

Article

Not peer-reviewed version

Research on Online Review Information Classification Based on Multimodal Deep Learning

[Jingnan Liu](#)*, [Yefang Sun](#), [Yueyi Zhang](#)*, Chenyuan Lu

Posted Date: 4 April 2024

doi: 10.20944/preprints202404.0346.v1

Keywords: Deep learning; Multimodality; Online review information; Sentiment analysis



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Research on Online Review Information Classification Based on Multimodal Deep Learning

Liu Jingnan ^{1,2,*}, Sun Yefang ^{2,3}, Zhang Yueyi ^{2,4,*} and Lu Chenyuan ^{2,5}

¹ Graduate Student at the College of Economics and Management, China Jiliang University, Research Area: Deep learning, Sentiment Analysis, Faculty of Economics and Management, China Jiliang University, Xiasha Higher Education Park, Hangzhou, Zhejiang Province, 310018

² Faculty of Economics and Management, China Jiliang University, Hangzhou, Zhejiang Province

³ Teaching Assistant at China Jiliang University College of Modern Science and Technology, Research Area: Quality Management; sunyefang21@163.com

⁴ Professor at the College of Economics and Management, China Jiliang University, Research Area: Quality Management, Faculty of Economics and Management, China Jiliang University, Xiasha Higher Education Park, Hangzhou, Zhejiang Province, 310018

⁵ Graduate student at the College of Economics and Management, China Jiliang University, Research Area: Quality Management; chenchenlulu06@126.com

* Correspondence: 13777574168@163.com (L.J.); 02a0702026@cjl.u.edu.cn (Z.Y.); Tel.: 13777574168 (L.J.); 13588117516 (Z.Y.)

Abstract: The incessant evolution of online platforms has ushered in a multitude of shopping modalities. Within the food industry, however, assessing the delectability of meals can only be tentatively determined based on consumer feedback encompassing aspects such as taste, pricing, packaging, service quality, delivery timeliness, hygiene standards, and environmental considerations. Traditional text data mining techniques primarily focus on consumers' emotional traits, disregarding pertinent information pertaining to the online products themselves. In light of these aforementioned issues in current research methodologies, this paper introduces the Bert BiGRU Softmax model combined with multimodal features to enhance the efficacy of sentiment classification in data analysis. Comparative experiments conducted using existing data demonstrate that the accuracy rate of the model employed in this study reaches 90.9%. In comparison to single models or combinations of three models with the highest accuracy rate of 7.7%, the proposed model exhibits superior accuracy and proves highly applicable to online reviews.

Keywords: deep learning; multimodality; online review information; sentiment analysis

1. Introduction

As e-commerce platforms are accompanied by features such as openness and transparency, fierce competition in the same category, product diversification, and review functions, consumers' perception and evaluation of goods are transferred from the staff in the store to the public view online, and users can share their views on goods and merchants at any time and any place, and express their own feelings. Users now have the ability to share their opinions on products and merchants, expressing their emotions at any time and from anywhere. Once a transaction is concluded, users' subjective product reviews convey a certain inclination of sentiment. These reviews serve a dual purpose as both "information providers" and "emotional influencers," significantly impacting merchants. Not only do they influence store reputation, but they also affect the long-term sales of their products. To enhance merchants' comprehension of sentiment tendencies within reviews, it becomes imperative to categorize and analyze product review information, thereby discerning whether the feedback leans towards positivity or negativity. This enables merchants to refine their stores and products based on the emotional orientation of the review data, ultimately augmenting user satisfaction, product sales, and store ratings.

The evaluation of review information is not solely influenced by subjective factors such as consumers' personal preferences, emotions, and personalities. It also exhibits a strong correlation with the objective factor of product quality itself. In order to conduct sentiment analysis on review data, the initial step involves extracting text features. Feature selection aims to extract pertinent information from raw attribute sets while reducing data dimensionality. These attribute sets can be derived based on dictionaries or commonly employed statistical methods [22,24].

Following feature extraction, the subsequent step in sentiment analysis entails sentiment categorization. This can be accomplished through various approaches, including machine learning, vocabulary-based methods, and deep learning techniques [26]. Sentiment analysis models employing polar vocabularies are frequently utilized for sentiment classification. However, the availability of sentiment vocabularies remains limited. Machine learning methods possess distinct advantages over sentiment dictionaries when it comes to nonlinear, high-dimensional pattern recognition problems.

Given that comments encompass a wealth of content, traditional neural network models may struggle to fully capture the complete context of a sentence or comment. Therefore, novel neural networks must employ different types of word embeddings. For instance, utilizing RNN variants like Bidirectional LSTM (BiLSTM) and Bidirectional GRU (BiGRU) proves beneficial for sentiment analysis tasks.

Traditional text data mining methodologies tend to overlook the relevant information pertaining to online products themselves, particularly the descriptive textual details regarding product quality and the potential disparities between depicted images and the actual unadorned goods. In our contemporary reality, individuals exist within a multimodal and interconnected environment, where various forms of information modalities, including text, speech, images, and videos, converge. Enriching linguistic expression methods through the utilization of multiple information modalities allows computers to better comprehend and interpret input data, leading to more precise and comprehensive output results [27].

Hence, when analyzing product reviews, it becomes essential to consider multimodal information in order to fully grasp consumers' emotions. This approach not only enhances users' genuine experiences and satisfaction but also improves merchants' sales performance. By delving into review information, merchants can tailor their products to align with consumers' current needs, thereby increasing relevance and appeal.

To enable AI to attain a deeper comprehension of the world, it is imperative to endow it with the capacity to learn, comprehend, and reason about multimodal information. Multimodal learning entails constructing models that enable machines to acquire knowledge from multiple modalities, allowing for effective communication and information transformation across each modality. Throughout the course of this research, feature representations were created using multimodal deep learning techniques, encompassing both image and text data.

In comparison to purely statistical models, machine learning models offer greater diversity and exhibit enhanced capability to capture the diverse features present in multimodal data. These models excel at approximating various nonlinear relationships, while showcasing superior adaptability.

2. Relevant Research

The research direction of the content of online review information is roughly divided into two categories: one is for the text mining of online review information; the other is from the multi-level of online review information, and the picture information in online review information has attracted a lot of attention in the academic community [16,17]. Traditional multi-scalar sentiment categorization methods mainly include two routes: sentiment dictionary-based and machine learning-based. Sentiment lexicon-based methods usually calculate sentiment polarity on a whole-sentence or sub-sentence basis and apply it to all facets involved in a sentence, but they cannot handle the complex mapping relationship between sentences and facets well [40-41]. The core of this approach is the application of affective lexicons, however, there are differences in the affective tendencies of affective words in different domains, making it difficult to generalize domain-specific affective lexicons to other domains [42-43].

D. Lahat et al. define multimodality as a qualification that can be used intensively compared to unimodality [34]. Zhao Liang considers multimodal data to be data obtained through different domains or perspectives for the same descriptive object, and calls each domain or perspective describing these data a modality [35]. The main objective of multimodal learning is to correlate and process multimodal data by building models.

