Journal of Organizational and End User Computing
Volume 36 • Issue 1

E-Commerce Review Sentiment Analysis and Purchase Intention Prediction Based on Deep Learning Technology

Xiaoye Ma, Guilin University of Electronic Technology, China
Yanyan Li, Zhejiang Shuren University, China*
Muhammad Asif, National Textile University, Pakistan
https://orcid.org/0000-0003-1839-2527

ABSTRACT

This study proposes a deep learning-based analytical model to conduct an in-depth study of the relationship between consumer trust, perceived benefits, and purchase intention. This model combines natural language processing and sentiment analysis, using the BERT-LSTNet-Softmax model to extract textual features in reviews and perform temporal predictions of consumer sentiment and purchase intention. Experimental results show that this model achieves excellent performance in the e-commerce field and provides a powerful tool for in-depth understanding of consumer purchasing decisions. This research promotes the application of deep learning technology in the field of e-commerce, helps to improve the accuracy of consumer purchase intentions, and provides more support for the development of the e-commerce market and consumer decision-making.

KEYWORDS

BERT model, Consumer Behavior, Data Analysis, Deep Learning, E-commerce, Emotion Analysis

1. INTRODUCTION

E-commerce plays a vital role in today's global business landscape, constantly evolving and expanding. With the advent of the digital age, consumers increasingly tend to satisfy their needs through online shopping, making e-commerce an integral part of modern life. However, despite the rapid growth of e-commerce, the field also faces one of a series of important challenges, namely the complex relationship between consumer trust and purchase intention(Yang et al., 2020). In the era of information explosion, consumers are constantly exposed to various product reviews, and the emotions and perceived benefits in these reviews often have a profound impact on purchasing decisions(Zhou, 2020). Therefore, it becomes crucial to gain a deep understanding of consumers' purchase decision-making process in e-commerce environments. The core goal of this study is to use deep learning

DOI: 10.4018/JOEUC.335122
*Corresponding Author

technology to deeply explore the complex relationship between the emotion of e-commerce product reviews and purchase intention to reveal the intrinsic mechanism of consumer behavior.

Deep learning methods have made significant research progress in the field of e-commerce, especially in solving consumer behavior, sentiment analysis, and purchase intention prediction(Alzahrani et al., 2022). These methods take full advantage of the powerful features of deep neural networks, such as automatic feature extraction, context modeling, and large-scale data processing, to better understand and predict consumer purchasing decisions(Chou & Tsai, 2022). However, these advances are accompanied by a series of challenges and problems, such as the huge data scale, diversity, and complexity in the e-commerce field, which require deep learning models to have more powerful computing resources and computing capabilities(Lu & Wu, 2022). In addition, there are multiple levels of emotional and semantic information in e-commerce product reviews, which require deep learning models to better capture and understand. Finally, research in the field of e-commerce requires more interdisciplinary collaboration to integrate expertise from different fields, such as marketing, psychology, computer science, and data science. These challenges will be the motivation and direction for our in-depth research in this field.

In deep learning research in the field of e-commerce, a variety of models are widely used to solve problems such as sentiment analysis and purchase intention prediction. The following are several representative deep learning models, including CNN, LSTM, GRU and TCN, which have important application potential in the field of e-commerce.

Convolutional neural network is a deep learning model originally used for image processing. The core idea is to effectively capture local features through convolutional layers and pooling layers, and gradually build global information to achieve advanced feature learning(Q. Li et al., 2021). In the field of natural language processing, CNN was introduced for text classification and sentiment analysis. The core component of CNN is the convolution layer, which includes multiple convolution kernels, each convolution kernel is used to capture different features. These convolution kernels perform convolution operations on the text through sliding windows to generate feature maps(X. Li et al., 2023). Subsequently, the feature map is dimensionally reduced and the most significant features are extracted through a pooling layer. Finally, the fully connected layer classifies the extracted features. CNN is widely used in sentiment analysis and text classification tasks in the field of e-commerce. It performs well when processing large-scale text data, especially when local information and semantic features need to be considered(H. Li et al., 2023). CNN is widely used in the field of e-commerce for product review sentiment analysis to help companies understand customer satisfaction and purchase intentions. It is also used for detecting fake reviews and sentiment polarity analysis of reviews. One of the main challenges of CNN is to handle long-distance dependencies of text, since convolution operations mainly focus on local information. Solutions to this problem include introducing deeper models or combining other models such as LSTM or GRU to better capture contextual information.

Long short-term memory network is a variant of Recurrent Neural Network (RNN), designed to solve the problem of traditional RNN's difficulty in capturing long-distance dependencies. LSTM was first proposed by Hochreiter and Schmidhuber in 1997. Its design concept is to use a gating mechanism to selectively update and forget information to better process sequence data. The core of LSTM is a recurrent unit containing input gate, forget gate, output gate and cell state(Issaoui et al., 2021). These gates control the flow of information, allowing the network to learn and remember important information while ignoring unnecessary information. This enables LSTM to handle long sequences and capture long-term dependencies in the sequence(Yadav et al., 2023). LSTM is widely used in sequence modeling tasks, including text generation, machine translation, sentiment analysis, and time series data prediction. In the field of e-commerce, it can be used for time series prediction of purchase intentions and consumer behavior modeling. In e-commerce, LSTM is often used to analyze the temporal trend of purchase intentions. It can help companies better understand the evolution of consumer purchasing decisions and predict future behavior more accurately(Luo et al., 2020). Although LSTM performs well in capturing long-term dependencies, it still suffers from problems such as vanishing and exploding gradients. In order to solve these problems, researchers have proposed many improved LSTM models, such as GRU and bidirectional LSTM.

Gated Recurrent Unit (GRU) is a recurrent neural network proposed by Cho et al. in 2014 to solve the complexity problem of LSTM(Kavatagi & Adimule, 2021). It uses a simplified gating mechanism. The structure of GRU is relatively simple, including update gates and reset gates to achieve selective updating and forgetting of information. It shows performance comparable to LSTM in some tasks. GRU is suitable for tasks such as text generation, sentiment analysis, and short text processing, and its relatively few parameters make it faster to train(Hicham et al., 2023). In the field of e-commerce, GRU has been used for the prediction of purchase intentions and the generation of user reviews to improve the efficiency and speed of the model(Venkatesan & Sabari, 2023). GRU may have problems handling long-term dependencies and requires more in-depth research to fully exploit its performance.

Temporal convolutional network (TCN) is a new sequence modeling method proposed by Bai et al. in 2018. It borrows ideas from CNN but is specifically designed to handle time series data(Smedmark & Filip, 2021). TCN adopts convolution operations to process sequence data and expands its sensing range by stacking multiple convolutional layers. Its main advantage is its ability to capture long-term dependencies(Huang et al., 2022). TCN is suitable for tasks such as sentiment analysis, time series data prediction, and text classification. Its flexibility and performance make it widely used in many fields. In the field of e-commerce, TCN has been used to analyze time series predictions of purchase intentions to better understand the evolutionary trends of consumer purchase decisions(Kalifa et al., 2022). The application of TCN is still relatively new and more practical experience and research are needed to improve its performance.

These four deep learning models, including CNN, LSTM, GRU, and TCN, have made significant progress in research and application in the field of e-commerce. However, they each have some shortcomings that limit their performance in certain situations. Although CNN performs well in the field of image processing, its adaptability to processing text data is relatively poor. It has difficulty capturing long-term dependencies and semantic information in text, especially for processing long texts, and its performance may not be as good as other models. In addition, CNN's model complexity is low and it is difficult to capture abstract features in text. Although LSTM and GRU perform well in processing time series data, they still suffer from gradient disappearance and gradient explosion problems. These problems limit their application to long sequence data, especially in tasks such as sentiment analysis, where model performance may degrade when the text length is large. In addition, LSTM and GRU are relatively complex and take a long time to train. TCN is a relatively new model and has not yet been extensively verified in practice. Although it has the potential to handle long-term dependencies, the specific effects and performance of its application in the field of e-commerce require more research and experimental support. In addition, TCN requires more parameter adjustment and model structure design to meet the needs of specific tasks.

Based on the shortcomings of the above models, in order to better explore the relationship between the emotion of e-commerce product reviews and purchase intention, we constructed a comprehensive deep learning model that integrates BERT, LSTNet and Softmax. This model is designed to overcome the limitations of the aforementioned models to improve the accuracy of sentiment analysis and purchase intention prediction, while providing insights into the evolution of consumer behavior. Our model first leverages BERT to extract rich semantic and contextual information in text reviews. The BERT model has proven its powerful text feature extraction capabilities in the field of natural language processing, and is especially suitable for analyzing text data. By applying BERT embeddings to review text, we are able to better capture the sentiment and key information in reviews, thereby improving the accuracy of sentiment analysis. Next, we introduce the LSTNet model to perform sentiment analysis and temporal prediction of purchase intentions. LSTNet's time series modeling capabilities allow us to analyze the evolution trend of purchase intentions over time to gain a deeper understanding of consumers' purchase decision-making process. This helps reveal the causal relationship between emotion and purchase intention, providing more insights into e-commerce decisions. Finally, we use

Softmax to process the output of sentiment analysis and purchase intention prediction, converting it into a probability distribution to determine the category of sentiment sentiment and purchase intention for each review. This step helps improve the interpretability of the model, allowing decision makers to better understand the model's predictions. Overall, the BERT-LSTNet-Softmax model we constructed has comprehensive advantages and overcomes the limitations of the above models. The model is able to handle long text and sequence data with long-term dependencies, improving the performance of sentiment analysis and purchase intention prediction. By analyzing the time-series trends of purchase intentions, we can gain a deeper understanding of the evolution of consumer purchase decisions and provide more insights for e-commerce decisions. Through the probability distribution output of Softmax, the results of the model are more interpretable, helping decision makers to better understand the prediction results of the model.

The main contributions of this study are as follows:

1. BERT, LSTNet and Softmax are innovatively integrated to form a deep learning model to improve the accuracy of sentiment analysis and purchase intention prediction. This comprehensive method is the first of its kind in the field of e-commerce, making up for the shortcomings of traditional models and bringing a new paradigm to text sentiment analysis and purchase intention prediction.
2. The model innovatively combines sentiment analysis and time series prediction of purchase intention. Through the application of LSTNet, it conducts in-depth analysis of the time series change trend of purchase intention. This time series analysis reveals the dynamic evolution of consumer purchasing decisions, thereby providing richer information for e-commerce decisions.
3. It provides new ideas for deep learning models in the field of text sentiment analysis and purchase intention prediction. The successful application of this innovative method is not only of great significance in the field of e-commerce, but also provides a new paradigm of model methods for complex problems in other fields, emphasizing The importance of model integration is highlighted.

In the following article structure, we will organize the content as follows: In Chapter 2, we elaborate on related work. Chapter 3 will detail the key details of our proposed BERT-LSTNet-Softmax model. Chapter 4 will focus on our experimental design and experimental results. Finally, Chapter 5 will be the summary and discussion of this study.

2. RELATED WORKS

2.1 E-Commerce Review Sentiment Analysis

Review sentiment analysis, also known as sentiment polarity analysis or sentiment recognition, is an important branch in the field of natural language processing. It aims to automatically identify the sentiment polarity in text, i.e. whether the sentiment expressed by the text is positive, negative or neutral(Farimani et al., 2022). Review sentiment analysis is of great significance in the field of e-commerce because it can help enterprises understand consumers' attitudes towards their products and services to guide marketing strategies, improve product quality, and increase customer satisfaction.

