Research on Short Video Hotspot Classification Based on LDA Feature Fusion and Improved BiLSTM

Linhui Li [1,2,3,†], Dan Dai [1,2,3,*,†], Hongjiu Liu [1,2,3], Yubo Yuan [1,2,3], Lizhong Ding [4] and Yujie Xu [1]

1 College of Mathematics and Computer Science, Zhejiang A&F University, Hangzhou 311300, China
2 Zhejiang Provincial Key Laboratory of Forestry Intelligent Monitoring and Information Technology, Hangzhou 311300, China
3 Key Laboratory of State Forestry and Grassland Administration on Forestry Sensing Technology and Intelligent Equipment, Hangzhou 311300, China
4 Agricultural and Forestry Technology Extension Center of Lin'an, Hangzhou 311300, China
* Correspondence: boatdriver@163.com
† These authors contributed equally to this work.

Abstract: Short video hot spot classification is a fundamental method to grasp the focus of consumers and improve the effectiveness of video marketing. The limitations of traditional short text classification are sparse content as well as inconspicuous feature extraction. To solve the problems above, this paper proposes a short video hot spot classification model combining latent dirichlet allocation (LDA) feature fusion and improved bi-directional long short-term memory (BiLSTM), namely the LDA-BiLSTM-self-attention (LBSA) model, to carry out the study of hot spot classification that targets *Carya cathayensis* walnut short video review data under the TikTok platform. Firstly, the LDA topic model was used to expand the topic features of the Word2Vec word vector, which was then fused and input into the BiLSTM model to learn the text features. Afterwards, the self-attention mechanism was employed to endow different weights to the output information of BiLSTM in accordance with the importance, to enhance the precision of feature extraction and complete the hot spot classification of review data. Experimental results show that the precision of the proposed LBSA model reached 91.52%, which is significantly improved compared with the traditional model in terms of precision and F1 value.

Keywords: feature fusion; self-attention; BiLSTM; hotspot classification

1. Introduction

Due to the influence of the COVID-19 epidemic, the traditional sales channels of agricultural products have been severely impacted, and the "short video + live broadcast" sales model with strong interaction and high participation is gradually replacing the traditional sales model. TikTok stands out among various short video platforms, on which 1.793 million special agricultural products have been sold nationwide [1]. When people watch short *Carya cathayensis* videos on the TikTok, they will have a topical discussion about the video content, including people's views on the video or *Carya cathayensis*-related content. The problem of how to discover the information contained in it and make use of the information to promote the realization of people's consumption behavior is what businesses are eager to solve at this stage.

User comments contain a large amount of user information and subjective emotional expression, but the sentences in user comments are short, context-dependent, colloquial, and noisy, which means it is difficult to classify the hot topics discussed by users. These comments need to be analyzed through data mining in order to understand the communication trend of video marketing and improve the quality of audience discussion hot spot analysis. Qi, J. [2] used the LDA model to achieve a good clustering analysis effect on online public opinion text data. Wu, D. [3] applied LDA model to short text clustering and feature

extraction, proving that LDA topic model plays an important role in text mining. Most current classification models are generally based on sentiment classification and large-scale datasets rough domain classification of online comments. Although contextual and local features are considered in the feature extraction of short texts, there are still some shortcomings in it. The LDA topic model can represent the topic information and global information contained in the text. LDA maps the higher-dimensional document-word vector space to the lower-dimensional document-topic and topic-word vector spaces. The potential semantic information of short text can be mined effectively. In order to further improve the classification accuracy of short text reviews, efficiently mine the opinions of individual product reviews, and learn the hot topics of discussion when people watching short videos and buy agricultural products. This paper chooses the *Carya cathayensis*, studies the topic hotspots contained in the comments by crawling the relevant comments on TikTok, and proposes a BiLSTM-self-attention hotspot classification model based on LDA feature fusion. The main contributions of this paper are as follows:

1. Data collection. Taking short videos containing *Carya cathayensis* as the main research object and TikTok platform as the main carrier, researchers crawled 16,282 comments data and built datasets based on them, enriching the research on the comment data of TikTok platform.
2. A feature fusion method based on the combination of LDA topic model and Word2Vec word vector is presented. Based on the word vector, this method integrated the topic features of the text, expanded the feature representation of the information of comments, and solved the problem that the sparse feature representation of short text is not obvious to some extent.
3. Construct a hot spot classification model based on BiLSTM-self-attention. Self-attention mechanism and BiLSTM were combined to extract features of comment information. Self-Attention can reduce the dependence of external parameters and make the model focus on the features of the text itself. The combination facilitated the extraction of more important features for classification.
4. Comparative experiment. The LDA-BiLSTM-self-attention model proposed in this paper (hereinafter referred to as LBSA model) was compared with convolutional neural network (CNN), LSTM, gated recurrent unit (GRU), BiLSTM, CNN-LSTM, GRU-CNN, LDA-CNN, LDA-BiLSTM, CNN-self-attention, BiLSTM-self-attention, and other models. Experimental results demonstrate that the proposed model can extract features more accurately and achieve a better classification effect.

2. Related Work

This section mainly describes related work from the following three aspects: short video research, topic models, and deep learning models.

2.1. Short Video Research

In recent years, many scholars at home and abroad have conducted research on the relationship between short video marketing and consumer behavior. Cheng et al. [4] studied that features of educational video communication, e.g., professionalism, familiarity, and attractiveness, in TikTok have positive effects on users' purchase intention, while interactivity does not. Jiaheng, Z et al. [5] found that the brevity and vividness of travel videos on TikTok have a positive impact on the reliability of the information. Its vividness, diversity, and playfulness have a positive impact on viewers' willingness to visit. Li [6] et al. took product discount, entertainment degree, and professional knowledge as independent variables and user trust and perceived value as mediating variables and constructed a mediation model to explore the direct and indirect effects of online celebrity characteristics on purchase intention. Gao, L [7] designed an influence prediction model based on the HetInf (heterogeneous neural network) to predict users' behavior. Hong, L [8] proposed to use reference factors such as the number of followers, likes and user works respectively, the quality of user work, behavior of following, commenting, and forwarding to measure

the influence of short video users comprehensively. According to the above research status, the model analysis of the relevant literature is rarely used in the short video research field, and studies mainly focus on multiple factors of the short video to assess the impact on consumer behavior and other aspects while the field of research on the comment content of short videos is still blank.

2.2. Topic Models

In the related research of natural language processing, topic classification has always been the key research object of domestic and foreign scholars, and the LDA topic model has been widely used as a mature method. Data sparseness is an obvious feature of short texts, which has always been considered to be the reason for the low classification precision of short texts. Therefore, in response to the low feature extraction efficiency of short texts, many scholars have improved the representation ability of short texts by methods of improving algorithms or model fusion. Shao, D.G [9] proposed a neural network topic model based on the framework of the Wasserstein Auto-Encoder, which is significantly better than general topic models on short text data sets. Luo, L. [10] combined LDA topic model with CNN with GRU and proposed an effective method to improve sentiment classification of short texts. Tan, X [11] added a deep pre-training task based on enhancing the BERT pre-training task and based on the LDA topic embedding optimization results, deeply integrated with the LDA model to dynamically present the fine-grained public sentiment of events. The above scholars use the combined model to enhance the ability of feature extraction and improve classification precision. In addition, the study found that ignoring the semantic importance of some feature words may lead to low classification precision. Therefore, Shao, D et al. [12] proposed a new and improved method of text representation by multiplying the "topic-document" matrix theta and the "word-topic" matrix phi, which achieved good results in the classification of Chinese news. Wang, B [13] proposed a strong feature dictionary method based on LDA and Information Gain model, in which the classification precision is improved by assigning more weight to the feature items. Zhou, W [14] proposed an unsupervised text representation method that uses the word embedding technique Word2Vec to obtain word vectors, which are then combined with the feature-weighted TF-IDF and the topic model LDA. This method not only improves the expressiveness of the vector space model, but also reduces the dimensionality of the document vector.