According to the research summary of related scholars, the main content of current multimodal research can be summarized as five levels of multimodal data representation, data mapping, data alignment, data fusion and co-learning [36]. Most of it revolves around cross-modal mapping between data. As the fields of computer vision and natural language processing continue to evolve, and as more large-scale datasets become available for research, multimodal data mapping methods continue to mature [37]. The mainstream multimodal data mapping method is: based on the existing mapping relationship, the existing multimodal data is first symbolized or vectorized, which is used as the input to the neural network, and then combined with the existing correspondences, mapped to another modal. After continuous training based on massive datasets, a cross-modal data mapping model with universal applicability is obtained. The framework of this multimodal data mapping approach is shown in Figure 1. One of the most common and widely used scenarios is image semantic recognition [38,39], which maps image modal data to textual modal data.

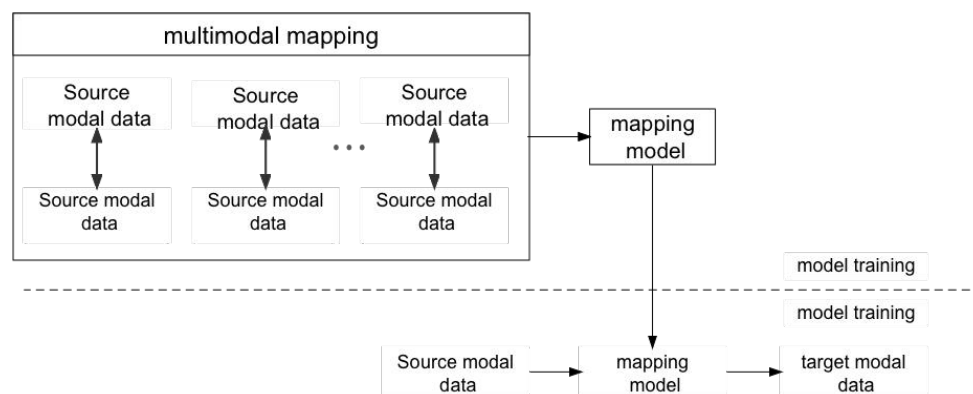


Figure 1. Multimodal data mapping approach.

Sentiment analysis of online product review information faces two major challenges: dimension mapping and sentiment word disambiguation. While the dimension mapping problem lies in the correct use of dimensions to map online review texts, sentiment word disambiguation refers to the situation where two or more dimensions exist for a sentiment word. Therefore, sentiment analysis of online reviews is considered as a multidimensional classification process [18-20].

For the construction of sentiment lexicon requires a lot of human intervention, and its completeness and accuracy have an important impact on the sentiment classification results. On the other hand, machine learning-based methods usually regard multi-scalar sentiment polarity recognition as a sequence labeling problem, which requires manually designing features and labeling them, and then using classifiers for training and learning. Common sequence labeling methods include conditional random fields, maximum entropy, plain Bayes, and support vector machines [44]. Despite the achievements of machine learning methods in multi-scalar sentiment classification, feature engineering is time-consuming and labor-intensive, and the classification results are highly dependent on feature quality.

Pang (2002) first applied the machine learning method of N-Gram to the field of sentiment analysis, and the experimental results showed that N-Gram achieved the highest classification accuracy of 81.9% [28]. Since feature selection affects the performance of machine learning methods, Abinash Tripathy (2016) analyzed online review comments through N-gram model combined with machine learning methods and the experimental results showed that SVM combined with unigram, bigram and trigram features obtained the best classification results [23].

Zheng Fei et al showed that a combination of LDA model and Word2Vec word vectors can be used to complete the modeling of comment text word vectors for sentiment classification in problems such as varying lengths of comment texts and non-uniformity of unit schemas [1]. Kim and Hovy (2004) applied synonyms and antonyms of WordNet dictionaries and hierarchical structures to analyze word vectors' sentiment tendencies [29]. Zhu Xiaoliang and other researchers solved the composition classification problem by using TextRank model to filter key sentence words and combining it with word embedding model to model documents. Aiming at the problem that Word2Vec model cannot identify the importance of special words and scene words in Chinese in the text [2]. Zhang Qian et al proposed to introduce the TFIDF model to weight the output word vector matrix, to get the weighted text vectorization model and classify it [3]. Yuting Yang et al researchers encoded text into high dimensional vectors with contextual semantic, sequential and sentiment information through Doc2Vec model and verified the effectiveness of document distributed representation approach [4].

Marco Guerini (2013) used Simple Bayes, K-Nearest Neighbors, Maximum Entropy, and Support Vector Machines to analyze reviews for sentiment propensity, with Support Vector Machines significantly outperforming the other methods with up to 83% accuracy on larger training sets [30]. Yunfei Shao et al researchers used LDA and TF-IDF to expand the input text features, and then further aggregated the features to form classification basis vectors by CNN, which improved the classification effect on news headline data [5,15]. Qu and Wang (2018) proposed a sentiment analysis model based on hierarchical attention networks with a 5% improvement in accuracy over recurrent neural networks [31]. Tao, Zhiyong et al researchers fused the bidirectional features of the BiLSTM model and used them for attention weight computation, achieving improved classification results on various benchmark datasets [6].

In particular, Duan Dandan et al human used a short text classification model based on BERT, which utilizes BERT's own sentence vector training to achieve automatic text classification [7,21]. Du Lin et al who obtained word vectors by inputting the text into the BERT model, and then input it into the BiLSTM containing self-attention in chronological order to realize the extraction and automatic classification of the text of Chinese medical records [8].

Trofimovich (2016) used LSTM (Long Short-Term Memory) to solve the problem of sentiment analysis in order to classify sentiment at the phrase level including linguistic rules such as negativity, intensity, and polarity, and trained on labeled text with BiLSTM for syntactic and semantic processing [32]. In the field of traditional machine learning classifiers, some of the literature uses Doc2Vec model and LDA model to obtain multi-channel text feature matrices and input the modeled text into SVM and LR for classification. By obtaining the final classification results through the voting mechanism among multiple classifiers, the model achieves excellent results on the short text classification problem [33]. Ge et al researchers used CNN network to extract features after representing the text with bag-of-words model and used SVM classifier to classify the adverse nursing events [9].

In the field of deep learning, H. Wang et al researchers obtained word embedding representations of documents and fed them into a two-channel classification model. The first channel is a three-layer CNN to extract local features. At the same time, the model fuses the input vectors obtained from the word embedding model with the output vectors of each layer of CNN to realize the reuse of the original features; the second channel is LSTM to obtain the context-associated semantics of the text. Finally, the vectors of the two channels are fused through a fully connected network to realize feature fusion, and the model achieves better results than the previous traditional model in the Sina news classification problem [10-12].

It is worth noting that the model uses a unidirectional LSTM, while Bi-LSTM is considered to be a better existence than unidirectional LSTM. LSTM models are used for phrase-level sentiment classification centered on regularization, which contains linguistics such as negativity, intensity, and polarity [13], so BiLSTM will be better for sentiment classification. Multi-channel Text CNN is able to obtain more adequate keyword aggregation than multi-layer Text CNN. Cho (2014) proposed Gated Recursive Units (GRU) to analyze dependency contexts, which showed significant improvements in

various tasks. Considering the multimodal, sparsely informative, highly unstructured, and word polysemous nature of review text data [14,25].

