applied sciences

MDPI

Article

Research on Online Review Information Classification Based on Multimodal Deep Learning

Jingnan Liu *, Yefang Sun, Yueyi Zhang * and Chenyuan Lu

College of Economics and Management, China Jiliang University, Hangzhou 314423, China; sunyefang21@163.com (Y.S.); chenchenlulu06@126.com (C.L.)
* Correspondence: 13777574168@163.com (J.L.); 02a0702026@cjlu.edu.cn (Y.Z.)

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 to be highly applicable to online reviews.

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

Citation: Liu, J.; Sun, Y.; Zhang, Y.; Lu, C. Research on Online Review Information Classification Based on Multimodal Deep Learning. *Appl. Sci.* **2024**, *14*, 3801. https://doi.org/10.3390/app14093801

Academic Editor: Christos Bouras

Received: 2 April 2024
Revised: 19 April 2024
Accepted: 22 April 2024
Published: 29 April 2024

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, the consumers' perceptions and evaluations 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. Kiran et al. (2020) considers that online shopping brings convenience, but the virtual nature of the e-commerce platform will result in online product introduction information not matching the real product, as well as the issues of poor product quality or after-sales service which cannot meet the needs of the consumers [1]. 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, the users' subjective product reviews convey a certain inclination of sentiment. Kaur and Sharmav (2023), Shuang et al. (2020), and Vijayaragavan et al. (2020) consider that these reviews serve a dual purpose as both "information providers" and "emotional influencers", significantly impacting merchants [2–4]. Not only do they influence a store's 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 toward 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. Yuting Yang et al. encoded text into high-dimensional vectors with contextual semantic, sequential, and sentiment information through the Doc2Vec model and verified the effectiveness of the document's distributed representation approach [5]. These attribute sets can be derived based on dictionaries or commonly employed statistical methods [6].

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 [7]. 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 the 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, wherein 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 [8].

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

In order 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. Sivakumar and Rajalakshmi (2022) showed that deep learning methods have achieved wide application in many areas such as image recognition, target detection, and network optimization, and are now also being integrated into sentiment analysis and traditional machine learning techniques. This integration has shown good results, especially in building sentiment vocabularies [9].

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. Related Works

This section presents a review of the field of sentiment categorization of online review information, highlighting the main areas of research, analytical methods, and conclusions, with a literature review to understand the contributions and applications of each method in their respective areas of research.

Research in the field of multimodal sentiment analysis has been extensively explored in an attempt to combine data of different modalities (e.g., text, image, audio, etc.) to improve the accuracy and comprehensiveness of sentiment analysis tasks. And previous research findings have shown that multimodal sentiment analysis methods that combine different modal data have better performance compared to single-modal methods. The combined use of multimodal data provides a richer and more comprehensive representation of sentiment and helps to improve the accuracy and robustness of sentiment classification. Research on multimodal sentiment analysis has used various approaches to integrate multimodal data, including feature-level fusion, model-level fusion, and attention mechanisms. Deep learning frameworks such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and attention mechanisms have been widely used for multimodal sentiment analysis tasks. The existing findings suggest that combining multimodal data can improve the performance of sentiment analysis, especially in terms of categorizing and understanding sentiments at a more granular level. Deep learning frameworks have shown good results in multimodal sentiment analysis, which helps to exploit the association and complementarity between different data modalities.

Pang (2002) first applied the machine learning method of N-Gram to the field of sentiment analysis, and the experimental results show that N-Gram achieved the highest classification accuracy of 81.9% [10]. Since feature selection affects the performance of machine learning methods, Abinash Tripathy (2016) analyzed online review comments through the N-gram model combined with machine learning methods, and the experimental results show that the SVM combined with unigram, bigram, and trigram features obtained the best classification results [11]. 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 [12,13]. 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 [14,15].

In particular, Duan Dandan et al. used a short text classification model based on BERT, which utilizes BERT's own sentence vector training to achieve automatic text classification [16,17]. 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, namely data mapping, data alignment, data fusion, and co-learning [18,19]. 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 [20]. The mainstream multimodal data mapping method is as follows: based on the existing mapping relationship, the existing multimodal data are first symbolized or vectorized, and are then 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 [21,22], which maps image modal data to textual modal data.

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 wherein two or more dimensions exist for a sentiment word. Therefore, sentiment analysis of online reviews is considered as a multidimensional classification process [23]. Kim and Hovy (2004) applied synonyms and antonyms of WordNet dictionaries and hierarchical structures to analyze word vectors' sentiment ten-

dencies [24]. Zhu Xiaoliang and other researchers solved the composition classification problem by using the TextRank model to filter key sentence words and combining it with a word embedding model to model documents. This was in consideration of the problem that the Word2Vec model cannot identify the importance of special words and scene words in Chinese in the text [25].

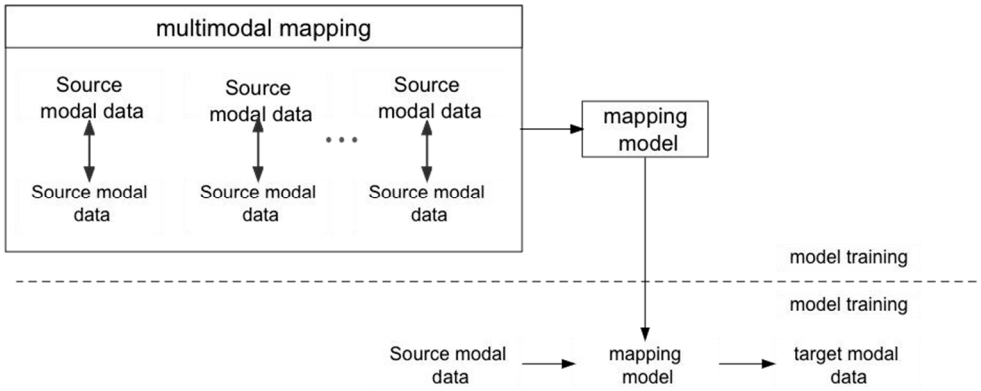

Figure 1. Multimodal data mapping approach.

The construction of a sentiment lexicon requires a significant amount 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 [26]. 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.

