applied sciences

MDPI

Article

AgriMFLN: Mixing Features LSTM Networks for Sentiment Analysis of Agricultural Product Reviews

Runmeng Liu [1,2], Haiyan Wang [1,2,*] and Yuanxi Li [1,2]

[1] School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China
[2] Engineering Research Center for Forestry-Oriented Intelligent Information Processing, National Forestry and Grassland Administration, Beijing 100083, China
* Correspondence: wanghaiyan2008@bjfu.edu.cn

Abstract: With the rapid development of the Internet, agricultural products have entered e-commerce platforms, and agricultural product reviews have become an important reference for consumers when purchasing agricultural products. However, due to the characteristics of different lengths, rich context-sensitive information, and multi-level information in the sentences of agricultural product reviews, the existing sentiment analysis methods cannot perform well enough to identify the sentiment tendency. To address this issue, we abstract the problem as a binary classification task to extract consumers' sentiment orientation by proposing a new method. This method utilizes an attention mechanism to assign different weights to different key information in the sentence, thereby extracting abundant semantic information from the sentence. The design of the long short-term memory (LSTM) gate can effectively solve the problem of extracting long sequences and context-related information. The proposed model achieves superior results on two agricultural product datasets compared to other baseline models, providing guidance for merchants to improve agricultural product quality and enhance customer satisfaction.

Keywords: sentiment analysis; agricultural product reviews; machine learning; LSTM; attention

check for updates

Citation: Liu, R.; Wang, H.; Li, Y. AgriMFLN: Mixing Features LSTM Networks for Sentiment Analysis of Agricultural Product Reviews. *Appl. Sci.* **2023**, *13*, 6262. https://doi.org/10.3390/app13106262

Academic Editor: José Salvador Sánchez Garreta

Received: 14 March 2023
Revised: 26 April 2023
Accepted: 16 May 2023
Published: 20 May 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the continuous promotion of agricultural industrialization, high-quality agricultural products are seeking a broader market. The traditional offline marketing of agricultural products has drawbacks, which means many characteristic agricultural products are limited and unable to enter a bigger market, resulting in a serious disconnect between production and sales. As the e-commerce industry continues to develop rapidly, it has opened new opportunities for the agricultural sector, enabling farmers to sell their products through online platforms. Consequently, an increasing number of consumers are opting to purchase agricultural goods online, and they can purchase the needed goods without leaving their homes owing to the wide variety and kinds of products available on e-commerce platforms [1]. When buying agricultural products online, consumers rely heavily on product reviews as a key reference. Unfortunately, there may be problems in the sales process of agricultural products, such as inconsistencies between descriptive information and actual agricultural goods, poor quality, inadequate after-sales service, and so on [2]. Hence, it is crucial to analyze these reviews accurately to provide consumers with trustworthy information when making purchasing decisions.

Sentiment analysis, sometimes referred to as text orientation analysis, or opinion mining, is the method of automatically determining the customer's emotional tendency from the given text [3]. Recently, sentiment analysis has gained primary attention in natural language processing (NLP) and text mining, covering various fields such as general products, politics, and society including movies, tourism, restaurants, and hotels [4–10]. Similarly, research in the field of agriculture should also take the importance of big data analysis into account, but the study of sentiment analysis in this field is still in its early stages.

Sentiment analysis for agricultural product reviews has important practical significance and plays a crucial role in the production, processing, sales, marketing, and consumer decision-making. They can help to explore consumers' attitudes and emotions toward agricultural products, optimize the quality and taste of agricultural products, improve the efficiency of agricultural product sales and circulation, and enhance the brand image and reputation of agricultural products.

Given the natural, diverse, and seasonal characteristics of agricultural products, and their regional and dispersed production, the review information on agricultural products has strong contextual relevance and complex language, making it difficult to assess sentiment tendencies. Due to the diversity of products and the complex language used in reviews, conducting sentiment analysis of agricultural product reviews is a complex task. The following difficulties need to be overcome in order to analyze the sentiment content of agricultural product reviews successfully:

1. Agricultural product reviews vary in length and often contain information regarding the origin, variety, and nutritional value of the agricultural products. This context is closely related to the sentiment tendency, making it a challenging task.
2. Agricultural product reviews often use more complex language to express emotions, such as the use of modifiers and metaphors. These language structures can be difficult to handle during the process of sentiment analysis.

To deal with the above challenges, we have proposed Mixing Features LSTM Networks for Sentiment Analysis of Agricultural Product Reviews (AgriMFLN).

1. For challenge one, we designed the LSTM layer to capture the sequence features of sentences and then conducted sentiment analysis. By enabling the model to capture the sequential nature of language and extract the emotional content from the input, the LSTM layer played a critical role in sentiment analysis of agricultural product reviews.
2. For challenge two, multi-head attention strategies were implemented to extract the semantic content of phrases. Complex semantic information can be extracted by fusing the vectors produced by the multi-head attention with the vectors produced by the embedding layer.

In conclusion, the contributions of this paper can be summarized as follows:

1. We have developed a model named 'AgriMFLN' for sentiment analysis of agricultural product reviews. This integrated deep learning model incorporates multi-level semantic information, timing sequencing information, and contextual information of sentences to provide a comprehensive analysis.
2. The model is capable of handling long series of agricultural product reviews effectively. The design of LSTM gates allows for the retention of previous information over an extended period, thereby increasing the neural network's long-term memory ability to better identify sentiment tendencies.
3. The multi-head attention strategy is able to learn the representation of text at different levels, resulting in a more effective representation of semantic information.
4. The proposed method has demonstrated superior performance compared to other baseline models when applied to two real datasets of agricultural product reviews.

2. Literature Review

The computational study of people's views, feelings, emotions, assessments, and attitudes toward entities including goods, services, organizations, issues, topics, events, and their attributes is known as sentiment analysis or opinion mining [11]. The term sentiment analysis was proposed by Nasukawa and Yi [12], but the task of sentiment analysis was first proposed by Pang and Lee [13].

Since the early 2000s, sentiment analysis has developed into one of the most active study topics in NLP [14]. Text sentiment analysis research is now being studied in a wide range of domains, including NLP, information extraction, information retrieval, text mining, ontology, and machine learning. It has become one of the top concerns in the field of NLP

during the past few years, with the discussed methods including the lexicon-based method and the machine learning method. The former is an unsupervised classification method, relying on the sentiment lexicon and a sentiment polarity score to achieve the calculation of sentence sentiment score to recognize emotional tendencies.

Wu et al. [15] proposed a method of building and expending a computer terminology dictionary based on the semantic dictionary WordNet. The advantage of this method is that the semantic structure of WordNet was preserved, and the automatic extension was constructed without manual intervention, which greatly reduced the workload. An approach to building a domain emotion dictionary was put out by Han et al. [16], which included mutual information to assign terms with part-of-speech (POS) tags in the lexicon. The method significantly improved the recognition performance of negative reviews. To help in sentiment analysis of social media content, Wu et al. [17] created the first sentiment dictionary of slang words, named 'SlangSD'. 'SlangSD' can easily be utilized as an extra emotion lexicon and can effectively improve the performance of even the most sophisticated casual text. Zargari et al. [18] suggested a hybrid approach for training a semi-manual sentiment dictionary. This method solved the heterogeneous effect of multiple intensifiers, increased the effectiveness of existing dictionaries in document polarity recognition, and took the relationship between emotion words and multiple intensifiers into account. Li et al. [19] investigated changes in Chinese public firms' attitudes toward environmental protection between 2018 and 2021 by using a sentiment analysis method based on a sentiment dictionary. The study examined the relationship between firms' carbon reduction attitudes and financial performance, finding that specific policies can increase positive attitudes toward carbon reduction, and attitudes toward ecological topics differ across industries.

