American Journal of Computer Science and Technology
2023; 6(2): 50-56
http://www.sciencepublishinggroup.com/j/ajcst
doi: 10.11648/j.ajcst.20230602.11
ISSN: 2640-0111 (Print); ISSN: 2640-012X (Online)

Chinese Text Sentiment Analysis Based on BERT-BiGRU Fusion Gated Attention

Huang Shufen, Liu Changhui, Zhang Yinglin

College of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan, China

Email address:
2230102421@qq.com (Huang Shufen), lch198@sohu.com (Liu Changhui)

To cite this article:
Huang Shufen, Liu Changhui, Zhang Yinglin. Chinese Text Sentiment Analysis Based on BERT-BiGRU Fusion Gated Attention. *American Journal of Computer Science and Technology*. Vol. 6, No. 2, 2023, pp. 50-56. doi: 10.11648/j.ajcst.20230602.11

Received: March 22, 2023; **Accepted**: April 18, 2023; **Published**: April 24, 2023

Abstract: To address the problem that Word2vec static encoding cannot give accurate word vectors about contextual semantics and cannot solve the problem of multiple meanings of words, we propose to use the BERT pre-training model as a word embedding layer to obtain word vectors dynamically; we introduce the gating idea to improve on the traditional attention mechanism and propose BERT-BiGRU-GANet model. The model firstly uses the BERT pre-training model as the word vector layer to vectorize the input text by dynamic encoding; secondly, uses the bi-directional gated recursive unit model (BiGRU) to capture the dependencies between long discourse and further analyze the contextual semantics; finally, before output classification, adds the attention mechanism of fusion gating to ignore the features with little relevance and highlight the key features with weight ratio features. We conducted several comparison experiments on the Jingdong public product review dataset, and the model achieved an F1 value of 93.06%, which is 3.41%, 2.55%, and 1.12% more accurate than the BiLSTM, BiLSTM-Att, and BERT-BiGRU models, respectively. It indicates that the use of the BERT-BiGRU-GANet model has some improvement on Chinese text sentiment analysis, which is helpful in the analysis of goods and service reviews, for consumers to select goods, and for merchants to improve their goods or service reviews.

Keywords: Sentiment Analysis, BERT Pre-training Model, BiGRU, Gated Attention

1. Introduction

With the rapid development of mobile Internet and intelligent technologies such as big data, online social networking has gradually penetrated people's daily life. By analyzing the sentiment of these text messages, it is possible to find out the users' sentiment towards a certain event or a certain product. When it comes to online shopping, sentiment analysis of these review data can help merchants not only adjust their products but also help users select the right products for them. Sentiment analysis plays an important role not only in e-commerce but also in analyzing popular trends in events and analyzing customer psychology, among other things. Nowadays, text sentiment analysis is still the most popular area of interest in natural language processing.

The most common way of text sentiment analysis is secondary sentiment classification, which divides the target text into positive categories with positive sentiment and negative categories with negative sentiment. Most of the existing sentiment analysis models use Word2vec static coding, but in the case of polysemous words, this approach cannot give different interpretations of polysemous words according to different contexts. To address this shortcoming, we use BERT pre-training model to dynamically encode the text, which effectively solves the problem of multiple meanings of words; we use a bi-directional gated recurrent network (BiGRU) to capture the semantic association between long sequences, which improves the training speed compared with BiLSTM model; the attention mechanism plays an important role in sentiment analysis work, but the traditional attention mechanism for each word The traditional attention mechanism assigns weights to each word, and the analysis accuracy decreases as the number of useless words increases, so we add gating to the attention mechanism to increase the weights of key features. the advantages of the BERT-BiGRU-GANet model are that the word vectors are dynamically acquired based on the semantics of the context, and the idea of gating is introduced in the attention mechanism to reduce redundant information.

