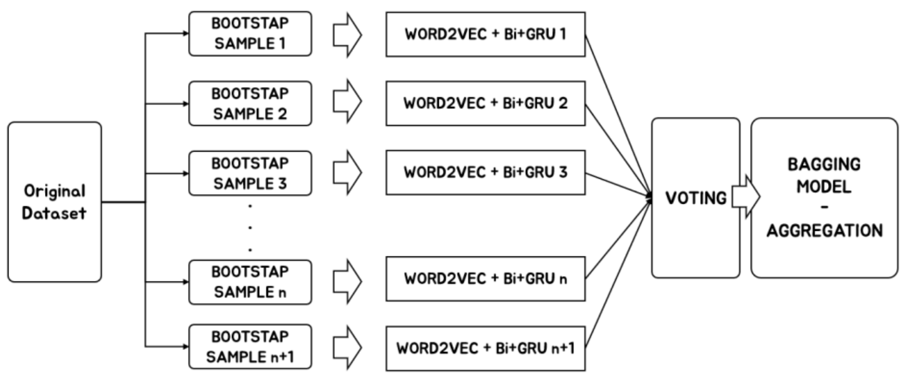

Figure 1. Proposal Model block diagram

Stage 1 is text preprocessing, stage 2 is data bootstrap, and stage 3 is sentiment analysis. First, data is collected, and then data purification steps are performed to remove all except Hangul. Then, tokenization is performed using a stemmer. In this process, unnecessary tokens are removed by designating stopwords such as 'like', 'when', 'of', 'and', etc. Now, integer encoding is performed on the data so that the computer can treat text as a number, and padding is performed to equalize sample lengths of different lengths. Bootstrap the data after padding.

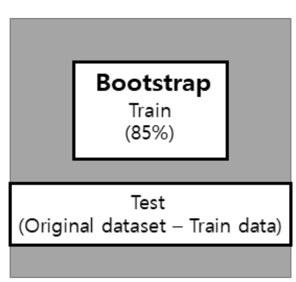

Figure 2. Proposal Model Bootstrapping Steps

The bootstrap process sets the training data to 85% of the experimental data as shown in <Figure 2>, and the rest except for the restored random sampled data from the training data is set as the validation data. As a result, the training data and the validation data do not overlap each other. Polarity detection that classifies positive/negative with the training data and validation data created in this way.

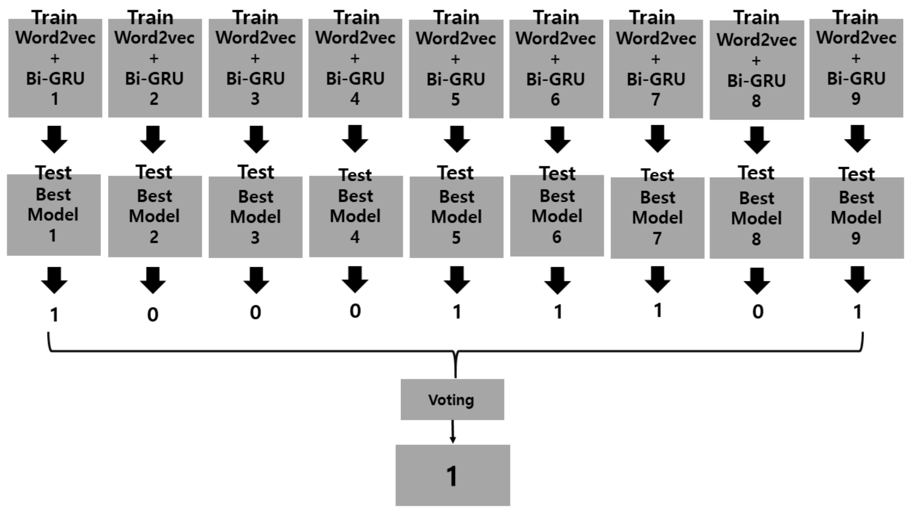

Figure 3. Suggestion model sentiment analysis stage

As shown in <Figure 3>, in the polarity detection stage, Word2vec training is performed on the bootstrapped training data first. Thereafter, each of the nine Bi-GRU-trained models is independently trained in parallel. Using the nine models learned in this way, the final model is trained by applying the majority voting method. To verify the validity of the trained model, validation data was used.