Improvements, such as enhanced feature representation and improved comprehensiveness of feature extraction, have been made to perform feature extraction in classification studies. However, there is still the problem of inconspicuous sparse feature representation. The LDA model uses an efficient probabilistic inference algorithm to process large-scale data. Still, the comment data are short, semantically sparse, and with insufficient co-occurrence information compared to large articles. Moreover, classifying them using only the LDA model will lead to inaccurate classification results.

2.3. Deep Learning Models

Long short-term memory (LSTM), BiLSTM, GRU, and CNN have been widely used in various natural language processing tasks, but each has its advantages and disadvantages. CNN [15] uses convolutional layers and max-pooling or max-timeout pooling layers to extract higher-level features, but easily ignores contextual semantic features. LSTM [16] can capture long-term dependencies between word sequences, but cannot take into account local features. The GRU [17] model can solve the problem of long distance dependence to a certain extent, but if the input data is too much, GRU will also cause the problem of long distance forgetting due to gradient disappearance. The BiLSTM [18] model can capture bidirectional dependencies and capture context features more accurately. Zhao [19] proposed a CNN-LSTM to improve text sentiment analysis, which takes account of both contextual and local features. Deng, L. [20] proposed a combined GRU-CNN model, which proved that the combined model had better classification effect than a single model. With

the rise of the attention mechanism, many scholars have added the attention mechanism to the LSTM to focus on important features to make up for the shortcomings of LSTM in feature extraction. Yang Jing [21] proposed an attention-based LSTM neural network that integrates social context knowledge and textual information through modeling. Wu Peng and Li Xiaotong [22] and Liu, G. [23] among others have also proposed a sentiment classification method based on attention and bi-directional long short-term memory (BiLSTM), and the effectiveness of the method has been demonstrated through experiments. Research on text classification using deep learning has mainly focused on how to extract features by better using neural networks, without considering that different words in a text contribute differently to the semantics and that different features have different effects on the classification model.

Therefore, this paper focuses on studying the comments under TikTok short videos to explore where consumers' concerns lie when watching short videos, enriching the study of comment data on the TikTok platform. A feature fusion method based on the combination of the LDA topic model and Word2Vec word vector is proposed in feature extraction. The method incorporates the topic features of the text based on the word vector, extends the feature representation of the comment information, and solves the problem of inconspicuous sparse feature representation of the short text to some extent. The fused features are input into a BiLSTM improved by self-attention, which reduces the dependence on external parameters and allows the model to focus on the text's features. Combining it with the BiLSTM facilitates better extraction of more important features for classification. After comparative experiments, the superiority of the method for topic classification was proven.

3. Models and Improvements

3.1. Overall Model Framework

In this paper, a bi-directional long and short-term memory network model based on feature fusion with the self-attentive mechanism optimization—LBSA model is proposed to enhance the feature representation of short texts and incorporate important thematic features. The improved model can retain the sparse features of short texts more completely and better extract the effective features of important texts. The LBSA model architecture mainly includes the following parts: pre-process, feature fusion, Bi-LSTM, Self-Attention, and softmax. The model structure is shown in Figure 1.

3.2. Feature Fusion Method Based on LDA Topic Model

3.2.1. LDA Topic Model

The LDA topic model was first proposed by BLEI [24] in 2003 as a model for document topic generation based on latent semantic analysis with a Dirichlet prior distribution. In the Bayesian network structure of the LDA topic model, for any text d, its topic distribution θ_d is:

$$\theta_d = \text{Dirichlet}\,(\vec{\alpha}) \tag{1}$$

where α is the hyperparameter of the topic distribution, it is a k-dimensional vector. For any topic k, the distribution of the subject term β_k is:

$$\beta_k = \text{Dirichlet}\,(\vec{\eta}) \tag{2}$$

where η is the hyperparameter of the topic distribution, and also a V-dimensional vector. V represents the number of all words in the vocabulary. For the nth word in any text d, the distribution Z_{dn} of its topic number obtained from the topic distribution θ_d is:

$$Z_{dn} = \text{multi}(\theta_d) \tag{3}$$
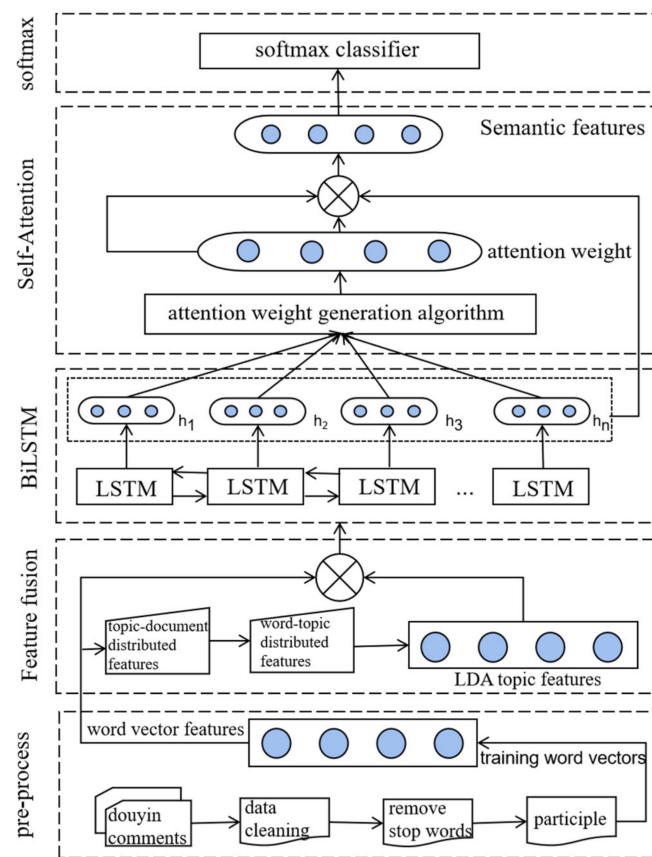
Figure 1. LBSA model structure diagram.

For this topic number, the probability distribution of the word W_{dn} can be obtained as:

$$W_{dn} = \text{multi}\left(\beta_{z_{dn}}\right) \tag{4}$$

The construction of the LDA topic model requires the setting of the number of topics. After the LDA topic model was proposed, perplexity [25] was initially used to determine the optimal number of topics. Under different numbers of topics, the smallest the perplexity is the optimal number of topics. However, the number of topics determined by this method is generally larger and the similarity between topics is also larger, which is more suitable for long text data. The text of the *Carya cathayensis* comments crawled in this paper is short and independently distributed, so it is not suitable for this method to determine the optimal number of topics. Coherence [26] is often used to evaluate the interpretability of the generated potential topics. When a topic is easy to interpret, the words ranking high under that topic will frequently co-occur in the corresponding text corpus. Therefore, a higher coherence index means that the number of topics is better. The Formula (5) of topic coherence [27] is as follows:

$$C\left(t; V^{(t)}\right) = \sum_{m=2}^{M} \sum_{l=1}^{m-1} \log \frac{D\left(v_m^{(t)}, v_l^{(t)}\right) + 1}{D\left(v_l^{(t)}\right)} \tag{5}$$

where is the number of M text, $D(v)$ denotes the frequency of occurrence of the phrase v in the text set, and $V^{(t)}$ denotes the M most likely words in the topic t.

