Hindawi
Mobile Information Systems
Volume 2021, Article ID 5568208, 14 pages
https://doi.org/10.1155/2021/5568208

Research Article

Multimodal Data Guided Spatial Feature Fusion and Grouping Strategy for E-Commerce Commodity Demand Forecasting

Weiwei Cai, Yaping Song, and Zhanguo Wei

School of Logistics and Transportation, Central South University of Forestry and Technology, Changsha 410004, China

Correspondence should be addressed to Zhanguo Wei; t20110778@csuft.edu.cn

Received 4 May 2021; Revised 24 May 2021; Accepted 1 June 2021; Published 12 June 2021

Academic Editor: Fazlullah Khan

Copyright © 2021 Weiwei Cai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

E-commerce offers various merchandise for selling and purchasing with frequent transactions and commodity flows. An accurate prediction of customer needs and optimized allocation of goods is required for cost reduction. The existing solutions have significant errors and are unsuitable for addressing warehouse needs and allocation. That is why businesses cannot respond to customer demands promptly, as they need accurate and reliable demand forecasting. Therefore, this paper proposes spatial feature fusion and grouping strategies based on multimodal data and builds a neural network prediction model for e-commodity demand. The designed model extracts order sequence features, consumer emotional features, and facial value features from multimodal data from e-commerce products. Then, a bidirectional long short-term memory network- (BiLSTM-) based grouping strategy is proposed. The proposed strategy fully learns the contextual semantics of time series data while reducing the influence of other features on the group's local features. The output features of multimodal data are highly spatially correlated, and this paper employs the spatial dimension fusion strategy for feature fusion. This strategy effectively obtains the deep spatial relations among multimodal data by integrating the features of each column in each group across spatial dimensions. Finally, the proposed model's prediction effect is tested using e-commerce dataset. The experimental results demonstrate the proposed algorithm's effectiveness and superiority.

1. Introduction

The e-commerce platform offers a wide range of commodities, with frequent purchases, transactions, and commodity flows. The dynamic and complex business environment has posed significant challenges for business decision-making. As a result, inventory management has become more complex, and supply chain costs have risen steadily [1, 2]. At the same time, the supply chain system, which includes businesses, upstream suppliers, and downstream customers, is becoming increasingly complicated. Logistics [3, 4] and marketing are interconnected and gradually integrated into this increasingly complex system, with the ultimate goal of satisfying customer marketing activities, as a result, putting the customer first, utilizing information technology to its full potential, and accurately forecasting consumer demand [5–8]. The consumer demand for goods is a critical link in an enterprise's supply chain process. The accuracy of e-commerce commodity demand forecasting determines its value, exceptionally reliable commodity demand forecasting, which is critical for e-commerce. Wrong or inaccurate forecasts can significantly impact product allocation and distribution, damaging e-commerce companies' decision-making efficiency and resource allocation. With a better understanding of consumer demands, companies can create better inventory plans with competitive prices and timely promotion plans. It improves customer satisfaction and service quality, lowers supply chain costs, and increases corporate profits and brand value. However, many factors will influence commodity demand forecasting, particularly for customers in different regions, and many uncertain influencing factors will lead to changes in demand. In addition, time series analysis [9–12] is commonly used in traditional sales forecasting techniques [13, 14]. Only historical sales data is used as a source of information. These techniques can be used with products

that have consistent or seasonal sales patterns. However, e-commerce platforms typically have many nonlinear, unstructured multimodal data, making traditional analysis and prediction difficult.

Fortunately, the rise of the mobile Internet, low-cost sensors, and low-cost storage has made obtaining large amounts of data more accessible. We can collect many other log data of e-commerce products over time, in addition to historical sales data. It includes consumer reviews, consumer portraits, page views (PV), search page views (SPV), user views (UV), search user view (SUV), selling price (PAY), user location (UL), and total merchandise sales (GMV). It provides a broad space for applying neural networks when combined with cheap computing power, especially the dramatic increase in GPU performance. Based on the above observations, this paper proposes spatial feature fusion and grouping strategies based on multimodal data and builds a neural network prediction model for e-commerce commodity demand. Initially, we consider the multimodal data of e-commerce products (such as historical orders, consumer reviews, and consumer portraits) while extracting different features. These features are order sequence features, consumer emotional features, and facial value features. Finally, we proposed a grouping strategy based on a bidirectional long-term, short-term memory network (BiLSTM). The network fully learns the contextual semantics of time series data while reducing the influence of other features on the local features of the group.

The main innovations of this article are as follows:

(1) This paper considers numerical data such as historical orders and text and image data such as consumer comments and portraits. Because nonlinear data like text and image are becoming more important in e-commerce prediction tasks, the analysis value increases. We can intuitively understand the customer's desire to buy a particular product using semantic sentiment analysis of consumer comments. We can depict the consumer portrait and understand the consumer's preferences by calculating the appearance level of the consumer portrait, which is useful for improving the prediction model's performance.

(2) This paper proposes a novel grouping strategy to consider both the long-distance dependence and the short-distance dependence in sequence data. It addresses the problem that the recurrent neural network only pays attention to the long-distance dependence in sequence data. Because of the short-distance dependence, when a significant distance separates two sets of features, their connection is weak, and less information is retained.

(3) This paper proposes a novel spatial dimension fusion strategy. It effectively obtains the deep spatial relations among multimodal data by integrating each column's features across spatial dimensions.

(4) The prediction effect of this model is verified using a dataset created by an e-commerce platform. The experimental results demonstrate the effectiveness and superiority of our algorithm.

The rest of the paper is organized as follows. In Section 2, related work is studied. In Section 3, the methodology is given. In Section 4, results and discussion are explained, followed by a conclusion in Section 5.

2. Related Work

This section discusses related work from various aspects to understand the problem addressed in this paper.

2.1. Forecast of Demand for Commodities. Demand forecasting [15, 16] is a critical component of the e-commerce supply chain and commodity inventory management. Improving the accuracy of forecasting results by studying the factors that influence demand is critical. The company's replenishment strategy and inventory cost reduction rely heavily on accurate forecasting results. Scholars divide forecasting methods into qualitative and quantitative forecasting based on forecasting. However, this article's research only looks at quantitative forecasting. The quantitative forecasting methods can be mainly divided into traditional time series forecasting, combined forecasting [17–19], and the current popular deep neural network methods.

(1) Traditional time series forecasting: the time series forecasting method is based on the continuous law of objective thing development. It is used to further speculate on the future development trend using historical data and statistical analysis. Models for time series forecasting t have been widely used in the economic field including the growth rate, moving average, exponential smoothing, random time series model, gray model, chaos, and fractal. Tsai et al. [20] proposed a hybrid time series model based on feature selection methods to predict stock prices in leading industries. The results show that the proposed model has better accuracy than the other listed models and provides convincing investment guidance. Mustafa and Yumusak [21] used seasonal time series methods to predict the natural gas demand in Sakarya Province, Turkey, and obtained considerable forecast results, with an average absolute percentage error of 15%. Recently, Maleki et al. [22] used a two-piece mixed normal distribution autoregressive time series model (called the TP-SMN-AR model) to predict confirmed and cured COVID-19 cases, with an average absolute percentage error of 1.6%. The result is helpful for disease control and resource allocation planning.