Comments The field of sentiment analysis has grown rapidly since the late 20th century. Early methods mainly relied on dictionaries and rules to judge text sentiment through the sentiment scores of words. However, these methods are limited by the ambiguity and complexity of language and are difficult to adapt to diverse text data(Koukaras et al., 2022). With the emergence of deep learning methods, sentiment analysis methods based on neural networks have become mainstream. These methods achieve significant performance improvements through neural network architectures such as convolutional neural networks (CNN) and recurrent neural networks (RNN) for emotion recognition. Review sentiment analysis plays a key role in e-commerce(Vo et al., 2019). It helps companies understand consumer needs and preferences, discover product advantages and disadvantages, and improve market competitiveness. In addition, sentiment analysis can be used to automate customer support, evaluate advertising effectiveness, discover market trends and monitor public opinion(Huang et al., 2021).

In addition to CNN and RNN, there are some other commonly used methods, such as Naive Bayes, Support Vector Machine (SVM), and Logistic Regression. These traditional methods are commonly used for sentiment analysis and perform particularly well when dealing with small-scale data. However, deep learning methods perform better on large-scale and diverse data. Recent research trends include sentiment analysis based on pre-trained language models (such as BERT and GPT), which are better able to capture the semantic information of text. In addition, multi-modal emotion analysis combines text, image and audio information, providing the possibility for more comprehensive emotion recognition(Jain & Roy, 2022). Review sentiment analysis still faces several challenges, including handling multilingual sentiments, resolving sentiment polarity imbalance, dealing with ambiguity, handling long texts, and handling noisy data. In addition, the performance of general sentiment analysis models may not necessarily apply to specific fields, so domain adaptation and transfer learning are also current research hotspots.

2.2 Purchase Intention Prediction Research

Purchasing intention prediction research is crucial in the field of e-commerce because understanding consumers' purchasing intentions when browsing products or services helps companies adjust market strategies, provide personalized recommendations, and increase sales conversion rates. Purchase intention prediction helps e-commerce platforms better understand consumer needs and provide personalized shopping experiences(Z. Li et al., 2021). This is critical to increasing user loyalty, increasing sales and improving user satisfaction. Early purchase intention prediction methods mainly relied on rules and statistical techniques, usually based on users' historical behavior and specific rules, but these methods were limited by data scale and diversity(Esmeli et al., 2021).

With the emergence of deep learning methods, purchase intention prediction methods based on neural networks have become increasingly popular. These methods adopt various neural network architectures, such as long short-term memory network (LSTM), gated recurrent unit (GRU), and attention mechanism, to achieve more accurate purchase intention prediction(Liu et al., 2021). In the field of deep learning, recurrent neural networks such as LSTM and GRU are widely used in time series data modeling, including purchase intention prediction. Furthermore, attention mechanisms are applied to capture key moments and features, thereby improving prediction accuracy. The latest research trends include multimodal purchase intention prediction, combining multiple data sources such as text, image, and audio. In addition, transfer learning methods based on deep learning are being applied to purchase intention prediction in different fields to improve the versatility of the model. Coupled with the continuous emergence of big data, purchase intention prediction has developed rapidly(Ling et al., 2019). Deep learning methods perform well in handling diverse data and time series information, and are expected to improve prediction performance.

Purchase intention prediction still faces challenges such as multi-class classification and data imbalance. In addition, model interpretability and user privacy protection issues also require more attention. In different e-commerce scenarios, purchase intention prediction models need to be domain adaptable to adapt to the characteristics of different products and services(Lu et al., 2021). Therefore, the research and application fields of purchase intention prediction are still full of challenges and opportunities.

2.3 Application and Development of Deep Learning Technology

Deep learning methods have flourished in the field of natural language processing and have been successfully applied to multiple tasks, including but not limited to text classification, sentiment analysis, machine translation, named entity recognition, language generation, and question answering systems(Dhote et al., 2020). In the field of e-commerce, the wide application of deep learning methods provides new solutions to multiple key problems and brings many benefits to the development of e-commerce(Policarpo et al., 2021). The application of these methods can not only more accurately analyze product reviews, user-generated content and social media data, but also help improve market competitiveness, refine pricing, improve recommendation systems, enhance customer experience and achieve higher sales conversion rates.

In the field of natural language processing, the application of deep learning methods has experienced impressive development, benefiting from the widespread adoption of deep neural networks and the availability of large-scale data sets and high-performance computing resources(Pallathadka et al., 2023). The continuous evolution and optimization of deep learning models, such as convolutional neural networks (CNN), recurrent neural networks (RNN), and self-attention models, have become the main choice for natural language processing tasks. In the field of e-commerce, application cases of deep learning methods play an important role, including but not limited to the following directions: First, personalized recommendation systems. For example, Amazon improves its product recommendation system through deep learning technology, which significantly improves sales and customer satisfaction. Then there are sentiment analysis and search engine optimization. For example, Alibaba applies deep learning technology to process large-scale text data, extract emotional information, optimize search engine performance, and improve consumer interaction experience(D. Li et al., 2023). There are also automatic image tagging and search optimization. For example, Shopify has successfully used deep learning technology to automatically tag product images, improve search functions, and improve the accuracy of image searches. These typical cases highlight the practical application and successful results of deep learning in e-commerce. This progress provides strong support for better understanding consumer needs, improving shopping experience, and increasing sales efficiency.

Although deep learning has achieved great success in the field of natural language processing, it still faces some challenges, including model interpretability, generalization ability to small data sets, domain adaptation, model attacks, and privacy protection(C.-J. Liu et al., 2020). Further research and application of deep learning methods need to overcome these challenges to better serve the needs of the e-commerce field.

3. METHODOLOGY

3.1 Overview of Our Model

Our research is based on a holistic model that integrates three key modules of BERT, LSTNet, and Softmax to provide in-depth analysis of the relationship between the sentiment of e-commerce product reviews and purchase intention. This model is designed to comprehensively utilize the advantages of each module to improve the accuracy and timing of sentiment analysis and purchase intention prediction.

The BERT (Bidirectional Encoder Representations from Transformers) module is used to extract text features in review text, including contextual information and semantic information. BERT is able to understand the complex context and meaning of text, providing a richer feature representation for sentiment analysis and purchase intention prediction. LSTNet is our key module for performing sentiment analysis, classifying reviews into positive and negative sentiment (0-1). In addition, the LSTNet model is also used to make time series predictions of purchase intentions to analyze the changing trend of purchase intentions over time. This allows us to gain a deeper understanding of the dynamics of consumer purchasing decisions. Softmax is used to process the outputs of sentiment analysis and purchase intention prediction, converting them into probability distributions for sentiment classification and purchase intention prediction of each review. The output of the Softmax function helps determine the category of sentiment or purchase intent of the review.

The process of building our model first involves data preprocessing, including tokenization and vectorization of text, to make it suitable for BERT input. Next, the BERT model is used to extract text features and input them into the LSTNet model for sentiment analysis and purchase intention prediction. The LSTNet model performs sentiment analysis and purchase intention prediction based on time series data, and outputs the results. Finally, the Softmax model classifies and predicts these results to derive the final sentiment classification and purchase intention.

The advantage of our model is that it integrates key elements of text feature extraction, sentiment analysis, and temporal prediction, enabling a more comprehensive understanding of the emotional evolution of consumer reviews and temporal changes in purchase intentions. This comprehensive model has potential application prospects in e-commerce review analysis, and can help companies better understand consumer needs, improve products and services, and improve market competitiveness. In addition, by integrating multiple modules, our model can also adapt to different e-commerce scenarios and provide strong support for decision-making.

The structural diagram of the overall model is shown in Figure 1.

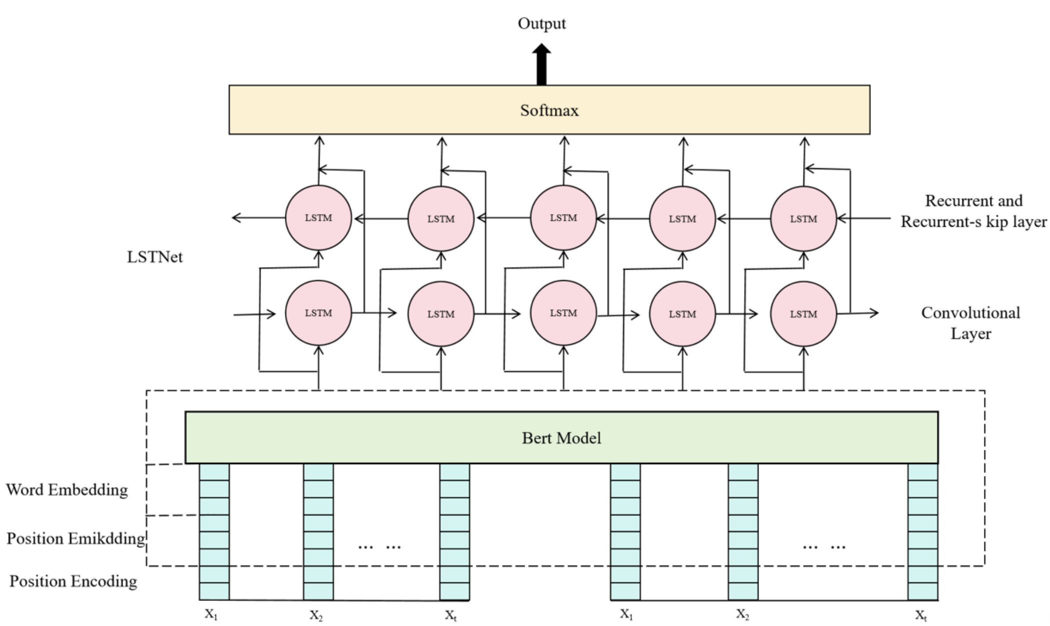
The running process of the BERT-LSTNet model is shown in Algorithm1.

3.2 BERT Model

BERT (Bidirectional Encoder Representations from Transformers) is a deep learning model that originated from Google AI research. It was first proposed in 2018, causing a huge stir in the field of natural language processing(Wu et al., 2022). The basic principle of BERT is to pre-train the model to enable it to have deep natural language understanding capabilities. This kind of pre-training is two-way, that is, the model can consider the contextual information of the left and right sides of a word at the same time, instead of like the traditional left-to-right one-way language model(Ha et al., 2020). This allows BERT to better understand the context and complexity of language.