In this paper, we propose the Bert BiGRU Softmax deep learning model with hybrid masking, comment extraction and attention mechanisms combined with multimodality. In order to improve the correctness of the results, this paper uses several models for image content recognition respectively, and finds that Squeeze Net performs better in terms of execution efficiency and accuracy, so Squeeze Net is finally adopted for image content recognition. The Bert BiGRU Softmax model extracts multidimensional product features from online reviews using the Bert model as an input layer; The bidirectional GRU model is used as a hidden layer to obtain the semantic code and compute the sentiment weights of the comments; finally, Softmax and the attention mechanism are utilized as an output layer to classify the positive or negative nuances.

3. Collection and Analysis of Data Sets

In this paper, the food category review information of the online platform was selected as the data source. With the sky-rocketing changes in the food business model, people's consumption habits are also quietly shifting. By clicking on your favorite food items on the mobile app, these food items are delivered on time and accurately to the designated area. It also includes the fact that people habitually look at the information about the online reviews of the restaurant before they go to spend their money in the restaurant before judging which store to go to.

However, with the rapid development of various platforms, the safety hazards of certain food products as reflected in the review information cannot be ignored. The occurrence of food safety incidents has serious harmful effects on consumers, takeaway platforms, food merchants and society beyond our imagination. Therefore, this paper aims to strengthen the food safety regulation of stores by the relevant authorities through the analysis of store reviews.

In reviews, users explicitly or implicitly rate the attributes of multiple items, including environment, price, food, and service. In this paper, four preprocessing steps were performed on the collected data to ensure the ethics, quality, and reliability of the reviews. Specifically, they include (1) user information (e.g., user ID, user name, avatar, and posting time) is deleted due to privacy considerations; (2) short comments with fewer than 50 Chinese characters and long comments with more than 1,000 Chinese characters are filtered out; (3) if the percentage of non-Chinese characters in a comment is more than 70%, the comment will be discarded; and (4) preprocessing of the data includes data cleansing, Chinese word splitting, de-duplication of words, etc.

In this paper, we use the Bert BiGRU Softmax deep learning model to perform sentiment analysis of online product quality reviews in terms of multiple dimensions such as service, taste, price, hygiene, packaging, and delivery time. The polarity is categorized into positive, neutral and negative dimensions.

4. Introduction to Multimodal Learning and Modeling

Multimodality refers to the different forms in which things are presented or experienced. Multimodality can be based on the human senses, including visual, auditory, and tactile modalities, each of which can represent human perception. The combination of multiple modal perceptions gives the complete human modal perception. At the same time multimodal can be used to represent different forms of data forms, can also be the same form of different formats, generally expressed as text, images, audio, video, mixed data.

Image content recognition has been an important research problem. With the continuous development of learning methods, the accuracy of image recognition is constantly improving, the original image consists of pixel matrix, the traditional edge recognition and other image recognition methods can only be divided by pixel blocks each time, the recognition effect is poor; and due to the existence of convolutional layer, the pixel matrix of the image after convolutional processing, it turns into a high-dimensional matrix of features, and transforms the simple pixel information into composite feature information. Through such a convolution operation, the computer is not only able

to recognize the basic edge detection, but also able to recognize shapes, such as circles, rectangles, etc., and then carry out continuous convolution, and ultimately realize the recognition of the object.

Light weighted based on Inception benefits. In this paper, we use the open-source tool Image AI based on Python language, which is based on ImageNet dataset for model training, integrating the mainstream ResNet50, DenseNet121, InceptionV3 and

Squeeze Net, which are four kinds of convolutional neural network-based deep learning models for image recognition, and also support the Customized model training is also supported. In order to improve the correctness of the results, this paper uses the above four models for image content recognition, and finds that Squeeze Net performs better in terms of efficiency and accuracy, so Squeeze Net is finally adopted for recognition.

Mathematical modeling for text is still an essential aspect in the field of natural language processing. There is a significant impact of how text is modeled on the effectiveness of downstream feature extraction models and classification models. Most of the common text modeling models are designed for English corpus, either question and answer corpus or comment corpus. However, in addition to the sparse information and highly unstructured characteristics of English commentary texts, Chinese commentary texts also have the problems of multiple meanings of words and non-uniformity of the smallest unit of expression. These problems usually result in limiting the classification effectiveness of traditional classification models on Chinese text.

At the same time, the information contained in the comment text includes not only textual information, but also image information. To textually model multimodal data expressed by these two kinds of information, there are not only the problems of word polysemy and textual representation granularity, but also the difficulties of acquiring picture information and performing sentiment tendency analysis.

The BERT model was proposed in 2018 (Pre-training of Deep Bidirectional Transformers for Language Understanding) as a milestone work in the field of pre-training, achieving the best current results on several NLP tasks and opening a new chapter. BERT is a pre-trained model on deep bi-directional transformers for language understanding, where transformer refers to a network structure for processing sequential data. BERT learns the semantic information of the text and applies it to tasks such as categorization, semantic similarity, etc. through outputs in the form of vectors. It is a pre-trained language model, i.e., it has been trained unsupervised on a large-scale corpus, and in using it we only need to train and update its parameters on this basis.

Unlike other language models, BERT is trained on unsupervised precisions, where information related to the left and right of the text is considered in each layer. Xia et al. argue that the supervised deep learning methodology approach relies on a large number of clean seismic records without ground-rolling noise as a reference label.The unsupervised learning method approach considers different temporal, lateral, and frequency features that distinguish the ground-roll noise from the real reflected waves in the seismic records before deep stacking. By designing the ground-roll suppression loss function, the deep learning network can learn the specific distribution characteristics of the real reflected waves in seismic records containing ground-roll noise.Bert's linguistic input representation consists of three components: word embeddings, segmentation embeddings, and position embeddings. The final embedding vector is a direct sum of the above three vectors.

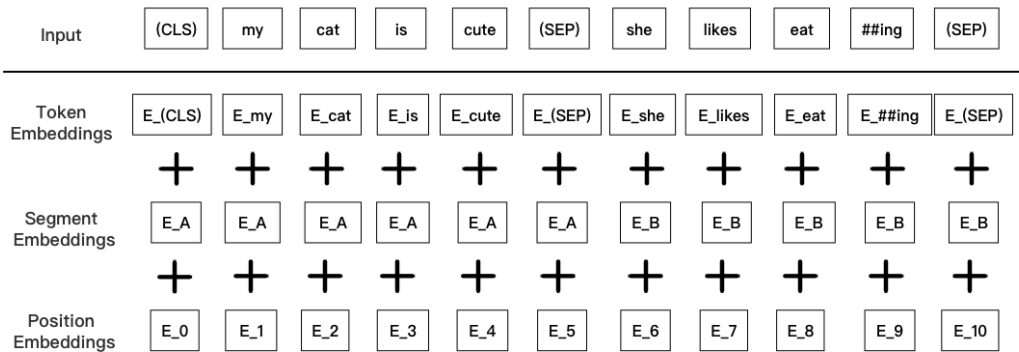
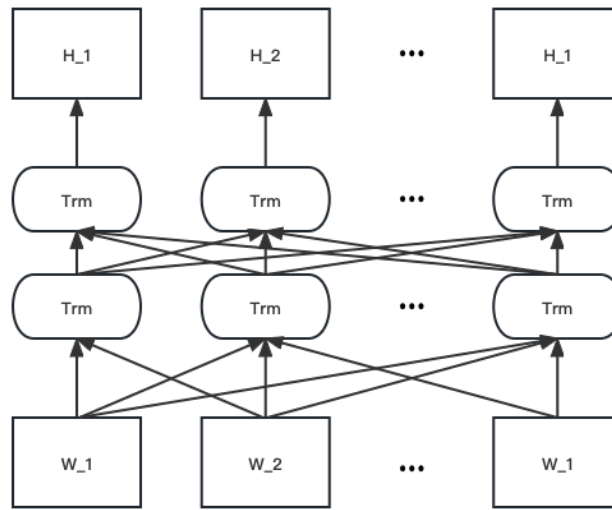