Traditional time series forecasting methods have proven to be simple and efficient when dealing with relatively simple linear data. They are widely used in all walks of life [23, 24]. However, the error is relatively large for data with a complex structure.

(2) Combination forecasting: a single forecasting method is difficult to manage for some relatively complex and challenging forecasting tasks. Its forecasting accuracy can be improved by combining a reasonable number of different methods in a scientific manner. Some scholars have studied combination forecasting methods extensively since J. N. Bates and C. W. J. Granger published "Combined Forecasting" in the 1870s. Huard et al. [25] constructed an e-commerce sales volume prediction model based on exponential smoothing and Holt linear trend methods. They proved the effectiveness of the model on the sales dataset provided by the e-commerce company Cdiscount. In order to obtain a more accurate sales forecast during the price war, Hsieh et al. [15] adjusted the seasonal index by using a simple moving average, removed the seasonal index from the sales volume, and then used the previous month's data subjected to regression analysis. Finally, a more accurate sales forecast is obtained in the price war. Bowen et al. [26] used two independent ARIMA-BP nonlinear combination models to predict sales in the next 5 days. They established a mean square error model to weigh the fitting and prediction results of the two predictions, which can better deal with the e-commerce merchandise sales forecast problem.

The combined forecasting model, in general, has clear advantages. For example, it can handle some relatively complex and difficult forecasting tasks. The combined forecasting method is better than a single method based on the traditional time series model. It is still inferior to a single method based on the traditional time series model. With complex unstructured multimodal data, however, the combined forecasting method is still difficult to use.

(3) Deep learning-based prediction model: with the booming development of the mobile Internet and the arrival of big data, the business of the e-commerce platform has become more complex with huge data. Nonlinear and unstructured data has become the most valuable data. Both traditional time series forecasting methods and combined forecasting methods have been unable to cope with the increasingly complex and challenging task of e-commerce demand forecasting. The mining and processing of nonlinear and unstructured data also have natural disadvantages. Fortunately, advances in computing power and the rise of deep learning [27–29] have given us new ways to solve this dilemma. Suchacka and Stemplewski [30] proposed a backpropagation neural network model to predict the purchases of active users in a Web store. Training and evaluation of the neural network are performed using user data reconstructed from server log data. The proposed deep neural network can achieve 99.6% and 87.8% recall rates, and the prediction results are effective and accurate. Giri et al. [31] converted clothing images into feature vectors, combined with historical sales data. They applied a backpropagation neural network to predict the sales of new products. The results show that the model performs well despite the small dataset. Sen and Lin [32] combined the LSTM method with sentiment analysis of consumer reviews. In the training stage, the sales data and comments captured from "Taobao" are preprocessed. The emotional level of the comments is analyzed in terms of "positive," "negative," and "confidence" to build a model to predict the short-term commodity demand in the e-commerce environment. The results show that the sentiment analysis of consumer comments has a great influence on the forecast results.

According to the review, the above deep learning prediction models differ from traditional linear numerical data such as order sales, images, text semantic understanding, and other unstructured multimodal data. It is becoming increasingly important in the e-commerce industry. As a result, using deep learning technology to create relevant predictive models in the e-commerce industry has become commonplace.

3. Methodology

Figure 1 is a flowchart of the overall architecture of our algorithm. The feature extraction methods for multimodal data are as follows: first, perform feature engineering on historical order data to obtain the data combination required for prediction. Second, the purchase desire's weight is calculated using natural language processing technology for text sentiment analysis on consumer review data. The face value of the consumer's portrait is calculated to match the product type. We use n sets of BiLSTM for deep feature extraction in the above three feature sequences and traditional and spatial feature fusion strategies to obtain feature spatial relationships. Finally, we obtain the forecast output of e-commerce commodity demand via the FC layer.

3.1. Feature Engineering. Feature engineering is a critical step in the data preprocessing stage that ensures the best possible feature data for the prediction task. This article first performs feature construction, feature selection, feature extraction, and feature processing on historical order data.

Feature construction: the selection of basic features and the derived features are the two main parts of the feature construction work in this paper. The following is the specific selection procedure:

To extract the basic features, first, select the basic characteristics that influence goods demand, such as the attributes of the commodity itself, sales volume, commodity market performance, and time. In general, ready-made nonattribute feature data required in the research can be extracted statistically when selecting basic features. Each goods ID has 20 basic features, which are counted and extracted in this paper.

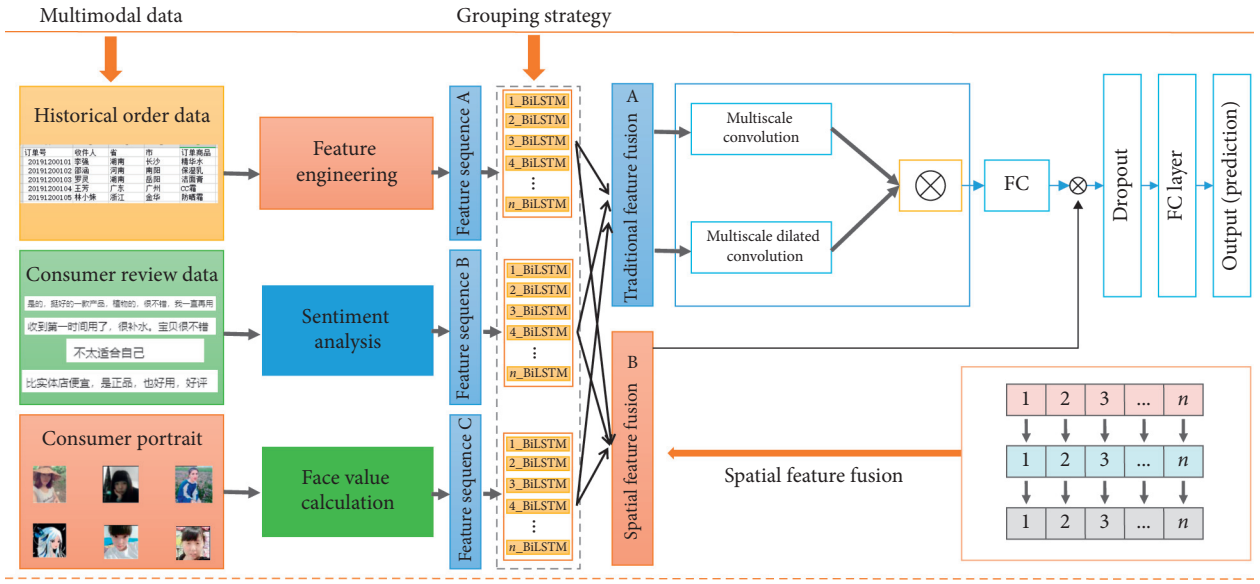

Figure 1: Schematic diagram of the overall architecture of our algorithm. ⊗ represents the feature fusion operation.