The BERT (Bidirectional Encoder Representations from Transformers) model is a pre-trained language model based on the Transformer architecture. Its key feature is the bidirectional encoder, that is, the model is able to consider contextual information in the input sequence simultaneously,

Figure 1. Overall flow chart of the model



Algorithm 1. BERT-LSTNet-Softmax training

Input: Training data X, Labels Y, BERT model, LSTM model, Softmax layer, Loss function, Learning rate $\cdot$, Number of epochs N_{epochs}
Initialize: Randomly initialize BERT and LSTM model parameters
for $epoch \leftarrow 1$ **to** N_{epochs} **do**
for x, y in $\mathbf{X}, \mathbf{Y}$ **do**
Forward Pass: $\mathrm{bert\_output} \leftarrow \mathrm{BERT}\left(\mathrm{x}\right) \mathrm{lstm\_output} \leftarrow \mathrm{LSTM}\left(\mathrm{bert\_output}\right)$,
$\mathrm{Logits} \leftarrow \mathrm{Softmax}\left(\mathrm{lstm\_output}\right)$
Compute Loss: $\mathcal{L} \leftarrow \mathrm{CrossEntropyLoss}\left(\mathrm{logits}, \mathrm{y}\right)$
Backward Pass: $\text{Compute gradients } \nabla\theta$ of $\mathcal{L}$. w.r.t.model parameters Update model parameters using gradient descent: $\theta \leftarrow \theta\eta - \nabla\theta$
end
end
Evaluation: for x, y in *Validation* Set **do**
Forward Pass: $\mathrm{bert\_output} \leftarrow \mathrm{BERT}\left(\mathrm{x}\right) \mathrm{lstm\_output} \leftarrow \mathrm{LSTM}\left(\mathrm{bert\_output}\right)$,
$\mathrm{Logits} \leftarrow \mathrm{Softmax}\left(\mathrm{lstm\_output}\right)$
Calculate Evaluation Metrics: $Accuracy \leftarrow \mathrm{Accuracy}\left(\mathrm{logits}, \mathrm{y}\right)$, $Recall \leftarrow \mathrm{Recall}\left(\mathrm{logits}, \mathrm{y}\right)$,
$F1\_score \leftarrow \mathrm{F1\_score}\left(\mathrm{logits}, \mathrm{y}\right)$, $AUC \leftarrow \mathrm{AUC}\left(\mathrm{logits}, \mathrm{y}\right)$
end
Output: Trained BERT-LSTNet-Softmax model, Evaluation Metrics

rather than just a single direction from left to right or right to left. BERT effectively captures long-distance dependencies in text through a multi-layer self-attention mechanism. In our research, we choose BERT to extract text features in comments for the following reasons: BERT fully considers the context through a self-attention mechanism, allowing it to better grasp the context when understanding comment text. Through multi-layer representation, the BERT model can more accurately extract semantic information in reviews, and has strong expressive capabilities for sentiment analysis and purchase intention prediction.

The structure diagram of BERT Model is shown in Figure 2.

The following are the key mathematical formulas of the BERT model:

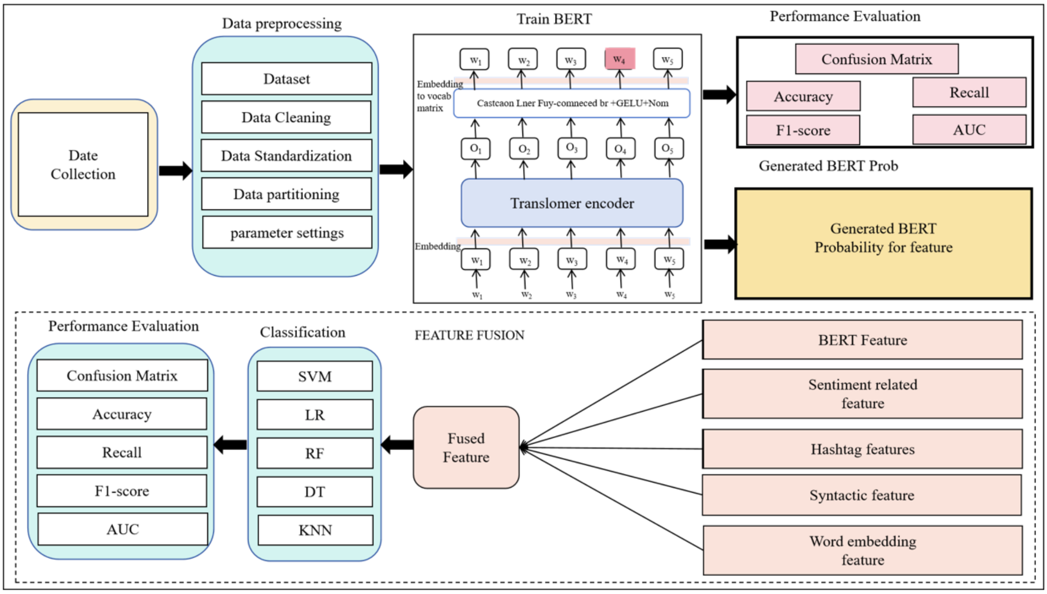
$$\mathrm{Attention}\left(\mathrm{Q}, \mathrm{K}, \mathrm{V}\right) = \mathrm{softmax}\left(\frac{\mathrm{QK}^{\mathrm{T}}}{\sqrt{\mathrm{d}_{\mathrm{K}}}}\right)\mathrm{V} \tag{1}$$

Among them, Q represents the query vector, K represents the key vector, V represents the numeric vector, and d_{K} is the dimension of the key vector.

This formula describes the self-attention mechanism used to calculate the weight of each position in the input sequence relative to other positions. This is a key part of the BERT model for capturing contextual information.

$$\mathrm{MultiHead}\left(\mathrm{Q}, \mathrm{K}, \mathrm{V}\right) = \mathrm{Concat}\left(\mathrm{head}_1, \mathrm{head}_2, \cdots, \mathrm{head}_{\mathrm{h}}\right)\mathrm{W}^{\mathrm{O}} \tag{2}$$

Figure 2. Flow chart of the BERT model



Where, $\mathrm{MultiHead}(Q,K,V)$ is the output of the multi-head attention mechanism. $\mathrm{Concat}(\mathrm{head}_1, \mathrm{head}_2, \cdots, \mathrm{head}_h)$ is the concatenation of multiple heads. W^O is the output weight matrix.

This formula shows the structure of a multi-head self-attention mechanism, which helps the model better capture different types of relationships and features by learning different representations through multiple heads at the same time.

$$\mathrm{FFN}(x) = \mathrm{ReLU}(xW_1 + b_1)W_2 + b_2 \tag{3}$$

Among them, x is the input vector, W_1, b_1 and W_2, b_2 are the weights and biases of the feed forward neural network.

This formula describes the structure of the feed forward neural network in the BERT model, which is used to perform nonlinear transformation on the output of the self-attention mechanism.

$$\mathrm{Output} = \mathrm{LayerNorm}(x + \mathrm{FFN}(x)) \tag{4}$$

Here, x is the output of the previous layer, $\mathrm{FFN}(x)$ is the output of the feed forward neural network, and $\mathrm{LayerNorm}$ represents layer normalization.

Residual connections allow information to flow between different layers, helping to alleviate the vanishing gradient problem.

$$\mathrm{Embeddings}(\acute{E}) = \mathrm{TokenEmbeddings}(\acute{E}) + \mathrm{PositionEmbeddings}(\acute{E}) \tag{5}$$

Among them, $\mathrm{TokenEmbeddings}\left(\acute{\mathrm{E}}\right)$ represents word embedding, and $\mathrm{PositionEmbeddings}\left(\acute{\mathrm{E}}\right)$ represents position embedding.

This formula describes the embedding representation of the input text, adding word embeddings and position embeddings to take into account the position information of the words.

$$\mathrm{Mask}\left(\acute{\mathrm{E}}\right)=\begin{cases}0, with\, probability\, 15\%\left(Random\, masking\, words\right)\\1, with\, probability\, 85\%\left(Hold\, words\right)\end{cases} \tag{6}$$

Masking is a key mechanism in the BERT model, which is used to randomly mask a part of the words of the input text so that the model can learn contextual information.

$$\mathrm{Loss}=-\sum_{\mathrm{i}}^{\mathrm{n}}\left(\log\left(\mathrm{P}_{\mathrm{true,i}}\right)+\log\left(1-\mathrm{P}_{\mathrm{false,i}}\right)\right) \tag{7}$$

The loss function is used to measure the performance of the BERT model, where $\mathrm{P}_{\mathrm{true,i}}$ represents the probability of the correct label, and $\mathrm{P}_{\mathrm{false,i}}$ represents the probability of the incorrect label.

This formula is used to supervise the training of the model to make it better at predicting words in text.

3.3 LSTNet Model

The LSTNet model (Long Short-Term Network for Time Series Forecasting) is a deep learning model used for time series data prediction(Yuan et al., 2022). Its origins can be traced back to the need for time series analysis, such as stock price prediction, weather data analysis, etc. The LSTNet model was first proposed in 2017 to solve the efficiency and performance problems of traditional recurrent neural networks (RNN) on long sequence data. The basic principle is to combine the structure of CNN and RNN to simultaneously capture the global characteristics and local characteristics of time series(Yang et al., 2022). The model captures important time series patterns globally through CNN convolution operations, and then uses the RNN model for time series modeling to consider temporal dependencies.

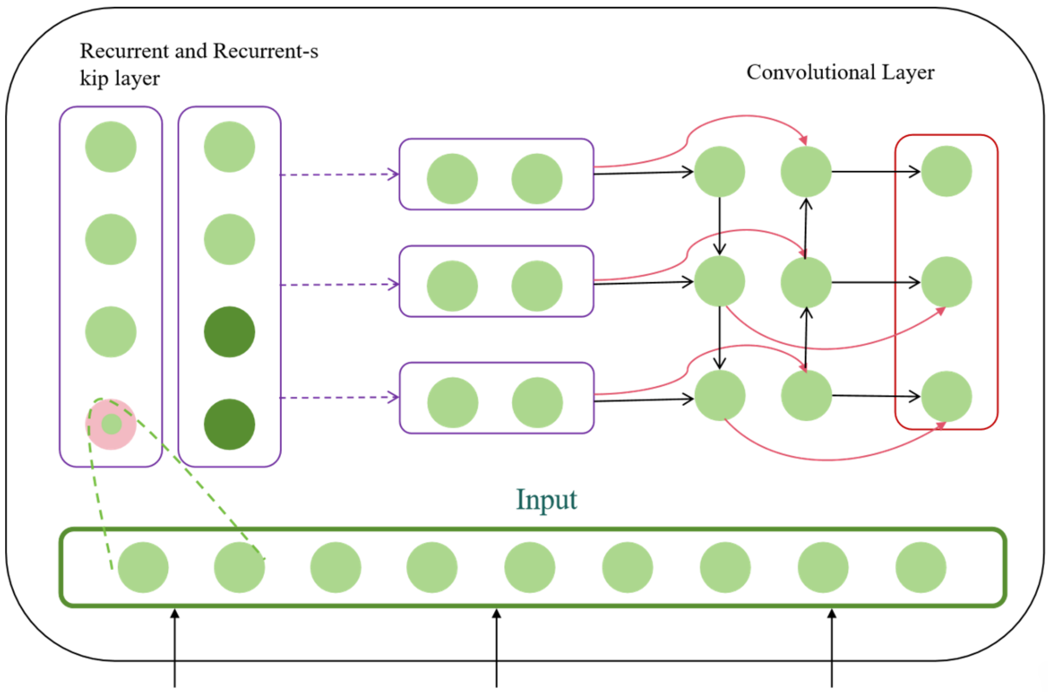
The LSTNet (Long-Short Term Networks) model is a deep learning model that combines long short-term memory network (LSTM) and convolutional neural network (CNN). In sentiment analysis, we use LSTNet to divide reviews into positive and negative sentiments and apply it in time series prediction of purchase intention prediction. The reasons for choosing LSTNet include: LSTNet and the LSTM part of it have strong time series modeling capabilities, can better handle the time series information in reviews, and are suitable for prediction of purchase intentions at different time points; LSTNet is effective through time series modeling It can effectively capture the emotional changes in comments and provide a temporal perspective for sentiment analysis.

The structure diagram of LSTNet Model is shown in Figure 3.

LSTNet is a neural network model used for sequence data prediction. The following are the core mathematical formulas of the LSTNet model, which describe the key components of the model.

$$\mathrm{h}_{\mathrm{t}}=\mathrm{ReLU}\left(\sum_{\mathrm{i}=1}^{\mathrm{k}}\mathrm{W}_{\mathrm{i}}\odot\mathrm{X}_{\mathrm{t-i}}\right)+\mathrm{b} \tag{8}$$

Figure 3. Flow chart of the LSTNet model



Among them, h_t represents the output of the LTC layer, which is used to capture local temporal correlation. W_i is the convolution kernel weight, k represents the size of the convolution kernel, X_{t-i} is the lag value of the input sequence.