2. Related Research

Currently, there are three main types of methods for text sentiment analysis: sentiment dictionary-based methods [1], traditional machine learning-based methods [2], and deep learning-based methods [3]. The main deep learning-based methods are recurrent neural networks [4] (RNN), convolutional neural networks [5] (CNN), long and short-term memory networks [6] (LSTM), and gated recurrent networks [7] (GRU). Google proposed a BERT pre-training model [8] for sentiment analysis and achieved good results. Hu and Liu [9] proposed a method using Mikolov et al [10] first proposed the word2vec technique. Xu et al [11] added positional encoding as well as an attention mechanism to the CNN model and experimented on the IMDB movie review dataset to get better classification results. Since RNN models experience gradient disappearance leading to training failure when analyzing long sequences [12], Cheng et al [13] constructed a hybrid CNN-BiGRU model to extract local and global features of text and demonstrated that the introduction of the attention mechanism in the hybrid model [14] can effectively improve the classification accuracy. To address the problem that fully connected neural networks cannot establish relevance for multiple related inputs, Vaswani et al [15] proposed a self-attention mechanism. Zhang [16] et al. combined CNN and LSTM methods to achieve better sentiment classification accuracy. The current stage of research mainly focuses on fine-tuned improvements based on pre-trained models, which can achieve better sentiment classification results. Plan et al [17] by adding the BiLSTM model to BERT pre-trained model, and by comparing experiments with the word2vec based BiLSTM model on three datasets, the BERT-BiLSTM model has better accuracy in sentiment analysis is higher. Yang, J. [18] et al. proposed to use the BERT model instead of word2vec to obtain word vectors and found that the text feature representation using the BERT model has a 5.56% improvement in accuracy over the classical word2vec word vectors.

The main work of this paper is as follows: first, we use BERT pre-training model to dynamically encode the text, which effectively solves the problem of multiple meanings of words; then, we use Bi-directional Gated Recurrent Network (BiGRU) to capture the semantic association between long sequences, which improves the training speed compared with BiLSTM model; finally, we add gating to the attention mechanism to increase the weight of key features. The advantages of the BERT-BiGRU-GANet model are the dynamic acquisition of word vectors based on the semantics of the context and the introduction of gating ideas in the attention mechanism to reduce redundant information.

3. Model Structure Design

The overall model framework of this paper is shown in Figure 1. After the text is pre-processed, the returned word vectors are different when input to the BERT pre-training model according to the different contexts of the input text, so that multiple meanings of a word can be distinguished; the dynamic word vectors obtained from the BERT pre-training model are input to the forward GRU and backward GRU to train the text features and further analyze the text semantics; the feature vectors obtained after the BiGRU model are used as input, and the Attention layer performs different weighting operations on these feature vectors, and the gating mechanism is introduced so that the features with very little relevance do not participate in the weighting operations, thus highly focusing the feature vectors with strong emotional tendency; finally, the sentiment classification is realized by Softmax.

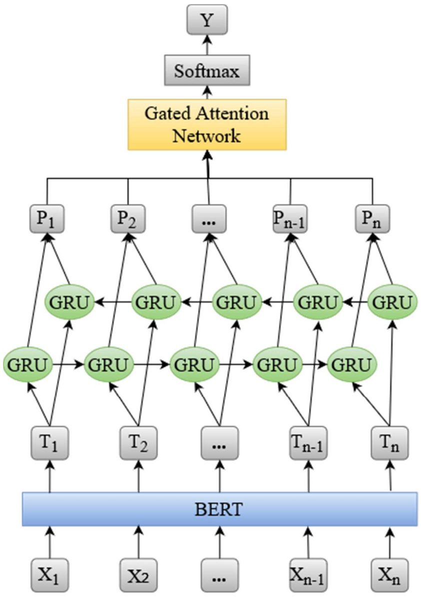

Figure 1. *BERT-BiGRU-GANet model.*

3.1. BERT Pre-Training Model

Word vectors, which represent the meaning of a word by its context, cannot represent multiple words using word2vec static encoding. For example, the word apple has different meanings in "I love apples" and "I like apples". However, if a large corpus defines an apple as a fruit when the model is trained, the apple in the second sentence will also be considered a fruit, which results in comprehension bias. To solve this drawback, BERT, a model that dynamically represents word vectors based on the semantics of text context, emerged.

