A Study of Sentiment Analysis Algorithms for Agricultural Product Reviews Based on Improved BERT Model

Ying Cao 1,2,†, Zhexing Sun 2,†, Ling Li 1,* and Weinan Mo 1,2

1 College of Mechanical and Electrical Engineering, Central South University of Forestry and Technology, Changsha 410004, China
2 Faculty of Modern Agricultural Engineering, Kunming University of Science and Technology, Kunming 650500, China
* Correspondence: liling@csuft.edu.cn
† These authors contributed equally to this work.

Abstract: With the rise of mobile social networks, an increasing number of consumers are shopping through Internet platforms. The information asymmetry between consumers and producers has caused producers to misjudge the positioning of agricultural products in the market and damaged the interests of consumers. This imbalance between supply and demand is detrimental to the development of the agricultural market. Sentiment tendency analysis of after-sale reviews of agricultural products on the Internet could effectively help consumers evaluate the quality of agricultural products and help enterprises optimize and upgrade their products. Targeting problems such as non-standard expressions and sparse features in agricultural product reviews, this paper proposes a sentiment analysis algorithm based on an improved Bidirectional Encoder Representations from Transformers (BERT) model with symmetrical structure to obtain sentence-level feature vectors of agricultural product evaluations containing complete semantic information. Specifically, we propose a recognition method based on speech rules to identify the emotional tendencies of consumers when evaluating agricultural products and extract consumer demand for agricultural product attributes from online reviews. Our results showed that the F1 value of the trained model reached 89.86% on the test set, which is an increase of 7.05 compared with that of the original BERT model. The agricultural evaluation classification algorithm proposed in this paper could efficiently determine the emotion expressed by the text, which helps to further analyze network evaluation data, extract effective information, and realize the visualization of emotion.

Keywords: sentiment analysis; deep neural network; BERT; S-LSTM

1. Introduction

With the development of the economy and Internet platforms, users are increasingly commenting on agricultural products, and the accumulation of agricultural product comments will inevitably cause issues of information overload. Agricultural producers need to efficiently obtain relevant information from the huge number of reviews. Compared with reviews in common text, online user reviews of agricultural products have the following characteristics: (1) domain-specificity with many specialized terms; (2) short sentences and little information; (3) obvious colloquialization of consumer expressions, common use of Internet terms and buzzwords, and a free writing style. Comments on agricultural products also show potential correlations with the user’s views on the different dimensions or attributes of the evaluation object. Therefore, it is necessary to use automated methods to extract the objects and user sentiment tendencies from the huge amount of user comments on agricultural products, and forming meaningful sentiment labels.

Sentiment Analysis (SA) is the process of classifying opinions and mining subjective texts with emotional overtones using natural language processing and text mining techniques [1]. Typically, the tasks studied in sentiment analysis can be divided into...
sentiment feature extraction, sentiment polarity classification, sentiment retrieval, generalization, etc. [2]. Zhang et al. [3] proposed a method based on fine-grained sentiment analysis and the Kano model to extract consumer demand for air purifier product attributes from online reviews. Shen et al. [4] used independent memory networks to learn user profiles and product information for review classification, resulting in a dual user and product memory network (DUPMN) model for sentiment analysis. Fu et al. [5] proposed a multifaceted sentiment analysis of Chinese online social commentaries (MSA-COSRs) to automatically discover the aspects discussed in Chinese social commentaries and the sentiments expressed by different aspects. Although Traditional Machine Learning and deep learning-based sentence classification methods show good performance in short text classification tasks, the static word vectors used in these models cannot represent multiple meanings of words in different contexts and cannot fully represent the fine-grained relationships between phrases and clauses.

In recent years, BERT [6] has become a state-of-the-art (SOTA) Natural Language Processing (NLP) model. It uses a pretraining task to obtain a priori semantic knowledge from a large unlabeled corpus and transfers this knowledge to the model to improve semantic feature extraction. With the classification function of BERT, Bedi et al. [7] developed a model to propose sentiment and complaint classifications based on energy-related tweets. Lin et al. [8] combined ensemble learning methods with text sentiment analysis to obtain a model based on BERT to identify harmful news, providing readers with a method to identify harmful news content. There are many asymmetric entity relationships in long sentences, so it is a challenge to construct comprehensive and reasonable entity features. In addition, sentiment analysis of Chinese is still a challenging task due to the lack of physical separation between Chinese words.

To this end, our study extracts sentiment tags from the comment corpus of agricultural products, focusing on the correlation between evaluation words and rating objects. We propose a sentiment analysis algorithm for product reviews based on an improved BERT model. This algorithm uses the BERT model to obtain context-dependent dynamic word vectors to solve multi-sense representation of words. Meanwhile, it extracts the feature vectors of phrases and inter-sentences in complex long sentences through word-level and sentence-level LSTM networks to explore the deep semantic information of sentences. We introduce the Focal Loss loss function to reduce the weight of easily classified samples, and the model is guided to focus on the samples with ambiguous emotion expressions and agricultural characteristic emotions, thus improving the model’s emotion classification accuracy.