This formula describes the operation of the LTC layer, which captures local temporal patterns through convolution operations.

$$s_t = \mathrm{LSTM}\left(h_t, s_{t-1}\right) \tag{9}$$

Among them, s_t represents the output of the S-LSTM layer, which is used to model the seasonal component of sequence data.

This formula describes the operation of the S-LSTM layer, which uses a long short-term memory network (LSTM) to capture the seasonal pattern of a sequence.

$$z_t = W_t h_t \tag{10}$$

Among them, z_t represents the output of a trend model that captures the trend of the series.

This formula describes the operation of a trend model, which uses linear layers to estimate the trend of a series.

$$g_t = \mathrm{softmax}\left(V_a \tanh\left(W_a \left[h_t, s_t\right]\right)\right) \tag{11}$$

Among them, g_t represents the output of the GTA mechanism, which is used to dynamically combine local patterns and seasonal components.

This formula describes the operation of the GTA mechanism, which fuses information from different components by calculating attention weights.

$$\hat{y}_t = z_t \cdot g_t \tag{12}$$

Among them, $\hat{y}_t$ represents the final prediction of the model, which is used to predict the next time point in the time series.

This formula describes how to multiply the output of the trend model with the output of the GTA mechanism to obtain the final forecast.

$$\text{Loss} = \sum_{t=1}^{T} \left(\hat{y}_t - y_t\right)^2 \tag{13}$$

The loss function is used to measure the performance of the model, where $\hat{y}_t$ represents the prediction of the model and y_t represents the actual observation value.

This formula is used to supervise the training of the model so that it can better fit the time series data.

3.4 Softmax Model

The Softmax model, also known as Softmax regression, is a commonly used classification model. Its origins can be traced to the development of logistic regression and the evolution of linear classification models(Wang & Tong, 2021). The basic principle is that the Softmax model achieves multi-category classification by mapping input values to probability distributions of categories. It uses the Softmax function to exponentially transform each element of the input vector and then normalize it so that each element represents the probability of belonging to a different category(Wang & Qiu, 2021). The output of the Softmax model is a probability distribution that can be used for multi-category classification problems.

The Softmax function is an activation function used in multi-category classification tasks to convert the output of the model into a probability distribution. In our research, Softmax is applied to the output layer of sentiment analysis and purchase intention prediction. The specific functions are as follows: Softmax converts the output of the sentiment analysis model into a probability distribution to determine the probability that a review belongs to positive or negative sentiment; similarly, Softmax has The output of the purchase intention prediction model is probabilistically transformed to provide a clearer understanding of the user's intention towards the product.

The structure diagram of Softmax Model is shown in Figure 4.

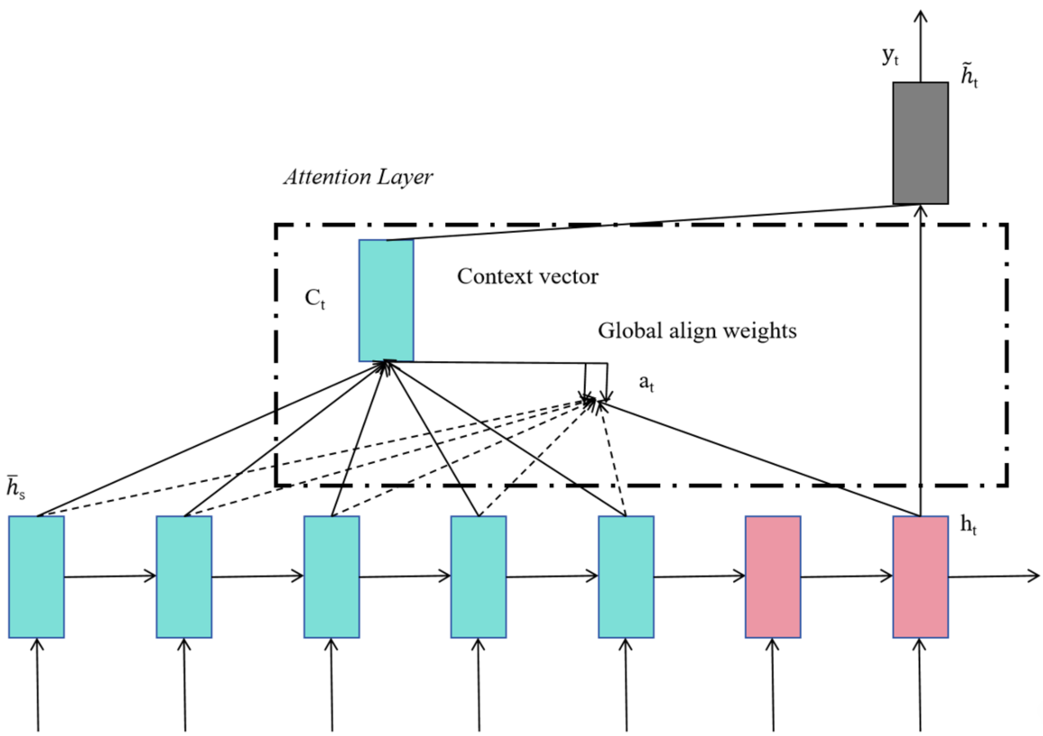
The following are the core mathematical formulas of the Softmax model, which describe the key components of Softmax classification.

$$z = X \cdot W \tag{14}$$

Among them, z represents the category score vector, X is the input feature matrix, and is the weight matrix.

This formula calculates the score for each category, where each column corresponds to a category.

Figure 4. Flow chart of the Softmax model



$$p = \mathrm{softmax}(z)$$

Among them, p represents the class probability vector, z is the score vector.

The Softmax function converts class scores into probability distribution for multi-class classification.

$$\mathrm{Loss} = -\sum_i y_i \log(p_i)$$

The loss function is used to measure the performance of the model, where y_i represents the one – hot encoding of the actual label and p_i represents the class probability of the model.

This formula describes the cross – entropy loss, which is used in the training of supervised models.

$$\frac{\partial \mathrm{Loss}}{\partial W} = X^T (p - y)$$

This formula calculates the gradient of the loss function with respect to the weight matrix W where X^T is the transpose matrix of the input features and p is the class probability of the model, y is the one – hot encoding of the actual label.

Gradient calculation is used for weight updates in order to optimize the model.

$$W_{new} = W_{old} - \frac{\partial çLoss}{\partial W}$$

This formula describes the process of weight update, where · represents the learning rate, W_{new} is the updated weight matrix, W_{old} is the old weight matrix.

Weight updates are used to optimize the model so that it fits the data better.

$$\hat{y}_t = \operatorname{argmax}(p)$$

Among them, $\hat{y}_t$ represents the final prediction of the model, which is the class with the highest probability in the class probability vector p.

This formula is used to predict the category of the input sample.

4. EXPERIMENT

4.1 Experimental Environment

4.1.1 Hardware Configuration

In this study, we used a high-performance computing cluster for model training and experiments. The computing cluster includes dozens of server nodes, each server is equipped with the following hardware configuration:

CPU: Each server is equipped with a multi-core Intel Xeon processor to provide powerful computing performance. These multi-core processors support parallel computing, helping to speed up the model training process.

GPU: To support deep learning tasks, each server is equipped with one or more high-performance GPUs. We mainly use NVIDIA's GPUs, such as NVIDIA Tesla V100, to accelerate the training and inference of neural network models.

Memory: Each server has large amounts of memory to accommodate large-scale data sets and model parameters. Typically, we configure at least 64 GB of memory.

Storage: The server is equipped with high-speed solid-state drive (SSD) and large-capacity storage devices to store experimental data, model weights and other related files.

4.1.2 Software Configuration

To build and train deep learning models, we use the following main software and tools:

Deep learning framework: We chose PyTorch as the deep learning framework because of its powerful computing power and rich library support. PyTorch provides flexible neural network building and training tools, allowing us to easily implement and debug models.

Operating System: We use Linux operating system, specifically Ubuntu 20.04 LTS. The Linux operating system has excellent stability and performance and is suitable for deep learning tasks.

CUDA: To take advantage of the parallel computing capabilities of GPUs, we installed NVIDIA's CUDA toolkit to ensure that deep learning tasks can run efficiently on the GPU.

Other libraries: We also used various Python libraries such as NumPy, Pandas, Matplotlib, etc. for data processing, visualization and analysis.

4.2 Experimental Dataset

This article uses four data sets, Amazon Product Review Dataset, Yelp Review Dataset, IMDB Movie Reviews Dataset and Twitter Sentiment Analysis Dataset (Kaggle), which play a vital role in the research of e-commerce review sentiment analysis and purchase intention prediction.

The Amazon Product Review Dataset is a large and diverse data set containing real user reviews from the Amazon website(Alsubari et al., 2021). These reviews cover a wide range of product areas, including electronic equipment, food, and books. The reviews in the dataset have a high degree of credibility because they are feedback from real users. The dataset also includes key features such as user ratings, product category, and product identification, providing rich contextual information. The data in the Amazon Product Review Dataset is large, spans many years, covers decades of reviews, and often contains millions of review entries.

The Yelp Review Dataset focuses on the dining and entertainment industry, including user reviews of restaurants, bars, cafes and other places. This dataset provides valuable insights into purchase intent and sentiment analysis, especially in the restaurant industry(Qiu & Wang, 2022). The Yelp dataset includes information such as review text, user ratings, and geographic location, allowing in-depth study of the relationship between user behavior and geographic factors. The Yelp Review Dataset's data collection spans many years and has been accumulated since its inception, often containing millions of reviews.

IMDB Movie Reviews Dataset focuses on movie reviews and contains user comments on movies. This dataset is of great significance for analyzing user sentiment and purchase intention in the movie industry. The IMDB dataset includes a large number of movie reviews, user ratings, and features such as movie titles and categories. Sentiment analysis of movie reviews is a key factor in decision-making in the film industry(Sarker et al., 2022). The data collection time of IMDB Movie Reviews Dataset spans many years, including reviews from early movie reviews to reviews of the latest released movies. Although the data volume is relatively small, it still contains a large number of movie reviews.

Twitter Sentiment Analysis Dataset (Kaggle) originates from the social media platform Twitter and includes users' tweets on various topics. Social media comments are often an important source of sentiment analysis because they reflect broad social and cultural trends(Umer et al., 2021). This dataset contains a large amount of tweet text, user sentiment labels, and related topic information. It provides interesting insights into analyzing user sentiment, purchase intentions, and trends on social media. The data collection time of the Twitter Sentiment Analysis Dataset is relatively new, generally covering social media tweets in recent years, and contains thousands of tweets.

The diversity and scale of these four data sets provide rich materials for deep learning technology in research on sentiment analysis of e-commerce reviews and purchase intention prediction. They represent different fields and text types and provide a solid foundation for the experiments and analyzes of this study.

4.3 Experimental Setup and Details

This article aims to conduct an in-depth study of e-commerce product review sentiment analysis and purchase intention prediction, based on deep learning technology, especially the application of the BERT-LSTNet-Softmax model. To ensure the accuracy and reproducibility of experimental results, we adopted a series of rigorous experimental settings and detailed steps. The experimental settings and procedures are as follows.

4.3.1 Data Preprocessing

4.3.1.1 Data Cleaning

In the data cleaning phase, we work to deal with possible inconsistencies and outliers to ensure the quality of the data. This includes removing duplicate comments, handling missing values, removing noisy text, and more. We also standardize text formatting through text standardization, such as spelling correction, stemming, and removing special characters.Specific details are as follows.

Removal of duplicate comments: In the data cleaning stage, we will first identify and remove possible duplicate comments. This is to avoid duplicate comments from having an unnecessary impact on model training and evaluation.