3.2.2. Fusion Method of Word Vector and LDA Topic Model

Yan, D, et al. [28] fused LDA document-topic distribution features with weighted Word2Vec word vector features to construct short text features in Weibo and performed

topic clustering based on the K-means algorithm, which exploited the LDA document-topic distribution features. Tang, H, et al. [29] used the LDA topic model word-topic distribution to calculate the topic similarity between words and their contextual words, which was incorporated into the word vector as topic semantic information. The above papers all use the documents generated by the LDA topic model to improve the word vector. In this paper, based on the above papers, the document-topic distribution and word-topic distribution of the LDA topic model are used, combined with the Word2Vec word vector to obtain the topic extension features, as follows.

First, the document-topic distribution characteristics can be obtained according to the trained LDA topic model, and then the document d_i under the maximum probability topic Z_{max} is selected. According to the topic Z_{max} under the topic word distribution, the top k subject words $(c_1, c_2, c_3, \cdots, c_k)$ and their probability $(p_1, p_2, p_3, \cdots, p_k)$ can be obtained, and the probability value should be normalized to calculate the weight of k words. The selected k in this paper is 30, as shown in Formula (6):

$$q_i = \frac{p_i}{\sum_{c=1}^{k} p_c} \tag{6}$$

where q_i denotes the normalized value of p_i and $(q_1, q_2, q_3, \cdots, q_k)$ is the weight size of the previous k words.

Second, the Word2vec model is used to obtain the word vector of each word $(X_1, X_2, X_3, \cdots, X_n)$, with the maximum probability topic Z_{max} in each document d_i as the key, and if the words in the document match the words under topic Z_{max}, the word vector of n words is weighted and summed to obtain the topic extension feature, as shown in Formula (7):

$$Y_i = \sum_{j=1}^{n} q_j \times X_j \tag{7}$$

where Y_i is the topic extension feature of document d_i. Finally, the Word2vec word vector features are stitched with the topic extension features as the input of the BiLSTM model. The specific process is shown in Figure 2 below. The method solves the problems of inconspicuous topic boundaries and sparse features caused by short comment contents and improves the precision of topic classification in this paper.

3.3. BiLSTM

The traditional recurrent neural network (RNN) has the problem of long-time dependence of sequence data when processing a large amount of data, that is, the problem of gradient dispersion when processing nodes with relatively distant time series. The memory cells in the LSTM [30] model can store the information of data with relatively long time series and transmit information through the gate structure with only a small amount of linear interactions. It reduces the loss of data, enabling the storage cells to effectively store the semantic information of longer sequences, which effectively solves such problems mentioned above. There are three gate structures in the LSTM: input gates i_t, forget gates f_t, and output gates O_t. At moment t, the input to the LSTM is x_t, h_{t-1}, and the output is h_t, y_t. First, read the input at moment t and use the sigmoid function to calculate the information lost by the forget gate cells:

$$f_t = \sigma \left(W_f \cdot [h_{t-1}, x_t] + b_f \right) \tag{8}$$

The input gate is used to determine which information to update in the cell. First, calculate the candidate information of the cell:

$$\widetilde{M_t} = \tan h \left(W_m \cdot [h_{t-1}, x_t] + b_m \right) \tag{9}$$
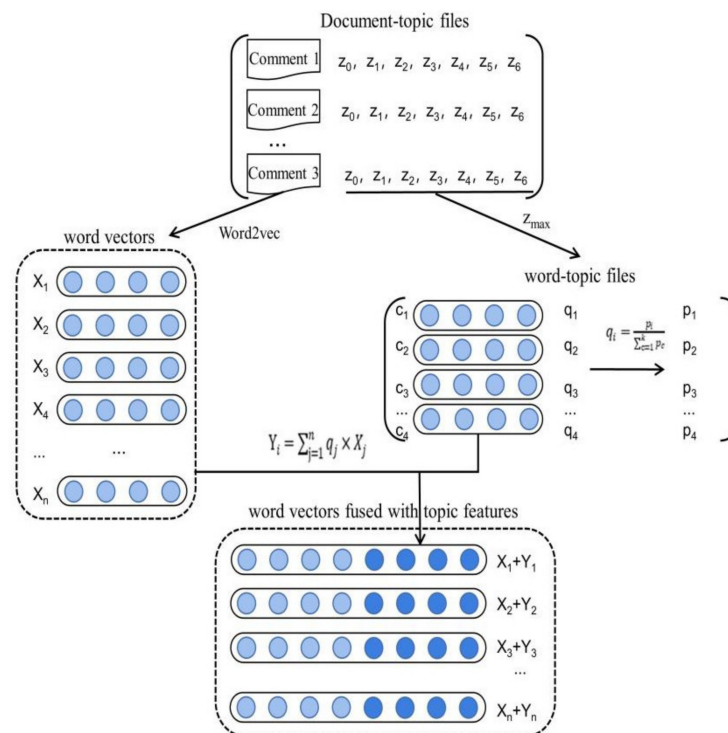
Figure 2. Flow chart of topic feature fusion.

Calculate what information to use for cell update:

$$i_t = \sigma \left(W_i \cdot [h_{t-1}, x_t] + b_i \right) \tag{10}$$

Use f_t multiplied by the old cell state M_{t-1} plus the new cell candidate information to complete the update of the cell information:

$$M_t = f_t \times M_{t-1} + i_t \times \widetilde{M_t} \tag{11}$$

Use the output gate to determine the cell state information to be output:

$$O_t = \sigma \left(W_o \cdot [h_{t-1}, x_t] + b_o \right) \tag{12}$$

Finally, use the tanh function to calculate the output cell information:

$$h_t = O_t \times \tan h M_t \tag{13}$$

where b_f, b_m, b_i, b_o, etc. represent the bias function, W_f, W_m, W_i, W_o, etc. represent weights, and σ represents the sigmoid function. BiLSTM concatenates the forward and backward outputs of the LSTM at moment t:

$$h_t = [\overrightarrow{h_t}; \overleftarrow{h}] \in R \tag{14}$$

In this paper, the BiLSTM takes the feature fusion vector of the previous layer as input and assigns random initialization to the weight W and bias function b. The LSTM unit calculates the values of forget gate f_t, input gate i_t, and output gate o_t based on the hidden layer information $h_t - 1$ of the previous moment and the input information x_t of the current moment. It then integrates them with the storage unit information $c_t - 1$ of the previous moment to obtain the unit output h_t of the current moment while updating the hidden layer information and the storage unit information as input to the LSTM unit of the next moment. Finally, the output vectors of the LSTM units of the previous and next moments are spliced as the output of the BiLSTM [31] with bi-directional semantic features.