The model structure of BERT is Seq2Seq, and the core is a Transformer encoder, the structure is shown in Figure 2, it is made by multi-layer superposition, Transformer is the structure of "encoder-decoder", but BERT only uses the encoder part as the model structure, as shown in Figure 3. This is shown in Figure 3 below. The input is transformed into a vector form by an input embedding layer and position encoding is added; after a multi-headed attention mechanism layer, the Add & Norm layer adopts a residual structure and performs a layer normalization operation to solve the phenomenon of gradient disappearance of the neural network; finally, the feedforward network is activated by a two-layer linear mapping and an activation function.

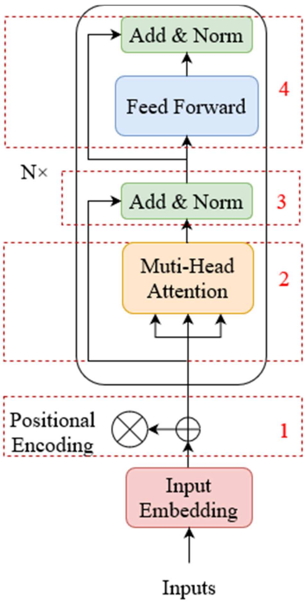

***Figure 2.** Transformer Encoder.*

BERT uses the Masked Language Model (MLM) to generate a deep bi-directional language representation and learn contextual semantic features to predict the masked words. The adjacent sentence prediction (NSP), on the other hand, achieves the determination of the position of two sentence segments by studying the association properties between them. The BERT model is jointly trained by these two tasks to make the output word vectors as comprehensive and accurate as possible and to provide better initial values for the subsequent models. The structure of the BERT model is shown in Figure 3.

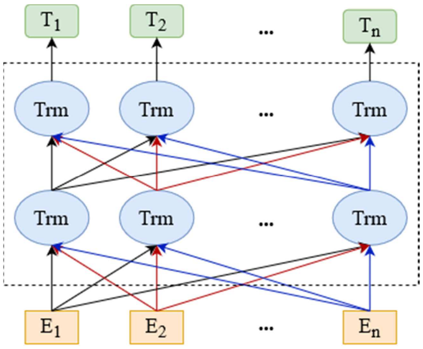

***Figure 3.** BERT Model.*

3.2. BiGRU Model

The GRU model has only two gates, the update gate and the reset gate, which is less computationally expensive and has better final results than LSTM model. The advantage of GRU model is that it can use the same gate for both forgetting and selection memory. Assuming that the input vector at moment t is x_t, the GRU is calculated as follows: z_t represents the update gate, r_t represents the reset gate, $\tilde{h}_t$ represents the hidden layer state, h_{t-1} and h_t represents the state of the hidden layer at moment $t-1$ and moment t, W is the weight, and σ is the sigmoid activation function.

$$z_t = \sigma\left(W_z \cdot \left[h_{t-1}, x_t\right]\right) \tag{1}$$

$$r_t = \sigma\left(W_r \cdot \left[h_{t-1}, x_t\right]\right) \tag{2}$$

$$\tilde{h}_t = \tanh\left(W_h \cdot \left[h_{t-1} * r_t, x_t\right]\right) \tag{3}$$

$$h_t = z_t * \tilde{h}_t + \left(1 - z_t\right) * h_{t-1} \tag{4}$$

However, GRU networks can only process text in one direction and can only rely on the information to the left of the word for prediction, thus causing the processing of text sentiment analysis to fall short of accurate judgments. In contrast, BiGRU consists of two networks, forward and backward, and can obtain information from both the front and back of a word, making the context more closely connected and the prediction structure closer to the actual results. The structure of the BiGRU model is given in Figure 4, which includes an input layer, a forward hidden layer, a backward hidden layer, and an output layer.