2. Related Work

2.1. Semantic Sentiment Analysis (SSA)

Sentiment analysis is generally defined as finding out the author’s opinion about an entity by analyzing people’s sentiments, opinions, attitudes, and emotions about specific elements (topics, products, individuals, organizations, and, etc.) [9]. Daniel et al. [10] proposed an unsupervised learning sentiment calculator based on a sentiment lexicon to analyze the sentiment polarity of events posted on the financial community of Twitter for financial transactions events. Popescu et al. [11] selected nouns and noun phrases that frequently appear in comments as candidate features and evaluated these features by calculating the mutual information value with a search engine. Penalver-Martinez et al. [12] adopted a semantic ontology approach to improve the efficiency of feature extraction and employed a vector analysis-based approach for the sentiment analysis of movie reviews. Zhao et al. [13] proposed an unsupervised model of “clustering first, extraction later” and then applied this approach to extract sentiment information.

The majority of research on sentiment analysis based on machine learning focuses on the optimization of feature extraction algorithms. Yu et al. [14] used the Naive Bayes algorithm (NB) to explore the relative importance and interrelatedness of social media and traditional media on the short-term stock performance of companies. Manek et al. [15] performed feature extraction based on the Gini index. Akhtar et al. [16] proposed a
cascade framework for the pruning and compression of feature selection. Meanwhile, they used a particle swarm optimization approach that integrates maximum entropy, conditional random field, and a support vector classifier for sentiment analysis. The effectiveness of this method was validated by feature extraction and sentiment analysis tasks in two different domains. Colace et al. [17] achieved satisfactory results with the Latent Dirichlet Allocation (LDA) based approach as an extraction tool for sentiment views in social networks and collaborative learning environments. In recent years, the development of artificial intelligence has led to the gradual application of deep learning-based methods for textual sentiment analysis. Giatsoglou et al. [18] proposed to put text documents in lexical, word embedding, and hybrid vector representation, and then used this to make a machine learning model for document sentiment polarity classification. This method can detect people’s opinions in different languages quickly and flexibly. Although all of the aforementioned methods are used for semantic sentiment analysis, they all have limitations.

2.2. BERT

BERT is a SOTA language representation model for a wide range of tasks [6]. It has been proven that the bidirectional training language representation model performs better in corpus feature extraction and context interpretation than the unidirectional language model, as shown in Figure 1. To achieve better vector representation, the Bert model uses a transformer structure to learn the contextual information of the input words, which integrates a multi-headed self-attentive mechanism to comprehensively mine the information in different locations under different subspaces and encode the information representation to each location. The main innovation of BERT is the use of a masked language model and next sentence prediction to capture word and sentence representations, respectively. Many researchers have applied a pre-trained BERT to improve the performance of semantic analysis tasks. For example, Lee et al. [19] used BERT as a word embedding method to integrate a BiLSTM network with an attention mechanism for medical text inference to extract deep semantic information from medical texts. Ohsugi et al. [20] proposed a conversational machine understanding model using BERT to encode each question and answer independently in a multi-round environment. Zhang et al. [21] proposed a BERT-JAM model for machine translation tasks and achieved good results. He et al. [22] proposed a multimodal fusion BERT model to explore time-dependent interactions between different modalities. The experimental results on a public dataset showed that this model achieves good performance.

Moreover, BERT can also be used to analyze other languages, such as Chinese [23], Arabic [24], French [25], etc. However, BERT can only obtain character-level representations before training [26]. For example, in Figure 2, the two sentences input into BERT have the same meaning. The token of the English sentence is a word, whereas the token of the Chinese sentence is a character.

![Figure 1. Structure of the BERT model.](image-url)
Li et al. [27] proposed a BERT-based sentiment analysis model for Chinese stock reviews. They used a BERT pre-trained language model to represent stock reviews at the sentence level and then fed the obtained feature vectors to a layer for classification. He et al. [28] proposed a new and accurate method that can automatically identify implicit relations in Chinese discourse using BERT and the proposed tree structure. Wan et al. [29] proposed a relation classification method (EC_BERT) based on the BERT model augmented with entity and entity context information. EC_BERT uses BERT to obtain a vector of sentence feature representations and combines the context utterance information of two target entities and feeds it into a simple neural network for relation classification. The method achieves better performance on the relation classification task than previous methods. These findings suggest that BERT can be used for sentiment analysis of agricultural products.