3.4. Self-Attention

The BiLSTM can output context-related semantic information but cannot highlight the importance of this information in the text. The attention mechanism is widely used in image processing, machine translation, sentiment analysis [32], etc. because it can extract important features of objects in sparse features. Adding an attention mechanism after the BiLSTM model can effectively reflect the importance of semantic information in text and enhance the feature expression in semantic information. The Attention function is essentially a mapping function consisting of multiple Query and Key-Value. The calculation process is mainly divided into the following three steps:

1. Calculate the similarity between the query and each key to obtain the corresponding weights, as shown in Formula (15):

$$f\,(Q,K) = QK^T \tag{15}$$

2. Normalize the weights with the softmax function, as shown in Formula (16):

$$a_i = \text{softmax}\,(f(Q,K)) \tag{16}$$

3. The weights and the corresponding key value are weighted and summed to obtain the final Attention value, as shown in Formula (17):

$$\text{Attention}(Q,K,V) = \sum a_i V \tag{17}$$

In this paper, the model is optimized using Self-Attention, which is an improvement of the attention mechanism that reduces the reliance of the network on external information and is better at capturing the relevance of the data or features. As shown in Formula (18):

$$\text{Attention} = \text{softmax}\left(\frac{X \cdot X^T}{\sqrt{d_k}}\right) X \tag{18}$$

In Formula (18), $X \in R^n$, X is the n-dimensional output vector of output unit, $\sqrt{d_k}$ is the regulatory factor, and this paper refers to the dimensionality of the input word vector. Its function is to avoid a situation where the softmax value is exactly 0 or 1 due to too large inner product of $X \cdot X^T$. Input the topic feature fusion vector into the BiLSTM-Self-Attention model, use the hidden layer state vector H obtained by the BiLSTM as the query, and use the hidden layer state vector hi output in each time step of the BiLSTM as the Key-Value. For example, Formulas (15–17) calculates it to obtain the weighted feature vector, which is finally input to the softmax layer for classification. The structure diagram of the BiLSTM-Self-Attention model is shown in Figure 3.

3.5. Softmax

A fully connected layer is added after the Self-Attention layer to map the weighted feature vectors output from the Self-Attention layer into the label space of the topic category, and a dropout layer is introduced to prevent overfitting caused by weight updates relying on only part of the features. Finally, the output results are input into the softmax layer for classification. Set the softmax cross loss function as the loss function of the model:

$$\text{loss} = (y, \hat{y}) = -\sum_{i=1}^{k} y_i \ln \hat{y}_i \tag{19}$$

where y is the probability of the true label, and $\hat{y}$ is the probability of the normalized label obtained after softmax calculation.
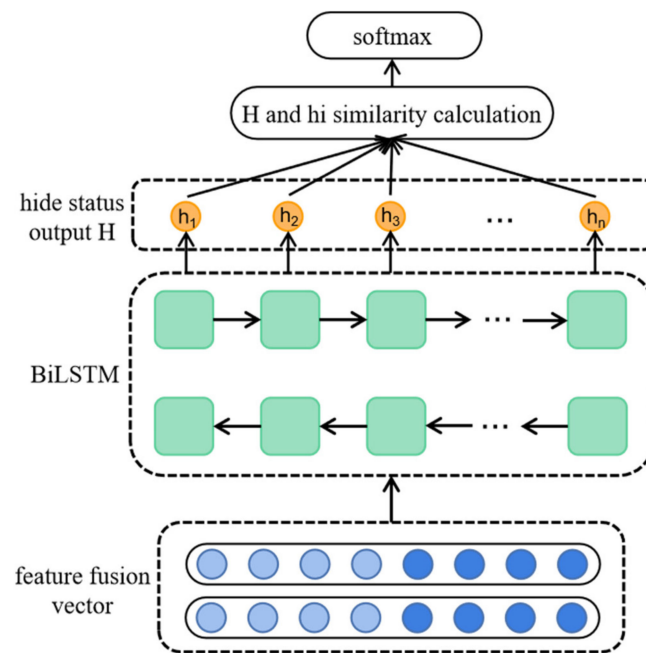
Figure 3. Structure of BiLSTM-Self-Attention model.

4. Experiments and Results

4.1. Evaluation Indicators

In this paper, the precision and recall rate are used to evaluate the classification effect of the model, and the *F*1 value is used to measure the overall classification effect. The formulas are as follows:

$$P = \frac{TP}{TP + FP} \tag{20}$$

$$R = \frac{TP}{TP + FN} \tag{21}$$

$$F1 = \frac{2PR}{P + R} \tag{22}$$

where *TP* represents the number of correct predictions as correct samples, *FP* represents the number of samples that are incorrectly predicted as the correct, and *FN* represents the number of samples correctly predicted as incorrect samples.

4.2. Data Acquisition and Processing

This experiment used the TikTok short video platform as the main research object, crawled the video comment data with *Carya cathayensis* products as the core, and obtained all comments in videos about *Carya cathayensis* from April 2021 to April 2022, a total of 16,282. The original comments contain a large number of comments, numbers, emoticons, and other noise unrelated to *Carya cathayensis*, which will affect the subsequent classification effect and needs to be removed. In addition, the comments crawled have many Spaces or useless symbols. If these symbols are retained, these symbols will be separated during word segmentation, which will lead to bad results. Therefore, it is necessary to delete the Spaces in the comments. As the main source of the dataset, 8242 comment data were left after eliminating irrelevant comments.

4.3. Splitting Words

For the original comment dataset, the exact mode of the jieba word-segment algorithm is employed in this paper to segment the data and load a custom domain dictionary to obtain better segmentation results. Combined with the stop word list of Harbin Institute of Technology and the stop word list of Baidu, a stop word dictionary was created to filter the

stop words that appear frequently but have little meaning for analysis to form the original corpus. The word segmentation results are shown in Table 1:

Table 1. Results of word segmentation.

Before the Segmentation of Words	After the Segmentation of Words
Do you have pecans without their skins on	Pecans without skin
You have many varieties. Can you mix them? I'll buy a few pounds	many varieties mix a few pounds
I've bought it many times. It's delicious	bought many times delicious
The order has been placed. When will it be delivered	Have ordered delivery
I placed an order, hoping to buy the same as yours. Don't be disappointed	placed an order hope buy disappointed
I just wanted to buy nuts, and when is it live?	want buy nuts live

4.4. Training Word Vectors

The traditional method of word vector transformation is one-hot encoding, which converts the text into a long vector containing only 1 and 0. It uses the size of the vocabulary as the dimension to represent the word vector, which can easily cause problems of high-dimensional sparsity and dimensionality explosion. The encoding does not take into account the semantic relationship of the context. Word2Vec [33] solves such a problem. The model performs unsupervised learning on large-scale corpora in the field. There are two main implementations: the CBOW model and the Skip-gram model. Skip-gram uses the conditional probability of one word to predict surrounding words, and the CBOW model works in the opposite way. In this paper, the Skip-gram model was used to generate word vectors and load the word embedding lookup table. The lookup table maps word vectors to word sequences from the original corpus and continuously learns and updates in the process, resulting in a more complete and authentic semantic vector representation of the target vocabulary. The word vector for text generation was 100 dimensions and the contextual word window was set to 5.

4.5. Topic Classification

4.5.1. Optimal Topic Number Confirmation

Input the trained Word2Vec word vector into the LDA topic model, set the appropriate prior distribution hyperparameters α (topic distribution), β (word distribution), and the number of topics and the number of iterations that affect the Gibbs sampling algorithm. This paper used the Gensim to perform LDA topic mining and Coherence to determine the optimal number of topics K. The parameter settings for building the LDA topic model are shown in Table 2.

Table 2. LDA topic model parameter settings.