Although prior knowledge has a certain reference value and significance for the methods based on sentiment dictionaries, it is dependent on the domain of the corpus, which is easily affected by the sentence structure of the text in the corpus, and the classification effect depends on the quality of the sentiment dictionary. This leads to instability in the sentiment analysis method, and the sentiment dictionary consumes huge resources for long-term maintenance. With the rapid development of modern networks, network data are growing at an unprecedented rate, so new words and expressions emerge one after another, and more and more old words with new meanings emerge, which makes it very difficult to maintain the sentiment dictionary to keep up with the pace of network updates.

To solve the shortcomings mentioned above, scholars have proposed a series of sentiment analysis methods based on machine learning algorithms, including classical machine learning methods and deep learning methods. The method based on deep learning has more advantages than the lexicon-based method mentioned above. Firstly, it achieves the goal of sentiment analysis by designing an algorithm training model. Therefore, it does not need to build a sentiment dictionary and manually extract semantic features, which saves manpower as well as material resources and does not require long-term maintenance. Secondly, the model adopts automatic extraction of sample features, which improves the scalability of the model and has a better effect than the lexicon-based method.

Nowadays, researchers are using more diversified and multi-level deep learning algorithms, and they are making the corresponding adjustments to the algorithm details such as the connection method and activation function of neurons. Deep learning first appeared in computer speech recognition and visual recognition and then was used in the field of NLP. It has become a popular technology in NLP and is now widely used. Many scholars have used the study of neural networks to successfully complete the sentiment analysis assignment.

Bernhard et al. [20] figured that classification performance can be improved through deep learning. They found that repeated neural networks and transfer learning are always better than classical machine learning. Additionally, they suggested utilizing transfer learning as an inductive method of transferring knowledge from related NLP tasks. Kim [21] proposed applying convolutional neural networks (CNN) to text sentiment analysis. The model used the Word2Vec model to convert text into word vector form, and then input the

word vector as one-dimensional data into CNN for classification learning using convolution kernels of different volumes.

An LSTM network is a unique kind of recurrent neural network (RNN) that is effective in dealing with lengthy sequences of data and discovering long-term dependencies. Zhao et al. [22] proposed a CNN-LSTM model applied to personality prediction, which used CNN to extract the spatial features and used the LSTM network for further extraction. The model paid more attention to emotion-related words or their corresponding target information during training. To accelerate the training of the model, six different parameter-reduced slim versions of the LSTM model were suggested by Gopalakrishnan et al. [23]. These variant models can be selected appropriately to reduce computational costs. Teng et al. [24] proposed a multi-dimensional topic classification model based on LSTM. The model constituted hierarchical multi-dimensional sequence computations and was composed of LSTM cell networks. The experimental findings revealed that the average precision of the proposed model ranged from 91% to 96.5%. Tang et al. [25] proposed a target-dependent long short-term memory (TD-LSTM) model, which used two LSTMs to model the semantic relationship between target words and their contextual words. Experimental results showed that incorporating target information into the LSTM can significantly improve classification accuracy and achieve better performance. Jelodar et al. [26] used the LSTM network to classify COVID-19-related comments. Their findings emphasized the significance of using the general public's opinions and relevant computational tools to comprehend COVID-19-related issues and guide related decision-making. Bhandari et al. [27] adopted a different approach to processing COVID-19-related comments. They proved the significance of stacked word embeddings in the deep learning model. The findings of the experiment also demonstrated the superior performance of stacked word embeddings in coping with the distinctive contextual semantic understanding from small tweets and the unbalancedness of the experimental dataset. Kai et al. [28] proposed a stacking model named word attention-based Bi-directional long short-term memory (Bi-LSTM) and CNN stacking model to analyze the emotional tendency of fans in regard to films. After the Bi-LSTM layer, the attention layer was included to give text keywords more weight so that the model could extract keywords more effectively. Experiments showed that the classification accuracy and recall were improved. Wang et al. [29] proposed a tree-structured regional CNN-LSTM model. This model can take both local information and long-term tendencies into consideration and achieve good classification results on different corpora. Chen et al. [30] proposed a mode combining CNN and regional LSTM(CNN-RLSTM). The results from various datasets demonstrated that the CNN-RLSTM model was better than the traditional model and the deep network model. Gao et al. [31] proposed a short-text aspect-based sentiment analysis method based on CNN and bidirectional gating recurrent unit (Bi-GRU). The model was more conducive to the extraction of deep features of the text to obtain a more complete text feature vector. The results showed that the classification effect can be improved, and the operation time can be reduced. Jiang et al. [32] proposed an aspect-based LSTM-CNN attention model for aspect-level sentiment classification. The model used CNN to identify local features and the attention mechanism to focus on the aspect information. The results showed that the model could effectively improve the accuracy of aspect-level sentiment classification. The combination of explicit knowledge from the external database with the implicit data in the LSTM model was highlighted by Ramaswamy et al. [33]. Their RecogNet-LSTM+CNN model with attention mechanism showed superior performance in aspect categorization and opinion classification. In 2021, Shobana and Murali [34] proposed an adaptive particle swarm optimization algorithm based on long short-term memory networks. They combined an opposition-based learning method with a particle swarm optimization algorithm to enhance weight parameters. As a result, the performance of the LSTM was improved. Kumar et al. [35] used a variety of embedding methods to address the multi-classification tasks in healthcare, such as determining from a patient's historical clinical records whether a specific health condition exists. Following a comparison of the results obtained using

classic machine learning and deep learning techniques, they also applied ensemble learning strategies to enhance the performance of the single model. Their experiments provided valuable references for the research of machine learning and deep learning in solving multi-classification tasks.