2.2 Experimental Data Sets

In this paper, Word2vec is implemented using Gensim, a library for natural language processing, and the sentiment classification model uses a model that combines Bi-GRU and bagging [6], [7], [8]. There are three datasets: small movie review data, large movie review data, and general review data. In this paper, small-scale movie review data is set as 'Dataset I', large-scale movie review data as 'Dataset II', and general review data as 'Dataset III'. Dataset I use 15,000 Naver movie reviews crawled using the Python library BeautifulSoup [9]. This is the data collection stage of the three-step process of sentiment analysis. Dataset I is somewhat smaller data than the papers described in the case of sentiment analysis. Dataset II uses 200,000 NSMC data used in the sentiment analysis case [10]. Dataset III uses 200,000 Naver shopping review data, which is general review data [11]. The three-review data are tested with a positive/negative ratio of 1:1. A summary of the experimental dataset is shown in <Table 1>. Dataset I is Naver movie review data crawled using the Python library BeautifulSoup.

Table 1. Experimental data

Sentiment analysis training and evaluation dataset	Number of Data
Dataset I	15,000 Naver movie reviews (Positive Negative Ratio 1:1) - 7500 positive data (8-10 points), - 7500 negative data (1-7 points)
Dataset II	200,000 NSMC data (Positive Negative Ratio 1:1) - 100,000 positive data, - 100,000 negative data
Dataset III	200,000 Naver Shopping review data (Positive Negative Ratio 1:1) - 100,000 positive data, - 100,000 negative data

The data consists of a total of 15,000 reviews, with a star rating out of 10, and data consisting of an integer from 1 to 10. In order to assign a star point as an emotional value label, 500 sample data from the total data are extracted and from which point a sentence containing a negative word appears. In this paper, tokenization is performed using the stemming analyzer Mecab [12]. The data were processed as refined sentences. When constructing the Word2vec model with the morphologically analyzed data, the Skip-Gram method was used, and the parameters for model building are shown in <Table 2>. Bi-GRU was used as the neural network model for sentiment analysis, and the parameters for model construction are shown in <Table 2>.

Table 2. Parameters for building the model

size: the number of dimensions of the vector *window: the number of words considered before/after training* *min_count: the minimum number of occurrences in the data* *workers: Number of processes for learning*			
activation function	Sigmoid: $S(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1}$		
loss function	Binary Cross Entropy: $BCE = \sum_{i=1}^{C'} t_i \, log\big(f(s_i)\big) = -t_1 \, log\big(f(s_1)\big) - (1-t_1) \, log\big(1-f(s_1)\big)$		
learning technique	RMSProp: $G_t = \gamma G_{t-1} + (1-\gamma)\big(\nabla_\omega J(\omega_t)\big)^2$ $\omega_{t+1} = \omega_t - \frac{\eta}{\sqrt{G_t+\epsilon}} \cdot \nabla_\omega J(\omega_t)$		
Regulatory Techniques	EarlyStopping, ModelCheckpoint		
Epochs	number 10	Batch size	size 60

3. EXPERIMENT RESULT

The experiment was compared by applying the existing model Bi-GRU and the Bagging-Bi-GRU proposed in this paper to Dataset I, Dataset II, and Dataset III <Table 3>. After the experiment, the precision, recall, and accuracy of the existing model and the proposed model were calculated using the model evaluation index. In addition, to verify the stability of the proposed model, the accuracy variance of the existing model and the proposed model was compared.