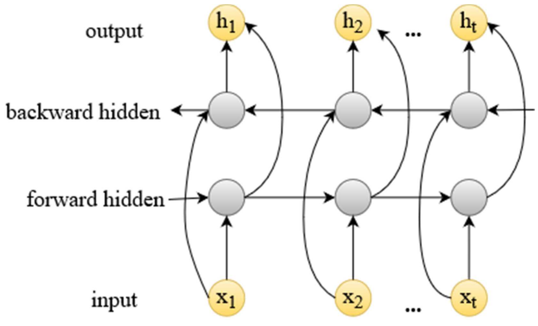

***Figure 4.** BiGRU Model.*

The specific formula of the BiGRU network structure is as follows: where W is the weight matrix of the hidden layer, x_t is the input at moment t, and b_t is the bias vector.

$$\vec{h}_t = GRU\left(x_t, \vec{h}_{t-1}\right) \tag{5}$$

$$\overleftarrow{h}_t = GRU\left(x_t, \overleftarrow{h}_{t-1}\right) \tag{6}$$

$$h_t = f\left(W_{\vec{h}_t}\vec{h}_t + W_{\overleftarrow{h}_t}\overleftarrow{h}_t + b_t\right) \tag{7}$$

3.3. Gated Attention Network

The traditional attention mechanism indicates whether the input information is important or not by calculating the average weighting of the input information, and generally focuses on the state sequence of the whole text sentence, and assigns smaller weights to even irrelevant word vectors, but in most cases, it is not necessary to focus on all words. Therefore, this paper improves in the attention layer by adding a gated attention network-based layer that can increase the weight proportion of key features. The GA-Net contains an auxiliary network and a backbone network, as shown in Figure 5, with the auxiliary network on the left and the attention network on the right.

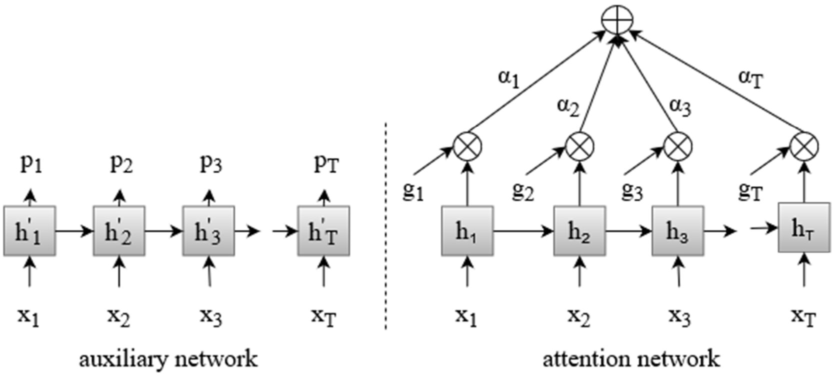

Figure 5. *GA-Net network.*

The GA-Net attention network differs from the traditional attention mechanism in that the GA-Net network comes with a gating mechanism. Gating g_1, g_2... g_t takes 1 to determine the inflow of information about the current state while taking 0 does not; the purpose of selectively activating part of the network is achieved, so that the model does not pay attention to the irrelevant information. And the auxiliary network part serves to generate binary gates for each position of the text to determine whether the position needs to be attended to. The output probability p_t of the auxiliary network is calculated as follows.

$$h'_t = BiGRU\left(h'_{t-1}, x_t\right) \tag{8}$$

$$p_t = sigmoid\left(Wh'_t\right) \tag{9}$$

The probability pt is used to determine the probability of a door opening using the Bernoulli distribution to generate binary gates in the auxiliary network. The gradient calculation for backpropagation cannot be used on the non-differentiable layers, making it difficult to train random networks with discrete variables. And the binary gate gt is gated by the Bernoulli function after gating is discrete values 0 and 1, which cannot be back-propagated by gradient descent to propagate the error. Therefore, this paper uses the Gumbel-softmax distribution instead of the Bernoulli distribution to solve this problem. It is a continuous distribution on the simplex (simplex) that can approximate the category samples, and its parameter gradient can be easily computed by the reparameterization trick. Experiments have shown that Gumbel-Softmax outperforms all one-sample gradient estimates on both Bernoulli and categorical variables [19]. The approximation of softmax for the unique heat vector is calculated as follows.