In 2017, the Google team [36] proposed a Transformer model for dealing with text sequence problems based on the attention mechanism, which completely abandoned the network structure of CNN and RNN. Compared with the traditional model, this model effectively reduced the calculation of the model and improved the efficiency of model parallelism. Ma et al. [37] proposed an interactive attention network (IAN) to handle both target and contextual semantic information in a special way. The model utilized an interactive attention mechanism to learn attention weights for both context and target and generated semantic representations for both. Through the design of this model, the IAN model could effectively represent the target and its contextual information, which was beneficial for fine-grained sentiment classification. In 2018, Radford et al. [38] proposed the generative pre-training (GPT) model. This model achieved language knowledge by pre-training on a diverse corpus and then could be transferred to solving specific tasks, such as semantic similarity assessment, question answering, and text classification. Devlin et al. [39] improved the Transformer model and proposed the bidirectional encoder representations from transformer (BERT) pre-training model. BERT was designed to pre-train deep bidirectional representations from unlabeled text, and it can be fine-tuned to deal with a wide range of tasks. Tan et al. [40] proposed a hybrid model, which integrated the robustly optimized BERT approach and LSTM for sentiment analysis. The robustly optimized BERT approach mapped the words into the meaningful word embedding space, while the LSTM model captured the long-distance contextual semantics effectively. Gao et al. [41] proposed a target-dependent BERT (TD-BERT) model in an aspect-level sentiment classification task, which implemented three target-dependent variants of the BERT model. Compared to traditional feature engineering methods, embedding-based models, and early BERT applications, this model achieved better performance. Zhang et al. [42] proposed a broad multitask transformer network (BMT-Net), which took advantage of both feature-based and fine-tuning methods. BMT-Net can learn robust contextual representation and achieve good results. To alleviate the reliance on annotated data, Gong et al. [43] proposed a text sentiment classification model based on the Transformer mechanism that combined knowledge distillation and text augmentation methods. They reduced the number of parameters based on the Transformer mechanism and solved the problem of low accuracy for the few-sample tasks. Lin et al. [44] proposed a multi-head self-attention transformation networks (MSAT) network for aspect-based sentiment tasks. The model used a multi-head target-specific self-attention mechanism to better capture global dependencies and introduced target-sensitive transformation to effectively address the issue of target-sensitive sentiment. Then, part-of-speech features were integrated into MSAT to capture the syntactic features of the sentence. The experiments showed that the proposed model achieved better effectiveness compared with several state-of-the-art methods.

In conclusion, scholars in different fields have constructed and improved classification models using both emotion lexicon-based and machine-based methods, to suit various application scenarios. The method based on sentiment lexicons requires the use of pre-constructed lexicons that contain different vocabulary for positive, negative, and neutral sentiment, which can be used to determine the sentiment orientation of the given text. The advantage of this method is that it is easy to understand and interpret, as it directly interprets specific vocabulary in the text. However, the method is limited by the quality of the sentiment lexicon and the application domain, as the lexicon may not include all sentiment words and cannot capture the context of the words, which may lead to errors. In addition, sentiment lexicons need to be updated regularly to reflect new vocabulary and new text data, which requires human and material resources.

In contrast, the deep learning-based sentiment analysis method can extract different features in the text, including words, phrases, syntactic structures, context, and contex-

tual information. This method does not require predefined sentiment lexicons, and it can extract sample features and automatically optimize and adjust the model during training, improving the model's scalability and having better performance than the sentiment dictionary-based approach. However, the deep learning-based approach also has drawbacks, such as taking a long time to train, requiring a large amount of training data for learning, and the difficulty of classification tasks when the training dataset is insufficient.

This paper draws inspiration from the success of multi-head attention in sentiment analysis, which has been proven effective in extracting information at multiple levels. Specifically, we have proposed a method seeking to improve upon the classical LSTM network by incorporating feature mixing techniques. The goal is to enable the model to capture a broader range of information and improve the accuracy of the evaluation process. By abstracting the agricultural product review evaluation as a sentiment analysis task and leveraging recent advancements in deep learning, the proposed approach represents a promising avenue for improving the efficiency and accuracy of agricultural product review evaluations.

3. Materials and Methods

To explain the proposed approach and procedure of this study, we have introduced each component of the model along with the explanation of the dataset and the data pretreatment steps in this section.

3.1. Problem Statement

In this manuscript, we aim to extract consumers' sentiment orientation from agricultural product reviews. We abstract the problem as a binary classification task, where consumers' sentiment toward agricultural products is classified as either positive or negative. Considering that agricultural product reviews vary in length and consumers may use complex language to express their sentiment orientation, we have proposed Mixing Features LSTM Networks to address this issue and evaluate it on two agricultural product datasets.

3.2. Experimental Datasets

This experiment was developed with the Baidu Paddle framework and trained with the GPU provided by Baidu AIStudio. The experimental datasets consist of two parts: one is from a public dataset obtained from GitHub, and the other is a private dataset obtained by scraping real agricultural product reviews from an e-commerce platform. Both datasets are Chinese, as explained in English in this paper.

3.2.1. The Public Dataset

The public dataset named 'online_shopping_10_cats' [45] is provided on GitHub. The public dataset consists of 62,774 comments, including 31,728 positive comments and 31,046 negative comments, and 10 categories, such as fruits, shampoo, tablets, books, mobile phones, etc. To achieve the sentiment analysis of agricultural products, this experiment extracted comments with the category of fruit from the dataset, amounting to 10,000. Among them, there are 5000 positive comments and 5000 negative comments. The field description is shown in Table 1.

Table 1. Field description.

Field	Description
Cat	Category: including tablets, clothes, computers, mobile phones, books, fruits, water heaters, shampoo, milk, hotels
Label	Emotional tendency, 1 means positive comment, 0 means negative comment
Review	Comments

3.2.2. The Private Dataset

Considering the limited data volume of the public dataset, we utilized web crawling techniques to automatically obtain review data from e-commerce platforms named 'Jing-Dong' and constructed our own private dataset. We first specified search keywords and filtering conditions to ensure that the obtained review data were relevant to the agricultural products, and then removed the reviews which were too short or invalid. Finally, we annotated the sentiment polarity of the product reviews. The product reviews obtained through web crawling have the characteristics of large data volume, high authenticity, and high timeliness, laying the foundation for the subsequent experiments. We obtained 14,222 product reviews, including 10,013 positive reviews and 4209 negative reviews.

Two datasets were used, respectively, for training and conducting testing. These comments were randomly divided into a training set and a test set, respectively, which are independent of each other, following a ratio of 8:2. The training sets were used to build the classification model, and then the model was used to test the data of the test set to evaluate the quality of the model. The dataset sample is shown in Table 2.

Table 2. Examples of comments.

Category	Comments	Label
Fruit	It's not fresh. When I received it, some parts were rotten, and the interior was also rotten. I don't recommend buying.	0
Fruit	Very delicious, very crispy, very sweet. I bought two boxes, my child likes it very much, the express is very fast, arrived at the same day, without any bad inside. Great, and I will come again.	1
Fruit	Good value for money. The fruit is not damaged and fresh. It can't be separated from the conscientiousness and responsibility of the delivery master.	1
Fruit	Apple is OK, sweet, crisp and tasty. Buy after eating. JD Express service is good.	1
Fruit	The actual product is very small, and the taste is not good. The color of the actual product is much worse than that published by the merchants, and the fruit shape is also bad.	0

Observing two datasets, we noticed that the length of agricultural product context text varies greatly. Some product reviews may simply provide a brief evaluation of the product, while others may have longer comments where consumers describe their purchasing experience with the agricultural product from multiple perspectives. These may include the fruit quality, the speed of delivery, the quality of service provided by the merchant, and the price of the agricultural product. These dimensions may all affect the emotional tendencies expressed by consumers in their comments.

Additionally, we found that complex language is frequently used in agricultural product reviews to express emotions. The context is strongly related to the product, and consumers may not directly say that the product is poor, but rather use modifiers and metaphors to describe its quality. For example, the sentence "I don't know if what I bought is jujube or pear. Can this kind of business last? Don't tell me that the Xinjiang pears are supposed to be this big. Luckily, I'm not a first-time buyer or eater. This is just outrageous" is a description of poor fruit taste and quality.