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 significant amount of attention in the academic community [27,28]. Zheng Fei et al. showed that a combination of the 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 [29]. Kim and Hovy (2004) applied synonyms and antonyms of WordNet dictionaries and hierarchical structures to analyze word vectors' sentiment tendencies [30]. Zhu Xiaoliang and other researchers solved the composition classification problem by using the TextRank model to filter key sentence words and combining it with a word embedding model to model documents. This was in consideration of the problem that the Word2Vec model cannot identify the importance of special words and scene words in Chinese in the text [31]. Zhang Qian et al. proposed to introduce the TFIDF model to weigh the output word vector matrix to obtain the weighted text vectorization model and classify it [32]. Yuting Yang et al. encoded text into high-dimensional vectors with contextual semantic, sequential, and sentiment information through the Doc2Vec model and verified the effectiveness of the document's distributed representation approach [33].

Du Lin et al. 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 [34]. Yunfei Shao et al. used LDA and TF-IDF to expand the input text features, and then further aggregated the features to form classification basis vectors with a CNN, which improved the classification effect on news headline data [35,36]. Qu and Wang (2018) proposed a sentiment analysis model based on hierarchical attention networks with a 5% improvement in accuracy over recurrent neural networks [37]. Tao, Zhiyong et al. fused the bidirectional

features of the BiLSTM model and used them for attention weight computation, achieving improved classification results on various benchmark datasets [38].

In particular, Duan Dandan et al. used a short text classification model based on BERT, which utilizes BERT's own sentence vector training to achieve automatic text classification [39,40]. 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 [41]. 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 [42]. In the field of traditional machine learning classifiers, some of the literature uses the 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 [43].

In the field of deep learning, H. Wang et al. 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 the CNN to realize the reuse of the original features; the second channel uses 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 [44–46].

It is worth noting that the model uses a unidirectional LSTM, while Bi-LSTM is considered to be a better option than unidirectional LSTM. LSTM models are used for phrase-level sentiment classification centered on regularization, which contains linguistics such as negativity, intensity, and polarity [47], 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 (GRUs) to analyze dependency contexts, which showed significant improvements in various tasks. This takes into consideration the multimodal, sparsely informative, highly unstructured, and word polysemous nature of text data review [48,49]. Basiri et al. (2021) presented a bidirectional convolutional neural network (CNN)–recurrent neural network (RNN) deep model for sentiment analysis of Twitter product reviews [50].

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.

There are some important studies in the current research area which is summarized as follows: first, there is a research gap on how to effectively integrate text and image data to improve the performance of sentiment analysis. Currently, although there have been studies on text sentiment analysis and image sentiment analysis, there is not enough research on how to effectively integrate these two different modal data for sentiment analysis. Such integration can help the model understand the content of the comments more comprehensively, thus improving the accuracy of sentiment classification.

Second, efficient models specifically designed for multimodal sentiment analysis tasks are still lacking. Most of the current research focuses on the sentiment analysis of single-

modal data (text or images), while there are relatively few sentiment analysis models for multimodal data. Therefore, how to design a multimodal sentiment analysis model that can fully utilize the information of text and image data to improve the accuracy of sentiment classification is a research direction that needs to be explored in depth.

Finally, there is still a relative lack of research on the reliability and scalability of multimodal sentiment analysis models applied to different domains and datasets. Understanding the applicability of multimodal sentiment analysis models on different domains and datasets, as well as the stability and reliability of their performance in real-world applications, are key research issues. Therefore, more in-depth research is needed to explore the application scope and generalization capability of multimodal sentiment analysis models.

Research question: How can multimodal data be processed using the multimodal Bert BiGRU Softmax model to achieve more accurate online review information classification to improve the consistency and accuracy of review information classification?

Objective: To realize the enhancement of online comment information classification based on multimodal deep learning to fill the identified gaps and improve the accuracy, consistency, and application scope of the classification.

3. Collection and Analysis of Data Sets

In this paper, the food review information of the Meituan online platform is selected as the data source. We chose the platform Meituan because it is a leading e-commerce platform for life services in China, which provides a wide range of services such as takeout, restaurant reservation, hotel booking, travel, movie tickets, bike sharing, and so on. Users can order, pay for, and book services online through Meituan. Meituan has a huge user base in China, covering both cities and villages, providing users with a convenient life service experience. Meituan has a large number of users in China, covering all ages and social groups. Users can enjoy fast and convenient services through the Meituan platform. Meituan's user experience is designed to be simple and easy to use, and the payment process is convenient, which is widely welcomed by users. Meituan has a large market share and influence in China's living service sector. As one of the largest e-commerce platforms for lifestyle services in China, Meituan's development has not only promoted the popularization of online consumption patterns, but also the digital transformation and enhancement of the service industry. Its convenient service model and rich product offerings are popular among Chinese consumers and have had a positive impact on China's lifestyle and consumption habits. 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 is worth noting that people habitually look at the information in the online reviews of the restaurant or store before they decide to spend their money there.

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) deletion due to privacy considerations; (2) filtering of short comments with fewer than 50 Chinese characters and long comments with more than 1000 Chinese characters; (3) discarding comments in which the percentage of non-Chinese characters is more than 70%; and (4) preprocessing of the data which includes data cleansing, Chinese word splitting, de-duplication of words, etc.

In addition, in order to avoid ethical issues and minimize potential bias during data processing and model training, it is necessary to eliminate bias in the data, including characteristics such as gender and ethnicity, and to ensure the diversity and transparency of the dataset. For the model training process, it is necessary to review the model's processing of the data and monitor the output of the model training at regular intervals, and to modify the model training process as soon as any bias occurs, as detailed below.

Integration of multimodal data from diverse datasets: ensure that the dataset is representative, covers diversity, and avoids the existence of specific groups or biases in the dataset in order to reduce the bias in the model training process. It is also important to integrate data of different modalities, such as text, images, etc., to synthesize multifaceted information and reduce the bias that may be brought by a single modality.

Balancing the dataset: address the problem of category imbalance to ensure that the samples of different categories are balanced to avoid the model bias. In the process of performing sentiment analysis of review messages, ensuring a balance between the number of reviews, positive, neutral, and negative sentiments is crucial for model training and performance, as well as for the results and rationality of the analysis. A balanced dataset not only helps the model to better learn the differences between various sentiment categories, but also improves classification accuracy. When dealing with multimodal data, it is therefore important to ensure that the number of comment sentiment categories is balanced to avoid sample imbalance which can lead to model training biased toward one category.