$$\hat{g}_t = \left[\hat{p}_t, i\right]_i = 0,1 \tag{10}$$

$$\hat{p}_t, i = \frac{\exp\left(\left(\log\left(p_t, i\right) + \varepsilon_i\right) / \tau\right)}{\sum_{j=0}^{1} \exp\left(\left(\log\left(p_t, j\right) + \varepsilon_j\right) / \tau\right)} \tag{11}$$

where ε_i is a random sample from Gumbel (0,1). The Gumbel-Softmax distribution tends to be uniquely hot when the temperature τ approaches 0. The attention mechanism weights with gating can be calculated by equation (12).

$$\alpha_t = \frac{g_t \cdot \exp e_t}{\sum_{t'=1}^{T} g_{t'} \cdot \exp e_{t'}}, \sum_{t=1}^{T} \alpha_t = 1 \tag{12}$$

4. Experimental Results and Analysis

4.1. Experimental Environment

This paper uses python 3.6 language for programming and pytorch 1.10 deep learning framework. A publicly available Jingdong shopping review dataset is used to validate the model architecture proposed in this paper for analyzing and verifying Chinese text sentiment analysis. After conducting several experiments on the BERT-BiGRU-GANet model proposed in this paper, some of the optimal parameter settings for the experiments were determined as shown in Table 1 below.

Table 1. *Model parameter setting.*

Parameters	Meaning	Value
gru_hidden	Number of GRU hidden layers	128
gru_hidden2	Number of hidden layer cells	64
learning_rate	Learning Rate Size	0.0005
batch_size	Number of batch training samples	128
dropout	Dropout parameters	0.5
hidden_size	Number of BERT hidden layers	768

4.2. Experimental Dataset

The dataset used in this paper is the publicly available Jingdong product reviews dataset on GitHub. The dataset contains 10 categories of books, tablets, cell phones, fruits, etc., with a total of 62,774 data, of which positive and negative reviews account for about half each, and the sentiment labels are divided into two categories [0,1], with negative sentiment labeled as 0 and positive sentiment labeled as 1. The positive and negative reviews of each category are randomly disordered and divided into the training set, test set, and validation set according to the ratio of 8:1:1, respectively. For the dichotomous classification problem, the experimental results were evaluated using Accuracy, Precision, Recall, and F1 values as the evaluation metrics.

4.3. Comparative Experimental Setup

The model proposed in this paper is compared with the following six models in the following comparative experimental setup.

(1) BiLSTM: BiLSTM network for obtaining word vectors using word2vec;
(2) BiGRU: BiGRU network for obtaining word vectors using word2vec;
(3) BiGRU-Att: Using word2vec to obtain word vectors and adding an attention mechanism to the BiGRU network;
(4) BiLSTM-Att [20]: Using word2vec to obtain word vectors and adding an attention mechanism to the BiLSTM network;
(5) BERT-BiGRU [21]: BERT model as word vector layer and BiGRU model as a hidden layer, using softmax for classification;
(6) BERT-BiGRU-Att: BERT model as the word vector layer, BiGRU model as the hidden layer, and then add the attention mechanism;
(7) BERT-BiGRU-GANet: The model proposed in this paper uses the BERT model as the word vector layer, the BiGRU model as the hidden layer, and adds a gating mechanism on the attention layer.

4.4. Experimental Results

In this paper, we compare in detail the models of each group and the accurate effect of this paper's model on Chinese text sentiment analysis under the same dataset. The experimental comparison results are shown in Table 2 below.

Table 2. *Experimental comparison results.*

Group	Model	Acc	P	R	F1
1	BiLSTM	89.66	89.65	89.67	89.66
2	BiGRU	90.01	90.01	89.99	90.03
3	BiLSTM-Att	90.52	90.60	90.46	90.50
4	BiGRU-Att	91.12	91.14	91.08	91.10
5	BERT-BiGRU	91.13	91.21	91.07	91.11
6	BERT-BiGRU-Att	91.95	91.98	91.92	91.94
7	BERT-BiGRU-GANet	93.07	93.06	93.07	93.06

From the results of the seven groups of experimental data shown in Table 2, it can be seen that: on the Jingdong product review dataset, the BERT-BiGRU-GANet model proposed in this paper performs the best among all the comparison experiments, achieving 93.07% and 93.06% results in terms of accuracy Acc and comprehensive index F1, which are both better than other comparison model groups. The experimental results of groups 1 and 2 show that BiGRU is more accurate than BiLSTM in practice, and the results of groups 2 and 3 reflect that adding an attention mechanism to the model can improve the classification accuracy of the model.