3.3. Research Methods

The AgriMFLN model proposed in this paper is advanced based on the LSTM model with the multi-head attention mechanism of the Transformer model. Since the Transformer model can focus more on the important information by using the self-attention mechanism,

the vector after embedding the text is merged with the vector generated by the Transformer, and then the fused vector is input to the LSTM layer.

3.3.1. LSTM Layer

LSTM [46] is consistent with the typical RNN framework, but it calculates the hidden state using distinct methods. LSTM can learn long-term dependence, solve the problem of gradient explosion and gradient disappearance in the process of long sequence training, and is suitable for classification and prediction tasks based on time series data, with better performance in longer sequences.

The LSTM model consists of a series of repeated memory units, each of which contains three gates with different functions. Each block includes one or more self-connected memory cells and three multiplicative units—the input, output, and forget gates—that continuously analogize the actions of writing, reading, and resetting the cells [47].

The structure of such a unit is shown in the following Figure 1.

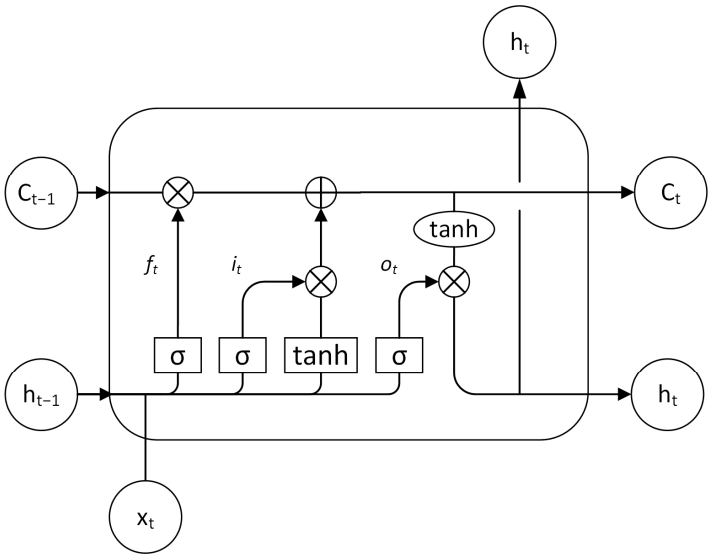

Figure 1. The structure of the LSTM unit.

The specific calculation formula is as follows:

The f_t is the forget gate:

$$f_t = \sigma\left(W^f x_t + U^f h_{t-1} + b_f\right) \tag{1}$$

The i_t is the input gate:

$$i_t = \sigma\left(W^i x_t + U^i h_{t-1} + b_i\right) \tag{2}$$

The C_t represents the state value of the current time in the memory cell:

$$C_t = f_t \times C_{t-1} + i_t \times \widetilde{C}_t \tag{3}$$

The o_t is the output gate:

$$o_t = \sigma(W^o x_t + U^o h_{t-1} + b_o) \tag{4}$$

The $\widetilde{C}_t$ represents the candidate memory cell status at the current time step:

$$\widetilde{C}_t = \tan \mathrm{h}(W^n x_t + U^n h_{t-1}) \tag{5}$$

h_t represents the hidden layer state at time t:

$$h_t = o_t \times \tan h(C_t) \tag{6}$$

W and U represent the weight matrix, b represents the deviation vector, σ represents the sigmoid activation function, and x_t represents the input vector at a time.

3.3.2. Transformer

Transformer is a machine learning model based on the attention mechanism, which was proposed in 2017. The Transformer model discards the sequential structure of traditional RNN and uses a mechanism based on self-attention, which has better parallel computing capability.

The encoder and decoder, respectively, depicted in the left and right halves of Figure 2, are implemented by the Transformer employing stacked self-attention and pointwise, fully connected layers. In this experiment, we only use the encoder module of the Transformer model.

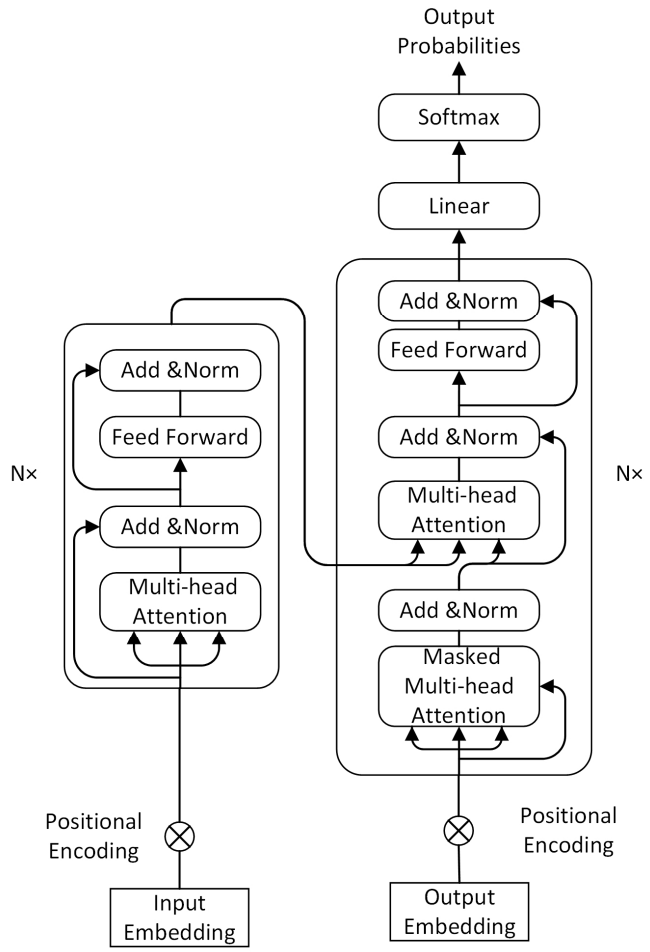

Figure 2. Structure of the Transformer.

First, the input text sequence is embedded to learn the feature of the text, similarly to other sequence transduction models. Since the original Transformer model is unable to capture sequential information, sine and cosine functions are used to add positional information, which is shown in Formulas (7) and (8). This method can not only obtain the

absolute position information of words but also the relative position information, which can help the text to better obtain the order of words.

$$PE_{(pos,2i)} = \sin\left(pos/10,000^{2i/d_{model}}\right) \tag{7}$$

$$PE_{(pos,2i+1)} = \cos\left(pos/10,000^{2i/d_{model}}\right) \tag{8}$$

The term *pos* refers to the position of the current character in the sentence and i refers to the dimension.

The core of the Transformer model is the attention mechanism, which can be described as mapping a Query and a set of Key-Value pairs to an output, where the Query, Keys, Values, and output are all vectors. The output is calculated as a weighted sum of the values, with each Value's weight determined by the Query's compatibility function with its corresponding Key, according to Formulas (9)–(12).

$$Q = XW_Q \tag{9}$$

$$K = XW_K \tag{10}$$

$$V = XW_V \tag{11}$$

$$X_{attention} = selfAttention(Q, K, V) \tag{12}$$

Scaled dot-product attention is implemented to calculate the self-attention score, according to Formula (13).