Parameter	Parameter Value
Hyperparameter α	50/k
Hyperparameter β	0.01
number of iterations I	30
Number of Topics k	7

The consistency was analyzed by setting the different numbers of topics to find the optimal number of topics among 8242 comments, and the specific results are shown in Figure 4. When K is taken as 7, the consistency value of the LDA topic model is the largest, and then Coherence increases.
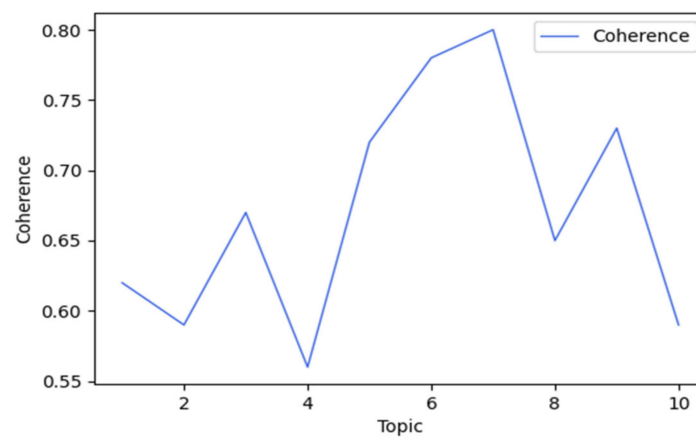
Figure 4. Topic coherence results.

To verify the effect of Coherence results, this paper analyzed the classification results by plotting the topic visualization rendering, calling pyLDAvis [34] to visually present the distance between topics and plotting pyLDAvis diagram with the number of topics as 7. As shown in Figure 5, the left side of the figure is based on the MDS algorithm to display the topic in the two-dimensional space, with each dot representing a topic and the numbers on the dot representing the topic serial numbers. On the right are the corresponding words under the topic, arranged in descending order of probability. When K = 7, the topic bubbles are evenly distributed in the graph without overlapping parts, which intuitively shows that the classification effect is better when the optimal topic number is 7.
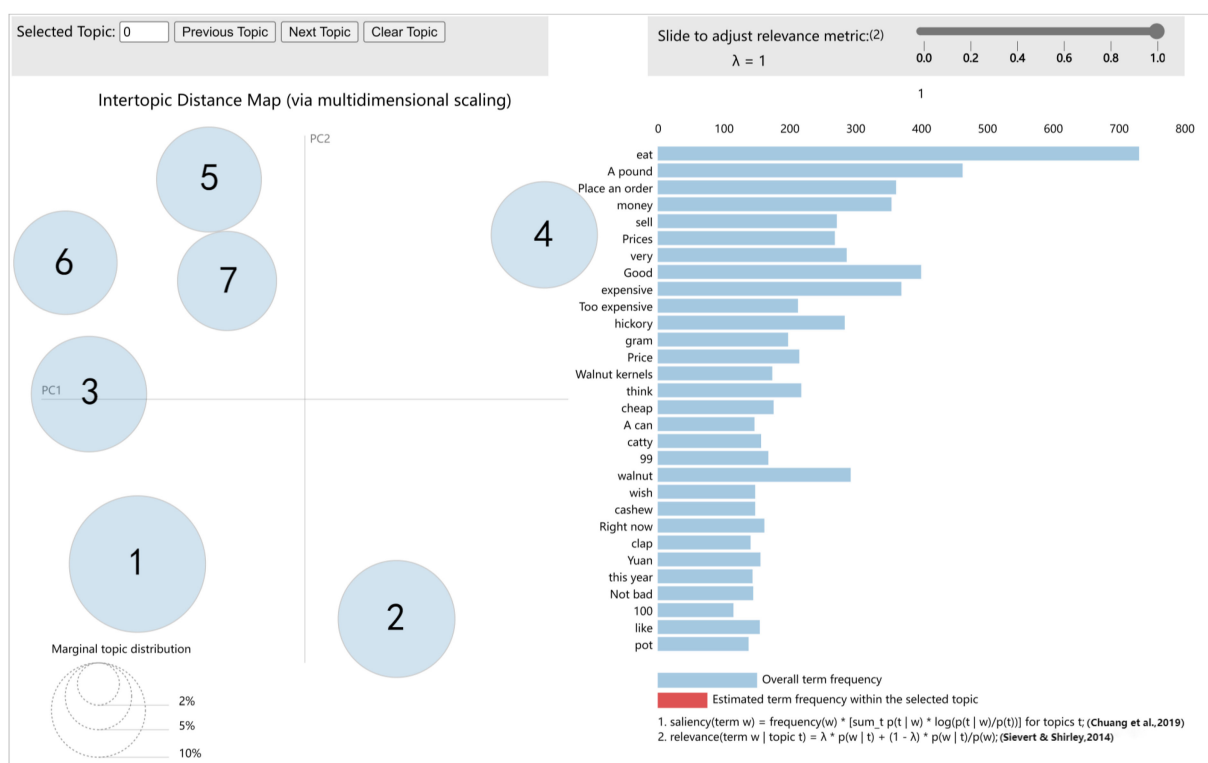
Figure 5. Best topic visualization. (In the figure, formula 1 refers to the literature [35], formula 2 refers to the literature [36].)

4.5.2. Classification Tags

When the optimal number of topics is 7, the LDA topic model trains the text to obtain a word-topic file, which includes topics and the corresponding words and their probabilities

under each topic. Select the top ten words in each topic with the highest contribution rate to summarize the topic. According to the word-topic file, the seven categories of topics were roughly summarized as "Buy", "Weight", "Delicious", "Taste", "Hand Peeled", "Quality", and "Price", and the comments of the data in this paper are manually classified according to these seven categories of tags. As shown in Table 3, the number of documents represents the number of comments belonging to this category, and the percentage represents the proportion of this category of texts in the total number of texts.

Table 3. Carya cathayensis review word-topic distribution and classification results.

Topic	Top 10 Probability Words	Tags	Number of Documents	Percentage (%)
Topic 0	Buy, order, hope, delicious, beat, delivery, support, want, have, received	Buy	1358	16.4
Topic 1	Money, 1catty, gram, pot, block, 100, net weight, yuan, catty, sell	Weight	1492	18.1
Topic 2	Delicious, buy, good, very, nice, really, repo, pecan, bought, really	Delicious	1309	15.8
Topic 3	Eat, pecan, buy, like, original taste, fragrant, taste, too sweet, want, salt and pepper	Taste	1098	13.3
Topic 4	Peel, eat, no, desiccant, like, no, shell, jar, difficult, hand	Hand Peeled	900	10.9
Topic 5	Walnut, pecan, buy, this year, eat, small, new goods, broken, Lin 'an, old goods	Quality	945	11.4
Topic 6	Price increase, expensive, buy, 99, too expensive, price, a little, now, yuan, before	Price	11.41	13.8

4.5.3. Topic Feature Integration

The document-topic file and the word-topic file are obtained through the LDA topic model calculation. Take the comment "This is delicious, how many grams of 49 yuan. When will the order be delivered now" as an example. The maximum probability topic Z_{max} corresponding to the comment in the document-topic file is topic 0, and the probability value of the top 30 words under topic0 is normalized to obtain the weight of the 30 words. In the sample comments, there are three words that match the words under topic 0 after the segmentation, namely "delicious", "order", and "delivery". The word vectors of these three words and their weights were multiplied as their topic extension features. Finally, the original word vectors of these three words were segmented with the topic extension feature vectors.