Transparency and interpretation: improve the transparency and interpretation of the model to ensure that the decision-making process of the model can be understood and interpreted, reducing potential bias and discrimination.

Review and monitoring: regularly review the model and the data processing process, monitor the model outputs, and identify potential biases and inequities in a timely manner and take corrective action.

Elimination of implicit bias and ethical review and moral considerations: review the dataset and features to eliminate possible implicit bias in them, such as sensitive features like gender, race, etc., and avoid unfairness in the model in these areas. At the same time, an ethical review should also be conducted to consider the possible social impacts of the model application and to ensure that the model design and application comply with ethical standards and laws and regulations.

The methods and measures in the above content can not only reduce potential bias and unfairness, but also ensure that ethical issues are paid attention to in the process of model training and data processing, improve the fairness and credibility of the model, and prevent the model from generating or expanding potential bias in the analysis.

For the Bert BiGRU Softmax model, the preprocessing steps and feature extraction techniques applicable to both textual and non-textual data have shown good adaptability in experimental simulations, but since most of the data randomly collected during data acquisition are textual and due to the space limitation of this article, more specific details on image processing will be demonstrated in future work. For the Bert BiGRU Softmax model, the input layer configuration is crucial to efficiently preprocess and extract features from both textual and non-textual data to provide key support for model training and performance.

Step 1: In the case of text data, preprocessing includes the application of tokenization, stemming/lemmatization, and word embeddings to convert text into numerical representations and capture semantic and emotional information.

Step 2: For image data, the preprocessing phase includes resizing, normalization, and augmentation techniques to ensure that the images are uniformly sized, the pixel values are normalized, and a diversity of augmented data is generated to improve generalization.

Step 3: In terms of feature selection and integration, multimodal feature integration is key. When combining text and image features, a parallel structure or a fusion structure can be used. The parallel structure processes text and image features separately and then

concatenates or sums the features; the fusion structure can utilize the attention mechanism or joint training to fuse text and image information to obtain a more comprehensive multimodal feature representation.

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 information can be used to represent different forms of data forms and can also be the same form of different formats, generally expressed as text, images, audio, video, or mixed data.

Text data can provide detailed semantic information, including sentiment vocabulary, emotional expressions, and contextual information, which helps to capture the emotional color and sentiment tendency in comments. Text features are transformed into numerical forms by natural language processing techniques (e.g., word embedding), which provide rich information about the content of the comments and play an important role in sentiment analysis. Image data contain visual information that conveys visual emotional content and emotional expression in comments, such as image mood, color, and emotional expression. Image features provide supplementation and enrichment of the comment content, which can help the model understand the emotional meaning of the comments more comprehensively. There is complementarity between text features and image features, and combining information from both modalities can provide a more comprehensive and multidimensional emotional expression.

The synthesis mechanism for integrating text features and image features is the key, and the features of the two modalities are fused by designing an effective feature fusion strategy. Methods such as parallel structure, fusion structure, and attention mechanism are used to combine text and image features dynamically so that the model can synthesize different modal information. By integrating text and image features, the model is able to more accurately capture the sentiment information in the comments and enhance the accuracy and robustness of sentiment analysis.

Improving the integration of multimodal data (e.g., text and images) by combining sophisticated approaches from different modalities can help improve the accuracy of product review classification. Combining information from text and image data can enhance the model's ability to understand and classify product reviews. The use of sophisticated multimodal integration methods can better capture the associations and interactions between different data modalities, thereby enhancing the performance of the model.

Specific methods include designing effective multimodal feature fusion strategies, optimizing model parameters using joint training, introducing an attention mechanism in the model to dynamically focus on key information parts, and designing complex multimodal neural network architectures to achieve the effective fusion of text and image data. Improved multimodal data integration through these complex methods can improve the accuracy and performance of the product review classification task, allowing the model to understand the review content more comprehensively and to classify and analyze it more accurately. This integrated approach of utilizing different data modalities helps to expand the range of model applications and enhance the effectiveness of practical applications.

Image content recognition has been an important research problem. With the continuous development of learning methods, the accuracy of image recognition is constantly improving. Due to the fact that the original image consists of a pixel matrix, and that 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 a convolutional layer, the pixel matrix of the image after convolutional processing 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 is beneficial. In this paper, we use the open-source tool Image AI based on Python language, which is based on the 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. 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. How text is modeled has a significant impact on the effectiveness of downstream feature extraction models and classification models. Most of the common text modeling models are designed for English corpuses, either question and answer corpuses or comment corpuses. 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, wherein 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 [51], 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, wherein 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, as shown in Figure 2.

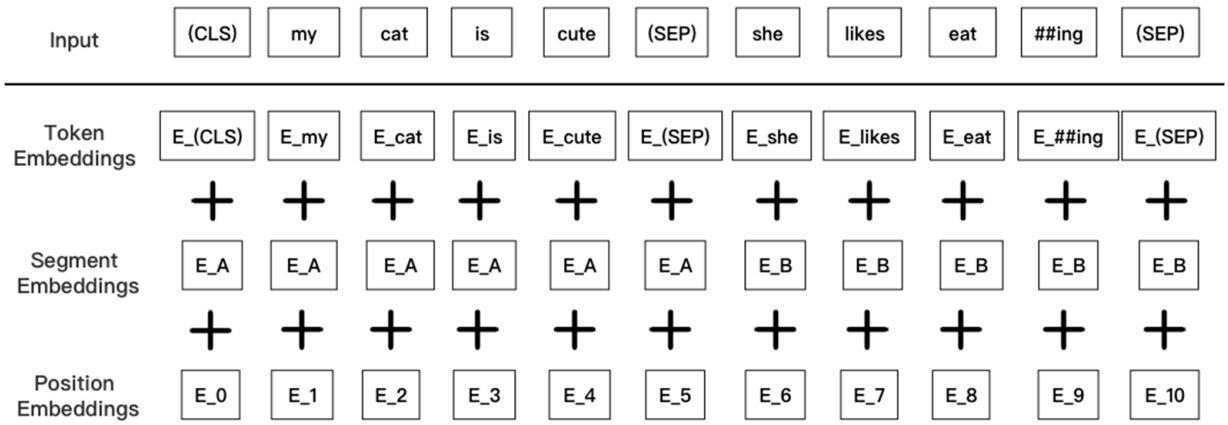

Figure 2. Structure of BERT input layer.