3. Method

3.1. LSTM

For intent classification tasks of long complex sentences, the LSTM network can filter information from the input units, and each input unit may have a different degree of influence on the final classification result. The network adds input gates, output gates, and forgetting gates for selecting information, and adds memory units for storing and updating information through different gates to the RNN. The updating process of the LSTM network is as follows: in time step \( t \), the forgetting gate, \( f_t \), first determines the information that can be forgotten based on the input, \( x_t \), at the current moment and the output, \( h_{t-1} \), from the previous moment, and the input gate, \( i_t \), determines the information that needs to be updated. For example, the network will create a new memory unit candidate, \( c_t \), through the \( \tanh \) function for words that have a high impact on intention, such as “好 (good)” and “便宜 (cheap)”. The updated memory units, \( c_t \), are filtered by combining the forgetting gate, \( f_t \), with the memory units, \( c_{t-1} \), of the previous moment to obtain the updated memory units to store the latest information. Meanwhile, the information flow is updated by the control gate and forgetting gate to generate the new output, \( h_t \).

As shown in Figure 3, the original LSTM is limited by the fact that the model only works on sequence-related data and cannot perform parallel operations. As for the information used in both directional streams in Bi-LSTM, the gradient still disappears after the sequence length exceeds a certain limit, which affects model learning accuracy. The sentence-level LSTM takes the sentence-level state as input, and performs parallel operations to improve feature extraction of text for complex emotional information in agricultural product evaluations.

Figure 3. The LSTM operation flow chart.
3.2. S-LSTM

Unlike BiLSTM, which encodes only one word at time, $t$ (time step), sentence-level LSTM encoding semantically encodes all words of the input text and the whole text at time, $t$. For a given agricultural product evaluation statement, $S = x_1, x_2, \ldots, x_n$, $x_i$ represents the $i$-th Chinese character and $n$ represents the length of the evaluation statement. The hidden vector of each word, $h_i$, and the hidden vector of the whole sentence, $d$, is combined. The LSTM state at moment, $t$, is formalized as, $H^t = \{h_0^t, h_1^t, h_2^t, \ldots, h_{n+1}^t, d^t\}$, which consists of each word, $x_i$, the word-level sub-state of, $h_i^t$, and the sentence-level state of the whole evaluation, $d^t$.

As for the initial state, $H^0, h_0^t = d^0 = h^0$. The cyclic time step, $t$, is used to model the information exchange from the previous moment state, $h^t-1$, to the current moment state, $h^t$. The transition consists of a word-level state transition and a sentence-level state transition.

Figure 4 presents the LSTM operation flow of the hidden state, $h_i^t$, whose update rules are shown in Equations (1)–(11):

\[
\xi_i^t = \left[ h_{i-1}^{t-1}, h_i^{t-1}, h_{i+1}^{t-1} \right] \tag{1}
\]

\[
\hat{h}_i^t = \sigma(W_i \xi_i^t + U_i x_i + V_i d_i^{t-1} + b_i) \tag{2}
\]

\[
\tilde{l}_i^t = \sigma(W_i \xi_i^t + U_i x_i + V_i d_i^{t-1} + b_i) \tag{3}
\]

\[
\tilde{r}_i^t = \sigma(W_i \xi_i^t + U_i x_i + V_i d_i^{t-1} + b_r) \tag{4}
\]

\[
\tilde{f}_i^t = \sigma(W_j \xi_i^t + U_j x_i + V_j d_i^{t-1} + b_f) \tag{5}
\]

\[
\tilde{s}_i^t = \sigma(W_c \xi_i^t + U_c x_i + V_c d_i^{t-1} + b_c) \tag{6}
\]

\[
o_i^t = \sigma(W_o \xi_i^t + U_o x_i + V_o d_i^{t-1} + b_o) \tag{7}
\]

\[
u_i^t = \tanh(W_u \xi_i^t + U_u x_i + V_u d_i^{t-1} + b_u) \tag{8}
\]

\[
l_i^t, l_i^t, r_i^t, f_i^t, s_i^t = \text{softmax}\left(\tilde{l}_i^t, \tilde{l}_i^t, \tilde{r}_i^t, \tilde{f}_i^t, \tilde{s}_i^t\right) \tag{9}
\]

\[
c_i^t = l_i^t \odot u_i^t + l_i^t \odot c_i^{t-1} + r_i^t \odot c_i^{t-1} + f_i^t \odot c_i^{t-1} + s_i^t \odot c_i^{t-1} \tag{10}
\]

\[
h_i^t = o_i^t \odot \tanh(c_i^t) \tag{11}
\]

![Figure 4. The LSTM operation flow of the hidden state, $h_i^t$.](image)

The connection, $\xi_i^t$, of forward-backward hidden vectors is used to construct the context window information, and the window size can control the degree of information exchange with neighboring words. The control gates, $l_i^t, l_i^t, r_i^t, f_i^t, s_i^t$ control the input,
$x_i$; the cell state of the content of the left text segment, $c_{i-1}^{l-1}$; the cell state, $c_{i}^{l-1}$, of the $i$-th text content; the cell state, $c_{i}^{l+1}$, of the right text segment; and the cell state, $c_{i}^{d-1}$, of the sentence-level cell, and normalizes this information. The transformation from the cell state, $c_{i}^{l}$, to the hidden state, $h_{i}^{l}$, of the output gate is indicated by $o_{i}^{l}$. The model parameters are $W_{x}, U_{i}, V_{s}$, and $b_{x}(\in \{i, o, l, r, s, u\})$, and $\sigma$ is the sigmoid function.