4.6. *Experimental Results*

4.6.1. Comparison of Model Results

The LDA topic extension features were obtained according to the fusion method mentioned in 1.2.2 above. Spliced with the word vector, they were input to the BiLSTM in this paper. The context related information was extracted using the BiLSTM. The weights were assigned by adding the self-attention mechanism within each channel, and finally input to the softmax classifier, and thus better classification results were obtained. To verify the classification effect of the model in this paper, the LBSA model was compared with other single models and combined models. The models compared with it were CNN, LSTM, GRU, BiLSTM, CNN-LSTM, GRU-CNN, LDA-CNN, LDA-BiLSTM, CNN-Self-Attention, and BiLSTM-Self-Attention. The CNN in the experiment adopted a 3-layer model architecture with word vector convolution windows set to 3, 4, and 5 respectively. The LSTM time step was set to 20, the vector dimension was 100, the number of cells was 128, the dropout was 0.5, the batch sample size was 32, the number of training rounds was 30, and the optimizer was Adam. The GRU the vector dimension was 100, the dropout was 0.5, the batch sample size was 32, the number of training rounds was 30. Due to the small data set, K-fold cross validation [37] was applied in the experiment. The whole data set is divided into k equal subsets, and the algorithm is repeated k times. One subset is randomly selected each time

as the test set, and the remaining sample set is used for training [38]. In this paper, k = 10 is selected. The average value of the results of multiple 10-fold cross-validation is taken as the final result of the model. After several experiments, the comparison results of single model are shown in Table 4, and comparison results of the combined models are shown in Table 5. The comparison of the accuracy rate of cross-validation of the combined model with ten folds is shown in Figure 6.

Table 4. Single model comparison results.

Model	Precision/%	Recall/%	F1/%
CNN [15]	87.66	86.61	87.13
LSTM [16]	86.43	85.12	85.77
GRU [17]	84.89	83.02	83.94
BiLSTM [18]	87.39	87.02	87.20

Table 5. Combined model comparison results.

Model	Precision/%	Recall/%	F1/%
CNN-LSTM [19]	87.61	86.23	86.91
GRU-CNN [20]	86.12	84.71	85.41
LDA-CNN	88.07	87.48	87.77
LDA-BiLSTM	88.92	86.36	87.62
CNN-Self-Attention	88.73	86.90	87.80
BiLSTM-Self-Attention	89.49	88.42	88.95
LBSA	91.65	90.17	90.90
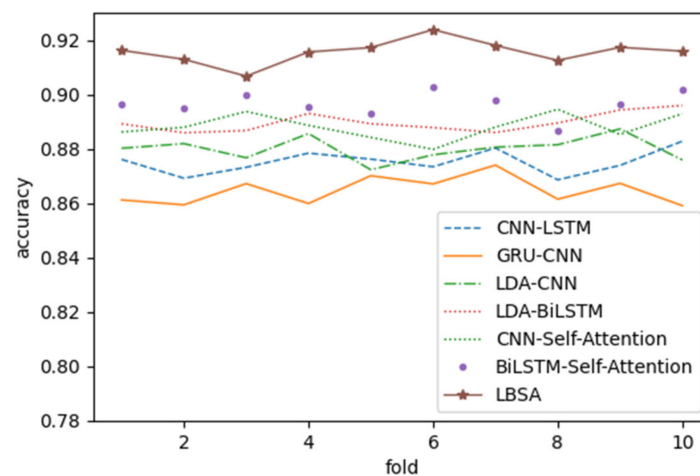
Figure 6. Comparison of Accuracy in 10-fold Cross-validation.

From the experimental results in Table 4, it can be seen that the single model achieved excellent results in the evaluation index results. Analysis of the experimental results led to the following conclusions:

1. Comparing the CNN with the LSTM, the performance of the CNN is significantly better than that of the LSTM. The reason for this is that the dataset crawled in this paper contains short texts targeted at the sales, quality, and price of videos about *Carya cathayensis*, which is not firmly contextualized and has sparse text features. The CNN is good at capturing local features in text, while LSTM is good at capturing contextual features, so the precision of the CNN was 1.23% higher than that of the LSTM.

2. Comparing the LSTM with the GRU, the accuracy of GRU is about 1.54% lower than that of LSTM. GRU is as effective as LSTM in capturing semantic associations between long sequences, but its structure and calculation are simpler than LSTM, and it may not be as effective as LSTM in processing small amounts of short text.

3. Comparing the BiLSTM with the CNN and the LSTM, the performance of the CNN is not much different from that of the BiLSTM. The BiLSTM doubled the computational complexity of the model by stitching the output of the LSTM in both directions at the same time, which improved the precision of the BiLSTM by approximately 0.96% compared with the LSTM.

From the experimental results in Table 5 and Figure 6, the accuracy of the combined model fluctuates but tends to remains fixed after 10-fold cross-validation. It can be seen that the feature fusion-based LBSA model proposed in this paper achieved excellent results in the evaluation index results with a precision rate of 91.65%, and the precision and recall rate improved by 2~5% compared with other models. Compared with other models, the following conclusions can be drawn:

1. Comparing the CNN-LSTM model and GRU-CNN model mentioned in the literature review with the LBSA model in this paper, the accuracy of LBSA is significantly higher than the two models. CNN-LSTM model and GRU-CNN model have good classification effect in other papers, but the classification effect in this paper is not so good. The reason may be that there are big differences between different data sets and different experimental parameters.

2. Comparing the LDA-CNN model with the LDA-BiLSTM model, the accuracy of the LDA-BiLSTM model is 0.85% higher than that of the LDA-CNN model. The accuracy of the feature fusion model is higher than that of the single model, which indicates that feature fusion has great influence on short text classification.

3. Comparing the CNN-Self-Attention model with the BiLSTM-Self-Attention model, the introduction of the Self-Attention mechanism on the basis of the CNN and the BiLSTM led to a significant improvement in the classification effect of the model. The precision was improved by 1.07% and 2.1% compared with the two models. The attention mechanism gave different weights to different text features, reducing the influence of noise on the classification, and helped to focus on the key features that affect the classification effect of short text, improving the classification model precision. The BiLSTM-Self-Attention model was slightly more effective than the CNN-Self-Attention model, so it was chosen to make further improvements.

4.6.2. Impact of Feature Fusion on the Model

In order to accurately characterize the features of short text data and further improve the performance of the BiLSTM-Self-Attention model, the LDA topic model with excellent performance was integrated into feature extraction. The topic extension features were obtained by merging the document-topic distribution and word-topic distribution of the LDA topic model with the Word2Vec word vector. The BiLSTM-Self-Attention model was compared with the LBSA model after feature fusion and the confusion matrix was drawn, as shown in Figure 7:

In Figure 7, (a) is the confusion matrix diagram of the BiLSTM-Self-Attention model test set results and (b) is the confusion matrix diagram of the LBSA model test set results after feature fusion. As can be seen from Figure 7, the classification effect of the LBSA model has been improved compared with the BiLSTM-Self-Attention model as a whole. For topic 0, the improved model reduced the number of comments that incorrectly classify topic 0 as topic 2, 3, and 6. For topic 4, the improved model has 125 correct classifications, 6 more than those in the BiLSTM-Self-Attention model, and the number of wrongly classified comments for topic 2 had changed from 6 to 3. Experiments demonstrate that the fusion of extended features based on the LDA topic model reduces the error rate caused by the inconspicuous feature boundaries, making it easier for the BiLSTM-Self-Attention model to detect important features of the text and better classification results.