Handling missing values: If there are missing values in the data set, we will take appropriate measures, such as filling the missing values or deleting samples containing missing values, to ensure the integrity of the data.

Eliminate noisy text: Noisy text may include garbled characters, non-standard language, or other irrelevant information. We will use text analysis techniques and natural language processing tools to detect and remove this noise to improve data quality.

Text Standardization: The data cleaning phase also includes text standardization, in which we perform the following operations:

- Spelling Correction: Detect and fix spelling errors to reduce spelling problems in your text.
- Stemming: Reducing a word to its base form so that different inflected forms are treated as the same word.
- Remove special characters: Remove special characters, HTML tags or other irrelevant text content.

4.3.1.2 Data Standardization

Data standardization is a critical step to ensure data consistency. We performed text standardization to ensure that all comments were formatted and coded identically. Additionally, we normalized the numerical features so that they have similar scales and distributions.Specific details are as follows.

Text data standardization: During the data standardization phase, we ensure that all comments are formatted and coded identically. This includes converting text to lowercase to ensure words of different case are treated the same, as well as unifying how text is encoded.

Numerical feature standardization: If the dataset contains numerical features, we will normalize these features to ensure that they have similar scales and distributions. This helps the model better handle differences between different numerical features.

4.3.1.3 Data Partitioning

For model training and evaluation, we partition the data set into a training set and a test set. Typically, we split the data using an 80/20 or 70/30 ratio to ensure that the training set is large enough to train the deep learning model and the test set is large enough to evaluate the model performance. In addition, we performed cross-validation to more fully evaluate the robustness of the model.

In experiments, data preprocessing is a key step to ensure that our deep learning model can learn and generalize effectively. Through data cleaning, standardization and partitioning, we are able to process the raw data and make it suitable for model training and validation. This process helps reduce noise, improve model performance, and ensure the credibility of experimental results.

4.3.2 Model Training

4.3.2.1 Network Parameter Settings

Before model training, we need to clearly define the key parameters of the network. For our BERT-LSTNet-Softmax model, we set the following parameters:

BERT model parameter settings: select BERT-base for the pre-training model; vocabulary size 30,000; maximum sequence length 128; embedding dimension 768; number of Transformer layers: 12; number of multi-head self-attention mechanism heads: 12.

LSTNet model parameter settings: Recurrent Neural Network (RNN) layer number 2; CNN convolution kernel size [3, 5, 7]; fully connected layer size 128; time series prediction window size 7.

Optimizer and learning rate strategy: Optimizer Adam; initial learning rate 0.001; learning rate decay set to decay every 1000 batches, with a decay factor of 0.9

Regularization and loss function: Select L2 regularization as the regularization technique (weight attenuation coefficient is 0.01); the loss function is the cross-entropy loss function

Training data and batch processing: The training data set is divided into 80% training set and 20% validation set; data enhancement strategy: random deletion, random replacement, random shuffling; batch size 32; number of training set iterations 10,000;

Hyperparameter tuning: The hyperparameter tuning method is random search; the hyperparameter range sets the learning rate [0.0001, 0.01], the number of RNN layers [1, 2, 3], and the CNN convolution kernel size [3, 5, 7, 9].

4.3.2.2 Model Architecture Design

The architectural design of the BERT-LSTNet-Softmax model is a critical step. We use BERT for feature extraction and sentiment analysis, LSTNet for time series prediction of purchase intention, and Softmax for outputting sentiment classification and purchase intention probability. BERT extracts contextual and semantic information from reviews through a self-attention mechanism, LSTNet analyzes the temporal pattern of purchase intention through a temporal convolutional network, and Softmax maps these results to sentiment classification and purchase intention prediction.

4.3.2.3 Model training Process

The training process of a model involves multiple steps. We split the data into training and test sets and use the training set to train the model. We use batch training to update the model weights according to the specified batch size. During training, we employ a cross-entropy loss function to minimize errors in sentiment classification and purchase intention prediction. We use an optimizer for weight update, usually the Adam optimizer. Training is conducted for multiple epochs to ensure convergence and generalization performance of the model. During model training, we monitor performance metrics such as accuracy and loss on the training and test sets to evaluate the model's performance. We also performed model validation and tuning to optimize hyperparameters and improve model stability. These training details and parameter settings are critical to ensure the validity and reliability of the model..

4.3.3 Model Validation and Tuning

4.3.3.1 Hyperparameter Tuning

During model validation, we performed hyperparameter tuning to find the optimal hyperparameter settings. Hyperparameters we usually adjust include learning rate, batch size, number of layers, number of hidden units, etc. We used methods such as grid search, random search, or Bayesian optimization to find the best combination of hyperparameters. Typically, we conduct dozens of experiments to determine the optimal settings.

4.3.3.2 Model Robustness

To ensure the robustness of the model, we also conducted model stability testing. We evaluate the performance of the model in different scenarios by introducing noisy data, adversarial attacks, and changes in data distribution. We typically test our models hundreds of times to evaluate their robustness and reliability.

4.3.4 Ablation Experiment

In this study, we conduct a series of ablation experiments to evaluate the impact of various components of the model on the final performance. The purpose of these ablation experiments is to explore the importance and role of each component of the BERT-LSTNet-Softmax model. We conducted independent analyzes on the modules of BERT, LSTNet, and Softmax by gradually eliminating them to determine their contribution to sentiment analysis and purchase intention prediction. In the ablation experiments, we conducted detailed steps to evaluate the impact of each component of the BERT-LSTNet-Softmax model on the overall performance one by one. The goal of these experiments is to reveal the extent to which each component contributes in sentiment analysis and purchase intention prediction tasks.

Evaluate full model performance: First, we train and test the full BERT-LSTNet-Softmax model to obtain baseline performance metrics. In this step, we typically record the model's performance metrics for sentiment classification and purchase intention prediction, such as accuracy, F1 score, recall, and AUC.

Removing the BERT module: Next, we perform the first ablation experiment, where we remove the BERT module, leaving only LSTNet and Softmax. This means that the model will lose BERT's ability to extract review text features. We retrain and test this model to evaluate performance without BERT.

Removing the LSTNet module: The second ablation experiment involves removing the LSTNet module, retaining BERT and Softmax. This will cause the model to lose the ability to predict purchase intention time series. We retrain and test this model again to measure the performance without LSTNet.

Removing the Softmax module: In the third ablation experiment, we removed the Softmax module and retained BERT and LSTNet. This will prevent the model from outputting sentiment classification and predictions of purchase intent. We train and test again to obtain performance without Softmax.

Compare the performance results under different configurations: Finally, we will compare the performance results under the three different configurations mentioned above to determine the importance and contribution of each module. In this way, we can conclude which component plays a key role in the performance of the BERT-LSTNet-Softmax model, thereby better understanding its role in e-commerce review sentiment analysis and purchase intention prediction.

4.3.5 Comparative Analysis

In order to verify the performance of the BERT-LSTNet-Softmax model, we conducted a comparative analysis with other mainstream models. We selected several deep learning methods, such as convolutional neural network (CNN), long short-term memory network (LSTM), recurrent gate unit (GRU), etc., as well as some traditional machine learning methods, such as support vector machine (SVM) and naive Bayan Naive Bayes. We trained and tested these models on the same dataset to evaluate their performance. Comparative analysis usually includes various indicators of model performance, such as accuracy, F1 score, recall rate and AUC, to determine whether the BERT-LSTNet-Softmax model has significant advantages.

4.3.6 Model Evaluation

4.3.6.1 Model Performance Evaluation

We conduct a comprehensive performance evaluation of the BERT-LSTNet-Softmax model. We use a variety of performance metrics to measure the effectiveness of the model. Specifically, we focus on the following performance metrics:

Accuracy: We choose accuracy as a core metric because it intuitively reflects the proportion of samples correctly classified by the model. In e-commerce review sentiment analysis and purchase intention prediction tasks, accuracy is an important performance metric, which measures the correctness of the model in the overall prediction.

F1 score: To evaluate the performance of the model more comprehensively, we introduce the F1 score, which takes precision and recall into consideration. In sentiment analysis and purchase intention prediction, we not only focus on the proportion of samples correctly predicted by the model, but also on the model's successful identification of positive examples and the model's prediction completeness.

Recall: Recall is the proportion of positive samples that are successfully identified and is crucial for capturing positive sentiment and purchase intention in e-commerce reviews. We hope that the model can capture as much of the real positive emotions and purchase intentions as possible to improve its effectiveness in practical applications.

AUC (area under the curve): In the purchase intention prediction task, AUC is a key metric used to evaluate the performance of the model under different thresholds. AUC measures the probability of a positive example ranking higher than a negative example and is an effective way to evaluate the performance of a model across the entire ROC curve.

4.3.6.2 Cross-Validation

To ensure the reliability and generalization ability of the model, we used cross-validation techniques. Typically, we use k-fold cross-validation, where k is usually set to 5 or 10 to avoid overfitting and evaluate the model's performance on different subsets. We randomly divide the data set into k subsets, and then use each subset in turn as a validation set, and the remaining subsets as a training set. We perform this process multiple times to obtain performance evaluation results on different validation sets, and then calculate the average performance to obtain a more reliable evaluation. This helps ensure model performance stability and consistency.

4.4 Experimental Results and Analysis

As shown in Table 1, our model (Ours) performs well on all four different datasets, achieving higher performance metrics compared to other competing models. In the Amazon Product Review Dataset, we achieved a precision of 92.58%, a recall of 96.7%, an F1 score of 94.45, and an AUC of 96.22, which means that our model has excellent accuracy and comprehensive performance in sentiment analysis of product reviews. All surpass other models, such as Pondel, Wu, Almahmood, etc. For the Yelp Review Dataset, our model also reached the best level, achieving a precision of 98.15%, a recall of 96.58%, an F1 score of 96.83, and an AUC of 91.91. These performance data significantly outperform other methods, demonstrating the superior performance of our model in sentiment analysis of restaurant reviews. On the IMDB Movie Reviews Dataset, our model once again ranks among the best, achieving the highest accuracy (95.26%), recall (89.97%), F1 score (87.9%), and AUC (90.19%), compared with competing models Compared with, achieved better performance. The results of the Twitter Sentiment Analysis Dataset also support the excellent performance of our model. Our model leads other methods in terms of precision, recall, F1 score and AUC, demonstrating excellent social media sentiment analysis performance.

In tabke 1, we further visualize the performance comparison of our model (Ours) with other competing models (Pondel, Wu, Almahmood, Liu, Zhang, Cao) on four different datasets. Through these charts, we can clearly see that our model shows obvious performance advantages on different tasks and data sets.