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, as shown in Figure 3. I In addition, we can add additional structures such as fully connected layers after the BERT model to perform fine-tuning operations for specific tasks, such as linguistic reasoning tasks like Q&A.

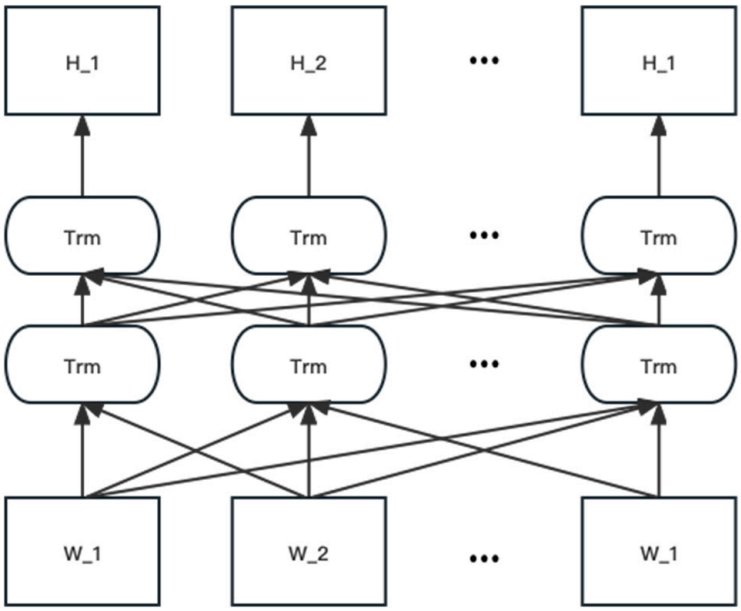

Figure 3. BERT model diagram.

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 Dc. For each word w_1 $(w_1, w_2, , w_m)$ and dimension dj in a review sentence, we assign a probability score that describes the probability $P(s)$ that word w_i belongs to class dj in an online product quality review (as in Equation (1)):

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

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

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

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

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).

$$Loss = i^{L(x^i_{[MASK]})F(x^i_{[MASK]})} \tag{4}$$

$$F\left(x^i_{[MASK]}\right) = \begin{cases} k\ if\ x\ \in\ R \\ 1\ if\ x\ \notin\ R \end{cases} \tag{5}$$

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 GRUs to pass information over multiple time periods in order to influence subsequent time periods. A 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, a GRU has a gate called update gate and reset gate rt. Using an input vector x and an output vector h_t, the model refines the information flow in the output-1 model by controlling ht. As with other types of recurrent network models, a 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 4.

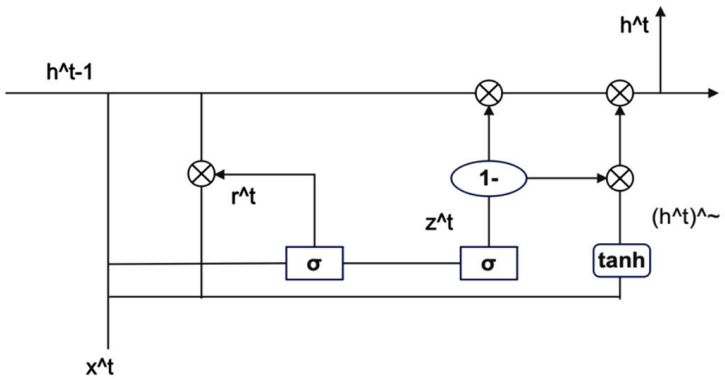

Figure 4. Structure of GRU.

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 the 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, and 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 $< x1, x2, \ldots, xt >$ (where xt denotes a concatenation of input features) and computes the corresponding hidden activations $< h1, h2, \ldots, ht >$. At the same time, a sequence of output vectors is generated from the input data $< y1, y2, \ldots, yt >$. At time t, the current hidden state is determined by three components: the input vectors $< x1, x2, \ldots, xt >$, 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).

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

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

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

$$h_t = w_t h_t + v_t h_t + b_t \tag{9}$$

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.

$$u_t = tanh(w_a h_t + b_a) \tag{10}$$

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

$$s_{it} = \sum_{i=1}^{n} a_{it} h_{it} \tag{12}$$

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

After completing the text vector representation, we also need to perform tasks such as classification and labeling, see Figure 5 below. At this point, the Softmax function can be used to map the text vectors to different classes of probability distributions (the processes represented by the arrows pointing to them in the figure below). 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.

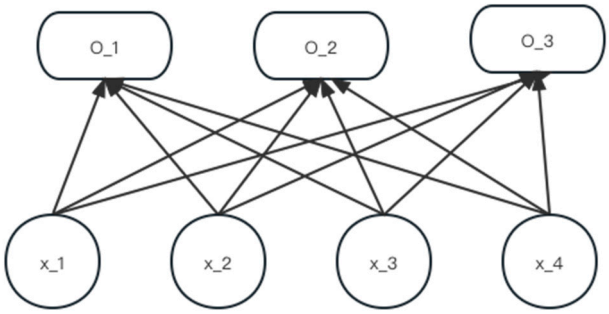

Figure 5. Softmax layer.

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.

BERT, as a pre-trained transformer model, has advantages in understanding text semantics and context, capturing rich semantic information, providing powerful text representation and improving the ability of semantic understanding. BiGRU, as a bi-directional recurrent neural network, is able to efficiently capture long-distance dependencies in text sequences, which is conducive to modeling text sequences with semantic and sentiment information, and has celebrated its sequence modeling ability. Softmax, as a classifier, is suitable for multi-category sentiment classification task, which can map the features extracted by BERT and BiGRU to different sentiment categories and realize the decision of sentiment classification, which greatly improves the classification ability of the model. As for multimodal feature fusion, by integrating text features extracted by BERT and sequence features captured by BiGRU, combined with image features, it can comprehensively utilize the information of text and image data to provide a more comprehensive multimodal feature representation. Optimizing and tuning the model and reasonably setting hyperparameters such as learning rate, batch size, dropout rate, etc., can improve the performance and generalization ability of the model through cross-validation or experimental adjustment.