The comparison experiments of group 5 and group 2 show that the word vector obtained by using the BERT pre-training model dynamically encoded as input is better than that obtained by using word2vec statically encoded-word The final experimental results of the model using the word vectors obtained from the BERT pre-training model as input is better than those using the word2vec static encoding, and the accuracy is 1.46% and 1.12% higher respectively; the model in group 6 introduces the attention mechanism based on group 5, which improves the classification ability of the model again; the model in group 7 is compared with the model in group 6, and the attention mechanism is improved to a gating-based attention mechanism, and the accuracy is improved by 1.12% as can be seen from the experimental comparison results.

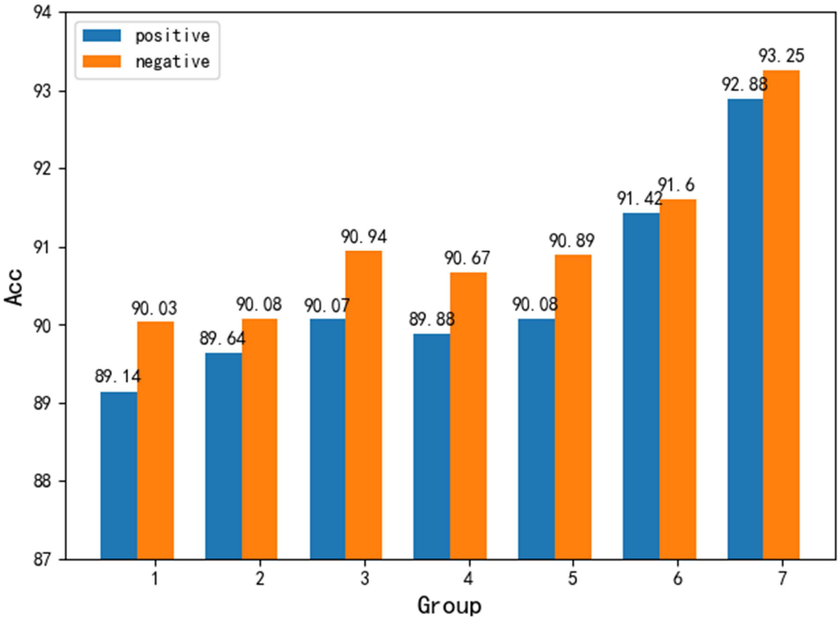

Figure 6. *The effect of each model on the analysis of different emotions.*

We also compare the effect of each of the above models on the sentiment analysis of different emotions, with F1 as the final evaluation index, and the experimental results are shown in Figure 6.

The comparison graph of the experimental results shows that the effect of the model on sentiment analysis is improving in all model groups, which is due to the continuous improvement of the model; the BERT-BiGRU-GANet model in group 7 proposed in this paper has the highest analysis effect on both types of sentiment, which confirms that this model has some improvement and enhancement effect on the previous model, and the effect of positive sentiment analysis reaches 92.88% and negative sentiment analysis reached 93.25%.

Since the dataset used in this paper has 10 categories, each category has more or less data volume, some categories have as much as 10,000 data volume and some have as little as 575. Therefore, sentiment analysis is performed for each category separately to compare the effect of the model on sentiment analysis for datasets with different data volumes. Because there are too many categories, this paper only compares group 7 BERT-BiGRU-Attention i.e., the BERT-BiGRU model based on the introduction of a general attention mechanism, and group 8 BERT-BiGRU-GANet i.e., the BERT-BiGRU model based on the addition of a fused gated attention mechanism, and conducts these two groups of models on each of the 10 categories Emotion analysis experiments were conducted to verify whether the attention mechanism of fusion gating proposed in this paper on the improvement of attention mechanism has some improvement effect.