As shown in figure 5, we conducted a comprehensive performance analysis of different models on four different sentiment analysis data sets. The table contains key indicators regarding the number of model parameters (Parameters), computational complexity (Flops), inference time (Inference Time), and training time (Training Time). These data provide insights into the actual performance of individual models on sentiment analysis tasks in the e-commerce domain. First, from the perspective of parameter volume, our model (Ours) has the lowest parameter volume on all data sets, which are 336.96M, 319.24M, 3.54G, and 5.34s respectively. This shows the advantage of our model in model complexity with a more lightweight architecture compared to competing models. In terms of computational complexity (Flops), our model also performs well, with Flops of 3.54G, 3.65G, 5.61G, and 5.61G respectively. In contrast, the Flops values of competing models are generally higher, which

Table 1(a). The comparison of different models in different indicators comes from Amazon product review dataset, Yelp review dataset, IMDB movie reviews dataset, and Twitter sentiment analysis dataset

Model	Datasets							
	Amazon Product Review Dataset				Yelp Review Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
Pondel(Pondel et al., 2021)	**90.7**	**90.52**	**89.18**	**91.65**	**86.68**	**84.14**	**88.59**	**87.74**
Wu(Wu & Yang, 2021)	95.41	92.2	85.19	90.97	88	89.24	86.5	85.8
Almahmood(Almahmood & Tekerek, 2022)	95.35	86.44	87.53	93.24	88.68	89.96	87.56	89.58
Liu(Y. Liu et al., 2020)	89.82	91.49	85.51	88.85	89.24	87.72	84.21	88.1
Zhang(Zhang, 2021)	87.65	90.33	88.13	90.48	94.24	87.32	89.84	93.56
Cao(Cao et al., 2022)	93.86	89.67	85.69	89.99	96.32	85.8	90.43	83.83
Ours	92.58	96.7	94.45	96.22	98.15	96.58	96.83	91.91

Table 1(b). The comparison of different models in different indicators comes from Amazon product review dataset, Yelp review dataset, IMDB movie reviews dataset, and Twitter sentiment analysis dataset

Model	Datasets							
	IMDB Movie Reviews Dataset				Twitter Sentiment Analysis Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
Pondel	86.33	88.61	86.45	93.22	93.48	89.09	87.92	84.12
Wu.	95.57	89.95	85.49	89.98	86.47	85.65	85.53	88.56
Almahmood	96.22	89.97	87.9	90.19	92.27	86.05	90.44	84.26
Liu	88.35	93.67	90.33	92.85	94.07	83.78	87.62	90.38
Zhang	95.26	85.64	91.15	90.26	91.65	87.96	83.99	90.98
Cao	85.75	83.92	88.02	85.56	95.5	91.26	85.99	84.63
Ours	94.64	97.4	97.45	93.35	93.06	94.35	98.15	94.24

shows that our model has a significant advantage in computational efficiency. Another key indicator is the inference time (Inference Time). Our model has achieved lower inference times on each data set, which are 5.34ms, 5.61ms, 5.61ms, and 5.34ms respectively. This means that our model can provide users with sentiment analysis results more quickly, making it more real-time in practical e-commerce applications. Finally, in terms of training time, our model also achieves performance unmatched by competitive models on all data sets. The training times are 326.27s, 319.24s, 3.65s, and 5.61s respectively, which means that our model can train and model iterate faster, improving work efficiency.

In order to present these performance comparisons more clearly, we further visualize these data, as shown in table 2. These visualizations highlight the significant advantages of our model over competing models in several key aspects, including parameter size, computational complexity, inference time, and training time.

As shown in Table 3, we conducted ablation experiments to compare the performance of different models on four sentiment analysis datasets. The table contains key indicators about accuracy, recall, F1 score, and AUC (area under the curve). These data are intended to evaluate the performance of individual models on sentiment analysis tasks. First, our model (Ours) achieved the highest accuracy on

Figure 5. Comparison of model performance on different datasets

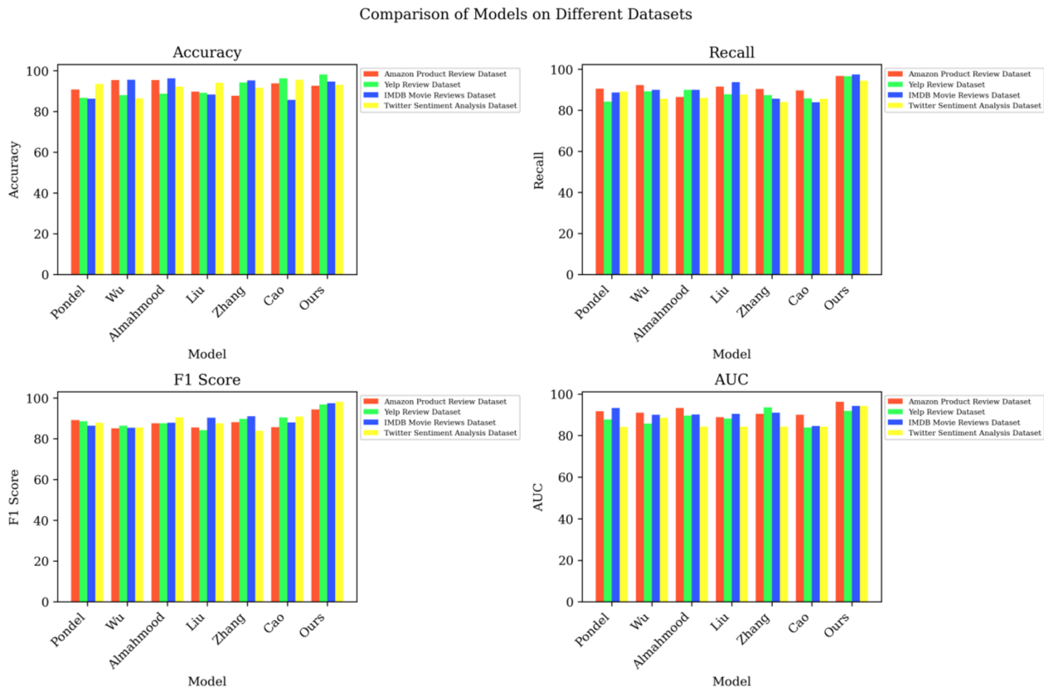


Table 2(a). The comparison of different models in different indicators comes from Amazon product review dataset, Yelp review dataset, IMDB movie reviews dataset, and Twitter sentiment analysis dataset

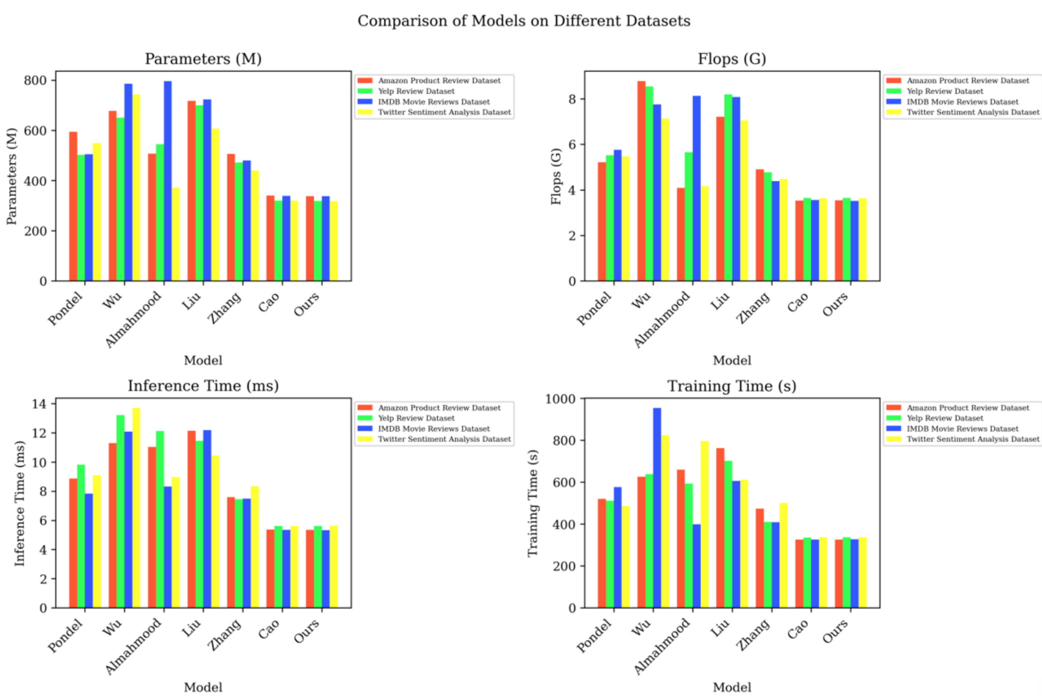
Model	Datasets							
	Amazon Product Review Dataset				Yelp Review Dataset			
	Parameters (M)	Flops (G)	Inference Time (ms)	Trainning Time (s)	Parameters (M)	Flops (G)	Inference Time (ms)	Trainning Time (s)
Pondel	593.93	5.21	8.87	520.14	501.83	5.53	9.82	511.25
Wu	677.86	8.78	11.29	625.48	650.68	8.55	13.21	638.98
Almahmood	506.61	4.09	11.03	659.72	543.99	5.66	12.11	593.05
Liu	717.27	7.21	12.13	762.44	700.19	8.20	11.46	701.33
Zhang	505.93	4.90	7.58	473.83	472.12	4.77	7.45	410.68
Cao	339.78	3.53	5.37	326.11	319.43	3.65	5.61	335.25
Ours	336.96	3.54	5.34	326.27	319.24	3.65	5.61	336.84

all datasets, 96.67%, 96.74%, 96.98%, 97.88% respectively. This highlights the superior performance of our model in terms of sentiment analysis accuracy compared to competing models. In addition, for the two key indicators of recall and F1 score, our model also achieved the highest values on all data sets. The recall rates are 96.19%, 93.94%, 95.88%, and 95.97% respectively, while the F1 scores are 93.76, 93.34, 92.92, and 92.98 respectively. This shows that our model not only has high accuracy, but also has good recall and F1 score while correctly classifying emotions. Finally, AUC serves as an indicator of the area under the curve and is used to evaluate the performance of the classification

Table 2(b). The comparison of different models in different indicators comes from Amazon product review dataset, Yelp review dataset, IMDB movie reviews dataset, and Twitter sentiment analysis dataset

Model	Datasets							
	IMDB Movie Reviews Dataset				Twitter Sentiment Analysis Dataset			
	Parameters (M)	Flops (G)	Inference Time (ms)	Trainning Time (s)	Parameters (M)	Flops (G)	Inference Time (ms)	Trainning Time (s)
Pondel	504.63	5.76	7.83	576.95	549.61	5.47	9.09	486.80
Wu	785.89	7.76	12.07	653.51	742.84	7.13	13.69	824.05
Almahmood	796.04	8.13	8.31	399.41	371.42	4.16	8.97	795.69
Liu	722.93	8.08	12.17	606.27	606.23	7.05	10.43	611.03
Zhang	480.26	4.39	7.49	409.18	439.24	4.47	8.34	500.64
Cao	338.45	3.56	5.34	325.78	319.91	3.63	5.60	337.59
Ours	337.50	3.52	5.32	326.63	318.05	3.63	5.63	337.42

Figure 6. Comparison of model performance on different datasets



model. Our model achieved the highest AUC values on all datasets, which were 95.45, 94.13, 94.16, 94.24 respectively. This once again highlights the superior performance of our model in sentiment analysis tasks.

To present these performance comparisons more clearly, we visualize these data, as shown in Figure 6. These visualization results further demonstrate and highlight the significant advantages of our model over competing models on multiple key metrics such as accuracy, recall, F1 score, and AUC.