In summary, the in-depth exploration of the model indicates that the combination of BERT, BiGRU, and Softmax leads to higher accuracy, avoids logical ambiguity in model validity, and enhances the understanding of model design and performance enhancement mechanisms.

When processing data and training models, attention also needs to be paid to the stability and performance capability of the research-designed models in the face of noisy or incomplete data, i.e., to ensure that the BERT BiGRU Softmax model is robust and noise-resistant, which implies the need to take a series of measures in the process of data processing, model design, and training, such as data cleansing, outlier handling, feature selection, regularization, integrated learning, etc. These methods are used to effectively reduce the noise and interference in the data, improve the generalization ability and applicability of the model, and thus ensure the robustness and reliability of the model in complex or noisy environments, as follows.

Step 1: Data cleaning and outlier handling: data cleaning is performed to identify and handle outliers and noise to reduce interference in the data. Robust statistical methods are used to deal with outliers, such as median instead of mean, or standardize the data using RobustScaler.

Step 2: Feature Selection as well as dimensionality reduction: select features that are robust and filter features that have a high impact on model predictions through feature

selection methods. Dimensionality reduction techniques such as principal component analysis (PCA) can be used to reduce data dimensionality and noise.

Step 3: Regularization and model complexity control: add regularization terms (e.g., L1, L2 regularization) to control model complexity, prevent overfitting, and improve noise immunity. Consider using simple models or integrated learning to reduce model complexity and enhance robustness.

Step 4: Integrated learning as well as model fusion: use integrated learning methods, such as random forests and gradient boosting trees, to combine the prediction results of multiple models and reduce the impact of noise on the model. Combining the prediction results of different models, model fusion is carried out through voting or weighted average to improve the robustness and noise resistance of the model.

Step 5: Cross-validation as well as model evaluation: use cross-validation techniques to evaluate the stability and generalization ability of the model and reduce the impact of noise on model performance. The average performance on different data subsets is considered in the model evaluation to improve the robustness and reliability of the model.

By combining the above methods and steps, it is possible to ensure that the model maintains stable performance and reliable prediction ability in the face of complex or noisy environments. The risk of data noise and model overfitting is greatly reduced, and the robustness and generalization ability of the model is improved.

The objective of sentiment analysis is to uncover the subjective emotional inclinations expressed by users toward products, as conveyed through online information. By leveraging deep learning techniques, sentiment analysis aims to establish connections between various features such as syntax, semantics, 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 toward goods.

5. Experiments and Analysis of Results

5.1. Experimental Environment

Jupyter Notebook, provided by Anaconda, is 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". 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 food rating dataset is presented in Table 2 below. "Very tasty", "affordable", and "good" have the highest scores. "Difficult to eat", "Inadequate", and "Speechless" had the lowest scores. 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.

Serial Number	Comment on the Content of the Message
55	Cake is very delicious, timely delivery, first-class service ah, the next opportunity to continue to buy.
119	Huge hard to eat, a salty and a sour.
268	Affordable price, good service attitude, the portion is super full, duck feet melt in the mouth, soft, very flavorful, super spicy, recommended!
529	Taste good, the portion is sufficient.
1146	Really speechless, the last two still think it can, today this pineapple bun head cream is stinky, completely inedible!
3785	The second time to eat, the taste is okay, just inside the Golden Harbor International, shopping tired can come to eat!
4216	Delivered 1 h, arrived at the things are cold.

The optimal ratio under the dataset of this study was determined by using a grid search method during the pre-experimental debugging process. Training, validation, and testing set splits typically follow 70–80% of the data for training, 10–15% for validation, and 10–15% for testing. It is important to ensure that the distribution of sentiment categories is similar across datasets to avoid sample bias. The number of epochs is 50, which determines the number of iterations to be performed on the training dataset, usually based on model convergence and computational resources, to ensure that the model has enough iterations to learn the patterns of the data and to be aware that too many epochs can lead to overfitting problems. The batch size of 128, the number of samples taken from the training data each time, affects the frequency of updating the model weights and the speed of convergence. Choosing the batch size appropriately helps the model to generalize better. A smaller batch size may result in a noisier training process, but the weights are updated more frequently, which helps it converge faster. Larger batch sizes may slow down the training but can take better advantage of GPU parallel computing to speed up the training process.

The problem of overfitting needs to be prevented in this process by early stopping, namely monitoring the model performance on the validation set and stopping training as soon as the performance no longer improves to avoid overfitting. Regularization, such as L1 or L2 regularization, is used to penalize the parameters of complex models to prevent overfitting. Dropout: Randomly dropping some neurons during the training process to reduce the complexity of the neural network and avoid overfitting the model to specific samples.

In this paper, we use a comparative model analysis approach to validate the classification effectiveness of the Bert BiGRU Softmax model designed in this paper for the multimodal sentiment classification task of line comment information.

When selecting a comparison model, consider the baseline model and current state-of-the-art methodology as a reference to ensure comparisons with industry levels. When configuring comparison models, harmonize hyperparameter settings and model architecture to ensure fair comparisons. In terms of dataset splitting, the same dataset should be used for training, validation, and testing, with a consistent data splitting strategy. In the model evaluation stage, the same evaluation metrics are selected for performance comparison, while statistical significance tests are performed to ensure reliable results. Finally, the results of the compared models are interpreted and analyzed to explore the reasons for the differences in the performance of the different models in order to fully assess the performance improvement of the proposed model. By following these steps, a fair and informative comparative analysis can be ensured to provide reliable assessment results for model development.

Based on the above requirements in this paper, the BiGRU model, CNN BiGRU model, BiGRU CNN model, BiGRU Attention model, Attention BiGRU CNN model, and Bert BiGRU Softmax model are selected for comparative analysis.