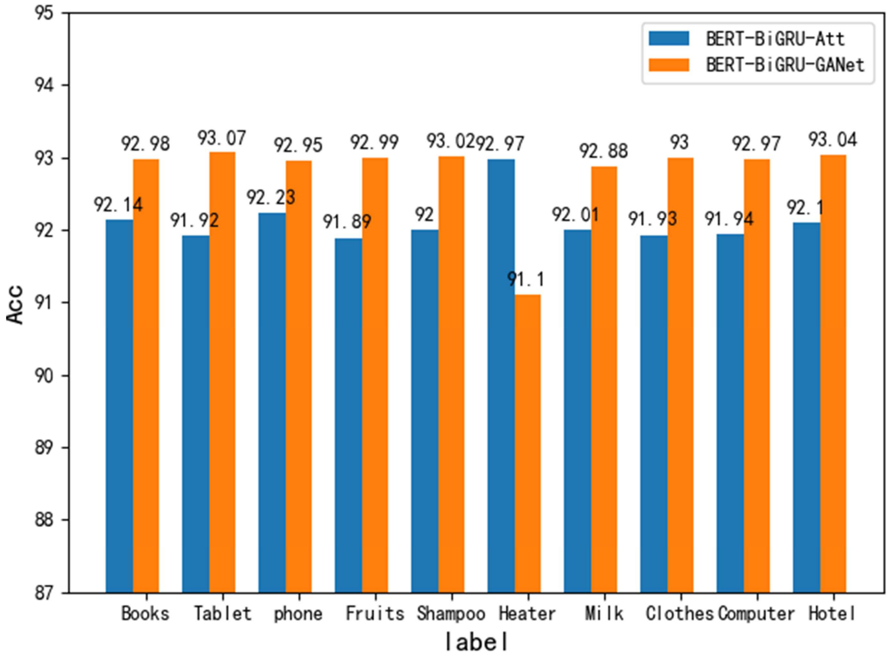

Figure 7. *Comparison of different categories of sentiment analysis.*

As can be seen from the above graph for the effect of sentiment analysis of categories with different amounts of data, the attention mechanism based on gating proposed in this paper outperforms the BERT-BiGRU model with an ordinary attention mechanism under all categories. It is verified that the introduction of gating in the attention mechanism can improve the sentiment analysis ability of the model, because the presence of gating can filter the unimportant information and pay more attention to the key information, and the final sentiment analysis results are more accurate. However, the sentiment analysis ability of the BERT-BiGRU-GANet model is weak under the category of Heaters (This category has a total of 575 data) with fewer data, which is the shortcoming of the model in this paper; how to optimize again to improve the sentiment analysis ability of this model in the dataset with very little data needs to be improved by subsequent research.

5. Conclusion

Text sentiment analysis is playing an increasingly important role in today's fast-paced business environment. It can be used not only to improve user experience, but also to help companies make marketing decisions and improve service quality. It is due to its unique advantages that text sentiment analysis is becoming more and more widely used in modern business environments.

In this paper, we propose to use the word vector obtained by dynamic encoding in the form of BERT pre-training model mask, which is closer to the contextual semantics to a certain extent; then combine with the BiGRU network structure to extract features; differ from the traditional attention mechanism of assigning weights to each word, we introduce the idea of gating in the attention mechanism to reduce the interference of redundant information, and designs the

BERT-BiGRU-GANet model. This model has a high accuracy rate compared with other models, and is useful in analyzing goods and service reviews, extracting evaluation objects and evaluation expressions, and identifying sentiment tendencies in reviews, which is beneficial to consumers in selecting goods and companies in improving goods or services. However, since we only use BERT as the word vector layer and the main network layer is BiGRU, how to perform hybrid model embedding on this basis is the research direction afterwards, and the comparative study with other optimized pre-trained model models is also the focus of the next work.

References

[1] Zhong Jiawa, Liu Wei, Wang Sili, Yang Heng. A review of text sentiment analysis methods and applications [J]. Data Analysis and Knowledge Discovery, 2021, 5 (06): 1-13.