Figure 2. Structure of Bert input layer.**Figure 3.** BERT model diagram.

In summary, for text classification tasks, the BERT model inserts a [CLS] symbol in front of the text and uses the output vector corresponding to this symbol as a semantic representation of the whole text for text classification. This notation has no obvious semantic information, so it can more "fairly" incorporate the semantic information of individual words or phrases in the text. In addition, we can add additional structures such as fully connected layers after the BERT model to perform fine-tune operations for specific tasks, such as linguistic reasoning tasks like Q&A.

By learning the distribution over the input text vectors, the Emotion Bert model can be efficiently used to learn feature extraction over variable-length sequences S . Given a review sentence S , we can directly obtain its category and the set of dimensions of the category D_c . For each word w_1 (w_1, w_2, \dots, w_m) and dimension d_j in a review sentence, we assign a probability score that describes the probability $P(s)$ that word w_i belongs to class d_j in an online product quality review (as in Equation (1))

$$P(s) = P(w_1, w_2, \dots, w_m) = \prod_{i=1}^m p(w_i | w_1, w_2, \dots, w_{i-1}) \quad (1)$$

Transformer adds sequence information to the sequence via word position embedding (PE) Formulas (2) and (3).

$$PE_{(pos, 2i)} = \sin(pos/10000^{2i/d_{model}}) \quad (2)$$

$$PE_{(pos, 2i+1)} = \cos(pos/10000^{2i/d_{model}}) \quad (3)$$

When d is 64, the text sequence is represented as 512 characters, the $2i$ is the even position in the given sequence of the input vector and $2i + 1$ is the odd position. When the transformer extracts the features $x_{[CLS]}$ and $x_{[MASK]}$ from the two special words $w_{[CLS]}$ and $w_{[MASK]}$ in the S -sequence, the Bert loss function considers only the prediction of the masked values and thus ignores the prediction of the non-masked words as shown in Equations (4) and (5).

$$Loss = i^{L(x_{[MASK]}^i)F(x_{[MASK]}^i)} \quad (4)$$

$$F(x_{[MASK]}^i) = \begin{cases} k & \text{if } x \in R \\ 1 & \text{if } x \notin R \end{cases} \quad (5)$$

GRU is a specific model of a recurrent neural network that performs machine learning tasks related to memory and clustering using connections through a series of nodes, which allows GRU to pass information over multiple time periods in order to influence subsequent time periods. GRU can

be considered as a variant of LSTM as both are similar in design and produce equally good results, both gate recursive units help in tuning the neural network input weights to solve the vanishing gradient problem. As a refinement of the recurrent neural network, GRU has a gate called update gate and reset gate r_t . Using an input vector x and an output vector h_t , the model refines the information flow in the output-1 model by controlling h_t . As with other types of recurrent network models, GRU with gated recurrent units can retain information over a period of time, which is why it is easiest to describe these techniques as "memory-centric" types of neural networks. In contrast, other types of neural networks without gated recurrent units typically do not have the ability to retain information. the structure of the GRU is shown in Figure.

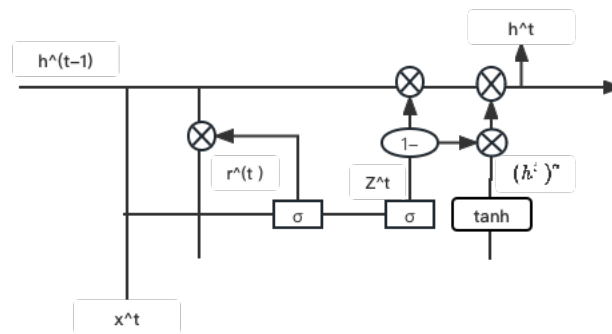


Figure 4. Structure of GRU.

BiGRU refers to Bidirectional Gated Recurrent Unit (BGRU), i.e., an additional reverse layer is added to the GRU. BiGRU can process both forward and reverse information of the input sequence, thus capturing features in the sequence more comprehensively and improving model performance. It can be used for a variety of tasks such as speech recognition, person name recognition, lexical annotation etc.

BiGRU has advantages over GRU such as bi-directionality, better performance, better handling of long sequences, and finer-grained feature representation. Therefore, BiGRU has become a very effective model in various sequence learning tasks. Since GRU retains only less state information and is prone to the problems of gradient vanishing or gradient explosion, its processing of long sequences may not be as effective as BiGRU.

BiGRU, on the other hand, introduces more state information through the inverse layer, which improves the handling of long sequences and reduces the risk of vanishing or exploding gradients. In combining BiGRU and Softmax, the input text sequence can be encoded using BiGRU, then the encoded result is passed to a fully connected layer, and finally classified using the Softmax activation function. Specifically, the output of BiGRU can be used as an input to the fully connected layer, and the output of the fully connected layer is then used for label prediction via Softmax. This combination can effectively improve the performance of text categorization, especially when facing complex text datasets.

The BiGRU model operates on a given sequence of input vectors $\langle x_1, x_2, \dots, x_t \rangle$ (where x_t denotes a concatenation of input features) and computes the corresponding hidden activations $\langle h_1, h_2, \dots, h_t \rangle$. At the same time, a sequence of output vectors is generated from the input data $\langle y_1, y_2, \dots, y_t \rangle$. At time t , the current hidden state is determined by three components: the input vectors $\langle x_1, x_2, \dots, x_t \rangle$, the forward hidden state and the backward hidden state. The reset gate (r_t) controls how much previous state information is ignored; the smaller the value of r_t , the more previous state information is ignored. The update gate (z_t) controls the extent to which the unit state receives new input information. The symbol \otimes represents elemental multiplication, σ denotes a sigmoid function, and \tanh denotes a hyperbolic tangent function. The hidden state h_t , update gate (z_t), and reset gate (r_t) of the BiGRU are computed by Equations (6)–(9).