Table 3. Comparison of experimental results for datasets I, II, and III

Comparison of experimental results for Dataset I			
Model	**Precision**	**Recall**	**Accuracy**
Word2vec+Bi-GRU	85%	84.15%	84.65%
Proposal (Word2vec + Bagging-Bi-GRU)	88% (3% increase)	89.79% (5% increase)	89% (4.3% increase)
Comparison of experimental results for Dataset II			
Model	**Precision**	**Recall**	**Accuracy**
Word2vec+Bi-GRU	83%	85.17%	84%
Proposal (Word2vec + Bagging-Bi-GRU)	87% (4% increase)	86.13% (5% increase)	86.97% (2.97% increase)
Comparison of experimental results for Dataset III			
Model	**Precision**	**Recall**	**Accuracy**
Word2vec+Bi-GRU	91.26%	92%	91.59%
Proposal (Word2vec + Bagging-Bi-GRU)	92.27% (1.01%increase)	93.21% (1.21% increase)	92.80% (1.39% increase)

4. CONCLUSION

In this paper, we compared the proposed model with Bi-GRU using Word2vec by applying it to sentiment analysis with Korean movie review data. A model suitable for Korean natural language processing was created by combining the neural network model and ensemble learning, and the study was conducted focusing on improving the performance of sentiment analysis on small data. The experiment was conducted with Dataset I (small movie review data), Dataset II (large movie review data), and Dataset III (general review data). As a result of the experiment, the performance of the proposed model was improved compared to the existing model for small movie review data, dataset I, with a precision of 3%, a recall rate of 5%, and an accuracy of 4.3%. The performance of the proposed model was improved compared to the existing model for the large-scale movie review data of Dataset II, with a precision of 4%, a recall rate of 0.96%, and an accuracy of 2.97%. Finally, the general review data of Dataset III has a precision of 1.01%, a recall rate of 1.21%, and an accuracy of 1.39%, which improves the performance of the proposed model compared to the existing model. In addition, the accuracy variance of the existing Bi-GRU model and the proposed model was compared. As a result of the experiment, the variance of the proposed model was reduced to 0.323 for Dataset I, 0.312 for Dataset II, and 0.282 for Dataset III, showing superior performance than the existing model. The purpose of this paper is to propose a Korean sentiment analysis model that improves accuracy better than the basic model and improves performance on small data. As a result, it can be seen that performance for large-scale data as well as small-scale data is improved. In addition, the performance for general review data was also improved. Although the proposed model has the disadvantage that it takes a long time, it showed that the performance improved compared to the existing model in all experiments, and it was experimentally confirmed that it provides versatility.

REFERENCES

[1] Pang, B. & Lee, L. "Opinion Mining and Sentiment Analysis", Foundations and Trends in Information Retrieval 2 (1-2), pp.1-35, 2008.
[2] Feldman, Ronen. "Techniques and applications for sentiment analysis." Communications of the ACM 56.4, pp. 82-89, 2013.
[3] Jain, Anmol, Aishwary Kumar, and Seba Susan. "Evaluating Deep Neural Network Ensembles by Majority Voting Cum Meta-Learning Scheme." Soft Computing and Signal Processing. Springer, Singapore, pp.29-37, 2022.
[4] Mustafin, Askar N., et al. "Using models of collective neural networks for classification of the input data applying simple voting." The Journal of Social Sciences Research", pp.333-339, 2022.
[5] Di Gennaro, Giovanni, Amedeo Buonanno, and Francesco AN Palmieri. "Considerations about learning Word2Vec." The Journal of Supercomputing", pp1-16, 2021.
[6] Keras, https://github.com/keras-team/keras
[7] Gensim word2vec, https://radimrehurek.com/gensim/
[8] Python Docs, https://docs.python.org/3.5/
[9] Github, https://github.com/e9t/nsmc
[10] Github, https://github.com/bab2min/corpus/tree/master/sentiment
[11] Na, Seung-Hoon, and Young-Kil Kim. "Phrase-based statistical model for korean morpheme segmentation and POS tagging." IEICE Transactions on Information and Systems 101.2, pp.512-522, 2021.
[12] Korean morpheme analyzer, Konlpy, https://www.konlpy.org/