[2] Hong Wei, Li Min. A review of research on text sentiment analysis methods [J]. Computer Engineering and Science, 2019, 41 (04): 750-757.

[3] Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules [J]. Advances in neural information processing systems, 2017, 30.

[4] Liu, L. W., Yu, S.. Recurrent neural network (RNN) and application research [J]. Science and Technology Perspectives, 2019 (32): 54-55. DOI: 10.19694/j.cnki.issn2095-2457.2019.32.022.2.

[5] S. Xiong, H. Lv, W. Zhao, and D. Ji, "Towards Twitter sentiment classification by multi-level sentiment-enriched word embeddings," Neurocomputing, vol. 275, pp. 2459-2466, 2018, doi: 10.1016/j.neucom.2017.11.023.

[6] Liu H L, He Y F. Seasonal attention in LSTM and its application to text sentiment classification [J/OL]. Systems Science and Mathematics: 1-19 [2023-03-14].

[7] Wang W, Sun YX, Qi QJ, Meng XF. A text sentiment classification model based on BiGRU-attention neural network [J]. Computer Application Research, 2019, 36 (12): 3558-3564. DOI: 10.19734/j.issn.1001-3695.2018.07.0413.

[8] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding [J]. arXiv preprint arXiv: 1810.04805, 2018.

[9] Minqing Hu, Bing Liu. Mining and summarizing customer reviews [P]. Knowledge discovery and data mining, 2004.

[10] Tomas Mikolov, Ilya Sutskever, Kai Chen 0010, Greg Corrado, Jeffrey Dean. Distributed Representations of Words and Phrases and their Compositionality. [J]. CoRR, 2013, abs/1310.4546.

[11] Xu Yizhou, Lin Xiao, Lu Li-Ming. A long text sentiment classification model based on hierarchical CNN [J]. Computer Engineering and Design, 2022, 43 (04): 1121-1126. DOI: 10.16208/j.issn1000-7024.2022.04.030.

[12] Lifu Wang, Bo Shen, Bo Hu, Xing Cao. Can Gradient Descent Provably Learn Linear Dynamic Systems? [J]. arXiv: 2211. 10582, 2022.

[13] Cheng Y, Sun H, Chen H, et al. Sentiment analysis using multi-head attention capsules with multi-channel CNN and bidirectional GRU [J]. IEEE Access, 2021, 9: 60383-60395.

[14] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron C. Courville, Ruslan Salakhutdinov, Richard S. Zemel, Yoshua Bengio. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. [J]. CoRR, 2015, abs/1502. 03044.

[15] Lima Luiz Renato, Godeiro Lucas Lúcio. Equity-premium prediction: Attention is all you need [J]. Journal of Applied Econometrics, 2022, 38 (1).

[16] X. Sun and C. Zhang, "Detecting anomalous emotion through big data from social networks based on a deep learning method," Multimedia Tools and Applications, vol. 79, no. 13-14, pp. 9687-9687, 2020, doi: 10.1007/s11042-018-5665-6.

[17] Peters M, Neumann M, Iyyer M, et al. Deep Contextualized Word Representations [J]. 2018.

[18] Phan, Huyen Trang, Ngoc Thanh Nguyen, et al. AspectLevel Sentiment Analysis Using CNN Over BERTGCN [J]. IEEE Access, 2022, 10: 110402-110409.

[19] Shao Nan. Research on discrete recommendation algorithm based on Gumbel-Softmax distribution [D]. University of Electronic Science and Technology, 2020. DOI: 10.27005/d.cnki.gdzku.2020.004485.

[20] Yang Xiuzhang, Wu Shuai, Ren Tianshu, Liu Jianyi, Song Jiwen, Liao Wenjing. Research on sentiment analysis of e-commerce reviews by integrating multi-headed attention mechanism and BiLSTM [J]. Information Technology and Informatization, 2022 (10): 5-9.

[21] Cui Jia-Bin. Research on text sentiment analysis based on BERT-BiGRU model [D]. Shanxi University, 2021. DOI: 10.27284/d.cnki.gsxiu.2021.000092.