The distinguishing factors obtained in this paper include all kinds of commodity attributes, price, market performance, commodity sales, and other characteristic data. We use the time sliding window method to deal with the demand and characteristics of commodities every week. One week (7 days) is taken as a window, in which the demands of each commodity in different areas are called labels. The working principle of the sliding window method is shown in Figure 2.

This paper uses the scaling method to deal with continuous numerical feature data because some historical e-commerce transaction data features have large values, such as the number of views and favorites. On the other hand, some have relatively small values with an extensive feature value range that is usually not conducive to the algorithm's convergence speed. As a result, this paper uses a scaling method to process this type of data that produces a mean value 0 and a variance 1, in addition, to increase the learning rate and then increase the speed of model training. Therefore, we standardize all features. The standardized formula is as follows:

$$\text{standardization}(x) = \frac{x - x_{\min}}{x_{\max} - x_{\min}}. \quad (1)$$

It is a one-hot encoding and distributed representation of attribute or category data. Because such characteristic values are discrete rather than continuous, and there is no sequential distinction between categories, one-hot coding is used. The dimension of characteristic data can be reduced, and data sparsity can be reduced by one-hot coding of characteristic values.

3.2. Sentiment Analysis

3.2.1. Data Collection. In this experiment, text sentiment analysis was used to analyze comments on skincare products that were crawled from an e-commerce platform as shown in Figure 3. In this figure, a business platform is used for goods as data has the advantage of each comment text having a corresponding star comment. Each text corresponding star represents the emotional tendency; we can put a week's comments as a negative tendency of data and five-star comments as a positive tendency of data, so that sentiment analysis can be aided further.

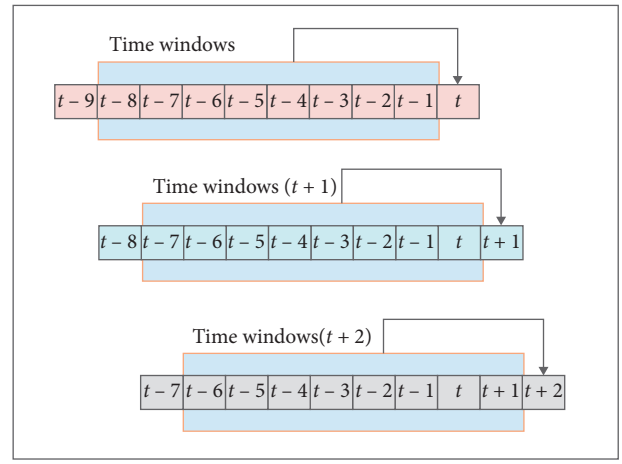

Figure 2: A schematic diagram of the time sliding window.

Each review has a review star, the content of the review, the product reviewed, and the review time, as shown in Figure 3. We can crawl review content and star ratings from related e-commerce platforms such as corpus and tags. However, each product's review data feature is that the data of positive reviews outnumber the data of negative reviews. If a product has a negative review, it should be avoided. When a product receives more than positive feedback, it will be removed from the shelves. To solve this problem, we crawl all of a product's negative review data and then crawl the corresponding positive reviews. This article crawled a total of 10,000 positive review texts, with the majority of them

Figure 3: A screenshot of comment data on an e-commerce platform.

Table 1: Data sample of consumer reviews.

Positive emotion text	Negative emotion text
I have been using this essence water and essence lotion for more than 2 years, and it is best to use it in autumn and winter	Not suitable for me
Buy again, good hydrating effect	The QR code cannot scan the product information, and the face does not feel hydrated when used, and it dries quickly, not as easy to use as other brands

focusing on everyday skincare products. The main characteristics of e-commerce review text, as shown in Table 1, are short text length. These are people's product opinions, which are usually only a few short sentences long. It is a serious case of colloquialization. People do not pay much attention to grammar when they make comments. They write more casually and do not adhere to strict grammatical and syntactic rules. Emotional semantics are difficult to grasp. The context in the text is sometimes relatively high, because the text is short and contains a lot of emotional information. It is possible that one or two words can determine the emotional tone of the entire text, making it difficult to assess the text's emotional tone.

Since the segmentation of wheat field plantation row images is a binary classification task, the numbers of vectors in primary caps and digit caps are both set to 2. The number of capsules in digit caps is also set to 2. In addition, this paper uses the ReLU function as the activation function of the network and uses the sigmoid function for classification.

3.2.2. Word Segmentation. The existing Chinese word segmentation tool JIEBA is used in this paper to perform Chinese word segmentation tasks. It employs a standard probabilistic language model word segmentation method. It can perform various tasks, including word part-of-speech tagging and keyword extraction from text data. The removal of stop words and vectorization of the text will be easier with good word segmentation.

Figure 4 shows a histogram of the number of corresponding texts based on the length of the text we receive. It can be seen that the majority of the e-commerce review texts are less than 300 words long, with only a few exceeding 200 words.

3.2.3. Text Vectorization. The word2vec model maps words to high-dimensional spaces with high efficiency. This model primarily uses text data's context information at a higher level. It employs neural networks to map all text data into a

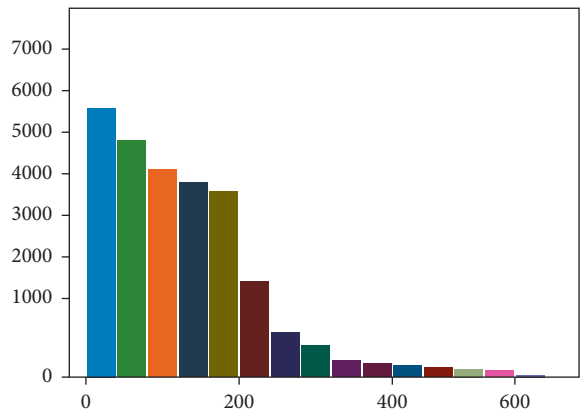

Figure 4: Statistical histogram of text data length.

more low-dimensional, practical, and dense real number matrix. The skip-gram model as shown in Figure 5 is used in this paper, and it consists of three layers: input, hidden, and output. The input of the input layer in the model is the one-hot form of the word $w(t)$. The hidden layer is not the same as the cbow model. It serves no purpose other than transferring data, and the output layer's purpose is to transfer probability. The highest n words' vector is output, and the model's result is calculated using softmax normalization.

