International Journal of Advanced Culture Technology Vol.9 No.2 86-90 (2021)
DOI https://doi.org/10.17703/IJACT.2021.9.2.86

IJACT 21-6-11

Sentiment analysis of Korean movie reviews using XLM-R

[1]Noo Ri Shin, [2]TaeHyeon Kim, [3]Dai Yeol Yun, [4]Seok-Jae Moon, [5]Chi-gon Hwang

[1]Graduate School of Smart Convergence KwangWoon University
[2]Department of Plasma Bioscience and Display, KwangWoon University
[3]Professor, Department of information and communication Engineering, Institute of Information Technology, Kwangwoon University, Seoul, 01897, Korea
[4]Professor, Institute of Information Technology, Kwangwoon University, Seoul, Korea
[5]Visiting Professor, Department of Computer Engineering, Institute of Information Technology, Kwangwoon University, Seoul, 01897, Korea
{4nchez, surowang, hibig10, msj8086, duck1052}@kw.ac.kr

Abstract

Sentiment refers to a person's thoughts, opinions, and feelings toward an object. Sentiment analysis is a process of collecting opinions on a specific target and classifying them according to their emotions, and applies to opinion mining that analyzes product reviews and reviews on the web. Companies and users can grasp the opinions of public opinion and come up with a way to do so. Recently, natural language processing models using the Transformer structure have appeared, and Google's BERT is a representative example. Afterwards, various models came out by remodeling the BERT. Among them, the Facebook AI team unveiled the XLM-R (XLM-RoBERTa), an upgraded XLM model. XLM-R solved the data limitation and the curse of multilinguality by training XLM with 2TB or more refined CC (CommonCrawl), not Wikipedia data. This model showed that the multilingual model has similar performance to the single language model when it is trained by adjusting the size of the model and the data required for training. Therefore, in this paper, we study the improvement of Korean sentiment analysis performed using a pre-trained XLM-R model that solved curse of multilinguality and improved performance.

Keywords: *Sentiment analysis, Transformer, BERT, XLM-R, Transfer learning, Fine tuning*

1. INTRODUCTION

Sentiment refers to a person's thoughts, opinions, and feelings toward an object. Sentiment analysis is a process of collecting opinions on a specific target and classifying them according to their emotions, and applies to opinion mining that analyzes product reviews and reviews on the web. Companies and users can understand public opinion about products and services and make choices about them. However, it is difficult to grasp a person's emotions contained in natural language. Even similar-looking sentences can be divided into positive and negative opinions with small changes, and their meanings can also vary depending on the speaker. [1-2]

As for the sentiment analysis technique, there was a technique using a dictionary in which words containing positives and negatives were constructed, but recently, it is analyzed using a machine learning technique. Techniques such as CNN (Convolutional Neural Network), RNN (Recurrent Neural Network), and LSTM (Long Short-Term Memory) using word embedding have also appeared.

Manuscript Received: April 19, 2021 / Revised: May 12, 2021 / Accepted: May 20, 2021
Corresponding Author: duck1052@kw.ac.kr
Tel*** - **** - ****
Visiting Professor, Department of Computer Engineering, Institute of Information Technology, Kwangwoon University, Korea

However, the technique of using a neural network after word embedding has a disadvantage in that one word has one vector value, and the context cannot be considered. For example, in Korean, '배' is used as the meaning of a person's body part, but it is also used as a means of fruit and transportation. Since the embedding vector has one vector value in the vector called '배', the two meanings cannot be distinguished. In addition, RNN models, such as LSTM, have a slower computational speed with longer sentences, and are difficult to properly represent relationships between distant words. In order to overcome this limitation, models using the Transformer structure emerge. A prime example is Google's BERT [5].

BERT (Bidirectional Encoder Representations from Transformers) used encoders of Transformer to demonstrate the model's bidirectionality using the MLM (Masked Language Model) and NSP (Next Sentence Prediction) training methods, and to disclose pre-trained models. Since its appearance, BERT has surpassed the existing SOTA (latest technology) algorithm in various natural language processing tasks, and various models such as RoBERTa (Robustly Optimized BERT Pretraining Approach) and ALBERT (A Lite BERT) that modified BERT appeared.

Among them, the Facebook AI team unveiled the XLM-R (XLM-RoBERTa), an upgraded XLM model. XLM-R solved the data limitation and the curse of multilinguality by training XLM with 2TB or more refined CC (CommonCrawl), not Wikipedia data. [6] This model showed that the multilingual model has similar performance to the single language model when it is trained by adjusting the size of the model and the data required for training. Also, like BERT, the pre-trained XLM-R model was also released. Pre-trained models have the advantage that they can be reused for various tasks with a small amount of data.