The BiGRU Attention model and Attention BiLSTM CNN model were selected 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 Algorithms	Batch File	Learning Rate	Dropout Rates
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	120	0.007	0.3
Current model	Bert BiGRU Softmax	Adam	128	0.006	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. It can be concluded from this that this paper's model performs optimally on the Chinese online review corpus.

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

The details of the configuration of the Bert BiGRU Softmax model are further analyzed: regarding the size of the model layers, we determine the layer size to be 200 based on the complexity of the dataset and the task; for the dropout rate, we determine the suffix rate to be 0.2 based on the proportion of neurons randomly disconnected during the training process to prevent overfitting; and in terms of the optimization algorithm, we choose the optimization algorithm for Adam due to its combination of the advantages of AdaGrad and RMSProp and its ability to adaptively adjust the learning rate. For the learning rate, we set the initial learning rate at 0.0006 based on the step size of the model each time the weights are updated during the training process.

Sentiment classification results for the CNN BiGRU model and Attention BiGRU CNN model are very similar, with Bert BiGRU Softmax outperforming the other five models, and the precision of the Attention BiGRU CNN model is 0.894, which is 1.5% lower than that of the Bert BiGRU Softmax model, while the F1 value is 0.4% lower than that of the Bert BiGRU Softmax model, and the recall is 0.1% lower than that of the Bert BiGRU Softmax model, as shown in Table 4.

The comparison of the different models revealed that the accuracy increased with the models' improvement. The comparison of the classification results of the simplest BiGRU model, the improved BiGRU Attention model, the Attention BiGRU CNN model, and the Bert BiGRU Softmax model revealed an increase in the accuracy of 6.2%, respectively, 3.4%, and 1.5%, 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 it 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.

Table 5 below shows the F1 values for each model for categorizing different aspects of sentiment. 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 overall 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.	1	0	0	1	0	1	1
190	The cake is very delicious, the delivery is timely, the service is first-class, 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

Table 6. *Cont.*

Classification Results								
Serial Number	**Comments**	**Taste**	**Price**	**Packaging**	**Service**	**Delivery Time**	**Hygiene**	**Environment**
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

Figures 6 and 7 show the categorization process and distribution of positive, neutral and negative ratings. After analyzing the data of consumer review information, the results show that about 44% of the review sentiment polarity is positive, 31% of the review sentiment polarity is negative, and 25% of the couple sentiment polarity is neutral. This percentage of sentiment polarity indicates that the majority of consumers were satisfied with the product or service and had a good consumer experience, while 31% of the consumers gave negative reviews. The number of negative reviews is small compared to that of positive reviews, but the small number of negative reviews provides a valuable reference for consumers when they are making choices and when businesses are making improvements in various areas. Positive reviews provide reference information for certain aspects that consumers value, for example, when consumers value the environment of the restaurant, the review information reflects the environmental information of the merchant and evaluates it, so that the weight of this review information in the decision-making process of consumers will be greater.

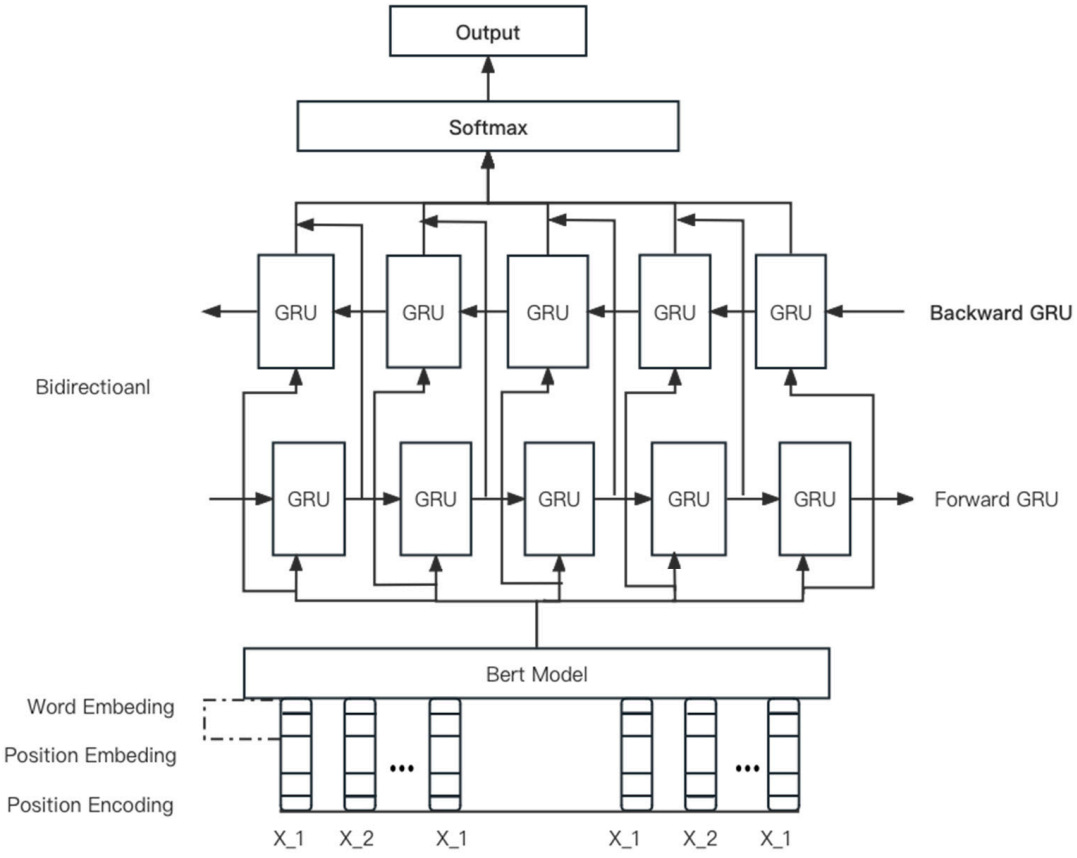

Figure 6. Bert BiGRU Softmax model.

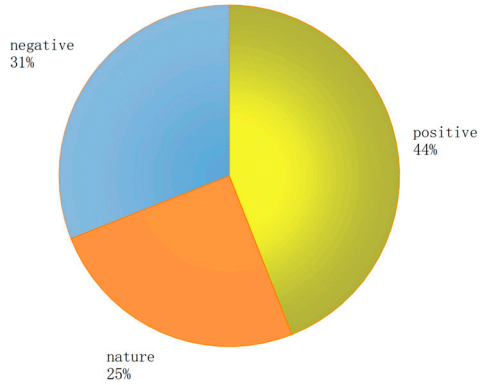

Figure 7. Distribution of the number of positive and negative evaluations.