Table 3(a). Ablation experiments on the BERT-LSTNet module comes from Amazon product review dataset, Yelp review dataset, IMDB movie reviews dataset, and Twitter sentiment analysis dataset

Model	Datasets							
	Amazon Product Review Dataset				Yelp Review Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
CNN-LSTM	86.3	92.52	83.86	92.95	85.69	92.8	89.08	89.38
Bert-BiGRU	92.84	87.87	89.53	87.99	95.42	90.63	86.1	89.49
Bi-LSTM	90.79	86.94	84.6	88.82	88.17	84.24	87.6	86.51
Ours	96.67	96.19	93.76	95.45	96.74	93.94	93.34	94.13

Table 3(b). Ablation experiments on the BERT-LSTNet module comes from Amazon product review dataset, Yelp review dataset, IMDB movie reviews dataset, and Twitter sentiment analysis dataset

Model	Datasets							
	IMDB Movie Reviews Dataset				Twitter Sentiment Analysis Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
CNN-LSTM	94.6	92.66	90.44	89.34	95.12	85.76	85.76	86.69
Bert-BiGRU	92.42	92.74	89.68	85.61	95.01	85.09	87.68	85.7
Bi-LSTM	92.72	91.75	90.9	90.66	87.23	91.3	90.07	90.09
Ours	96.98	95.88	92.92	94.16	97.88	95.97	92.98	94.24

As shown in figure 7, we conducted cross-module ablation experiments to compare the performance of different optimizers (Adam, RMSprop, Nadam and our method) on different data sets. The table includes key indicators such as accuracy (Accuracy), number of parameters (Parameters), floating point operations (Flops), inference time (Inference Time), and training time (Training Time). In table 4, these data are used to evaluate the performance of individual optimizers on sentiment analysis tasks. First of all, our method has excellent performance in accuracy, reaching 94.87%, which has obvious advantages over the accuracy of other optimizers (Adam, RMSprop, Nadam) (86.74%, 85.82%, 92.8%). This shows that our method is better able to achieve correct classification in sentiment analysis tasks. Second, our method performs well in terms of the number of parameters and floating point operands. The number of parameters is 8.48M, and Flops is 92.64G. This shows that our method has relatively small model size and computational cost while maintaining high accuracy, which is very important for sentiment analysis tasks under resource constraints. Furthermore, our method also performs well in inference time and training time, which are 92.64ms and 92.27s respectively. This means that our model can perform inference and training in a shorter time, improving efficiency.

Finally, we visualize these performance comparisons as shown in figure 8. These visualizations highlight the significant advantages of our method over other optimizers in terms of accuracy and computational efficiency.

5. CONCLUSION AND DISCUSSION

In this study, we introduce an e-commerce review sentiment analysis and purchase intention prediction model based on deep learning technology, namely the BERT-LSTNet-Softmax model. Our research

Figure 7. Comparison of model performance on different datasets

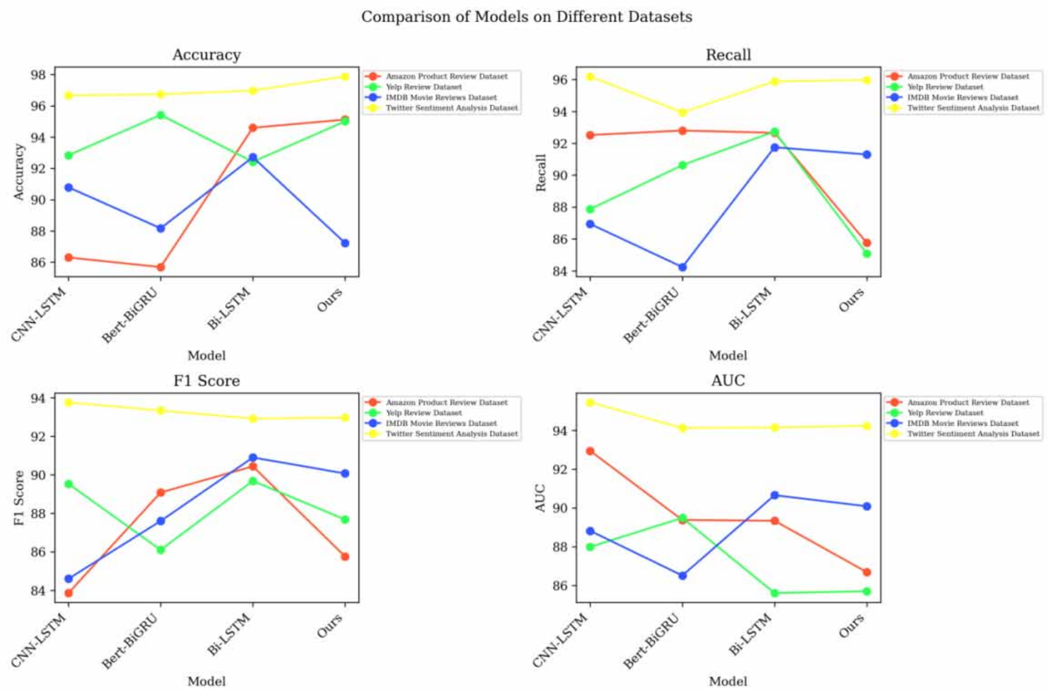


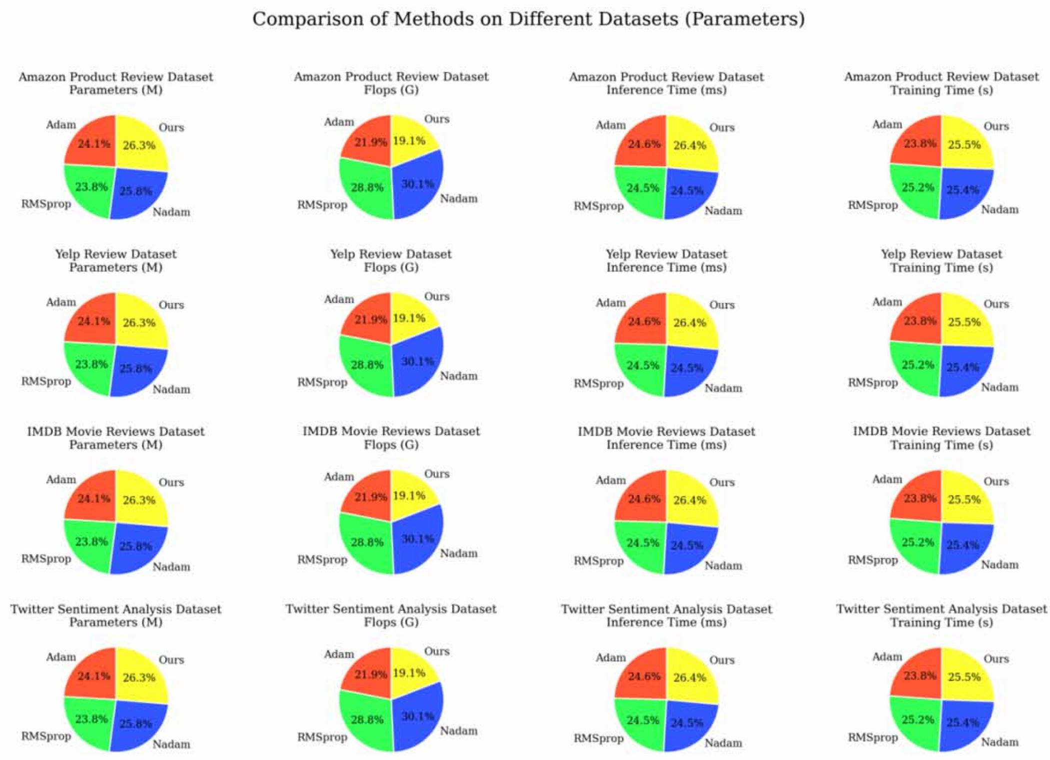
Table 4(a). Ablation experiments on the cross module using different datasets

Model	Datasets							
	Amazon Product Review Dataset				Yelp Review Dataset			
	Parameters (M)	Flops (G)	Inference Time (ms)	Trainning Time (s)	Parameters (M)	Flops (G)	Inference Time (ms)	Trainning Time (s)
Adam	86.74	9.71	84.35	86.3	90.37	8.43	86.45	84.23
RMSprop	85.82	8.11	84.01	89.13	95.79	11.09	84.19	84.05
Nadam	92.8	11.23	86.15	92	87.13	11.59	85	89.7
Ours	94.87	8.48	92.64	92.27	91.22	7.36	94.05	92.66

Table 4(b). Ablation experiments on the cross module using different datasets

Model	Datasets							
	IMDB Movie Reviews Dataset				Twitter Sentiment Analysis Dataset			
	Parameters (M)	Flops (G)	Inference Time (ms)	Trainning Time (s)	Parameters (M)	Flops (G)	Inference Time (ms)	Trainning Time (s)
Adam	87.9	10.79	86.39	92.59	86.72	9.51	86.86	85.1
RMSprop	86.94	10.75	86.26	84.21	96.41	11.64	87.46	89.96
Nadam	86.69	10.71	86.07	84.45	91.49	9.64	90.5	90.76
Ours	93.04	8.55	93	92.47	91.91	8.29	88.89	91.14

Figure 8. Comparison of model performance on different datasets



looks at providing insights into the link between the sentiment of e-commerce product reviews and purchase intent. Through a review of the experimental results, we first clarify the research question, which is how to effectively analyze the sentiment of e-commerce reviews in order to predict consumers' purchase intentions. Next, we introduced the BERT-LSTNet-Softmax model in detail, where BERT is responsible for text feature extraction, LSTNet is used for sentiment analysis and purchase intention prediction, and Softmax is used for output processing. We comprehensively describe the model design and training process, and conduct extensive experiments on multiple datasets with satisfactory results. Experiments have proven that the BERT-LSTNet-Softmax model performs well in e-commerce review sentiment analysis and purchase intention prediction tasks, providing a powerful tool for a deeper understanding of consumer purchase decisions.

Despite the remarkable success of the BERT-LSTNet-Softmax model, we honestly point out that it has some limitations. First, model training on large-scale data sets requires relatively long time and high computational cost. Secondly, the interpretability of the model needs to be improved, especially in terms of sentiment analysis and the specific decision-making process of purchase intention prediction. These challenges will be a key focus of our future research.

For future work, we plan to further improve the performance and scalability of the BERT-LSTNet-Softmax model in multiple aspects. First, we will delve into more pre-trained language models and sentiment analysis methods to improve the performance level of the model. Secondly, we plan to expand research application scenarios, including but not limited to online advertising recommendation and product recommendation, to broaden the application scope of the model. Most importantly, we hope to promote the application of deep learning technology in the field of e-commerce through this research, provide consumers with more accurate product and service choices, and promote the healthy development of the e-commerce market.

ACKNOWLEDGMENTS

Thanks to the support of 2023 Guangxi college young and middle -aged teachers' scientific research basic ability improvement project.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

FUNDING AGENCY

This paper is funded by the 2023 Guangxi college young and middle -aged teachers' scientific research basic ability improvement project: Research on the formation mechanism and path of rural e-commerce live broadcasts on Guangxi's specialty agricultural product brands under the background of rural revitalization.(2023KY0192)

REFERENCES

Almahmood, R. J. K., & Tekerek, A. (2022). Issues and Solutions in Deep Learning-Enabled Recommendation Systems within the E-Commerce Field. *Applied Sciences (Basel, Switzerland)*, *12*(21), 11256. doi:10.3390/app122111256

Alsubari, S. N., Deshmukh, S. N., Al-Adhaileh, M. H., Alsaade, F. W., & Aldhyani, T. H. (2021). Development of integrated neural network model for identification of fake reviews in E-commerce using multidomain datasets. *Applied Bionics and Biomechanics*, *2021*, 2021. doi:10.1155/2021/5522574 PMID:33953796

Alzahrani, M. E., Aldhyani, T. H., Alsubari, S. N., Althobaiti, M. M., & Fahad, A. (2022). Developing an intelligent system with deep learning algorithms for sentiment analysis of E-commerce product reviews. *Computational Intelligence and Neuroscience*, *2022*, 2022. doi:10.1155/2022/3840071 PMID:35669644

Cao, Y., Shao, Y., & Zhang, H. (2022). Study on early warning of E-commerce enterprise financial risk based on deep learning algorithm. *Electronic Commerce Research*, *22*(1), 21–36. doi:10.1007/s10660-020-09454-9

Chou, H.-H., & Tsai, F.-S. (2022). Technology-enabled mobilization in the emergence of a value co-creating ecosystem. [JOEUC]. *Journal of Organizational and End User Computing*, *34*(1), 1–17. doi:10.4018/JOEUC.312855

Dhote, S., Vichoray, C., Pais, R., Baskar, S., & Mohamed Shakeel, P. (2020). Hybrid geometric sampling and AdaBoost based deep learning approach for data imbalance in E-commerce. *Electronic Commerce Research*, *20*(2), 259–274. doi:10.1007/s10660-019-09383-2

Esmeli, R., Bader-El-Den, M., & Abdullahi, H. (2021). Towards early purchase intention prediction in online session based retailing systems. *Electronic Markets*, *31*(3), 697–715. doi:10.1007/s12525-020-00448-x

Farimani, S. A., Jahan, M. V., Fard, A. M., & Tabbakh, S. R. K. (2022). Investigating the informativeness of technical indicators and news sentiment in financial market price prediction. *Knowledge-Based Systems*, *247*, 108742. doi:10.1016/j.knosys.2022.108742

Hicham, N., Karim, S., & Habbat, N. (2023). Enhancing Arabic Sentiment Analysis in E-Commerce Reviews on Social Media Through a Stacked Ensemble Deep Learning Approach. *Mathematical Modelling of Engineering Problems, 10*(3).