$$Z_t = (w_z[h_{t-1}, x_t]) \quad (6)$$

$$r_t = (w_r[h_{t-1}, x_t]) \quad (7)$$

$$h_t = GRU(x_t, h_{t-1}) \quad (8)$$

$$h_t = w_t h_t + v_t h_t + b_t \quad (9)$$

Softmax functions are widely used in tasks such as text categorization, sentiment analysis, and machine translation. Often, we need to represent a piece of text as a vector to facilitate subsequent computation and analysis. One of the common ways to represent text vectors is to use the Word Embedding technique to map each word to a low-dimensional vector of real numbers, and then transform the entire text into a fixed-length vector through some aggregation or transformation operations. In the following algorithm, w_a represents the weight matrix of the attention function, \tanh refers to the hyperbolic tangent function, y represents the sentiment analysis results, and b_a represents the corresponding bias of the output layer.

$$u_t = \tanh(w_a h_t + b_a) \quad (10)$$

$$\alpha_t = \frac{\exp(u_t^T u_a)}{\sum_t \exp(u_t^T u_a)} \quad (11)$$

$$s_{it} = \sum_{i=1}^n a_{it} h_{it} \quad (12)$$

$$y = \text{softmax}(w_a s_{it} + b_a) \quad (13)$$

After completing the text vector representation, we also need to perform tasks such as classification and labeling. At this point, the Softmax function can be used to map the text vectors to different classes of probability distributions. Specifically, in natural language processing tasks, a neural network model is usually used as a classifier, with text vectors as inputs, and after several layers of fully connected layers and nonlinear activation functions, the outputs are finally mapped to individual categories using the Softmax function, and the probability values of each category are calculated. Ultimately, we can consider the category with the largest probability value as the category to which the text belongs.

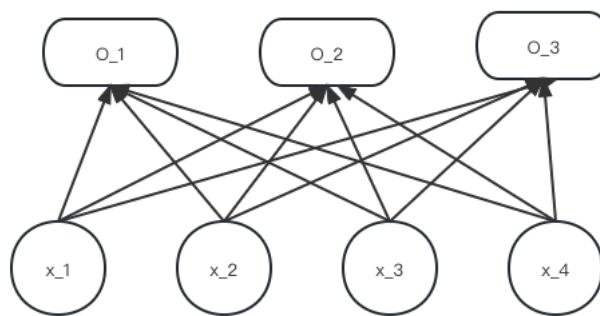


Figure 5. Softmax layer.

This paper investigates the Bert BiGRU Softmax model for sentiment analysis of online product quality reviews. The sentiment Bert model is used as an input layer for feature extraction in the preprocessing stage. The hidden layer of the bi-directional GRU performs dimension-oriented sentiment classification by using bi-directional long and short-term memory and selective recursive units to maintain the long-term dependencies inherent in the text regardless of length and number of occurrences.

The output layer of Softmax calculates sentiment polarity by merging to smaller weighted dimensions according to the attraction mechanism. The output layer of Softmax calculates sentiment polarity by merging to smaller weighted dimensions according to the attraction mechanism. Figure 6 shows the structure of the Bert BiGRU Softmax model as follows.

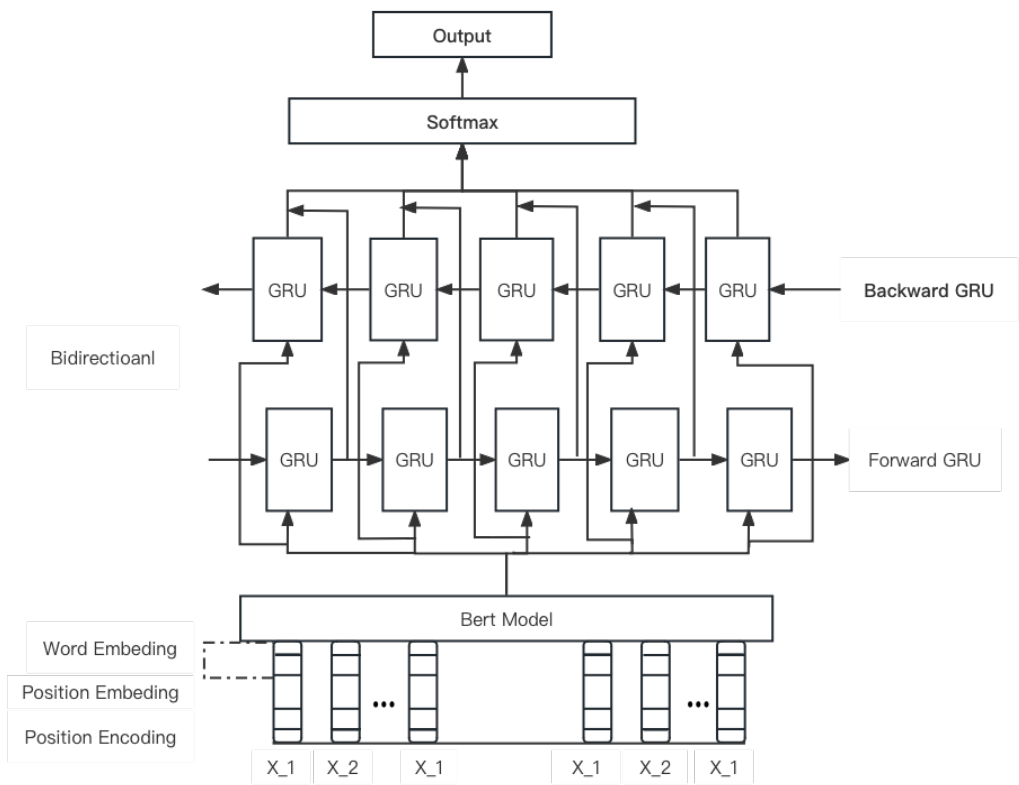


Figure 6. Bert BiGRU Softmax model.

The objective of sentiment analysis is to uncover the subjective emotional inclinations expressed by users towards products, as conveyed through online information. By leveraging deep learning techniques, sentiment analysis aims to establish connections between various features such as syntax, semantics, and emoticons, and sentiments. It involves categorizing user-generated content into positive, negative, or neutral sentiments, thereby enabling a better understanding of users' opinions and attitudes towards goods.

5. Experiments and Analysis of Results

5.1. Experimental Environment

Jupyter Notebook provided by Anaconda as the main development tool. The programming language used in this paper is python, the deep learning framework is Tensorflow, and the Keras toolkit based on TensorFlow is used as the main building tool for neural network models. Specific parameters are shown in Table 1.

Table 1. Experimental environment parameters.

Experimental environment	Environment Configuration
Operating system	windows11
IDLE	Jupyter Notebook 6.4.11+pycharm2022
TensorFlow	tensorflow_gpu-1.14.0

Keras	2.3.1
GPU	GTX1660Ti (6G)
Python	3.9

5.2. Experimental Process and Analysis

This paper crawls and analyzes a large dataset of 150 predefined dimensions from 500,000 online reviews of food products from Meituan, Hungry’s and other online sites that cover almost all aspects of different products with positive and negative polarities, assigning dimensions to "delivery", "hygiene" and "service". "hygiene" and "service". We use deep learning models of RNN, LSTM, GRU, BiGRU, BiLSTM, Bert BiLSTM, and Bert BiGRU Softmax for sentiment analysis of online product quality ratings in terms of hygiene, service, and price dimensions. The table below shows the dataset on food ratings. "Very tasty", "affordable" and "good" have the highest scores. "Difficult to eat", "Inadequate" and "Speechless" had the lowest scores. High scores for "Recommended" and "Timely" are the higher comments in the review messages. Additionally, "tasty" scored the highest of all sentiments, suggesting that taste has the most influence on a product's sales in food platforms.