Table 7 and Figure 8 refer to the final results of data processing. The data were categorized into precision, recall and F1 values for positive, neutral and negative attitudes, and the results from Table 7 and Figure 8 show that the precision of positive attitudes is higher than the precision of negative attitudes. Negative attitude has the highest three values and neutral attitude has the lowest three values.

Table 7. Classification results.

Metrics	Positive	Neutral	Negative
Precision	0.9432	0.9377	0.9812
Recall	0.9225	0.9162	0.9805
F1 score	0.9328	0.9274	0.9711

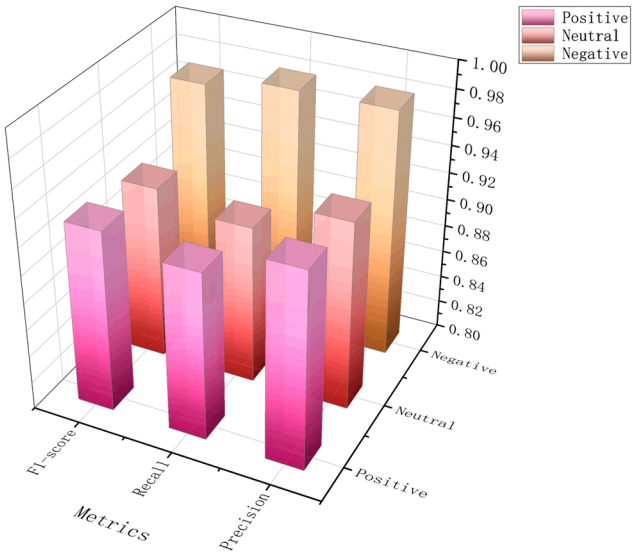

Figure 8. Different indicators obtained from the proposed model.

6. Discussion and Conclusions

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 sold on digital platforms. These evaluations encapsulate diverse dimensions including user encounters, product attributes, and satisfaction with services rendered. They serve as invaluable points of reference for both manufacturers and platform administrators. By 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 the 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 manuscript. Y.S.: optimized it. Y.Z.: methodology. C.L.: collected part of the data. All authors have read and agreed to the published version of 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 Statement: The dataset and code used in this article are available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Kiran, R.; Kumar, P.; Bhasker, B. Oslcfit (organic simultaneous LSTM and CNN Fit): A novel deep learning based solution for sentiment polarity classification of reviews. *Expert Syst. Appl.* **2020**, *157*, 113488. [CrossRef]
2. Vijayaragavan, P.; Ponnusamy, R.; Aramudhan, M. An optimal support vector machine based classification model for sentimental analysis of online product reviews. *Future Gener. Comput. Syst.* **2020**, *111*, 234–240. [CrossRef]
3. Shuang, K.; Yang, Q.; Loo, J.; Li, R.; Gu, M. Feature distillation network for aspect-based sentiment analysis. *Inf. Fusion* **2020**, *61*, 13–23. [CrossRef]
4. Kaur, G.; Sharma, A. A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. *J. Big Data* **2023**, *10*, 5. [CrossRef] [PubMed]
5. Yang, Y.T.; Wang, M.Y.; Tian, X. Sina Microblog Sentiment Classification Based on Distributed Representation of Documents. *J. Intell.* **2016**, *35*, 151–156.
6. Shao, Y.F.; Liu, D.S. Classifying Short-texts with Class Feature Extension. *Data Anal. Knowl. Discov.* **2019**, *3*, 60–67.
7. Qu, Z.; Wang, Y.; Wang, X. A Hierarchical attention network sentiment classification algorithm based on transfer learning. *J. Comput. Appl.* **2018**, *38*, 3053–3056.
8. Tao, Z.; Li, X.; Liu, Y.; Liu, X. Classifying Short Texts with Improved-Attention Based Bidirectional Long Memory Network. *Data Anal. Knowl. Discov.* **2019**, *3*, 21–29.
9. Sivakumar, S.; Rajalakshmi, R. Context-aware sentiment analysis with attention-enhanced features from bidirectional transformers. *Soc. Netw. Anal. Min.* **2022**, *12*, 104. [CrossRef]
10. Pang, B.; Lee, L.; Vaithyanathan, S. Thumbs up? Sentiment classification using machine learning techniques. *arXiv* **2002**, arXiv:cs/0205070.
11. Tripathy, A.; Agrawal, A.; Rath, S.K. Classification of sentiment reviews using n-gram machine learning approach. *Expert Syst. Appl.* **2016**, *57*, 117–126. [CrossRef]
12. Mekel, D.; Frasincar, F. ALDONA: A hybrid solution for sentence-level aspect-based sentiment analysis using a lexicalised domain ontology and a neural attention model. In Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, Limassol, Cyprus, 8–12 April 2019.
13. Thet, T.T.; Na, J.C.; Khoo, C.S.G. Aspect-based sentiment analysis of movie reviews on discussion boards. *J. Inf. Sci.* **2010**, *36*, 823–848. [CrossRef]
14. Ain, Q.T.; Ali, M.; Riaz, A.; Noureen, A.; Kamran, M.; Hayat, B.; Rehman, A. Sentiment Analysis Using Deep Learning Techniques: A Review. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 424–433.
15. Liu, H.; Chatterjee, I.; Zhou, M.; Lu, X.S.; Abusorrah, A. Aspect-Based Sentiment Analysis: A Survey of Deep Learning Methods. *IEEE Trans. Comput. Soc. Syst.* **2020**, *7*, 1358–1375. [CrossRef]
16. Wang, H.T.; Song, W.; Wang, H. Text classification method based on hybrid model of LSTM and CNN. *Small Microcomput. Syst.* **2020**, *41*, 1163–1168.
17. Wu, P.; Ying, Y.; Shen, S. Research on the classification of netizens' negative emotions based on bidirectional long and short-term memory model. *J. Intell.* **2018**, *37*, 845–853.
18. Zhang, H.T.; Wang, D.; Xu, H.L.; Sun, S.Y. Research on microblog opinion sentiment classification based on convolutional neural network. *J. Intell.* **2018**, *37*, 695–702.
19. Wang, S.; Chen, Y.; Yi, Z. A Multi-Scale Attention Fusion Network for Retinal Vessel Segmentation. *Appl. Sci.* **2024**, *14*, 2955. [CrossRef]
20. Xu, S.K.; Zhou, Z.Y. Sentiment classification model and application of WeChat tweets based on multi-scale BiLSTM-CNN. *Intell. Sci.* **2021**, *39*, 130–137.
21. 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. *Intell. Sci.* **2021**, *39*, 15–22.
22. Cho, K.; van Merrienboer, B.; Bahdanau, D. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv* **2014**, arXiv:1409.1259.
23. Zhao, H.; Liu, Z.; Yao, X.; Yang, Q. A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach. *Inf. Process. Manag.* **2021**, *58*, 102656. [CrossRef]
24. Kim, S.; Hovy, E. Determining the sentiment of opinions. In *Proceedings of the COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics, Geneva, Switzerland, 23–27 August 2004*; pp. 1367–1373.
25. 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. *Comput. Appl. Softw.* **2019**, *36*, 220–226.
26. Xu, F.; Pan, Z.; Xia, R. E-commerce product review sentiment classification based on a nave Bayes continuous learning framework. *Inf. Process. Manag.* **2020**, *57*, 102221. [CrossRef]
27. Duan, D.D.; Tang, J.S.; Wen, Y.; Yuan, K.H. Chinese Short Text Classification Algorithm Based on BERT Model. *Comput. Eng.* **2021**, *47*, 79–86.
28. Guerini, M.; Gatti, L.; Turchi, M. Sentiment analysis: How to derive prior polarities from SentiWordNet. *arXiv* **2013**, arXiv:1309.5843.
29. Zheng, F.; Wei, D.; Huang, S. Text Classification Method Based on LDA and Deep Learning. *Comput. Eng. Des.* **2020**, *41*, 2184–2189.