In Figure 5, the input vector x represents the one-hot encoding of a Chinese word and the corresponding output vector $\{y_1, y_2, \ldots, y_C\}$. The i^{th} row of the weight matrix W between the input and the hidden layer represents the weight of the i^{th} word in the vocabulary. The objective function of the skip-gram model is

$$L = \frac{1}{T}\sum_{t=1}^{T}\sum_{-n\le j\le n}^{T} \log P\left(w_{t+j}|w_t\right). \tag{2}$$

Finally, the $soft\,\max$ layer is used to make the final sentiment analysis classification decision. The output of $soft\,\max$ represents the relative probability between different emotion categories. Suppose that the sentiment label is $e \in \{1, 2, \ldots, A\}$, and there are a total of A values, which represent A sentiment categories. For sample x, the calculation equation of the conditional probability is as follows:

$$p(e = a|x) = \text{softmax}\left(w_a^T x\right) = \frac{\exp\left(w_a^T x\right)}{\sum_{a=1}^{A}\exp\left(w_a^T x\right)'}, \tag{3}$$

where w_a represents the weight vector of the a-th sentiment category.

3.3. Face Value Calculation. The facial value calculation is used to determine the skin type of the consumer. As far as known, there are five skin types, namely, normal, dry, oily, mixed, and sensitive skin. This paper builds a CNN model to evaluate facial appearance and gives the skin quality classification results.

As shown in Figure 6, the network model here is VGG16, and the corresponding pretraining model is used. The K-fold cross-validation method is used in the training process. The

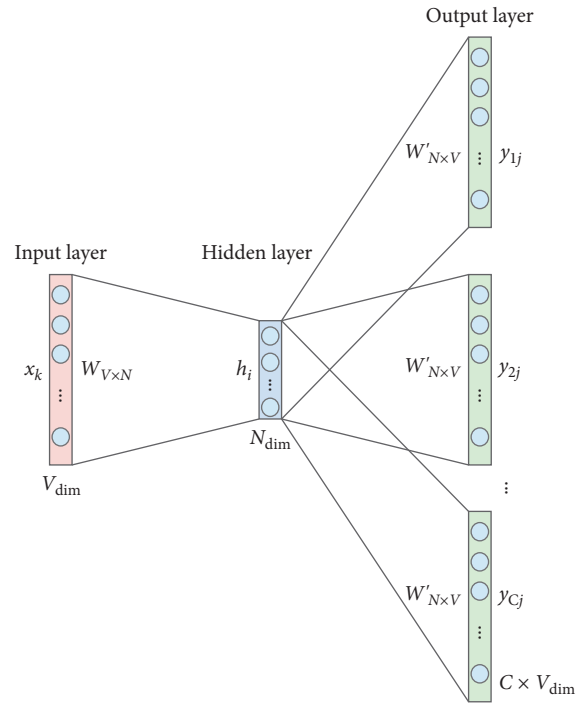

Figure 5: Schematic diagram of skip-gram.

specific idea is to divide the dataset into K parts, one of which is used as the verification set, and the remaining K−1 parts as the training set. The cross-validation is repeated K times. Each piece of data will become a validation set. Finally, the average value will be taken as the accuracy rate.

3.4. Grouping Strategy. The most important feature of a recurrent neural network is making predictions by combining current and previous feature information. The goal is to better preserve the information between the features in the sequence when we only need to consider the most recent part of the information. When a greater distance separates two groups of features, the connection between them is weaker, and thus less information is retained. This situation not only lowers the final prediction accuracy, but also increases the model's computational complexity. As a result,

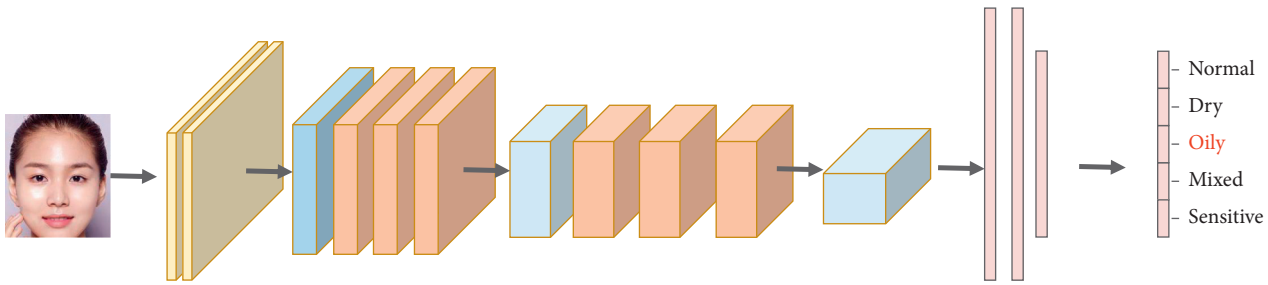

Figure 6: Schematic diagram of facial value calculation and skin quality evaluation model.

the GBL grouping sequence strategy was created. The following is the grouping strategy's calculation equation:

$$G = \{g_1, g_2, \ldots, g_n\}, \tag{4}$$

where G represents the overall feature sequence, g represents the local feature sequence, and n represents the number of groups.

To extract feature sequences A, B, and C from multimodal data, the calculation equation is as follows:

$$\begin{aligned} G_A &= \{g_{a,1}, g_{a,2}, \ldots, g_{a,n}\}, \\ G_B &= \{g_{b,1}, g_{b,2}, \ldots, g_{b,n}\}, \\ G_C &= \{g_{c,1}, g_{c,2}, \ldots, g_{c,n}\}, \end{aligned} \tag{5}$$

where G_A represents the extracted feature sequence of historical order data, G_B represents the extracted feature sequence of consumer comment emotions, and G_C represents the extracted feature sequence of facial value and skin type.

The three groups of extracted feature sequences G_A, G_B, and G_C are grouped to weaken the context information between the features with weaker relations and enhance the context information between the features with stronger relations. Based on the grouping strategy, we introduced the BiLSTM network, the purpose of which is to fully obtain the contextual information between the features in each group of sequences. At the same time, the BiLSTM network can better obtain context information within a certain sequence step. The calculation equation of the grouping strategy based on the BiLSTM network is as follows:

$$\begin{aligned} O_{A,n} &= \mathrm{LSTM}_{\mathrm{forward}}(g_{a,n}) \otimes \mathrm{LSTM}_{\mathrm{backward}}(g_{a,n}), \\ O_{B,n} &= \mathrm{LSTM}_{\mathrm{forward}}(g_{b,n}) \otimes \mathrm{LSTM}_{\mathrm{backward}}(g_{b,n}), \\ O_{C,n} &= \mathrm{LSTM}_{\mathrm{forward}}(g_{c,n}) \otimes \mathrm{LSTM}_{\mathrm{backward}}(g_{c,n}), \end{aligned} \tag{6}$$

where $\{g_{a,n}, g_{b,n}, g_{c,n}\}$ represents the input local, new feature sequence; $\{O_{A,n}, O_{B,n}, O_{C,n}\}$ represents the local, new feature output sequence; n represents the sequence number of each group; and $\otimes$ represents the forward and backward feature fusion strategy of LSTM.

3.5. Spatial Feature Fusion. Through the grouping strategy in the previous section, we have fully obtained the local context information of each group. However, to ensure the integrity of the contextual information of the features of the entire multimodal data, we use multiscale traditional convolution and multiscale cavity convolution. Its purpose is to achieve the spatial dimension fusion of different sets of features. In order to more intuitively understand the difference between spatial dimension fusion and traditional fusion, we have made a detailed explanation through Figure 7.