At moment, $t$, the update of the sentence-level state of the whole text, $d_{t}$, is shown in Equations (12)–(18).

$$h_{i} = \text{avg}(h_{0}^{l-1}, h_{1}^{l-1}, \ldots, h_{n+1}^{l-1})$$

(12)

$$f_{i}^{d} = \sigma(W_{d}d_{t-1}^{l-1} + U_{d}h_{i} + b_{d})$$

(13)

$$f_{i}^{l} = \sigma(W_{f}d_{t-1}^{l-1} + U_{f}h_{i}^{l-1} + b_{f})$$

(14)

$$o_{i}^{l} = \sigma(W_{o}d_{t-1}^{l-1} + U_{o}h_{i} + b_{o})$$

(15)

$$f_{0}^{l}, \ldots, f_{n+1}^{l}, f_{d}^{l} = \text{softmax}(f_{0}^{l}, \ldots, f_{n+1}^{l}, f_{d}^{l})$$

(16)

$$c_{d}^{l} = f_{d}^{l} \odot c_{d}^{l-1} + \sum_{i=0}^{n} f_{i}^{l} \odot c_{i}^{l-1}$$

(17)

$$d_{t}^{l} = o_{i}^{l} \odot \tanh(c_{d}^{l})$$

(18)

The training model adds multiple forgetting gates to the existing output, input, and loop gates of the original LSTM model, as shown in Figure 5. This way the model can receive more feature information and adjust the state of the time step for deep feature learning; thus, improving the model’s feature learning ability.

![Figure 5. The LSTM operation flow of the sentence-level state, $d_{t}$.](image)

### 3.3. Improved BERT

The main function of the input layer is to convert the agricultural comment text into a format that can be recognized by the model. In the input layer, the input of position vectors, word vectors, and segment vectors are stitched together: (1) Word vector: the model does not perform word separation for Chinese text and directly uses a single word as the base unit of the text. The word vector contains the main feature information. (2) Position vector: compared with recurrent neural networks and short- and long-term memory networks, BERT cannot determine the position of input words by its model structure alone. For example, “I can’t like Gochujang rice” and “I may not like Gochujang rice” express different...
emotional tendencies. So, this paper adds location vectors to remember word position.

(3) Segment vector: used to distinguish two input texts to meet the needs of different tasks.

The BERT model outputs both sentence-level and character-level feature vectors. As shown in Figure 6, the leftmost output of the model (CLS) represents the sentence vector that contains the semantics of the whole sentence, and the remaining parts represent the vector output of each character. The BERT model consists of a twelve-layer Transformer network, which mainly uses the Encoder part of the Transformer network. In the Encoder part, the attention mechanism is used to calculate the Input-Output relationship and to learn features that cannot be learned by the shallow network through a deeper network level.

Figure 6. The structure of the improved BERT model.

The structure of each Transformer in the BERT model is shown in Figure 7. The Transformer consists of several overlapping units, and each of them is composed of two parts: a self-attentive mechanism and a feedforward neural network. The Self-Attention mechanism helps the current node to focus on the current word and obtain the semantics of the context. Each of these sub-layers incorporates a residual network and a normalization operation.
Figure 7. Transformer coding unit structure.

The output of each sub-layer is represented as follows:

\[
\text{sub-layer\_output} = \text{LayerNorm}(x + (\text{SubLayer}(x)))
\]  

The sub-layers within the unit are designed with residual connections to achieve information transfer between layers. The input to the Encoder is an embedded representation of a word in a sentence. The Self-attention layer allows the Encoder to view the information before and after each word as it encodes. The output of the Encoder is passed through another layer, where Add means the input and output of the Self-attention layer are added, and Norm means the output of the Self-attention layer is normalized. In this way, the output of the Self-attention layer has a fixed mean of 0 and a standard deviation of 1. The normalized list of vectors is then passed into a fully connected feed-forward neural network. The main module in the Encoder section is Self-attention, and its core idea is to calculate the interrelationship between each word in a sentence and all the words in that sentence, and then use these interrelationships to adjust the weight of each word. The new expression of each word contains not only the semantics of the word, but also its relationship with other words, so the vector obtained by this method has a more global expression than the traditional word vector. Multiple Self-Attention mechanisms are used for parallel computation, and the outputs are multiplied by a random initialized matrix. The final result is outputted by a linear transformation.