30. Go, A.; Bhayani, R.; Huang, L. *Twitter Sentiment Classification Using Distant Supervision*; CS224N Project Report; Stanford University: Stanford, CA, USA, 2009; pp. 1–12.
31. Kim, Y. Convolutional neural networks for sentence classification. *arXiv* **2014**, arXiv:1408.5882.
32. Wu, Z.H.; Chen, Y.J. Genetic Algorithm Based Selective Neural Network Ensemble. In Proceedings of the 17th International Joint Conference on Artificial Intelligence, Seattle, WA, USA, 4–10 August 2001; Volume 2, pp. 797–802.
33. 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*; Springer: Singapore, 2015; pp. 489–495.
34. Du, L.; Cao, D.; Lin, S.Y.; Qu, Y.Q.; Ye, H. Extraction and automatic classification of Chinese medical records based on BERT and Bi-LSTM fusion attention mechanism. *Comput. Sci.* **2020**, *47*, 416–420.
35. 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. *Libr. Intell. Work* **2019**, *63*, 99–108.
36. Ko, A.R.; Sabourin, R.; de Souza Britto, A. Combining Diversity and Classification Accuracy for Ensemble Selection in Random Subspaces. In Proceedings of the 2006 IEEE International Joint Conference on Neural Network Proceedings, Vancouver, BC, Canada, 16–21 July 2006; pp. 2144–2151.
37. Trofimovich, J. Comparison of neural network architectures for sentiment analysis of Russian tweets. In Proceedings of the Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2016", Moscow, Russia, 1–4 June 2016.
38. Tsai, C.Y.; Chen, C.J. A PSO-AB Classifier for Solving Sequence Classification Problems. *Appl. Soft Comput.* **2015**, *27*, 11–27. [CrossRef]
39. Lahat, D.; Adali, T.; Jutten, C. Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects. *Proc. IEEE* **2015**, *103*, 1449–1477. [CrossRef]
40. Zhang, Q.; Gao, Z.M.; Liu, J.Y. Research on short text categorization of microblogs based on Word2vec. *Inf. Netw. Secur.* **2017**, *6*, 57–62.
41. Zhao, L. Research on Multimodal Data Fusion Algorithm. Master's Thesis, Dalian University of Technology, Dalian, China, 2018.
42. Qian, Q.; Huang, M.; Lei, J.; Zhu, X. Linguistically regularized LSTMs for sentiment classification. *arXiv* **2016**, arXiv:1611.03949.
43. Ge, X.W.; Li, K.X.; Cheng, M. Text Classification of Nursing Adverse Events Based on CNN-SVM. *Comput. Eng. Sci.* **2020**, *42*, 161–166.
44. 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. *Ecol. Mag.* **2019**, *38*, 2236–2242.
45. Guo, L.J.; Peng, X.; Li, Z.H.; Zhang, M. A Chinese language dependent syntactic treebank for multi-domain and multi-source text Syntactic Tree Library Construction for Multi-Domain and Multi-Source Texts. *J. Chin. Lang. Inf.* **2019**, *33*, 34–42.
46. Zheng, Y. Methodologies for cross-domain data fusion: An overview. *IEEE Trans. Big Data* **2015**, *1*, 16–34. [CrossRef]
47. Kennedy, J.; Eberhart, R.C. A discrete binary version of the particle swarm algorithm. In Proceedings of the 1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation, Orlando, FL, USA, 12–15 October 1997; Volume 5, pp. 4104–4108.
48. Chandra, A.; Chen, H.; Yao, X. Trade-off between Diversity and Accuracy in Ensemble Generation. In *Multi-Objective Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 429–464.
49. Cai, Q.P.; Ma, H.Q. A Fine-grained Sentiment Analysis Model for Product Reviews Based on Word2Vec and CNN. *Libr. Intell. Work* **2020**, *64*, 49–58.
50. Basiri, M.E.; Nemati, S.; Abdar, M.; Cambria, E.; Acharya, U.R. ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model for sentiment analysis. *Future Gener. Comput. Syst.* **2021**, *115*, 279–294. [CrossRef]
51. 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. [CrossRef]