Therefore, in this paper, we study Korean sentiment analysis using a pre-trained model of XLM-R Base that solves the curse of multilinguality and improved performance. The data used in the study is the public data NSMC (Naver Sentiment Movie Corpus) [7] in Korean. In the XLM-R, experiments for other languages such as XNLI exist, but there is no experiment using Korean. Therefore, in this study, we conduct Korean sentiment analysis using the XLM-R model and study a method to improve performance at the same time. In addition, the pre-trained model is fine-tuned to find the optimal model and compare the performance with existing models. As a result of the experiment, the performance improved by up to 3.63% compared to the previous model.

2. RELATED RESEARCH & TECHNIQUES

2.1. Related Research

Pre-training is a method of constructing similar data for a problem to be solved and training it in advance, and has the advantage of being able to train by putting large unlabeled data. Natural language processing research has been studied in a variety of ways, from pretrained word embeddings to pretrained transformer-based language models, and greatly improved the technology. In particular, transformer-based language models showed good performance in the field of natural language processing. The Korean language field also conducted various studies using this method. In [3], performed Korean natural language processing using BERT. By comparing the existing multilingual BERT and the BERT trained with the large-capacity Korean corpus, Korean natural language processing tasks such as Named Entity Tagging, Sentiment Analysis, Dependency Parsing and Semantic Role Labeling were performed. In [4], created a lightweight KR-BERT model that reflected the characteristics of Hangul and compared it with the existing Korean models, and processed the Korean natural language task.

2.2. Techniques

2.2.1 XLM-R (XLM-RoBERTa)

XLM-RoBERTa [6] is an upgraded version of the existing model, XLM, and is inspired by RoBERTa. As the existing XLM model is also under-tuned like BERT, it was determined that simple improvement of the learning process of unsupervised MLM leads to much better performance. Training data was trained with large-

scale refined CC, not Wikipedia, Unlike the previous method, it used a 250k token Sentence-Piece-based vocabulary. As a result, it showed better performance compared to the existing multilingual models, and in particular, the performance for low-resource languages (Swahili and Urdu) was far superior to that of the existing models. In addition, it was shown that the multilingual model performs similarly to the single language model when training by adjusting the size of the model and data required for training.

3. XLM-R FINE-TUNING

The model used in the experiment is XLM-R Base (named XLM-R). The structure of the model is 12 layers, the number of neurons in the hidden layer is 768, the attention layer is 12, and the dropout ratio is 0.1. Figure 1 shows the model structure used in this study. When fine tuning the model, the existing hyper parameters of XLM-R were left as they were, and a layer was added or the prune method was applied to the model. The added layers are BiLSTM and BiGRU (L = 1, H = 128, D = 0.1 are applied equally to both layers), and a 20% global prune is applied to the 12 attention layers of XLM-R.

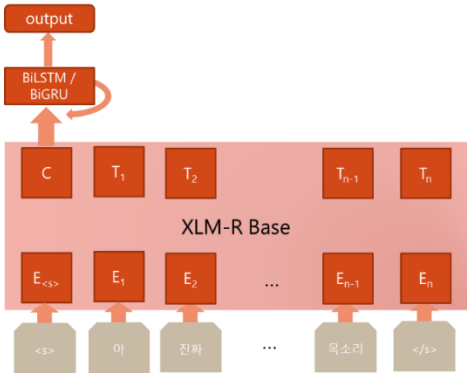

Figure 1. XLM-R Model Structure

The data used for sentiment analysis is NSMC. NSMC consists of 150k training data and 50k test data, labeled positive (1) and negative (0). For model validation in the experiment, training data was divided into training data and validation data at a ratio of 9:1. Therefore, the training data has 135,000.

The experiment was conducted in Google's Colaboratory, and the pre-trained model was taken from the Hugging Faces library. First, Korean sentiment analysis was performed with XLM-R to confirm the results, and then the experiment was performed by adding the prune method and layer. Table 1 is a table of settings used in the experiment. The training epoch was set to 4 times, the maximum sentence length was 128, and the batch size was 32. AdamW was used as the learning algorithm, and the learning rate was set to 2e-5, and epsilon was set to 1e-8. Gradient clipping was applied to prevent the exploding gradient problem.