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{13}$$

Then, the multi-head attention mechanism connects multiple self-attention mechanisms to enable the model to learn related information in different subspaces, according to Formula (14). Through different Q, K, and V matrices, the model's ability to focus on different positions is increased by the multiple attention mechanism, which also provides the self-attention layer with several "representation subspaces", improving the ability to learn different meanings of words. Compressing multiple Q, K, and V matrices into a weight matrix can obtain the output of the multi-head attention mechanism, that is, the input of the feedforward network.

$$MultiHead(Q, K, V) = Concat(head_1, \ldots, head_h)W^O \tag{14}$$

where $head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$.

The encoder and decoder block also contain a fully connected feed-forward network, according to Formula (15).

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \tag{15}$$

In Transformer, data passing through the attention layer and *FFN* layer will be processed by an *Add & Norm*. The *Add & Norm* process refers to a step where a residual connection is added to the output of a sub-layer, followed by layer normalization. This process helps in preserving information from the original input while mitigating the vanishing gradient problem and improving training efficiency.

3.3.3. The Proposed Approach

The structure diagram of the model proposed in this paper is shown in Figure 3.

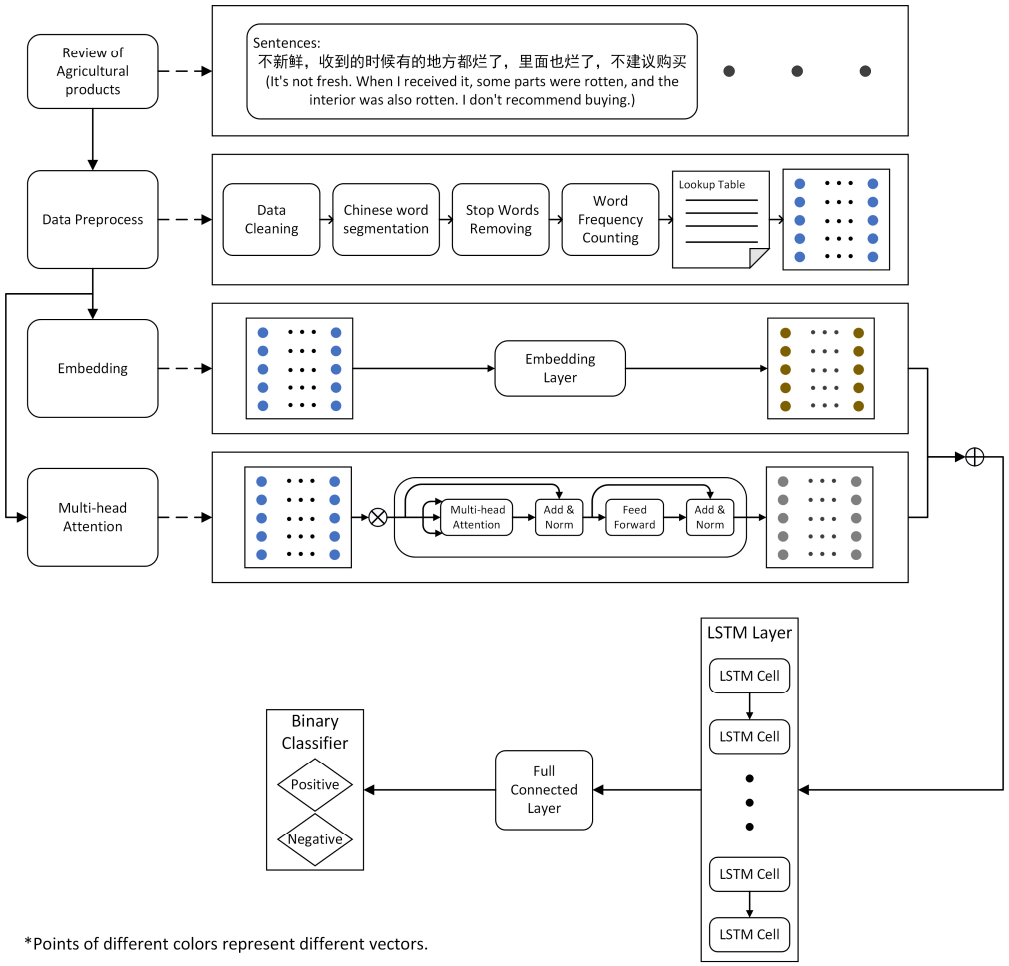

Figure 3. The Flowchart of AgriMFLN.

The flowchart of the model is described as follows:

(1) Data Acquisition and Preprocessing

Data acquisition and preprocessing are crucial steps in building natural language processing models. In this process, the raw data needs to be cleaned by removing noise and invalid data, so that the processed data are cleaner and more reliable. Next, the Jieba word segmentation tool for Chinese words is used to tokenize the text data into individual words, and then remove stop words and perform word frequency analysis to better understand the characteristics and structure of the text data.

(2) Embedding Layer

The embedding process is an important step in converting text data into vector representations. In this process, initial vectors are randomly generated and continuously updated during model training to effectively express the semantic information of the text. The embedding layer is used to convert the input discrete feature vectors into dense vector representations to facilitate the neural network's understanding of the input meaning. The main role of the embedding layer is to learn how to map input discrete features to continuous low-dimensional vector representations while maintaining the distance between similar features.

In this experiment, the embedding process of words is included as part of the input vector for the LSTM layer, which maximally preserves the original structure of classical LSTM neural networks to handle text with long sequence features. After embedding and the process of back-propagation, each word is represented by an array of specific

dimensions, and the semantic relationship between words is also contained in the matrix generated by word embedding.

(3) Multi-Head Attention Layer

Conducting the multi-head attention mechanism to extract semantic information from text aims to improve the performance of the model in processing sequential data. In the multi-head attention mechanism, the input vector is first split into multiple heads, and then each head undergoes independent linear transformation and attention calculation. Finally, the output vectors of each head are concatenated together and transformed linearly to obtain the final output vector. The benefit of this multi-head attention mechanism is that it allows the model to focus on different aspects of the input sequence simultaneously, thus better capturing the internal relationships within the text sequence.

In this experiment, considering the multi-level information in agricultural product text, such as consumers' evaluation of their shopping experience from different aspects, may affect the overall sentiment of the sentence. Therefore, the design of the multi-head attention mechanism can assign different weights to information at different levels, allowing the model to simultaneously focus on different aspects of the sentence such as grammar and semantics, to better understand the meaning of the text and effectively express the semantic information of the sentence.

(4) Feature Fusion Layer

As the multi-head attention mechanism can better focus on the important semantic information of text, this model proposes a feature fusion layer that combines the vectors generated from text embedding with those from multi-head attention and then inputs the fused vector to the LSTM layer for sentiment analysis of agricultural product reviews.