Table 2. Commentary information of takeaway platforms on received goods.

Cake is very delicious, timely delivery, first-class service ah, the next opportunity to continue to buy.
Huge hard to eat, a salty and a sour.
Affordable price, good service attitude, the portion is super full, duck feet melt in the mouth, soft, very flavorful, super spicy, recommended!
Taste good, the portion is sufficient.
Really speechless, the last two still think it can, today this pineapple bun head cream is stinky, completely inedible!
The second time to eat, the taste is okay, just inside the Golden Harbor International, shopping tired can come to eat!
Delivered 1 hour, arrived at the things are cold.

In order to verify the classification effect of the Bert BiGRU Softmax model designed in this paper in the multimodal sentiment classification task of online review information, BiGRU model, CNN BiGRU model, BiGRU CNN model.

BiGRU Attention model, and Attention BiLSTM CNN were selected as models as competing models for comparison experiments on the same corpus. Attention-BiLSTM CNN are used as competing models for comparison experiments on the same corpus. The above five comparison models are also tuned for hyperparameters using grid search and tri-fold cross-validation. The specific parameter settings of all comparison models are shown in Table 3.

Table 3. Comparison of model hyperparameter settings.

Model	Optimization function	Batch file	Learning rate	Discard rate
-------	-----------------------	------------	---------------	--------------

Comparison model	BiGRU	Adagrad	64	0.012	0.4
	CNN BiGRU	RMSProp	192	0.009	0.2
	BiGRU CNN	Adam	96	0.006	0.2
	BiGRU Attention	Adagrad	64	0.014	0.2
	Attention BiGRU CNN	RMSProp	128	0.007	0.3
Current model	Bert BiGRU Softmax	RMSProp	120	0.012	0.2

The classification effect of each model is evaluated based on the metrics of Precision, Recall and F1 value. Table 4 gives the overall sentiment classification performance of each model on different evaluation metrics. The Attention BiGRU CNN model designed in this study outperforms the other five compared models in terms of precision rate and F1 value, and the recall rate is slightly lower than that of the Attention BiLSTM CNN model and higher than that of the other models. Taken together this paper's model performs optimally on the Chinese online review corpus. Further analysis:

Table 4. Comparison of results.

Model		Precision	Recall	F1
Comparison model	BiGRU	0.847	0.811	0.822
	CNN BiGRU	0.832	0.827	0.826
	BiGRU CNN	0.857	0.822	0.829
	BiGRU Attention	0.875	0.821	0.829
	Attention BiGRU CNN	0.894	0.827	0.841
Current model	Bert BiGRU Softmax	0.909	0.828	0.845

CNN BiGRU model and Attention BiGRU CNN model have very similar emotion classification effects, and Bert BiGRU Softmax outperforms the other five models, which indicates that the combination of the attention mechanism, bi-directional GRU, and BERT is more effective in capturing faceted emotion features, and the Attention BiGRU CNN model has 0.2% and 0.4% higher precision rate and 0.1% lower recall rate than Bert BiGRU Softmax model. Both are comparable for multisection sentiment categorization, but the former is slightly better than the latter.

According to the comparison of different models it is found that with the improvement of the model the accuracy rate is increasing, the comparison of classification results from the simplest BiGRU model improved BiGRU Attention model, Attention BiGRU CNN model and Bert BiGRU Softmax model found that the precision rate is improved by 3.4%, 1.5% and the recall rate is improved by 0.7% and 0.6% respectively.

This is due to the fact that the model that does not incorporate the attention mechanism cannot effectively differentiate the emotional information corresponding to different facets, and is susceptible to the emotional information of other facets when judging the emotional polarity of each facet. After adding the attention mechanism, the model can assign heterogeneous weights to different

features to emphasize the corresponding emotional information in each scene, reduce the inter-scene influence and the interference of irrelevant factors, and effectively improve the emotional classification effect of the model.

The following table demonstrates the F1 values for sentiment classification for different facets for each model. It can be seen that the models are relatively ineffective in categorizing the emotions of environment and packaging, with F1 values below 0.8, indicating that the semantic information of these two facets is more complex and more difficult to capture accurately than other facets. In addition, the Bert BiGRU Softmax model was designed in this study to optimize the F1 value in four facets: hygiene, taste, service, and price. The related results further indicate that the introduction of the attention mechanism and the combination of the BiGRU model and the BERT model can effectively improve the model's ability to categorize multi-dimensional emotions.

Table 5. Comparison of F1 values for multisection sentiment classification results.

Comparison model	Comparison model					Current model
Facet	BiGRU	CNN BiGRU	BiGRU CNN	BiGRU Attention	Attention BiGRU CNN	Bert BiGRU Softmax
Flavor	0.883	0.854	0.862	0.845	0.841	0.858
Price	0.832	0.840	0.826	0.843	0.853	0.856
Packaging	0.791	0.840	0.854	0.841	0.855	0.851
Service	0.848	0.852	0.862	0.855	0.830	0.844
Delivery time	0.832	0.841	0.826	0.843	0.853	0.854
Hygiene	0.903	0.894	0.898	0.909	0.913	0.923
Environment	0.758	0.787	0.777	0.759	0.778	0.788

Table 6 gives the multisection sentiment classification results of the Bert BiGRU Softmax model for 10 randomly selected sample comment data and the corresponding predicted probability values for this result. As can be seen, comment number 1179 should have a sentiment polarity of 0 on the service scale, and the model misclassifies it as -1. In addition to this, the emotion classification results of the sample comments are basically correct across all facets, indicating that the model in this paper performs well in the multi-faceted emotion polarity classification task.

Table 6. Partial results of Bert BiGRU Softmax multisection sentiment classification.

Classification results								
serial number	Comments	taste	price	packaging	service	delivery time	hygiene	environment
80	The environment is very good, the service is very warm, the flavor is very good, the fish is also very fresh, my son and I eat so full, the buns and dumplings are very tasty yo.	1	0	0	1	0	1	1
190	The cake is very delicious, the delivery is timely, the service is first-class ah, the packaging is also OK, next time there is an opportunity to continue to buy.	1	0	1	1	1	0	0
270	It's awful. One is salty and the other is sour.	-1	0	0	0	0	0	0
320	There were actually worms in the rice, so disgusting, I don't want to eat it anymore! I don't want to eat it!	-1	0	0	-1	0	-1	0
569	The flavor is still good, and the group buy is really cheaper.	1	1	0	0	0	0	0
772	The owner is warm, the restaurant is clean, the taste is good, next time to come again	1	0	0	1	0	1	1
1020	It's good, it's good. I've bought it several times, it's cheap, and the service is good.	1	1	0	1	0	0	0
1179	Let's not talk about the other flavors. It's the diarrhea that's real.	0	0	0	-1	0	-1	0
2980	I clearly ordered two orders of thin rice, can't you see? I'll never buy it again.	0	0	0	-1	0	0	0
5678	I've always eaten mutton noodles in her house, the flavor is very authentic and delicious highly recommended!	1	0	0	0	0	0	0

6. Discussion and Conclusion

Online reviews are consumer evaluations and feedback on products sold on online platforms. These reviews record multimodal information such as user experience, product characteristics and service satisfaction, and are important references for both manufacturers and platform operators. By analyzing customer feedback from reviews, vendors and platform operators can learn how the products they produce and sell are performing in the marketplace, and at the same time be able to continually improve their product designs and optimize their service processes based on user needs and expectations, thereby improving product quality and maintaining customer relationships.