(**a**)   (**b**)
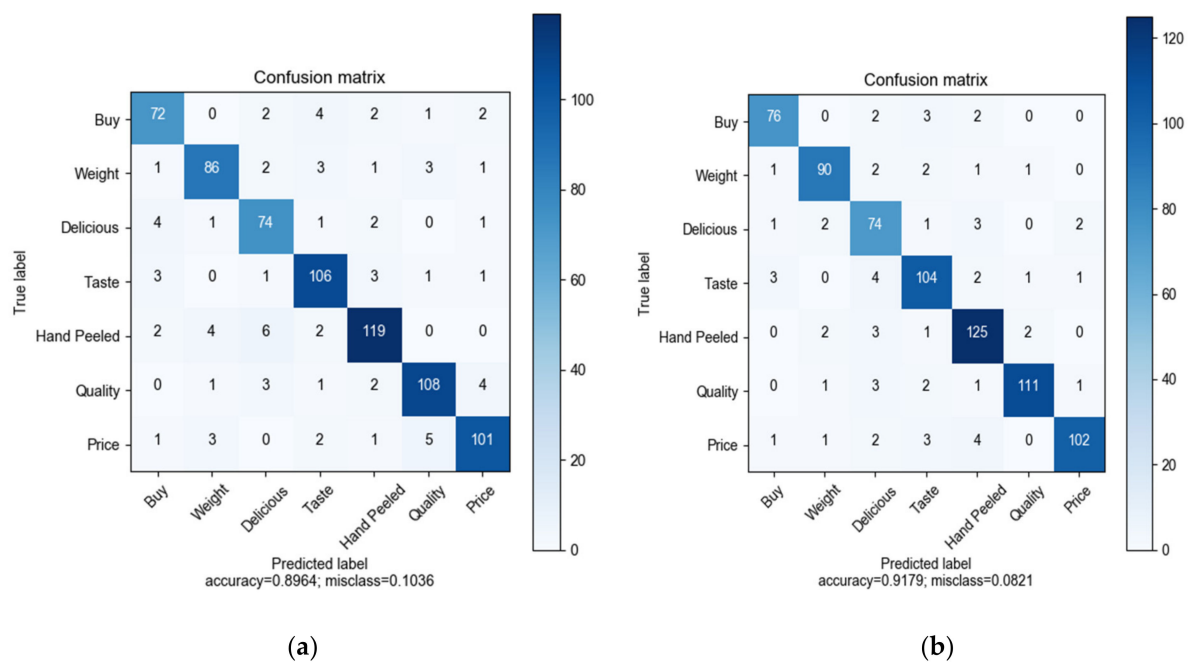
Figure 7. Confusion matrix comparison chart. (**a**) the confusion matrix diagram of the BiLSTM-Self-Attention. (**b**) the confusion matrix diagram of the LBSA.

5. Conclusions

To address gap in the research field of short video comment content and the problems of sparse short text data and incomplete feature extraction, this paper crawled all the comments in videos about *Carya cathayensis* from April 2021 to April 2022 on the TikTok platform as the research object, combined the LDA topic model with the Word2Vec method, and thus proposed a feature fusion method based on short text topic vector. After that, the fused word vectors were input into the BiLSTM for backward and forward text semantic modeling to obtain the high-level feature expressions of text sequences. The Self-Attention was used to assign certain weights to important features to reduce the influence of features of noise. Considering that the weight update would lead to reliance on some features causing model overfitting, the dropout layer was added to randomly deactivate some parameters and finally input to the softmax classifier to achieve the purpose of classification for short video comment hotspots. The specific findings of the study are as follows:

1. A study using TikTok short video comments as a dataset was proposed. Crawling comments in TikTok short video about *Carya cathayensis* as the original data for short video research expands the ideas for short video marketing research. It fills the gap in the field of short video comment research.

2. The method of feature fusion based on the combination of the LDA topic model and Word2Vec word vector was proposed. The method incorporated the topic features of the text based on the word vector, extended the feature representation of the comment information, and solved the problem of inconspicuous sparse feature representation of short text to some extent.

3. The short text classification model based on the feature fusion–LBSA model was proposed. It had a better classification effect for the short video comments of *Carya cathayensis*, with a precision of 91.65%, which was 2~5% higher than CNN, LSTM, GRU, BiLSTM, CNN-LSTM, GRU-CNN, LDA-CNN, LDA-BiLSTM, CNN-Self-Attention, and BiLSTM-Self-Attention models. The research of this method can help merchants quickly extract the discussion hotpots of consumers in videos from many comments and provide new decision information for *Carya cathayensis* merchants when posting videos and selling *Carya cathayensis*.

The feature fusion rules set in this paper are simple, and the impact of potential topics on the classification effect is not considered. In addition, the current data set has a single domain, and each comment only contains one kind of hotspot. Therefore, the next research direction is to enhance the comprehensiveness of theme feature fusion, expand the data set, expand the data set in different fields, and identify multiple hot spots in a review.