Figure 7(a) represents the traditional feature fusion strategy. G_1 represents the context information of the local features generated by the first set of features through the BiLSTM network. G_2 is directly spliced behind G_1, while G_3 is spliced behind G_2, and so on. The new feature sequence generated after $m-1$ splicing includes completeness of the context information of the entire feature sequence. Then, deep features are mined through convolution and hole convolution. The calculation equation is as follows:

$$\begin{aligned} \mathrm{MSC} &= \sum_{k=3}^{j} \mathrm{conv}\left(\left(x_{i-1} * w_{i-1} + b_{i-1}\right) + \left(x_i * w_i + b_i\right) + \left(x_{i+1} * w_{i+1} + b_{i+1}\right)\right), \\ \mathrm{MSDC} &= \sum_{k=3,d=2}^{3,l} \mathrm{conv}\left(\left(x_{i-3} * w_{i-3} + b_{i-3}\right) + \left(x_i * w_i + b_i\right) + \left(x_{i+3} * w_{i+3} + b_{i+3}\right)\right), \end{aligned} \tag{7}$$

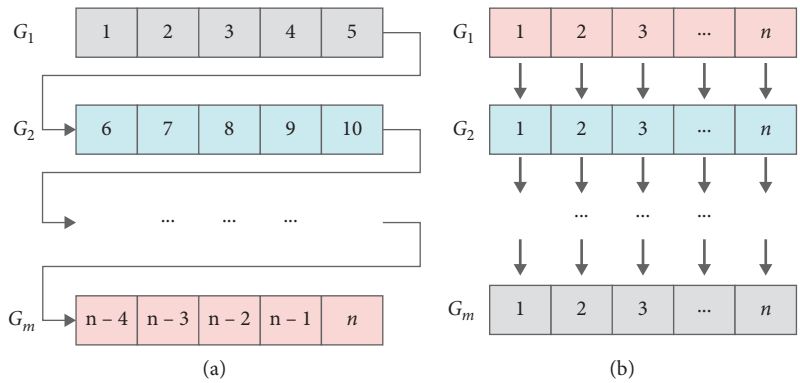

Figure 7: (a) Traditional feature fusion strategy. (b) Spatial dimensional feature fusion strategy.

where MSC represents the operation result of multiscale convolution. MSDC represents the operation result of multiscale dilated convolution. x represents the characteristics of the sample. i represents the i-th feature in the sample. w represents the weight coefficient; b stands for bias; k is the size of the convolution kernel; j is the number of convolution kernels; d represents the scale of dilated convolution expansion; and l represents the number of dilations of dilated convolution.

Figure 7(b) shows the feature fusion strategy of the spatial dimension. This strategy integrates the features of each column in each group across spatial dimensions. We fused the first feature of G_1 with the first feature G_2 until the first feature G_m. At the same time, we used traditional convolution and dilated convolution to mine the features of the same spatial dimension to generate deep features of the spatial dimension. The calculation equation is as follows:

Table 2: Hyperparameter setting.

Type	Hyperparameter
Optimizer	Adam
Learning rate	0.001
β_1	0.9
β_2	0.999
Epsilon	$1e-08$
Decay	$3e-8$

$$\begin{aligned} \mathrm{MSC}^* &= \sum_{k=3}^{j} \mathrm{conv}\left(\left(x_{i-n} * w_{i-n} + b_{i-n}\right) + \left(x_i * w_i + b_i\right) + \left(x_{i+n} * w_{i+n} + b_{i+n}\right)\right), \\ \mathrm{MSDC}^* &= \sum_{k=3,d=2}^{3,l} \mathrm{conv}\left(\left(x_{i-3n} * w_{i-3n} + b_{i-3n}\right) + \left(x_i * w_i + b_i\right) + \left(x_{i+3n} * w_{i+3n} + b_{i+3n}\right)\right), \end{aligned} \tag{8}$$

where n represents the length of each group of local features; MSC^* represents the operation result of multiscale convolution in the spatial dimension; and MSDC^* represents the operation result of multiscale dilated convolution in the spatial dimension.

4. Experiments and Results

4.1. Dataset. The dataset in this article uses historical sales data, consumer review data, and consumer portrait data collected by skincare product e-commerce platform. The dataset contains historical information of 200 products over more than a year, a total of more than 20,000 pieces of data information. Through data cleaning and feature engineering, this paper constructs a training set and a test set that can be used for neural networks.

4.2. Hyperparameter Settings. The main parameters are shown in Table 2: *lr* means that the learning rate is 0.01; bata_1 means that the exponential decay rate of the first-order moment estimation is 0.9; bata_2 means the second-order moment. The estimated exponential decay rate is 0.999. Epsilon is set to $1e-08$. Decay indicates that the learning rate decay value is $3e-8$ after each parameter update. The experiments with all the algorithms were performed on a computer equipped with a single NVIDIA GTX1080 GPU (8 GB).

4.3. Evaluation Criteria. For prediction problems, it is necessary to establish prediction performance evaluation indicators to verify the feasibility and accuracy of the prediction model, considering that e-commerce commodity demand forecasting is generally for the purchase and inventory replenishment of e-commerce companies. The forecast error of the demand for selling a larger number of commodities has a greater impact than selling after commodities under the equivalent error. Therefore, the error

selected in this paper should consider the error between the predicted value and the true value and consider the ratio between the error and the true value.

Mean square error (MSE): this indicator is the square of the difference between the real quantity and the predicted quantity and then summed and averaged. The calculation equation is as follows:

$$\mathrm{MSE} = \frac{1}{N}\sqrt{\sum_{i=1}^{N}\left(y_i - \widehat{y}_i\right)^2}. \tag{9}$$

Root mean square error (RMSE): this indicator is the square root calculation of the ratio of the square sum of the difference between the real quantity and the predicted quantity to the number of observations. This is used to measure the deviation between the predicted quantity and the real quantity. The calculation equation is as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \widehat{y}_i\right)^2}. \tag{10}$$

Mean absolute error (MAE): this metric is used to average absolute error. This value more accurately reflects the current state of the forecast error, i.e., the difference between the actual quantity and the forecast. The following is the calculation formula:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \widehat{y}_i\right|. \tag{11}$$

Mean absolute percentage error (MAPE): this metric considers the difference between the predicted and actual value. It also computes the ratio between the predicted error and the true value at the same time. The following is the calculation equation:

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \widehat{y}_i}{y_i}\right| \times 100\%, \tag{12}$$

where TP is true positives (the number of wheat pixels that are correctly detected), FP is false positives (the number of wrongly detected pixels as wheat), and FN is false negatives (the number of wheat pixels that are wrongly detected).

4.4. Experimental Results. Since the final output of the model is a probability distribution, in order to be able to obtain the predicted value of each tested product, this article uses a sampling method to output the predicted value. It selects $u_{i,t,m}$ corresponding to the largest $p_{i,t,m}$ as the predicted output value of the product, and part of the tested product demand. The prediction result of the quantity is shown in Figure 8.