3.4. Loss Function

This paper conducts sentiment analysis research on agricultural product review data, and there are sentiment feature word differences in agricultural product review data. The review data containing common sentiment feature words and with positive and negative sentiment tendency categories occupy most of the data set. Their sentiment features are more prominent and easier to be correctly classified, making the model biased toward easily classified samples that dominate in number during training. To address this issue, this paper introduces the Focal Loss function to reduce the weight of easily classified samples and focus the model more on fuzzy sentiment expression and sample data containing...
sentiment feature words in the agricultural domain. The cross-entropy formula for bi-classification is shown in Equation (20):

\[
CE(p, y) = \begin{cases} 
- \log p, y = 1 \\
- \log (1 - p), y \neq 1 
\end{cases}
\]  

(20)

In the above formula, \( CE \) refers to cross-entropy, \( p \) denotes the prediction probability, and \( y \) denotes the label value.

The Focal Loss function is obtained by improving the cross-entropy loss function for data imbalance, as shown in Equations (21) and (22).

\[
p_t = \begin{cases} 
-p, y = 1 \\
-(1 - p), y \neq 1 
\end{cases}
\]  

(21)

\[
Focal Loss = (1 - p_t)^\gamma \log(p_t)
\]  

(22)

This paper validates \( \gamma = 2 \) in the experiment. When \( \gamma = 0 \), Focal Loss becomes the cross-entropy loss function. When a large value is assigned to \( p_t \) and a small weight is assigned to the easily classified samples, the Focal Loss function reduces the loss value of the easy classification, making its contribution to the overall loss reduced. In this paper, multiclassification sentiment analysis is performed for sentiment, and the modified cross-entropy loss function is shown in Equation (23), which meets the requirements of the multiclassification task.

\[
FL = - \sum_{i=0}^{n} (1 - p_p)^\gamma p_t \times \log(p_t)
\]  

(23)

In the above formula, \( p_t \) denotes the true sample label, \( p_p \) denotes the predicted category of the model output, and \( n \) is the number of categories.

4. Experiments

4.1. Data Acquisition and Preprocessing

The review data obtained from different sources had some differences. To ensure the comprehensiveness and representativeness of information sources, the agricultural product reviews were obtained from four channels: social media, news websites, e-commerce websites, and offline surveys. The data were processed with comment de-duplication, mechanical compression, de-word processing, and semantic comment deletion to construct a clean, original corpus.

In this paper, the product review data is annotated with manual sentiment categories, and sentiment is categorized into positive, negative, and neutral sentiment reviews. The introduction of the categories is presented in Table 1:

| ID | Category            | Description                                         |
|----|---------------------|------------------------------------------------------|
| C1 | Positive emotions   | The text contains positive emotional tendencies       |
| C2 | Neutral emotions    | The text does not contain emotional tendencies        |
| C3 | Negative emotions   | The text contains negative emotional tendencies       |

To ensure the objectivity and accuracy of the labeled data, the same batch of data is labeled twice, and the labeling results of the two groups of data are verified against one another. In this way, 6152 positive sentiment data items, 1873 neutral sentiment data items, and 7925 negative sentiment data items were obtained.

The data were randomly sorted; 70% were taken as the training set, 15% as the validation set, and 15% as the test set. The model parameters were continuously adjusted.
and optimized according to the classification effect on the validation set. Finally, the model was tested on the test set.

In this paper, the original dataset, balanced dataset, and data-enhanced dataset were constructed to investigate the performance of the BERT model on different types of datasets (The data set parameters as shown in Table 2):

1. As imbalance in a dataset affects the optimal performance of the model, a balanced dataset needs to be approximated by balancing the corpus. To reduce the influence of the imbalanced dataset on the BERT model and achieve optimal results, a balanced dataset was built based on the minimum amount of sentiment data among the three categories, and 2000 data items were randomly selected from each of the three sentiment categories to build the balanced dataset.

2. The sentiment labeling results indicated that the proportion of sentiment categories in the agricultural product review data set was unbalanced. To conduct our experiment and observe our model perform in a real environment, we assumed that the proportion of data categories in the data set was consistent with practice and constructed the original data set based on this proportion.

3. To investigate the effect of the improved BERT model to handle the problem of sentiment feature word discrepancy and to investigate whether this problem was affected by the balance of the dataset, we constructed a data-enhanced balanced dataset by adding a few class samples based on the balanced dataset using the back-translation method.

Table 2. Data set parameters.

| Data Set Name       | Positive Category Samples | Neutral Category Samples | Negative Category Samples | Sample Proportion |
|---------------------|---------------------------|--------------------------|---------------------------|-------------------|
| Balanced data set   | 2000                      | 2000                     | 2000                      | 1:1:1             |
| Raw data set        | 6152                      | 1873                     | 7925                      | 3.3:1:4.2         |
| Enhanced data set   | 5320                      | 5320                     | 5320                      | 1:1:1             |

4.2. Experimental Environment

The computing hardware used in the experiment was RTX 2080Ti (NVIDIA, 11GB), and the deep learning framework was TensorFlow. The specific configuration of the experimental environment is shown in Table 3.