Huang, C., Han, Z., Li, M., Wang, X., & Zhao, W. (2021). Sentiment evolution with interaction levels in blended learning environments: Using learning analytics and epistemic network analysis. *Australasian Journal of Educational Technology*, *37*(2), 81–95. doi:10.14742/ajet.6749

Huang, J., Chen, Q., & Yu, C. (2022). A New Feature Based Deep Attention Sales Forecasting Model for Enterprise Sustainable Development. *Sustainability (Basel)*, *14*(19), 12224. doi:10.3390/su141912224

Issaoui, Y., Khiat, A., Bahnasse, A., & Ouajji, H. (2021). An advanced LSTM model for optimal scheduling in smart logistic environment: E-commerce case. *IEEE Access : Practical Innovations, Open Solutions*, *9*, 126337–126356. doi:10.1109/ACCESS.2021.3111306

Jain, S., & Roy, P. K. (2022). E-commerce review sentiment score prediction considering misspelled words: A deep learning approach. *Electronic Commerce Research*, 1–25. doi:10.1007/s10660-022-09582-4

Kalifa, D., Singer, U., Guy, I., Rosin, G. D., & Radinsky, K. (2022). Leveraging World Events to Predict E-Commerce Consumer Demand under Anomaly. *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. ACM.

Koukaras, P., Nousi, C., & Tjortjis, C. (2022). Stock market prediction using microblogging sentiment analysis and machine learning. Telecom.

Li, H., Lu, Y., Zhu, H., & Ma, Y. (2023). A Novel AB-CNN Model for Multi-Classification Sentiment Analysis of e-Commerce Comments. *Electronics (Basel)*, *12*(8), 1880. doi:10.3390/electronics12081880

Li, Q., Li, X., Lee, B., & Kim, J. (2021). A hybrid CNN-based review helpfulness filtering model for improving e-commerce recommendation Service. *Applied Sciences (Basel, Switzerland)*, *11*(18), 8613. doi:10.3390/app11188613

Li, X., Li, Q., & Kim, J. (2023). A Review Helpfulness Modeling Mechanism for Online E-commerce: Multi-Channel CNN End-to-End Approach. *Applied Artificial Intelligence*, *37*(1), 2166226. doi:10.1080/08839514.2023.2166226

Li, Z., Zhou, X., & Huang, S. (2021). Managing skill certification in online outsourcing platforms: A perspective of buyer-determined reverse auctions. *International Journal of Production Economics*, *238*, 108166. doi:10.1016/j.ijpe.2021.108166

Ling, C., Zhang, T., & Chen, Y. (2019). Customer purchase intent prediction under online multi-channel promotion: A feature-combined deep learning framework. *IEEE Access : Practical Innovations, Open Solutions*, *7*, 112963–112976. doi:10.1109/ACCESS.2019.2935121

Liu, C.-J., Huang, T.-S., Ho, P.-T., Huang, J.-C., & Hsieh, C.-T. (2020). Machine learning-based e-commerce platform repurchase customer prediction model. *PLoS One*, *15*(12), e0243105. doi:10.1371/journal.pone.0243105 PMID:33270714

Liu, Y., Lu, J., Yang, J., & Mao, F. (2020). Sentiment analysis for e-commerce product reviews by deep learning model of Bert-BiGRU-Softmax. *Mathematical Biosciences and Engineering*, *17*(6), 7819–7837. doi:10.3934/mbe.2020398 PMID:33378922

Liu, Y., Tian, Y., Xu, Y., Zhao, S., Huang, Y., Fan, Y., Duan, F., & Guo, P. (2021). TPGN: A Time-Preference Gate Network for e-commerce purchase intention recognition. *Knowledge-Based Systems*, *220*, 106920. doi:10.1016/j.knosys.2021.106920

Lu, C.-W., Lin, G.-H., Wu, T.-J., Hu, I.-H., & Chang, Y.-C. (2021). Influencing factors of cross-border e-commerce consumer purchase intention based on wireless network and machine learning. *Security and Communication Networks*, *2021*, 1–9. doi:10.1155/2021/8388480

Lu, C.-Y. J.-Y., & Wu, H.-T. (2022). A Hierarchical Clustering Federated Learning System Based on Industry 4.0. [JOEUC]. *Journal of Organizational and End User Computing*, *34*(1), 1–16. doi:10.4018/JOEUC.313194

Luo, Y., Ma, J., & Li, C. (2020). Entity name recognition of cross-border e-commerce commodity titles based on TWs-LSTM. *Electronic Commerce Research*, *20*(2), 405–426. doi:10.1007/s10660-019-09371-6

Pallathadka, H., Ramirez-Asis, E. H., Loli-Poma, T. P., Kaliyaperumal, K., Ventayen, R. J. M., & Naved, M. (2023). Applications of artificial intelligence in business management, e-commerce and finance. *Materials Today: Proceedings*, *80*, 2610–2613. doi:10.1016/j.matpr.2021.06.419

Policarpo, L. M., da Silveira, D. E., da Rosa Righi, R., Stoffel, R. A., da Costa, C. A., Barbosa, J. L. V., Scorsatto, R., & Arcot, T. (2021). Machine learning through the lens of e-commerce initiatives: An up-to-date systematic literature review. *Computer Science Review*, *41*, 100414. doi:10.1016/j.cosrev.2021.100414

Pondel, M., Wuczynski, M., Gryncewicz, W., Lysik, L., Hernes, M., Rot, A., & Kozina, A. (2021). Deep Learning for Customer Churn Prediction in E-Commerce Decision Support. BIS, Qiu, J., & Wang, S. (2022). A deep matching model for detecting reviews mismatched with products in e-commerce. *Applied Soft Computing*, *129*, 109624.

Sarker, K. U., Saqib, M., Hasan, R., Mahmood, S., Hussain, S., Abbas, A., & Deraman, A. (2022). A Ranking Learning Model by K-Means Clustering Technique for Web Scraped Movie Data. *Computers*, *11*(11), 158. doi:10.3390/computers11110158

Umer, M., Ashraf, I., Mehmood, A., Kumari, S., Ullah, S., & Sang Choi, G. (2021). Sentiment analysis of tweets using a unified convolutional neural network-long short-term memory network model. *Computational Intelligence*, *37*(1), 409–434. doi:10.1111/coin.12415

Venkatesan, R., & Sabari, A. (2023). Deepsentimodels: A Novel Hybrid Deep Learning Model for an Effective Analysis of Ensembled Sentiments in E-Commerce and S-Commerce Platforms. *Cybernetics and Systems*, *54*(4), 526–549. doi:10.1080/01969722.2022.2148510

Vo, A.-D., Nguyen, Q.-P., & Ock, C.-Y. (2019). Sentiment analysis of news for effective cryptocurrency price prediction. *International Journal of Knowledge Engineering*, *5*(2), 47–52. doi:10.18178/ijke.2019.5.2.116

Wang, S., & Qiu, J. (2021). A deep neural network model for fashion collocation recommendation using side information in e-commerce. *Applied Soft Computing*, *110*, 107753. doi:10.1016/j.asoc.2021.107753

Wang, X., & Tong, Y. (2021). Application of an emotional classification model in e-commerce text based on an improved transformer model. *PLoS One*, *16*(3), e0247984. doi:10.1371/journal.pone.0247984 PMID:33667262

Wu, P., & Yang, D. (2021). E-commerce workshop scheduling based on deep learning and genetic algorithm. *International Journal of Simulation Modelling*, *20*(1), 192–200. doi:10.2507/IJSIMM20-1-CO4

Wu, X., Magnani, A., Chaidaroon, S., Puthenputhussery, A., Liao, C., & Fang, Y. (2022). A Multi-task Learning Framework for Product Ranking with BERT. Proceedings of the ACM Web Conference 2022. ACM.

Yang, L., Li, Y., Wang, J., & Sherratt, R. S. (2020). Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. *IEEE Access : Practical Innovations, Open Solutions*, *8*, 23522–23530. doi:10.1109/ACCESS.2020.2969854

Yang, X., Yang, Y., Su, J., Sun, Y., Fan, S., Wang, Z., Zhang, J., & Chen, J. (2022). Who's Next: Rising Star Prediction via Diffusion of User Interest in Social Networks. *IEEE Transactions on Knowledge and Data Engineering*, *35*(5), 5413–5425. doi:10.1109/TKDE.2022.3151835

Yuan, J., Li, Z., Zou, P., Gao, X., Pan, J., Ji, W., & Wang, X. (2022). Community Trend Prediction on Heterogeneous Graph in E-commerce. Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining. ACM.

Zhou, L. (2020). Product advertising recommendation in e-commerce based on deep learning and distributed expression. *Electronic Commerce Research*, *20*(2), 321–342. doi:10.1007/s10660-020-09411-6

Muhammad Asif is currently working as Director, Graduate Studies and Reserach and Tenured Associate Professor of Computer Science. He also served as Chairman of the Department of Computer Science at National Textile University, Faisalabad, till November 04, 2020. Before this, he was a research scholar in the Computer Science and Information Management Department at the Asian Institute of Technology, Thailand. He received his MS and Ph.D. from AIT in 2009 and 2012 on HEC foreign Scholarship. Over the time, he was a visiting researcher at the National Institute of Information Tokyo, Japan, where he worked on the design and development of a data collection system for infectious diseases through modern technologies. He has also worked on a very important project of the Pakistan Air Force (PAF) related to Pakistan's Air Traffic Control System. He was mainly involved in the architectural design of the systems as a software engineer. Asif is actively involved in teaching research and administration activities at the national textile University Faisalabad. He has Two PhD students who graduated in his credit and fifty-five plus MS and BS level students. He has won funded projects of more than Three Hundred (300) million from National and International funding agencies. He enjoys more than 200 impact factors from his research publications in top-class computer science journals and allied domains. He is also serving as Associate Editor of many top journals of computer science, namely, IEEE Access (IF 3.37), PlosOne (IF 3.752), Peerj Computer Science (IF 2.41), CMC (IF 3.77), Frontiers in Artificial Ingelligence (IF 4.0) and Frontiers in Communications and Networks. He is a reviewer of several reputed journals and authored a number of research papers in reputed journals and conferences. He is also a permanent member of the Punjab Public Service Commission (PPSC) as an advisor and program evaluator at the National Computing Education Accreditation Council (NCEAC) Islamabad.