Different from the LSTM neural network, which consists of an embedding layer, LSTM layer, and output layer, this experiment modifies the input vector of the LSTM layer. The input vector of the LSTM layer is composed of two parts: one part comes from the encoder part of the Transformer. The sentence vector is first input into the position encoding module of the Transformer for position encoding and then input into the encoder module of the Transformer to extract semantic information of the sentence. Combined with the attention mechanism, the input is transformed into a vector that the machine can easily learn. Thus, the vector input into the LSTM layer contains detailed information extracted by the attention mechanism, which cannot be represented by the classical LSTM model. The other part comes from the classical LSTM neural network's embedding layer. After embedding, each word is represented by an array with a specific dimension, and the semantic relationship between words is also contained in the matrix generated by word embedding. Similarly, the sentence contains semantic information. The results of the embedding and the encoder are fused together and then input into the LSTM network.

(5) LSTM Layer

Given the superior performance of LSTM neural networks in processing long sequences and extracting contextual features, and the fact that agricultural product reviews have varying lengths and rich contextual information, this experiment implements an LSTM layer for time-series text modeling of agricultural product reviews to learn long-term dependencies in the text.

The advantage of an LSTM network is its ability to automatically learn long-term dependencies in the input sequence and control the flow of information more finely through gate mechanisms. The fused vector is fed into the LSTM layer for modeling long sequences of text and extracting complex semantic information in agricultural product reviews. This process can help us better understand the relationships between words in the text and thus analyze the meaning and sentiment of text data more effectively.

(6) Fully Connected Layer

The fully connected layer is connected to perform binary classification. This process helps us determine which category the text data belongs to, whether it is positive or negative, and obtain a predicted sentiment for the sentence.

Finally, to evaluate the performance of the model, we adopt performance metrics to assess and compare it with the baseline model. This process can help us better understand the strengths and weaknesses of the model, and identify areas for improvement, to better apply it in practical scenarios.

4. Results

In this section, we have described the experimental parameter settings, performance metrics, and results. The proposed model has been compared with baseline models on two agricultural product datasets, and the results have been explained.

4.1. Parameter Settings

This experiment uses Python 3.7 for coding. The parameter settings in the experiment are shown in Table 3.

Table 3. The parameter settings of the experiment.

Parameters	Values
Batch Size	128
Maximum Sequence Length	128
Learning Rate	0.01
Dropout Rate	0.2
Multiple Attention Heads	4
Epoch Number	50
Hidden Size	256
Embedding Size	256

4.2. Performance Metrics

Four evaluation metrics are used in this paper to evaluate the performance of the classification model: *Accuracy*, *Recall*, *Precision*, and *F1-score*.

The calculation formulas are shown in Formulas (16)–(19). *TP* (True Positive): refers to the number of positive samples that are correctly classified as positive by the model; *FP* (False Positive): refers to the number of negative samples that are incorrectly classified as positive by the model; *TN* (True Negative): refers to the number of negative samples that are correctly classified as negative by the model; *FN* (False Negative): refers to the number of positive samples that are incorrectly classified as negative by the model.

(1) *Accuracy*

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

(2) *Precision*

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

(3) *Recall*

$$Recall = \frac{TP}{TP + FN} \tag{18}$$

(4) *F1-score*

$$F1\text{-}score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{19}$$

4.3. Results

Accuracy, *Precision*, *Recall*, and *F*1 of the proposed model are calculated in the experiment, and the Transformer model, LSTM model, and Bi-LSTM model are used as the contrast experiments. The results of the experiment on the public dataset are shown in Table 4.

Table 4. The results of the experiment on the public dataset.

Model	Accuracy	Precision	Recall	F1
LSTM	90.21%	89.71%	90.83%	90.27%
Bi-LSTM	90.99%	92.27%	89.48%	90.85%
Transformer	91.20%	97.48%	84.58%	90.57%
AgriMFLN	92.92%	92.49%	93.54%	92.97%

The results of the experiment on the private dataset are shown in Table 5.

Table 5. The results of the experiment on the private dataset.

Model	Accuracy	Precision	Recall	F1
LSTM	94.82%	94.96%	97.90%	96.41%
Bi-LSTM	95.10%	97.45%	95.60%	96.52%
Transformer	95.03%	94.37%	98.90%	96.58%
AgriMFLN	96.06%	96.41%	98.10%	97.25%

The *Accuracy* of the LSTM model is 90.21% and 94.82%, on the public and private datasets respectively, which shows that the LSTM network can extract sequence features in sentences and has some effect when applied to the classification of agricultural product reviews. The *Accuracy* of the Bi-LSTM model is 90.99% and 95.10%, indicating that the classification efficiency and performance of the Bi-LSTM model are better than the single LSTM model on sequential text. The *Accuracy* of the Transformer model is 91.20% and 95.03%, indicating that the model based on multi-head attention will achieve better sentiment analysis than the model based on historical information of the sequence.

On the public dataset, the AgriMFLN model proposed in this paper is superior to other methods, with an *Accuracy* of 92.92%, which is 2.71, 1.93, and 1.72 percentage points higher than the LSTM model, Bi-LSTM model, and Transformer model, respectively, which achieved good results. The *Recall* is 93.54%, which is 2.71, 4.06, and 8.96 percentage points higher than the LSTM model, Bi-LSTM model, and the Transformer model, respectively, while its *Precision* is slightly lower than the other models. *F*1-*score*, as a comprehensive indicator, balances the impact of *Precision* and *Recall* and comprehensively evaluates the advantages and disadvantages of a classifier. The *F*1-*score* of the model proposed in this experiment is 92.97%, which is 2.70, 2.12, and 2.40 percentage points higher than the LSTM model, Bi-LSTM model, and Transformer model, respectively.

On the private dataset, the proposed model achieved an *Accuracy* of 96.06%, which is 1.24, 0.96, and 1.03 percentage points higher than the LSTM, Bi-LSTM, and Transformer models, respectively. The *Precision* is 96.41%, which is 1.45 and 2.04 percentage points higher than the LSTM and Transformer models, respectively. The *Recall* is 98.10%, which is 2.50 percentage points higher than the Bi-LSTM model. The *F*1-*score* is 97.25%, which is 0.84, 0.73, and 0.67 percentage points higher than the three baseline models, respectively.

Overall, the results of the experiment show that the AgriMFLN model proposed in this paper performs better than the other models on both public and private datasets. All evaluation metrics of the model proposed are generally higher than the LSTM model, Bi-LSTM model, and Transformer model, except for the *Precision*, which shows that the method proposed in this experiment is superior to other models in agricultural product datasets and has good performance to agricultural product texts. These results demonstrate

the effectiveness of the AgriMFLN model in accurately classifying agricultural data and its potential for practical applications in the field.

5. Discussion

Embedding is a crucial process in deep learning. In classical LSTM neural network models, the process of word embedding involves randomly initializing vectors, which are then continuously updated during training. In comparison to classical LSTM networks, this study has improved the model by incorporating feature fusion into the word embedding process. The model first uses the multi-head attention strategy of the Transformer model to extract the semantic information of the sentence. The attention mechanism allows the model to focus on specific features related to the target during training and can capture both global and local connections. The multi-head attention mechanism allows the model to concentrate on information from different representation subspaces at different positions, thereby extracting word vectors containing richer feature information. The generated vectors from the multi-head attention mechanism are fused with the word vectors generated during the embedding process to increase the information content of the vectors and avoid the random generation of word vectors in traditional LSTM neural networks. The resulting vectors are then input into the LSTM layer to capture the sequence features of the sentence and perform sentiment analysis.