It can be seen from Figure 8 that the model in this paper has a good fit between the predicted value of the short-term demand for e-commerce commodities and the actual value. The predicted value is extremely close to the actual value on the whole. In order to demonstrate the effectiveness of our model, we randomly select 8 products in the test set for an 8-week prediction test and use RMSE and MAPE to evaluate the error results quantitatively. The results are shown in Table 3.

It can be seen from Table 3 that, for the 6 randomly selected goods Good_ID, the predicted RMSE fluctuates between 2.03 and 3.48, and the MAPE fluctuates between 1.27% and 1.62%. The average value of RMSE is 2.6891, and the average value of MAPE is 1.41%, indicating that the prediction errors of the 6 randomly selected commodities in the prediction results are relatively stable, so this also fully proves the effectiveness of the model in this paper.

4.5. Ablation Experiment of Multimodal Data. In this section, we conducted an ablation experiment of multimodal data, segmented. We combined the data of the three modalities to observe the influence of each part on the experimental results. *A* represents historical order data, *B* represents consumer review data, and *C* represents consumer portrait data. The experimental results are shown in Table 4.

As shown in Table 4 and Figure 9, when considering three types of multimodal data simultaneously, the prediction model achieves the best prediction results. Secondly, it can also be found that data error considering any two modes is lower than that of single-mode data. Therefore, this proves the effectiveness of using three modes of data simultaneously in this paper.

4.6. Ablation Experiment of Feature Fusion Strategy. Since the model in this paper adopts traditional feature fusion and spatial feature fusion strategies, to deeply analyze the impact of the above two strategies on the experimental results, the feature fusion ablation experiment is conducted. We assume that *T* represents the traditional feature fusion measurement, and *S* represents the spatial feature fusion. The experimental results are shown in Table 5.

It can be seen from Table 5 that the combination of traditional feature fusion and spatial feature fusion strategy achieves the best prediction effect. Meanwhile, it is also found that the spatial feature fusion strategy is superior to the traditional one. Therefore, it proves the effectiveness of the spatial feature fusion strategy.

4.7. Comparative Experiments. To further verify the effectiveness and superiority of this model, this section applies other prediction methods for verification and comparison with the same dataset. The comparison model mainly selected ARIMA and MLP-LSTM.

From Tables 6 and 7, we can comprehensively compare the performance of the three models on the commodity demand forecasting task in 6. We find that the three models have different performances for different commodities in terms of RMSE and MAPE. However, from an overall point of view, the prediction error fluctuations of the 6 test commodities in this model are smaller than those of the ARIMA model and MLP-LSTM, indicating that it has better accuracy.

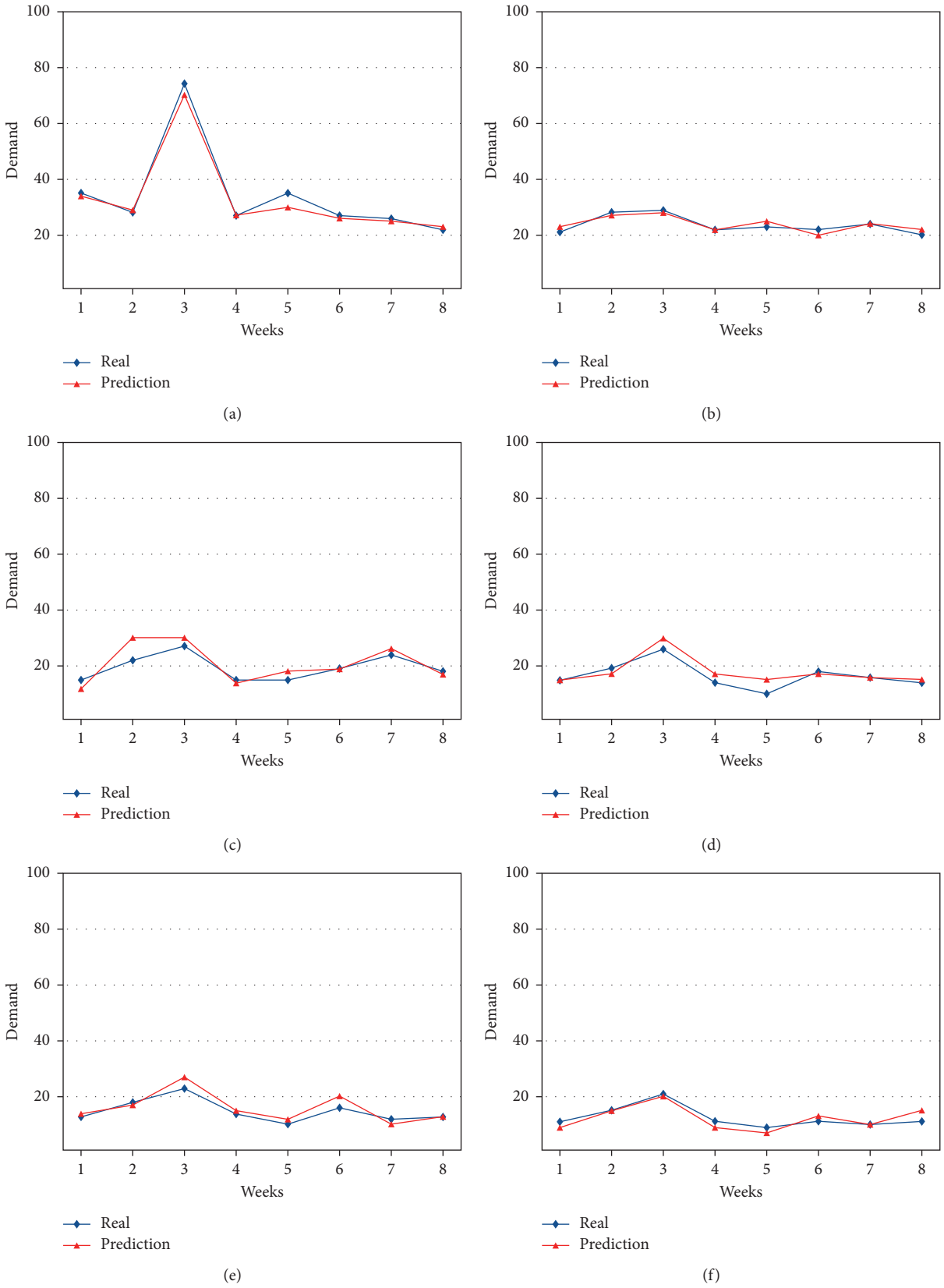

Figure 8: Comparison results of forecast results of commodity demand on the test dataset. (a) Test commodity 1; (b) test commodity 2; (c) test commodity 3; (d) test commodity 4; (e) test commodity 5; (f) test commodity 6.

Table 3: Forecast error of commodity demand.