Table 3. Experimental environment.

| Projects                | Configuration          |
|-------------------------|------------------------|
| Operating Platforms     | CUDA 10.1              |
| Operating System        | Windows 11             |
| Memory                  | 16 GB                  |
| Python Versions         | Python 3.8.1           |
| TensorFlow Versions     | TensorFlow 1.16.2      |

4.3. Model Parameter Setting

The relevant parameters of the model are shown in Table 4.

Table 4. Model parameter settings.

| Parameter Name                                      | Parameter Value |
|----------------------------------------------------|-----------------|
| Maximum sentence length                            | 128             |
| Batch size                                         | 64              |
| Transformer layers                                 | 12              |
| Transformer number of hidden neurons               | 768             |
| Learning rate                                      | 0.0001          |
| Maximum number of iterations                       | 30              |
| Optimization methods                               | Adam            |
| Number of hidden layer nodes of LSTM network       | 200             |
4.4. Experimental Results

In this experiment, Precision ($P$), Recall ($R$), and F1 values were used as evaluation indexes. Precision ($P$) refers to the proportion of the number of correctly classified sentences to the number of all sentences predicted, and Recall ($R$) refers to the proportion of the number of correctly classified sentences to the number of true sentences for a particular classification. In practice, Precision ($P$) and Recall ($R$) are contradictory measures, where a high recall rate may result in a low accuracy rate, and vice versa. The F1 value is a combination of precision and recall, and the higher the $F1$ value, the better the model performance.

As this paper focused on a multi-category classification task, the classification performance of the model was evaluated by the average value of each metric for each category.

\[
\text{Precision} = \frac{TP}{TP + FP} \quad \text{(24)}
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \quad \text{(25)}
\]

\[
F1 = \frac{2 \times P \times R}{P + R} \quad \text{(26)}
\]

The $F1$ and loss rate changes of BERT, BERT-BiLSTM, and our method on both the original dataset and the data-enhanced balanced dataset in 30 iterations are shown in Figure 8.

Experiments were conducted on the balanced dataset, and the BERT model proposed in this paper was compared with the Convolutional Neural Network for Text Classification (TextCNN) and traditional machine learning models, including support vector machine (SVM), k nearest neighbors (KNN), and BAYES. The experimental results are shown in Table 5 and Figure 9.
Table 5. Comparison between the BERT model and machine learning models on the balanced dataset.

| Models    | Indicators | C1     | C2     | C3     | F1   | Value |
|-----------|------------|--------|--------|--------|------|-------|
| BERT      | P          | 91.26  | 84.25  | 88.64  |      | 87.57 |
|           | R          | 93.84  | 81.23  | 86.33  |      |       |
|           | F1         | 92.53  | 82.71  | 87.47  |      |       |
| TEXTCNN [30] | P     | 86.42  | 74.21  | 71.44  |      | 76.36 |
|           | R          | 89.54  | 62.50  | 75.21  |      |       |
|           | F1         | 87.95  | 67.85  | 73.28  |      |       |
| SVM [31]  | P          | 84.55  | 60.50  | 73.10  |      | 73.46 |
|           | R          | 88.73  | 63.21  | 70.85  |      |       |
|           | F1         | 86.59  | 61.84  | 71.96  |      |       |
| KNN [32]  | P          | 62.45  | 68.83  | 68.42  |      | 63.76 |
|           | R          | 74.28  | 57.25  | 54.88  |      |       |
|           | F1         | 67.85  | 62.51  | 60.91  |      |       |
| BAYES [33]| P          | 67.22  | 48.85  | 58.42  |      |       |
|           | R          | 78.14  | 51.24  | 53.47  |      |       |
|           | F1         | 72.27  | 50.02  | 55.84  |      |       |

Figure 9. Comparison of experimental results of BERT and machine learning methods.

Experiments were conducted on both the original dataset and the data-enhanced balanced dataset to compare the improved BERT and BERT-BiLSTM models. Whether this method is effective in classifying the sentiment of agricultural products remains to be verified. The experimental results are shown in Table 6 and Figure 10.
Table 6. Comparison of experimental results of the improved BERT* on both the original dataset and the data-enhanced balanced dataset.

