

Using RoBERTa-BiGRU and improved Stacking Ensemble Learning for Sentiment Classification

Lan You¹, Zhengyi Guan¹, Zeng Han¹, Zhang JiaHe¹, Jia Jing¹, Qingxi Peng², Fanyu Han¹, Keumseok Peter Koh^{3,*}

(1. School of Computer Science and Information Engineering, Hubei University, 430062 Wuhan, P.R. China;

(2. Wuhan Donghu University, School of Computer Science, Wuhan 430212, China;

(3. Computer Science and Deputy Director, the Naval Academy Research Institute in France, Lanveoc-Poulmic, BP 600, 29240 Brest, France;)

(Correspondence: claramunt@ecole-navale.fr)

Abstract: Along with the development of information technology, comments on social networks have become a major way for the public to express their attitudes and standpoints on social events. Accurately identifying the sentiment orientation of social texts is of great value for public opinions control, and social stability maintenance. Since the traditional sentiment recognition models only focus on mining plain semantics on reviews, many problems including poor classification effect and limited generalization ability still existed. With regard to the shortcomings, this paper proposed a deep sentiment semantic recognition model based on Bidirectional Encoder Representations from Transformers and Bidirectional Gated Recurrent Unit (BERT-BiGRU) multi-model ensemble learning. Firstly, the contextual semantic feature representation of the review text is obtained through the BERT pre-training model. And then combined with the BiGRU to extract the deep nonlinear feature vectors, so as to achieve the optimal sentiment recognition results under the single model. Next, to obtain a model with stable effect and better performance on the multi-aspect, we combine multiple sentiment classifiers with superior performance and differentiation on the BERT series pre-training models using the Staking ensemble learning algorithm. Finally, integrate the deep features of each base model adopting Categorical Boosting (CatBoost) algorithm to achieve the final optimization. The experimental results on two public datasets (COV19 and ChnSenti) show that the model proposed in this paper has superior sentiment recognition effects than other traditional models.

Key words: sentiment recognition; the BERT pre-training model; the bidirectional GRU; ensemble learning; deep feature

1 Introduction

Sentiment recognition is the gordian technique of artificial intelligence, which is to perceive and understand human sentimental intention through text, image, and other media from the perspective of the machine. Due to the huge amount of unstructured text comments generated at any moment, which usually contain the subjective sentimental intention of the user, nowadays social networking sites are becoming an important platform for people to follow current events and share their personal views. The sentiment recognition of social comments is of great significance to public opinion control [1], commercial marketing [2][3], social governance [4], and has become one of the research hot spots among the field of natural language processing in recent years.

Social network site commends have typical features such as rich sentimental meaning and different text lengths. Hence, how to judge sentimental polarity from texts of different lengths is a key issue that the current sentiment recognition system urgently needs to solve [6]. The traditional Word2Vec (Word to Vector) or GloVe (Global Vectors) word vector pre-training language model could, to some extent, learn the context information of words. However, there are deviations in the semantic model processing of the same words in different contexts with different meanings. As is shown in two phrases, "Do you like the area where you are living?" and "The city covers an area of 30,000 square kilometres.", in which the word "area" represents the region in the former example, whereas it stands for the meaning of 'proportion' in the latter example. After using Word2Vec word vector model for dimensionality reduction, it is found that the word "area" coincidences in two-dimensional space, which indicates that the "area" vector expressed by Word2Vec in two examples is consistent. Therefore, the polysemy of word is difficult to be recognized in traditional language models. In this paper, we adopt the BERT model to solve this problem by dynamically adjusting the semantic features of word vectors. Moreover, The BiGRU was used for

semantic coding to enhance the semantic expression of the text and mine the deeper semantic information of long-distance contextual sentiment in online text.

In the current research, most models have made great progress in predicting text sentiments through machine learning or deep learning [5-7]. However, most of the models are single model. Because of its randomness, the single model may perform well in one aspect, but there is insufficient generalization ability. Therefore, this paper makes full use of the differences between models and adopts integrated learning as well as voting strategies to fuse multiple models, aiming to train a stable sentiment classification model with balanced performance in all aspects. Consequently, a deep sentimental recognition model based on BERT integrated network is proposed in this paper to solve the problems of insufficient extraction of deep sentimental features of text and limited generalization ability. The main contributions are as follows:

(1)The BERT pre-training model is used to replace the word embedding layer of the traditional model to obtain the implicit semantic word vector representation of the text, and then the deep semantic features of the context are extracted through the bidirectional GRU, so as to improve the feature extraction ability of the model for the text sentiment.

(2)Using a stacking strategy to integrate the results generated by multiple excellent and differentiated base classifiers to obtain a stable and balanced sentiment classifier, which effectively improves the generalization ability of the model.

(3)The results of multi-group comparative experiments on public datasets(COVID-19 and ChnSenti) demonstrate that the proposed model has achieved better sentiment classification effect and stronger generalization ability.

5 Conclusion and Future Work

Concerning the research hotspot of sentiment recognition in social network text, this paper proposes a deep sentiment semantic recognition model based on Bert-BigrU multi-module integration learning. It can efficiently extract polysemous words and deep sentimental semantics, a remarkable improvement compared with the traditional language model. Meanwhile, to improve the model's generalisation ability, this paper adopts ensemble learning based on the comprehensive analysis of the variance as well as the deviation of the model to observe the performance of different models on different parameters and data sets. Thus, making the base model has the ability of error correction to each other effectively and obtain better ensemble results. Moreover, by designing experiments on two and three classification corpora, it is shown that the BERT-BiGRU model considerably outperforms the state of the art sentiment recognition models in multiple evaluation metrics.

In the future, we would like to consider using richer data to mine sentimental semantics, such as emoticons, pictures, or video semantics to assist in mining deep sentimental semantics.