Table 1. Hyper Parameters

Hyper Parameters	XLM-R
Epoch	4
Learning Rate	2e-5
Epsilon	1e-8
Gradient Clipping	1.0
Max Length	128
Batch Size	32
Global prune	20%
Dropout	0.1

In order to compare the results of the sentiment analysis using the proposed models, the evaluation indexes of Precision, Recall, and Accuracy were used. These indicators are frequently used in classification problems, and the formula of the proposed evaluation indicator is as follows. Equation (1) is the proportion of what the model classifies as True, which is actually True. Equation (2) is the ratio of what the model predicts to be true out of what is actually true. Equation (3) is the ratio of correct answers from the whole. Simply, regardless of True or False, it is just a percentage of correct answers, and the closer to 1, the better.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

4. EXPERIMENT RESULT

Table 2 is a comparison result of fine-tuning models in this paper and existing sentiment analysis models. The results of BERT (Multilingual) and BERT (Morpheme-tag) were taken from [3], BERT (Multilingual) is Google's multilingual BERT, and BERT (Morpheme-tag) is a model learned in Korean. The results of KoBERT (SKT), KorBERT (ETRI), and KR-BERT were obtained in [4], and the above models were trained in Korean. The sentiment analysis result using the XLM-R model was better than the existing models.

Among the fine-tuned models in this paper, XLM-R+prune is a model that globally prunes the L1 norm weight part of the attention (key, query, value) of the encoder part of the XLM-R model. XLM-R+BiLSTM is a model that performs sentiment analysis by applying the context vector generated through the XLM-R model to Bi-LSTM. XLM-R+prune+BiLSTM is a model that performs sentiment analysis by applying the context vector generated through the XLM-R model to which prune is applied to Bi-LSTM. XLM-R+BiGRU is a model that performs sentiment analysis by applying the context vector generated through the XLM-R model to Bi-GRU. XLM-R+prune+BiGRU is a model that performs sentiment analysis by applying context vector generated through XLM-R model to which prune is applied to Bi-GRU.

Comparing the results, the model that applied prune to XLM-R and combined Bi-GRU showed the best performance. In the precision and recall part, it also showed better performance than other fine-tuning models. Comparing with the existing models, the accuracy is up to 3.63% different from the existing models. By cutting unnecessary parts of the learning model with prune and combining BiGRU to increase the size of the model, the performance could be improved.

Table 2. Performance Comparison

Model	Precision (%)	Recall (%)	Accuracy (%)
BERT(Multilingual) [3]	-	-	87.43
BERT(Morpheme-tag) [3]	-	-	86.57
KoBERT (SKT) [4]	-	-	89.01
KorBERT (ETRI) [4]	-	-	89.84
KR-BERT [4]	-	-	89.38
XLM-R	89.88	89.80	89.84
XLM-R+prune (our)	90.08	90.03	90.06
XLM-R+BiLSTM (our)	90.04	90.04	90.04
XLM-R+prune+BiLSTM (our)	90.16	90.13	90.15
XLM-R+BiGRU (our)	90.17	90.14	90.15
XLM-R+prune+BiGRU (our)	**90.23**	**90.20**	**90.20**

5. CONCLUSION

In this paper, Korean sentiment analysis was conducted using the pre-learning model XLM-R, which supplemented the problems found in the multilingual support model. As a result, it showed better performance compared to the pre-trained model in single language. However, a pre-learning model supporting multiple languages must use more Vocabs than a pre-learning model specialized for a single language, and this can increase the size of the model and increase the learning time and worsen the learning result. Therefore, when pre-train in a single language, such as RoBERTa, focusing on learning methods or learning with more data sets, performance is expected to improve further.

REFERENCES

[1] Y.T. Oh, M.T. Kim, and W.J. Kim, "Korean Movie-review Sentiment Analysis Using Parallel Stacked Bidirectional LSTM Model," Journal of KIISE 46.1, 45-49, 2019. doi: 10.5626/JOK.2019.46.1.45

[2] G.Y. Kim, and C.K. Lee, " Korean Movie Review Sentiment Analysis Using Convolutional Neural Network," In: Proc. of the KIISE Korea Computer Congress, 747-749, 2016.

[3] K.H. Park, S.H. Na, J.H. Shin, and Y.K. Kim, "BERT for Korean Natural Language Processing: Named Entity Tagging, Sentiment Analysis, Dependency Parsing and Semantic Role Labeling," The Korean Institute of Information Scientists and Engineers, 584-586, 2019.

[4] S.A. Lee, H.S. Jang, Y.M. Baik, S.Z. Park, and H.P. Shin, "A Small-Scale Korean-Specific BERT Language Model," Journal of KIISE 47.7, 682-692, 2020. doi: 10.5626/JOK.2020.47.7.682

[5] J. Devlin, M.W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv preprint arXiv:1810.04805, 2018.

[6] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzman, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised Cross-lingual Representation Learning at Scale," arXiv preprint arXiv:1911.02116, 2019.

[7] L. Park, Naver sentiment movie corpus v1.0, https://github.com/e9t/nsmc