Online reviews encompass consumer appraisals and assessments of merchandise vended on digital platforms. These evaluations encapsulate diverse dimensions including user encounters, product attributes, and contentment with services rendered. They serve as invaluable points of reference for both manufacturers and platform administrators. Through scrutinizing customer feedback embedded within these reviews, vendors and platform operators can glean insights into the reception and efficacy of their offerings in the market. Consequently, they can embark on a journey of perpetual refinement in terms of product conception, whilst optimizing service procedures to align with user requisites and anticipations. This endeavor ensures enhanced product excellence and the preservation of enduring customer rapport.

Through meticulous examination of customer feedback derived from reviews, vendors and platform operators are equipped with the means to ascertain the performance of their merchandise within the marketplace. Simultaneously, this practice allows them to continuously refine their product designs and optimize service processes in accordance with user needs and expectations. This concerted effort not only enhances product quality but also fosters enduring customer relationships.

The proposed model comprises several integral components: foremost, the utilization of the Bert model as a feature extractor facilitates the extraction of semantic representations from the input layer's textual comments, thereby capturing nuanced semantic relations between words and contextual information. Subsequently, the integration of the BiGRU model, augmented with an attention mechanism, assumes the role of a hidden layer, facilitating the acquisition of high-dimensional semantic coding encompassing attention probabilities at the input layer and textual sequences of contextual information. Finally, the Softmax classifier is employed to categorize all comments into sentiment polarity prediction and trend classification tasks, thus enabling effective recognition and analysis of users' affective tendencies and emotional attitudes.

Comprehensive experiments were conducted on a substantial review dataset, comparing the performance of the proposed Bert BiGRU Softmax model against other advanced models such as CNN BiGRU, BiGRU, and BiGRU Attention. The experimental findings unequivocally demonstrate that the Bert BiGRU Softmax model exhibits superior performance and accuracy, thereby effectively augmenting the precision of sentiment analysis pertaining to online product quality reviews.

Furthermore, we delve into the integration of domain expertise and human experiential insights into the training process of the model, aiming to cater more effectively to diverse industries and contextual nuances. Looking ahead, there is ample room for further refinement and optimization of the model. This can be achieved by amalgamating multimodal data and leveraging a plethora of information sources to conduct extensive and profound investigations into sentiment analysis. Such endeavors hold promise for uncovering novel breakthroughs and innovations, ultimately enhancing the performance of the model in real-world scenarios.

Author Contributions: J.L.: wrote the main manuscript style. Y.S.: optimized it. Y.Z.: methodology and C.L.: collected part of the data. All the authors read the manuscript.

Funding: This work was supported by the National Social Science Fund of China under Grant No.18BJY033.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data availability The data set and code used in this article are available upon request from the corresponding author.

Conflicts of Interest The authors declare that there are no competing interests related to the content of this article.

References

1. M. Guerini, L. Gatti, M. Turchi.: Sentiment analysis: How to derive prior polarities from SentiWordNet. arXiv preprint arXiv: 1309.5843(2013)
2. Y. KIM.: Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882(2014)
3. Zhou, Z.H., Wu, J.X., Jiang, Y., et al.: Genetic Algorithm Based Selective Neural Network Ensemble. In: Proceedings of the 17th International Joint Conference on Artificial Intelligence, Seattle, WA, 2001, vol.2, pp.797-802(2001)
4. Tama, B. A., Rhee, K. H.: A Combination of PSO-Based Feature Selection and Tree-Based Classifiers Ensemble for Intrusion Detection Systems. In: Advances in Computer Science and Ubiquitous Computing pp 489-495 Singapore: Springer, 2015,10.1007/978-981-10-0281-6_71(2015)
5. Zhang, L., Wang, L.W., Huang, B., Liu, Y.T.: A Sentiment classification model and experimental study of multi-scale convolutional neural network microblog comments based on word vectors. In: Library and Intelligence Work. 2019,63(18):99-108, 10.13266/j.issn.0252-3116.2019.18.012.(2019)
6. Tsai, C.Y., Chen, C.J.: A PSO-AB Classifier for Solving Sequence Classification Problems. In: Applied Soft Computing, vol. 27, pp. 11-27, Feb. 2015 10.1016/j.asoc.2014.10.029(2015)
7. D. Lahat, T. Adali and C. Jutten.: Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects. In: Proceedings of the IEEE, vol. 103, no. 9, pp. 1449-1477, Sept. 2015,10.1109/JPROC.2015.2460697(2015)
8. Zhao, L.: Research on Multimodal Data Fusion Algorithm. In: Dalian: Dalian University of Technology (2018)
9. Xu, X.B., Chen, L., W, C.L.: A multi-sensor data fusion algorithm based on the unified representation of multi-source heterogeneous information. In: Journal of Henan University: Natural Science Journal of Henan University: Natural Science, 2005 (3): 67-71(2005)
10. Tian, G., Han, L., Zhao, Y.H.: The application of multi-source data fusion for real-life three-dimensional modeling in land consolidation. Application of multi-source data fusion and three-dimensional modeling in land consolidation. In: Ecology Magazine, 2019, 38 (7): 2236-2242(2019)
11. Guo, L.J., Peng, X., Li, Z.H., et al.: A Chinese language dependent syntactic treebank for multi-domain and multi-source text Syntactic Tree Library Construction for Multi-Domain and Multi-Source Texts. In: Journal of Chinese Language and Information, 2019, 33(2): 34-42(2019)
12. Zheng, Y.: Methodologies for cross-domain data fusion: an overview. In: IEEE Transactions on Big Data, 2015, 1 (1): 16-34.16-34(2015)
13. J. Kennedy and R. C. Eberhart.: A discrete binary version of the particle swarm algorithm. In: 1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation, Orlando, FL, USA, 1997, pp. 4104-4108 vol.5, 10.1109/ICSMC.1997.637339(1997)
14. Chandra, A., Chen, H., Yao, X.: Trade-off Between Diversity and Accuracy in Ensemble Generation. In: Multi-objective Machine Learning pp 429-464 2006,10.1007/3-540-33019-4_19(2006)
15. A. H. .-R. Ko, R. Sabourin and A.: Combining Diversity and Classification Accuracy for Ensemble Selection in Random Subspaces. In: The 2006 IEEE International Joint Conference on Neural Network Proceedings, Vancouver, BC, Canada, 2006, pp. 2144-2151, 10.1109/IJCNN.2006.246986(2006)
16. B. Pang, L. Lee, S. Vaithyanathan.: Thumbs up? sentiment classification using machine learning techniques. In: Proc. 2002 Conf. on Empirical Methods in Natural Language Process., USA, (2002),79-86, 10.48550/arXiv.cs/0205070(2002)
17. A. Tripathy, A. Agrawal, S. K. Rath.: Classification of sentiment reviews using n-gram machine learning approach, Expert Systems with Applications, 57 (2016), 117-126, 10.1016/j.eswa.2016.03.028(2016)
18. Zheng, F., Wei, D.H., Huang, S.: Text classification method based on LDA and deep learning. In: Computer Engineering and Design. 2020, 41(8): 2184-2189(2020)
19. S. Kim, E. Hovy.: Determining the sentiment of opinions. In Proc. Conf. Comput. Linguis. Geneva, Switzerland, (2004), 1367-1373(2004)
20. Zhu, X.L., Shi, Y.D.: Automatic Classification Model of Composition Material in Primary School Based on Text Rank and Character-level Convolutional Neural Network. In: Computer Application and Software.2019, 36 (1): 220-226(2019)
21. Zhang, Q., Gao, Z.M., Liu, J.Y.: Research on short text categorization of microblogs based on Word2vec. In: Netinfo Security. 2017(1): 57-62(2017)
22. Yang, Y.T., Wang, M.Y., T, X.Y., et al.: Sina Microblog Sentiment Classification Based on Distributed Representation of Documents. In: Journal of Intelligence. 2016, 35(2): 151-156(2016)
23. M. Guerini, L. Gatti, M. Turchi.: Sentiment analysis: How to derive prior polarities from Senti Word Net, arXiv preprint arXiv: 1309.5843(2013)