| Models         | Indicators | C1    | C2    | C3    | F1 Value |
|----------------|------------|-------|-------|-------|----------|
| BERT           | P          | 90.31 | 74.25 | 83.52 | 82.81    |
|                | R          | 91.25 | 73.23 | 84.13 |          |
|                | F1         | 90.78 | 73.73 | 83.82 |          |
| LSTM           | P          | 76.36 | 75.13 | 75.56 |          |
|                | R          | 77.53 | 74.65 | 76.32 | 75.92    |
|                | F1         | 76.94 | 74.89 | 75.94 |          |
| BiLSTM + Attn  | P          | 79.63 | 78.63 | 77.54 |          |
|                | R          | 78.65 | 77.69 | 75.96 | 78.01    |
|                | F1         | 79.14 | 78.16 | 76.74 |          |
| AOA-LSTM       | P          | 83.65 | 82.56 | 82.26 |          |
|                | R          | 82.14 | 81.23 | 80.25 | 82.01    |
|                | F1         | 82.89 | 81.89 | 81.24 |          |
| CapsNet        | P          | 82.23 | 83.46 | 81.32 |          |
|                | R          | 81.25 | 80.67 | 80.56 | 81.57    |
|                | F1         | 81.74 | 82.04 | 80.94 |          |
| BERT*          | P          | 90.42 | 77.21 | 84.22 |          |
|                | R          | 92.91 | 74.55 | 85.12 | 84.06    |
|                | F1         | 91.65 | 75.86 | 84.67 |          |
| BERT-LSTM [34] | P          | 92.14 | 88.31 | 87.10 |          |
|                | R          | 91.51 | 86.45 | 86.85 | 88.72    |
|                | F1         | 91.82 | 87.37 | 86.97 |          |
| BERT-BiLSTM [35]| P         | 92.58 | 90.12 | 87.55 |          |
|                | R          | 88.86 | 89.41 | 88.10 | 89.42    |
|                | F1         | 90.68 | 89.76 | 87.82 |          |
| Improved BERT* | P          | 92.89 | 90.45 | 88.64 |          |
|                | R          | 92.26 | 89.41 | 86.24 | 89.86    |
|                | F1         | 92.25 | 89.92 | 87.42 |          |

* indicates that FL is used as the loss function.

In this study, the proposed improved BERT model segmented texts according to Chinese language characteristics, which was different from the traditional method of using sentiment dictionaries, relying on the text vectors derived from the context to solve the problem of sentiment analysis for agricultural product evaluation. Model performance was evaluated using the F1 value metric, and BERT achieved good performance in comparison with traditional machine learning methods (the F1 value improved by more than 11). In the comparison experiment on deep learning methods, the F1 value of the BERT model with modified loss function was improved by 1.25 compared with that of the original model, whereas the F1 value of the BERT-LSTM model reached 88.72. The F1 value of the SOTA BERT-BiLSTM model reached 89.42, whereas the F1 value of the improved BERT model using sentence-level LSTM was improved by 7.05 compared with that of the original BERT model. The F1 value of the improved BERT model using sentence-level LSTM was improved by 7.05 compared with that of the original BERT model, and the F1 value was improved by 0.44 compared with that of the BERT-BiLSTM model. Sentence-level LSTM was used to divide the text vector obtained from the pre-trained model, and text features were extracted to solve the problem of LSTM and Bi-LSTM that causes the semantic association between contexts to gradually weaken with the passing of states [36]. Sentence-level LSTM method was used to encode agricultural product evaluation sentences to construct sentence-level states of the entire text with global information. The experimental results indicate that the model has good prospects for applications in the sentiment analysis field for agricultural product evaluation.
Figure 10. Comparison of experimental results of the improved BERT model.

5. Discussion

From the above experiments, we found that the improved BERT model outperformed other similar natural language processing models in performance evaluation indicators such as accuracy, recall, and F1-score, and had advantages in sentiment analysis for agricultural product evaluation. At present, there are many studies on sentiment analysis of Chinese reviews [21,35,36]. We selected methods with source code to test the performance of their proposed method using our dataset. The results are shown in Table 7.

Table 7. Comparison between the improved BERT model and previously proposed models using the balanced dataset.

| Models                        | Indicators | C1   | C2   | C3   | F1 Value |
|-------------------------------|------------|------|------|------|-----------|
| BiGRU + Attention [37]        | P          | 91.26| 86.85| 91.53| 87.75     |
|                               | R          | 90.63| 87.94| 90.33|           |
|                               | F1         | 88.94| 85.39| 88.93|           |
| Improved BiLSTM-CRF [38]      | P          | 89.42| 84.67| 86.76| 86.03     |
|                               | R          | 88.54| 82.01| 84.83|           |
|                               | F1         | 88.98| 83.32| 85.78|           |
| BERT + FC [21]                | P          | 92.55| 86.50| 88.20| 89.18     |
|                               | R          | 91.37| 88.53| 87.94|           |
|                               | F1         | 91.96| 87.50| 88.07|           |

The experimental results show that the performance of our model in sentiment analysis of agricultural product evaluation is state-of-the-art, and can better complete the task of sentiment classification.
As reviewers and consumers have common buyer interests, consumers are more inclined to believe the information provided by reviewers than product information provided by merchants. Online reviews reflect emotional tendencies, satisfaction, and product information. Such reviews can help consumers eliminate some doubts, reduce their risk perception, and enhance their willingness to buy. Therefore, for enterprises, starting from the emotional tendencies of users, more accurately excavating the advantages and disadvantages of products and sales plays an important role in improving consumers’ willingness to buy and increasing online sales of agricultural products.