The experimental results in the datasets of agricultural product reviews used in this study show that the method proposed in this experiment is about 1–3 percentage points higher than the LSTM model, Bi-LSTM model, and Transformer model, respectively, in *Accuracy*, *Recall*, and *F*1-*score*, which are scores superior to other models. This model not only takes advantage of the processing ability of LSTM neural networks to sequence information, but also makes full use of the excellent ability of multi-head attention mechanisms to represent semantic information, and it has achieved good results.

In addition, we also have found that the *Accuracy* of the two datasets differed by approximately five percentage points. Possible reasons for this may include differences in total data volume or differences in the ratio of positive to negative samples. We will further explore this issue in the future.

6. Conclusions

In this paper, we have proposed a new model named Mixing Features LSTM Networks for Sentiment Analysis of Agricultural Product Reviews (AgriMFLN). Considering that agricultural product reviews are characterized by varying lengths, strong contextual relevance, and complex semantic information, current methods face challenges in effectively and accurately extracting the rich semantic information to identify the sentiment tendencies of agricultural product reviews. To address these shortcomings, we have proposed a method based on the improved LSTM model with the attention strategy. We argue that the attention mechanism can identify the target area that requires more attention and can allocate additional resources accordingly, which is suitable of this scenario of sentiment analysis of agricultural product reviews. This enables the model to focus on the relevant information while ignoring irrelevant details, thereby gathering more detailed and meaningful information about the target. In this study, we implement the multi-head attention mechanism to extract multi-level semantic information in sentences. The resulting semantic features are then fused with the Embedding layer. The LSTM network is trained using these inputs to classify product reviews, which is effective in dealing with the long sequence information and contextual relationships in agricultural product review texts.

The proposed method has shown superior performance when compared to existing benchmark models and shows that it is effective for sentiment analysis of agricultural product reviews. Conducting sentiment analysis on agricultural product reviews can help consumers accurately understand the quality and taste of agricultural products. At the same time, it can also help production enterprises better grasp market demand, adjust product quality, and taste characteristics, and improve the market competitiveness of

their products, so that they can enhance their brand image and reputation of agricultural products. Therefore, sentiment analysis of agricultural product reviews has important practical significance and value. The findings of this study have practical implications for e-commerce platforms and agricultural product manufacturers, providing guidance for improving product quality and customer satisfaction.

Due to time constraints, it should be noted that although the model proposed in this experiment achieved good results on two datasets, these results were only tested on the specific dataset used in this experiment and have not been tested on a more extensive dataset of agricultural product reviews, or even on language data from other fields. In the future, it is necessary to further expand the scale of the experiment and increase testing on different fields and types of corpora, to verify the model's generalization ability, reliability, and practicality. In addition, we have noticed that the efficacy of using prior knowledge has been proven [48], this will also be one of the future research directions of this article.

Author Contributions: Conceptualization, R.L.; methodology, R.L.; software, R.L. and Y.L.; validation, R.L.; formal analysis, R.L.; investigation, R.L.; writing—original draft preparation, R.L.; writing—review and editing, R.L. and Y.L.; supervision, H.W.; project administration, H.W.; funding acquisition, H.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant number 32071775) and the Development Research Centre of Beijing New Modern Industrial Area (grant number JD2021001).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used to support the findings of this study are available from GitHub.

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

References

1. Liang, R.; Wang, J.Q. A Linguistic Intuitionistic Cloud Decision Support Model with Sentiment Analysis for Product Selection in E-commerce. *Int. J. Fuzzy Syst.* **2019**, *21*, 963–977. [CrossRef]
2. Ji, P.; Zhang, H.-Y.; Wang, J.-Q. A Fuzzy Decision Support Model with Sentiment Analysis for Items Comparison in e-Commerce: The Case Study of http://PConline.com. *IEEE Trans. Syst. Man Cybern. Syst.* **2018**, *49*, 1993–2004. [CrossRef]
3. Zeng, D.; Dai, Y.; Li, F.; Wang, J.; Sangaiah, A.K. Aspect based sentiment analysis by a linguistically regularized CNN with gated mechanism. *J. Intell. Fuzzy Syst.* **2019**, *36*, 3971–3980. [CrossRef]
4. Thet, T.T.; Na, J.-C.; Khoo, C.S. Aspect-based sentiment analysis of movie reviews on discussion boards. *J. Inf. Sci.* **2010**, *36*, 823–848. [CrossRef]
5. Park, E.; Kang, J.; Choi, D.; Han, J. Understanding customers' hotel revisiting behaviour: A sentiment analysis of online feedback reviews. *Curr. Issues Tour.* **2018**, *23*, 605–611. [CrossRef]
6. Alaei, A.R.; Becken, S.; Stantic, B. Sentiment Analysis in Tourism: Capitalizing on Big Data. *J. Travel Res.* **2019**, *58*, 175–191. [CrossRef]
7. Ali, F.; Kwak, D.; Khan, P.; El-Sappagh, S.; Ali, A.; Ullah, S.; Kim, K.H.; Kwak, K.-S. Transportation sentiment analysis using word embedding and ontology-based topic modeling. *Knowl.-Based Syst.* **2019**, *174*, 27–42. [CrossRef]
8. Nakayama, M.; Wan, Y. The cultural impact on social commerce: A sentiment analysis on Yelp ethnic restaurant reviews. *Inf. Manag.* **2018**, *56*, 271–279. [CrossRef]
9. Tran, T.; Ba, H.; Huynh, V.N. Measuring hotel review sentiment: An aspect-based sentiment analysis approach. In Proceedings of the Integrated Uncertainty in Knowledge Modelling and Decision Making: 7th International Symposium, IUKM 2019, Nara, Japan, 27–29 March 2019; Proceedings 7. Springer International Publishing: Berlin, Germany, 2019; pp. 393–405.
10. Karami, A.; Shah, V.; Vaezi, R.; Bansal, A. Twitter speaks: A case of national disaster situational awareness. *J. Inf. Sci.* **2019**, *46*, 313–324. [CrossRef]
11. Liu, B. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*; University of Illinois: Chicago, IL, USA, 2015.
12. Nasukawa, T.; Yi, J. Sentiment analysis: Capturing favorability using natural language processing. In Proceedings of the International Conference on Knowledge Capture, DBLP, Sanibel Island, FL, USA, 23–25 October 2003.
13. Pang, B.; Lee, L. Opinion Mining and Sentiment Analysis. *Found. Trends® Inf. Retr.* **2008**, *2*, 1–135. [CrossRef]
14. Zhang, L.; Wang, S.; Liu, B. Deep learning for sentiment analysis: A survey. *arXiv* **2018**, arXiv:1801.07883. [CrossRef]