Good_ID	RMSE	MAPE (%)	MSE	MAE
192	2.3979	1.32	1.75	5.75
73	3.2787	1.50	2.25	10.75
106	3.4820	1.62	12.12	2.62
213	2.6457	1.41	7.00	2.00
50	2.3184	1.36	1.88	5.38
239	2.0311	1.27	1.75	5.25

Table 4: Errors in the forecast of demand for two test commodities.

Modal type	Commodity 1 (Good_ID: 192)		Commodity (Good_ID: 73)	
	RMSE	MAPE (%)	RMSE	MAPE (%)
A	11.2562	6.25	10.6542	5.75
B	10.2542	5.22	11.1320	6.94
C	8.2545	5.21	12.1212	4.69
A + B	7.9856	3.25	7.0982	3.99
B + C	4.1542	2.76	3.9956	3.36
A + C	4.5698	2.27	3.4785	2.65
Ours	**2.3979**	**1.32**	**2.3184**	**1.36**

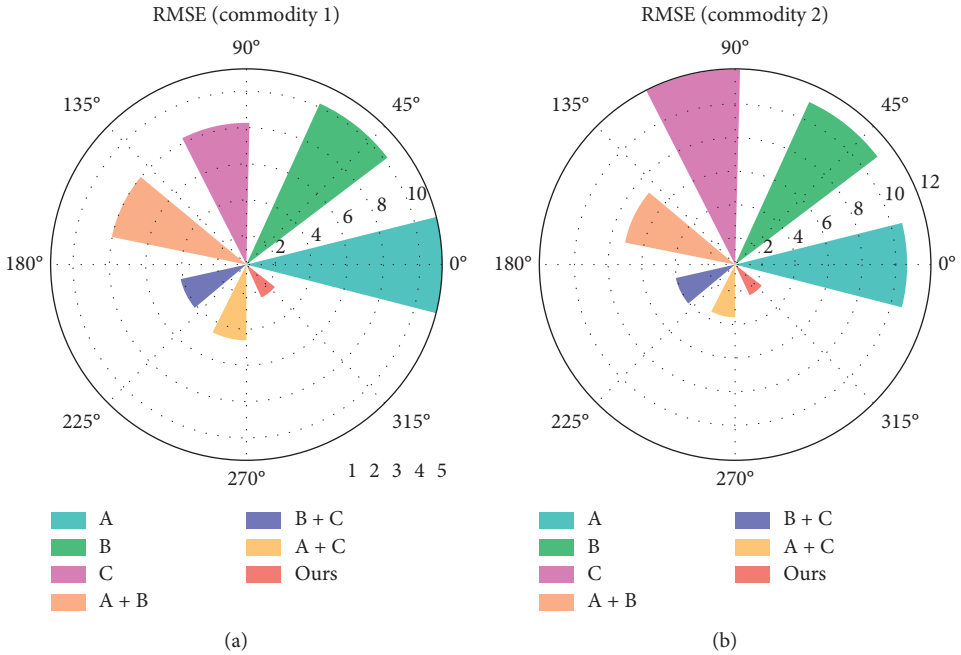

Figure 9: Continued.

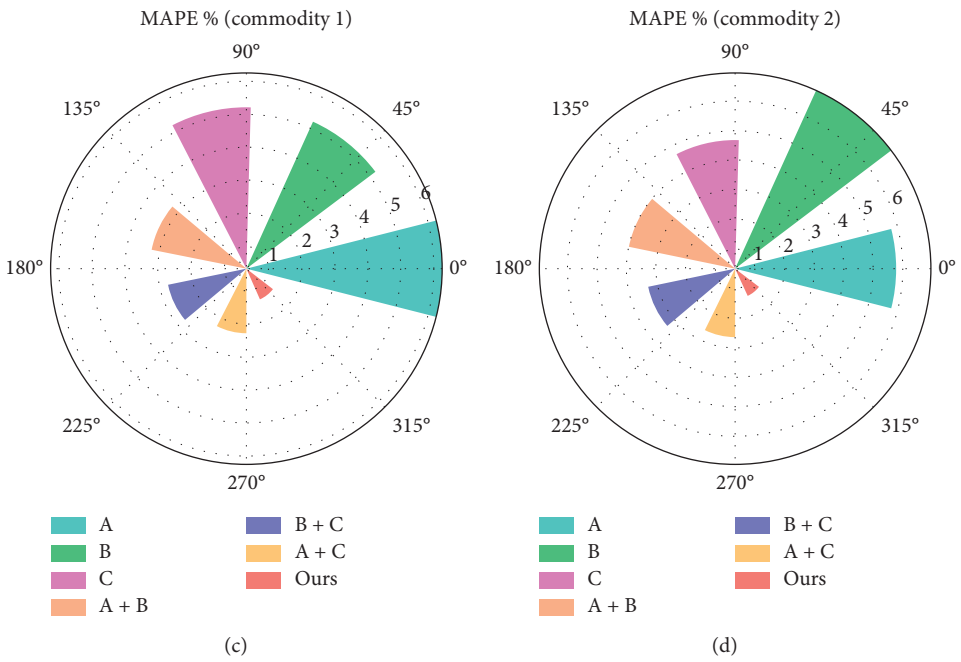

Figure 9: RMSE and MAPE of two test commodities demand forecasts.

Table 5: Results of feature fusion ablation experiment.

Modal type	Commodity 1 (Good_ID: 192)		Commodity 2 (Good_ID: 73)	
	RMSE	MAPE (%)	RMSE	MAPE (%)
T	16.3585	5.96	13.4856	6.11
S	12.2987	4.21	9.3654	5.94
$T+S$ (ours)	**2.3979**	**1.32**	**2.3184**	**1.36**

Table 6: RMSE comparative experimental results of different methods.

Methods	Good_ID: 192	Good_ID: 73	Good_ID: 106	Good_ID: 213	Good_ID: 50	Good_ID: 239
ARIMA	26.659	25.3666	30.458	19.2565	17.0025	22.3695
MLP-LSTM	22.0145	17.3695	25.2365	21.5625	20.6953	15.6526
Ours	**2.3979**	**3.2787**	**3.4820**	**2.6457**	**2.3184**	**2.0311**

Table 7: MAPE comparative experimental results of different methods.

Methods	Good_ID: 192 (%)	Good_ID: 73 (%)	Good_ID: 106 (%)	Good_ID: 213 (%)	Good_ID: 50 (%)	Good_ID: 239 (%)
ARIMA	8.25	6.96	7.99	5.74	6.14	9.25
MLP-LSTM	3.65	6.59	9.25	2.96	3.48	5.11
Ours	**1.32**	**1.50**	**1.62**	**1.41**	**1.36**	**1.27**

5. Conclusion

For e-commerce companies, accurate and reliable e-commerce commodity demand forecasting is essential. This paper proposes a spatial feature fusion and grouping strategy based on multimodal data. It establishes a neural network prediction model for e-commerce commodity demand. First of all, the ablation experiment proved the positive influence of multimodal data on the prediction task. It indicates that consumer reviews and consumer portraits are important factors influencing in demand forecasting. In addition, we also found that the feature relationships between the three modal data are not independent. However, there are closely related relationships, which we call spatial relationships. The superiority of spatial feature fusion is proved through ablation experiments. Finally, the e-commerce product dataset generated by the e-commerce platform is used to test the prediction effect of the proposed model. The experimental results prove the effectiveness and superiority of the algorithm.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

All the authors do not have any possible conflicts of interest.