6. Conclusions and Future Work

In this study, we proposed a sentiment analysis algorithm based on the improved BERT model to obtain sentence-level feature vectors of agricultural product evaluations containing complete semantic information. The obtained feature vectors were fed into a classifier for sentiment classification. The output layer of BERT was connected to the sentence-level LSTM (S-LSTM) network, and the features were hierarchically extracted by the LSTM. Additionally, this paper introduces the Focal Loss function that reduces the weight of easy-to-classify samples in model training and guides the model to focus on feature word samples with more ambiguous sentiment expression and agricultural characteristic sentiment. Compared with other task-specific and general task-independent pre-training models, the improved BERT model performs better for Chinese emotion analysis tasks.

In future research, we aim to apply this novel method to different key domains to validate generalizability and test the model on multilingual datasets to observe its performance in other languages. In the current study, we used a clean original corpus for sentiment analysis of agricultural product evaluation. However, in the actual online detection process, some merchants publish false positive reviews that do not accurately reflect real purchase experiences, by brushing orders and establishing a service [39], creating a high-quality store image to attract consumers, in order to increase sales, and even going to competitors’ stores to deliberately fabricate malicious and slanderous bad reviews. These fake reviews seriously interfere with consumers’ purchasing decisions. Therefore, finding a method to detect and delete fake online reviews and improve the quality of data has become an urgent problem that needs to be solved. At the same time, agricultural product reviews often include information in the forms of text, images, and videos. The next step of research should focus on the multimodal data fusion of online review text, images, and videos, and establishing a multimodal data mining model. We also plan to test this model further using data sets in different languages (e.g., Arabic, English, French, etc.) and observing how the model performs in other languages.

Author Contributions: Y.C. and L.L. conceived the idea of this paper; Z.S. designed and W.M. completed the experiments; W.M. made modifications to the experiments; Y.C. and Z.S. wrote and revised the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Hunan Province Key R&D Program: Research and the development of tea oil rapid identification technology and equipment (Grant No. 2022NK2048); General project of Natural Science Foundation of Hunan Province (Grant No. 2020JJ4142).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data required to reproduce these findings cannot be shared at this time as the data is also part of an ongoing study.

Acknowledgments: The authors would like to thank all the reviewers who participated in the review.

Conflicts of Interest: The authors declare no conflict of interest.
29. Wan, Y.; Sun, L.Y.; Zhao, P.; Wang, J.F.; Tu, S. Relation Classification Based on Information Enhanced BERT. *J. Chin. Inf. Process.* 2021, 35, 69–77.

30. Chen, T.; Chen, Y.; Lv, M.; He, G.; Zhu, T.; Wang, T.; Weng, Z. A Payload Based Malicious HTTP Traffic Detection Method Using Transfer Semi-Supervised Learning. *Appl. Sci.* 2021, 11, 7188. [CrossRef]

31. Rizwan, A.; Iqbal, N.; Ahmad, R.; Kim, D.-H. WR-SVM Model Based on the Margin Radius Approach for Solving the Minimum Enclosing Ball Problem in Support Vector Machine Classification. *Appl. Sci.* 2021, 11, 4657. [CrossRef]

32. Choi, S. Combined kNN Classification and Hierarchical Similarity Hash for Fast Malware Detection. *Appl. Sci.* 2020, 10, 5173. [CrossRef]

33. Oh, H.-J.; Syifa, M.; Lee, C.-W.; Lee, S. Land Subsidence Susceptibility Mapping Using Bayesian, Functional, and Meta-Ensemble Machine Learning Models. *Appl. Sci.* 2019, 9, 1248. [CrossRef]

34. Xu, X.; Tian, K. A Novel Financial Text Sentiment Analysis-Based Approach for Stock Index Prediction. *Quant. Tech. Econ.* 2021, 38, 124–145.

35. Leng, J.; Wang, D.; Ma, X.; Yu, P.; Wei, L.; Chen, W. Bi-level artificial intelligence model for risk classification of acute respiratory diseases based on Chinese clinical data. *Appl. Intell.* 2022, 1–18. [CrossRef]

36. Hsieh, Y.; Zeng, X. Sentiment Analysis: An ERNIE-BiLSTM Approach to Bullet Screen Comments. *Sensors* 2022, 22, 5223. [CrossRef]

37. Zhou, L.; Bian, X. Improved text sentiment classification method based on BiGRU-attention. *J. Phys. Conf. Ser.* 2019, 1345, 032097. [CrossRef]

38. Miao, Y.L.; Cheng, W.F.; Ji, Y.C.; Zhang, S.; Kong, Y.L. Aspect-based sentiment analysis in Chinese based on mobile reviews for BiLSTM-CRF. *J. Intell. Fuzzy Syst.* 2021, 40, 8697–8707. [CrossRef]

39. Feng, J.Y.; Wu, D.D.; Wang, B.; Wang, Z.; Mu, W.S. Online Comments Analysis and Its Application Research Progress in E-commerce of Fresh Agricultural Products. *Trans. Chin. Soc. Agric. Mach.* 2021, 52, 504–512.