Author Contributions: Conceptualization, L.L. and D.D.; methodology, L.L. and D.D.; software, L.L.; validation, L.L.; formal analysis, L.L., H.L, Y.Y., L.D. and Y.X.; investigation, L.L., D.D., H.L., Y.Y., L.D. and Y.X.; data curation, L.L.; writing—original draft. L.L.; resources, D.D. and H.L.; writing—review & editing, D.D. and H.L.; visualization, D.D., H.L. and Y.Y.; supervision, D.D. and H.L.; project administration. D.D. and H.L.; funding acquisition, D.D., H.L. and L.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Youth Science Foundation Project under grant number 42001354, the Ministry of Education Humanities and Social Sciences Research Program Fund Project under grant number 18YJA630030 and 21YJA630054, and the Basic Public Welfare Research Program of Zhejiang Province under grant number LGN18C130003.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Douyin Data Report. Available online: https://xw.qq.com/amphtml/20220111A0AWP600 (accessed on 11 January 2022).
2. Qi, J.; Xun, L.; Zhou, X.; Li, Z.; Liu, Y.; Cheng, H. Micro-blog user community discovery using generalized SimRank edge weighting method. *PLoS ONE* **2018**, *13*, e0196447. [CrossRef]
3. Wu, D.; Yang, R.; Shen, C. Sentiment word co-occurrence and knowledge pair feature extraction based LDA short text clustering algorithm. *J. Intell. Inf. Syst.* **2020**, *56*, 1–23. [CrossRef]
4. Cheng, F.E.I.; Seo, L.J. The Impact of the Educational Influencer Characteristics of the Short Video App Tik Tok on the Intention to Purchase Online Knowledge Content. *J. Brand Des. Assoc. Korea* **2021**, *19*, 77–94.
5. Jiaheng, Z.; Choi, K. The Effect of Tourism Information Quality of TikTok on Information Reliability and Visit Intention: Focusing on Moderating Effects of Homogeneity. *Northeast. Asia Tour. Res.* **2022**, *18*, 1–21.
6. Li, Z.X.; Zhang, S.M.; Bin Liu, H.; Wu, Q.T. Study on the Factors Influencing Users' Purchase Intention on Live-Streaming E-Commerce Platforms: Evidence from the Live-Streaming Platform of TikTok. *J. China Stud.* **2021**, *24*, 25–49. [CrossRef]
7. Gao, L.; Wang, H.; Zhang, Z.; Zhuang, H.; Zhou, B. HetInf: Social Influence Prediction With Heterogeneous Graph Neural Network. *Front. Phys.* **2022**, *9*, 787185. [CrossRef]
8. Hong, L.; Yin, J.; Xia, L.-L.; Gong, C.-F.; Huang, Q. Improved Short-video User Impact Assessment Method Based on PageRank Algorithm. *Intell. Autom. Soft Comput.* **2021**, *29*, 437–449. [CrossRef]
9. Shao, D.; Li, C.; Huang, C.; An, Q.; Xiang, Y.; Guo, J.; He, J. The short texts classification based on neural network topic model. *J. Intell. Fuzzy Syst.* **2022**, 1–13. [CrossRef]
10. Luo, L.-X. Network text sentiment analysis method combining LDA text representation and GRU-CNN. *Pers. Ubiquitous Comput.* **2019**, *23*, 405–412. [CrossRef]
11. Tan, X.; Zhuang, M.; Lu, X.; Mao, T. An Analysis of the Emotional Evolution of Large-Scale Internet Public Opinion Events Based on the BERT-LDA Hybrid Model. *IEEE Access* **2021**, *9*, 15860–15871. [CrossRef]
12. Shao, D.; Li, C.; Huang, C.; Xiang, Y.; Yu, Z. A news classification applied with new text representation based on the improved LDA. *Multimedia Tools Appl.* **2022**, *81*, 21521–21545. [CrossRef]
13. Wang, B.; Huang, Y.; Yang, W.; Li, X. Short text classification based on strong feature thesaurus. *J. Zhejiang Univ. Sci. C Comput. Electron.* **2012**, *13*, 649–659. [CrossRef]
14. Zhou, W.; Wang, H.; Sun, H. A Method of Short Text Representation Based on the Feature Probability Embedded Vector. *Sensors* **2019**, *19*, 3728. [CrossRef]
15. Wang, J.; Li, Y.; Shan, J.; Bao, J.; Zong, C.; Zhao, L. Large-Scale Text Classification Using Scope-Based Convolutional Neural Network: A Deep Learning Approach. *IEEE Access* **2019**, *7*, 171548–171558. [CrossRef]
16. Jang, B.; Kim, M.; Harerimana, G.; Kang, S.U.; Kim, J.W. Bi-LSTM Model to Increase Accuracy in Text Classification: Combining Word2vec CNN and Attention Mechanism. *Appl. Sci.* **2020**, *10*, 5841. [CrossRef]
17. Yu, S.; Liu, D.; Zhu, W.; Zhang, Y.; Zhao, S. Attention-based LSTM, GRU and CNN for short text classification. *J. Intell. Fuzzy Syst.* **2020**, *39*, 333–340. [CrossRef]

18. Xie, J.; Chen, B.; Gu, X.; Liang, F.; Xu, X. Self-Attention-Based BiLSTM Model for Short Text Fine-Grained Sentiment Classification. *IEEE Access* **2019**, *7*, 180558–180570. [CrossRef]
19. 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]
20. Deng, L.; Ge, Q.; Zhang, J.; Li, Z.; Yu, Z.; Yin, T.; Zhu, H. News Text Classification Method Based on the GRU_CNN Model. *Int. Trans. Electr. Energy Syst.* **2022**, *2022*, 1–11. [CrossRef]
21. Yang, J.; Zou, X.; Zhang, W.; Han, H. Microblog sentiment analysis via embedding social contexts into an attentive LSTM. *Eng. Appl. Artif. Intell.* **2021**, *97*, 104048. [CrossRef]
22. Wu, P.; Li, X.; Ling, C.; Ding, S.; Shen, S. Sentiment classification using attention mechanism and bidirectional long short-term memory network. *Appl. Soft Comput.* **2021**, *112*, 107792. [CrossRef]
23. Liu, G.; Guo, J. Bidirectional LSTM with attention mechanism and convolutional layer for text classification. *Neurocomputing* **2019**, *337*, 325–338. [CrossRef]
24. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
25. Principe, V.A.; Vale, R.G.D.S.; de Castro, J.B.P.; Carvano, L.M.; Henriques, R.A.P.; Lobo, V.J.d.A.e.S.; Nunes, R.D.A.M. A computational literature review of football performance analysis through probabilistic topic modeling. *Artif. Intell. Rev.* **2022**, *55*, 1351–1371. [CrossRef]
26. Korenčić, D.; Ristov, S.; Šnajder, J. Document-based topic coherence measures for news media text. *Expert Syst. Appl.* **2018**, *114*, 357–373. [CrossRef]
27. Mimno, D.M.; Wallach, H.M.; Talley, E.; Leenders, M.; McCallum, A. Optimizing Semantic Coherence in Topic Models. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh, UK, 27–31 July 2011.
28. Yan, D.; Mei, X.; Yang, X.; Zhu, P. Research on Microblog Text Topic Clustering Based on the Fusion of Topic Model and Word Embedding. *J. Mod. Inf.* **2021**, *41*, 67–74. [CrossRef]
29. Tang, H.; Wei, H.; Wang, Y.; Zhu, H.; Dou, Q. Text Semantic Enhancement Method Combining LDA and Word2vec. *Comput. Eng. Appl.* **2022**, *58*, 135–145.
30. Du, J.; Vong, C.-M.; Chen, C.L.P. Novel Efficient RNN and LSTM-Like Architectures: Recurrent and Gated Broad Learning Systems and Their Applications for Text Classification. *IEEE Trans. Cybern.* **2020**, *51*, 1586–1597. [CrossRef]
31. Zeng, Y.; Zhang, R.; Yang, L.; Song, S. Cross-Domain Text Sentiment Classification Method Based on the CNN-BiLSTM-TE Model. *J. Inf. Process. Syst.* **2021**, *17*, 818–833.
32. Li, W.; Qi, F.; Tang, M.; Yu, Z. Bidirectional LSTM with self-attention mechanism and multi-channel features for sentiment classification. *Neurocomputing* **2020**, *387*, 63–77. [CrossRef]
33. Wang, Y.; Zhu, L. Research on improved text classification method based on combined weighted model. *Concurr. Comput. Pr. Exp.* **2020**, *32*, 5140. [CrossRef]
34. Chehal, D.; Gupta, P.; Gulati, P. RETRACTED ARTICLE: Implementation and comparison of topic modeling techniques based on user reviews in e-commerce recommendations. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *12*, 5055–5070. [CrossRef]
35. Chuang, J.; Manning, C.D.; Heer, J. Termite: Visualization Techniques for Assessing Textual Topic Models. In Proceedings of the International Working Conference on Advanced Visual Interfaces, Capri Island, Italy, 21–25 May 2012.
36. Sievert, C.; Shirley, K.E. LDAvis: A method for visualizing and interpreting topics. In Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces at the Association for Computational Linguistics, Baltimore, MD, USA, 27 June 2014.
37. Dave, V.; Singh, S.; Vakharia, V. Diagnosis of bearing faults using multi fusion signal processing techniques and mutual information. *Indian J. Eng. Mater. Sci.* **2020**, *27*, 878–888. [CrossRef]
38. Bolourchi, P.; Moradi, M.; Demirel, H.; Uysal, S. Feature Fusion for Classification Enhancement of Ground Vehicle SAR Images. In Proceedings of the 2017 UKSim-AMSS 19th International Conference on Computer Modelling & Simulation (UKSim), Cambridge, UK, 5–7 April 2017; pp. 111–115. [CrossRef]