24. Shao, Y.F., Liu, D.S.: Classifying Short-texts with Class Feature Extension. In: Data Analysis and Knowledge Discovery. 2019, 3(9): 60-67, 10.11925/infotech.2096-3467.2018.1423(2018)
25. Cai, Q.P., Ma, H.Q.: A Fine-grained Sentiment Analysis Model for Product Reviews Based on Word2Vec and CNN. In: Library and Intelligence Work .2020,64(6):49-58(2020)
26. Z. Qu, Y. Wang, X. Wang.: A Hierarchical attention network sentiment classification algorithm based on transfer learning, J. Comput. Appl., 38 (2018), 3053–3056(2018)
27. Tao, Z.Y., L. X.B., Liu, Y., et al.: Classifying Short Texts with Improved-Attention Based Bidirectional Long Memory Network. In: Data Analysis and Knowledge Discovery. 2019, 3(12): 21-29, 10.11925/infotech.2096-3467.2019.0267(2019)
28. Duan, D.D., Tan, J.S., Wen, Y., et al.: Chinese Short Text Classification Algorithm Based on BERT Model. In: Computer Engineering, 2021, 47(1): 79-86(2021)
29. A. Go, R. Bhayani, L. Huang.: Twitter sentiment classification using distant supervision. In: CS224N Project Report, Stanford, 1 (2009), 1–12(2009)
30. Du, L., Cao D., Lin, S.Y., et al.: Extraction and automatic classification of Chinese medical records based on BERT and Bi-LSTM fusion attention mechanism. In: Computer Science. 2020, 47 (S2): 416-420, 10.1109/BIBM49941.2020.9313359(2020)
31. J. Trofimovich.: Comparison of neural network architectures for sentiment analysis of Russian tweets. In Proc. Int. Conf. Dialogue 2016, RGGU(2016)
32. Qian, Q., Huang, M., Lei, J., Zhu, X.: Linguistically regularized LSTMs for sentiment classification. In Proc. 55th Ann. Meet. Assoc. Comput. Ling., 1 (2016), 1679–1689, 10.48550/arXiv.1611.03949 (2016)
33. Ge, X.W., Li, K.X., Cheng, M.: Text Classification of Nursing Adverse Events Based on CNN-SVM. In: Computer Engineering & Science. 2020, 42(1): 161-166(2020)
34. Wang, H.T., Song, W., Wang, H.: Text classification method based on hybrid model of LSTM and CNN. In: Small Microcomputer Systems. 2020, 41(6): 1163-1168(2020)
35. Wu, P., Ying, Y., Shen, Si.: Research on the classification of netizens' negative emotions based on bidirectional long and short-term memory model. In: Journal of Intelligence. 2018,37(8):845-853(2018)
36. Zhang, H.T., Wang, D., Xu, H.L., Sun, S.Y.: Research on microblog opinion sentiment classification based on convolutional neural network. In: Journal of Intelligence. 2018,37(7):695-702(2018)
37. Xu, S.K., Zhou, Z.Y.: Sentiment classification model and application of WeChat tweets based on multi-scale BiLSTM-CNN. In: Intelligence Science. 2021,39(5):130-137, 10.1016/j.neucom.2016.02.077(2016)
38. Fan, H., Li, P.F.: Sentiment analysis of short texts based on Fast Text word vectors and bi-directional GRU recurrent neural network--Taking the text of Weibo comments as an example. In: Intelligence Science. 2021,39(4):15-22(2021)
39. K. Cho, B. van Merriënboer, D. Bahdanau.: On the properties of neural machine translation: encoder-decoder approaches. In Proce. Eighth Workshop Syntax, Semant. Struct. Stat. Trans., Doha, Qatar, (2014), 103–111, 10.48550/arXiv.1409.1259(2014)
40. Mekel D , Frasinicar F .ALDONA: a hybrid solution for sentence-level aspect-based sentiment analysis using a lexicalised domain ontology and a neural attention model[C]//the 34th ACM/SIGAPP Symposium.ACM, 2019.DOI:10.1145/3297280.3297525.
41. Thet T T , Na J C , Khoo C S G .Aspect-based sentiment analysis of movie reviews on discussion boards[J].Journal of Information Science, 2010, 36(6):823-848.DOI:10.1177/0165551510388123.
42. Tul Q , Ali M , Riaz A ,et al.Sentiment Analysis Using Deep Learning Techniques: A Review[J].International Journal of Advanced Computer Science and Applications, 2017, 8(6).DOI:10.14569/IJACSA.2017.080657.
43. Liu H , Chatterjee I , Zhou M C ,et al.Aspect-Based Sentiment Analysis: A Survey of Deep Learning Methods[J].IEEE Transactions on Computational Social Systems, 2020, PP(99):1-18.DOI:10.1109/TCSS.2020.3033302.
44. Xu F , Pan Z , Xia R .E-commerce product review sentiment classification based on a nave Bayes continuous learning framework[J].Information Processing & Management, 2020, 57(5):102221.DOI:10.1016/j.ipm.2020.102221.
45. Xia, J.; Dai, Y. An Unsupervised Learning Method for Suppressing Ground Roll in Deep Pre-Stack Seismic Data Based on Wavelet Prior Information for Deep Learning in Seismic Data. Appl. Sci. 2024, 14, 2971.
46. Wang, S.; Chen, Y.; Yi, Z. A Multi-Scale Attention Fusion Network for Retinal Vessel Segmentation. Appl. Sci. 2024, 14, 2955.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.