15. Wu, X.; Zhang, W.; Zhang, X.; Xu, J. A Method to Build and Expand the Domain Dictionary Automatically Based on WordNet. In Proceedings of the International Conference on Trustworthy Computing and Services, Heraklion, Greece, 24–26 August 2015.
16. Han, H.; Zhang, J.; Yang, J.; Shen, Y.; Zhang, Y. Generate domain-specific sentiment lexicon for review sentiment analysis. *Multimedia Tools Appl.* **2018**, *77*, 21265–21280. [CrossRef]
17. Wu, L.; Morstatter, F.; Liu, H. SlangSD: Building, expanding and using a sentiment dictionary of slang words for short-text sen-timent classification. *Lang. Resour. Evaluat.* **2018**, *52*, 1–15. [CrossRef]
18. Zargari, H.; Zahedi, M.; Rahimi, M. GINS: A Global intensifier-based N-Gram sentiment dictionary. *J. Intell. Fuzzy Syst. Appl. Eng. Technol.* **2021**, *40*, 11763–11776. [CrossRef]
19. Li, C.; Li, L.; Zheng, J.; Wang, J.; Yuan, Y.; Lv, Z.; Wei, Y.; Han, Q.; Gao, J.; Liu, W. China's Public Firms' Attitudes towards Environmental Protection Based on Sentiment Analysis and Random Forest Models. *Sustainability* **2022**, *14*, 5046. [CrossRef]
20. Bernhard, K.; Suzana, I.; Mathias, K.; Feuerriegel, S.; Prendinger, H. Deep learning for affective computing: Text-based emotion recognition in decision support. *Decis. Support Syst.* **2018**, *115*, 24–35.
21. Kim, Y. Convolutional Neural Networks for Sentence Classification. *arXiv* **2014**, arXiv:1408.5882.
22. Zhao, J.; Lin, J.; Liang, S.; Wang, M. Sentimental prediction model of personality based on CNN-LSTM in a social media environment. *J. Intell. Fuzzy Syst.* **2021**, *40*, 3097–3106. [CrossRef]
23. Gopalakrishnan, K.; Salem, F.M. Sentiment Analysis Using Simplified Long Short-term Memory Recurrent Neural Networks. *arXiv* **2020**, arXiv:2005.03993.
24. Teng, F.; Zheng, C.; Li, W. Multidimensional topic model for oriented sentiment analysis based on long short-term memory. *J. Comput. Appl.* **2016**, *36*, 2252–2256.
25. Tang, D.; Qin, B.; Feng, X.; Liu, T. Effective LSTMs for Target-Dependent Sentiment Classification. In Proceedings of the 26th International Conference on Computational Linguistics, Osaka, Japan, 26 December 2015; pp. 3298–3307.
26. Jelodar, H.; Wang, Y.; Orji, R.; Huang, H. Deep Sentiment Classification and Topic Discovery on Novel Coronavirus or COVID-19 Online Discussions: NLP Using LSTM Recurrent Neural Network Approach. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 2733–2742. [CrossRef] [PubMed]
27. Bhandari, A.; Kumar, V.; Thien Huong, P.T.; Thanh, D.N. Sentiment Analysis of COVID-19 Tweets: Leveraging Stacked Word Embedding Representation for Identifying Distinct Classes Within a Sentiment. In Proceedings of the International Conference on Artificial Intelligence and Big Data in Digital Era, Ho Chi Minh, Vietnam, 18–19 December 2021; Springer: Cham, Switzerland, 2022.
28. Kai, S.; Li, Y.; Zhao, H. Word attention-based Bi-LSTM and CNN ensemble for Chinese sentiment analysis. *Comput. Sci. Appl.* **2020**, *10*, 312–324.
29. Wang, J.; Yu, L.-C.; Lai, K.R.; Zhang, X. Tree-Structured Regional CNN-LSTM Model for Dimensional Sentiment Analysis. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2019**, *28*, 581–591. [CrossRef]
30. Chen, K.; Ke, W. A hierarchical neural model for target-based sentiment analysis. *Concurr. Comput. Pract. Exp.* **2021**, *33*, e6184. [CrossRef]
31. Gao, Z.; Li, Z.; Luo, J.; Li, X. Short Text Aspect-Based Sentiment Analysis Based on CNN + BiGRU. *Appl. Sci.* **2022**, *12*, 2707. [CrossRef]
32. Jiang, M.; Zhang, W.; Zhang, M.; Wu, J.; Wen, T. An LSTM-CNN attention approach for aspect-level sentiment classification. *J. Comput. Methods Sci. Eng.* **2019**, *19*, 859–868. [CrossRef]
33. Ramaswamy, S.; Chinnappan, J. RecogNet-LSTM+CNN: A hybrid network with attention mechanism for aspect categorization and sentiment classification. *J. Intell. Inf. Syst.* **2022**, *58*, 379–404. [CrossRef]
34. Shobana, J.; Murali, M. Adaptive particle swarm optimization algorithm based long short-term memory networks for sentiment analysis. *J. Intell. Fuzzy Syst. Appl. Eng. Technol.* **2021**, *40*, 10703–10719. [CrossRef]
35. Kumar, V.; Recupero, D.R.; Riboni, D.; Helaoui, R. Ensembling Classical Machine Learning and Deep Learning Approaches for Morbidity Identification from Clinical Notes. *IEEE Access* **2020**, *9*, 7107–7126. [CrossRef]
36. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, arXiv:1706.03762.
37. Ma, D.; Li, S.; Zhang, X.; Wang, H. Interactive Attention Networks for Aspect-Level Sentiment Classification. *arXiv* **2017**, arXiv:1709.00893. [CrossRef]
38. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. Improving language understanding by generative pre-training. **2018**. *Preprint*.
39. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2018**, arXiv:1810.04805.
40. Tan, K.; Lee, C.; Anbananthen, K.; Lim, K.M. RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis with Transformer and Recurrent Neural Network. *IEEE Access* **2022**, *10*, 21517–21525. [CrossRef]
41. Gao, Z.; Feng, A.; Song, X.; Wu, X. Target-Dependent Sentiment Classification With BERT. *IEEE Access* **2019**, *7*, 154290–154299. [CrossRef]
42. Zhang, T.; Gong, X.; Chen, C.L.P. BMT-Net: Broad Multitask Transformer Network for Sentiment Analysis. *IEEE Trans. Cybern.* **2021**, *52*, 6232–6243. [CrossRef]
43. Gong, X.; Ying, W.; Zhong, S.; Gong, S. Text Sentiment Analysis Based on Transformer and Augmentation. *Front. Psychol.* **2022**, *13*, 906061. [CrossRef]

44. Lin, Y.; Wang, C.; Song, H.; Li, Y. Multi-Head Self-Attention Transformation Networks for Aspect-Based Sentiment Analysis. *IEEE Access* **2021**, *9*, 8762–8770. [CrossRef]
45. GitHub. Available online: https://github.com/SophonPlus/ChineseNlpCorpus/tree/master/datasets/online_shopping_10_cats (accessed on 26 May 2022).
46. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
47. Graves, A. *Supervised Sequence Labelling with Recurrent Neural Networks*; Springer: Berlin, Germany, 2012. [CrossRef]
48. Kumar, V.; Recupero, D.R.; Helaoui, R.; Riboni, D. K-LM: Knowledge Augmenting in Language Models Within the Scholarly Domain. *IEEE Access* **2022**, *10*, 91802–91815. [CrossRef]

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.