References

[1] Y. Yu, X. Wang, R. Y. Zhong, and G. Q. Huang, "E-commerce logistics in supply chain management: practice perspective," *Procedia CIRP*, vol. 52, pp. 179–185, 2016.

[2] M. B. Murtaza, V. Gupta, and R. C. Carroll, "E-marketplaces and the future of supply chain management: opportunities and challenges," *Business Process Management Journal*, vol. 10, no. 3, pp. 325–335, 2004.

[3] L. Huang, G. Xie, J. Blenkinsopp, R. Huang, and H. Bin, "Crowdsourcing for sustainable urban logistics: exploring the factors influencing crowd workers' participative behavior," *Sustainability*, vol. 12, no. 8, Article ID 3091, 2020.

[4] L. Huang, G. Xie, W. Zhao, Y. Gu, and Y. Huang, "Regional logistics demand forecasting: a BP neural network approach," *Complex & Intelligent Systems*, pp. 1–16, 2021.

[5] D. Adebanjo and R. Mann, "Identifying problems in forecasting consumer demand in the fast moving consumer goods sector," *Benchmarking: an International Journal*, vol. 7, no. 3, pp. 223–230, 2000.

[6] A. R. Pinjari and C. Bhat, "Computationally efficient forecasting procedures for Kuhn-Tucker consumer demand model systems: application to residential energy consumption analysis," *Journal of Choice Modelling*, vol. 39, Article ID 100283, 2021.

[7] V. Sillanpää and J. Liesiö, "Forecasting replenishment orders in retail: value of modelling low and intermittent consumer demand with distributions," *International Journal of Production Research*, vol. 56, no. 12, pp. 4168–4185, 2018.

[8] T. Y. Kim, R. Dekker, and C. Heij, "Spare part demand forecasting for consumer goods using installed base information," *Computers & Industrial Engineering*, vol. 103, pp. 201–215, 2017.

[9] J. Liu, C. Liu, L. Zhang, and Y. Xu, "Research on sales information prediction system of e-commerce enterprises based on time series model," *Information Systems and e-Business Management*, vol. 18, pp. 823–836, 2019.

[10] G. Chniti, H. Bakir, and H. Zaher, "E-commerce time series forecasting using LSTM neural network and support vector regression," in *Proceedings of the International Conference on Big Data and Internet of Thing, BDIOT2017*, pp. 80–84, London, UK, December 2017.

[11] H. Fahmy, "How technological emergence, saturation, and rejuvenation are re-shaping the e-commerce landscape and disrupting consumption? A time series analysis," *Applied Economics*, vol. 53, no. 6, pp. 742–759, 2021.

[12] D. Wei, P. Geng, L. Ying, and L. Shuaipeng, "A prediction study on e-commerce sales based on structure time series model and web search data," in *Proceedings of the 26th Chinese Control and Decision Conference (2014 CCDC)*, pp. 5346–5351, Changsha, China, 31 May-2 June 2014.

[13] C.-W. Chu and G. P. Zhang, "A comparative study of linear and nonlinear models for aggregate retail sales forecasting," *International Journal of Production Economics*, vol. 86, no. 3, pp. 217–231, 2003.

[14] T. Van Calster, F. V. D. Bossche, B. Baesens, and W. Lemahieu, Profit-oriented Sales Forecasting: A Comparison of Forecasting Techniques from a Business Perspective. Econometrics, 2020.

[15] P.-H. Hsieh, "A study of models for forecasting E-commerce sales during a price war in the medical product industry," in *Proceedings of the 21st International Conference on Human-Computer Interaction, HCI International 2019*, pp. 3–21, Orlando, FL, USA, July 2019.

[16] M. Li, S. Ji, and G. Liu, "Forecasting of Chinese E-commerce sales: an empirical comparison of ARIMA, non-linear autoregressive neural network, and a combined ARIMA-NARNN model," *Mathematical Problems in Engineering*, vol. 2018, Article ID 6924960, 12 pages, 2018.

[17] Z. Liu, P. Jiang, L. Zhang, and X. Niu, "A combined forecasting model for time series: application to short-term wind speed forecasting," *Applied Energy*, vol. 259, Article ID 114137, 2020.

[18] L. Xiao, J. Wang, Y. Dong, and J. Wu, "Combined forecasting models for wind energy forecasting: a case study in China," *Renewable and Sustainable Energy Reviews*, vol. 44, pp. 271–288, 2015.

[19] G. J. Chen, K. K. Li, T. S. Chung, H. B. Sun, and G. Q. Tang, "Application of an innovative combined forecasting method in power system load forecasting," *Electric Power Systems Research*, vol. 59, no. 2, pp. 131–137, 2001.

[20] M.-C. Tsai, C.-H. Cheng, M.-I. Tsai, and H.-Y. Shiu, "Forecasting leading industry stock prices based on a hybrid time-series forecast model," *PLoS One*, vol. 13, no. 12, Article ID e0209922, 2018.

[21] M. Akpinar and N. Yumusak, "Year ahead demand forecast of city natural gas using seasonal time series methods," *Energies*, vol. 9, no. 9, Article ID 727, 2016.

[22] M. Maleki, M. R. Mahmoudi, D. Wraith, and K.-H. Pho, "Time series modelling to forecast the confirmed and recovered cases of COVID-19," *Travel Medicine and Infectious Disease*, vol. 37, Article ID 101742, 2020.

[23] A. Cujia, D. Agudelo-Castañeda, C. Pacheco-Bustos, and E. C. Teixeira, "Forecast of PM10 time-series data: a study case in Caribbean cities," *Atmospheric Pollution Research*, vol. 10, no. 6, pp. 2053–2062, 2019.

[24] S. Hajifar, H. Sun, F. M. Megahed, L. A. Jones-Farmer, E. Rashedi, and L. A. Cavuoto, "A forecasting framework for predicting perceived fatigue: using time series methods to forecast ratings of perceived exertion with features from wearable sensors," *Applied Ergonomics*, vol. 90, Article ID 103262, 2021.

[25] M. Huard, R. Garnier, and G. Stoltz, "Hierarchical robust aggregation of sales forecasts at aggregated levels in e-commerce, based on exponential smoothing and holt's linear trend method," 2020, https://arxiv.org/pdf/2006.03373.pdf.

[26] T. Bowen, Z. Zhe, and Z. Yulin, "Forecasting method of e-commerce cargo sales based on ARIMA-BP model," in *Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, pp. 133–136, Dalian, China, June 2020.

[27] X. Ning, Y. Wang, W. Tian, L. Liu, and W. Cai, "A biomimetic covering learning method based on principle of homology continuity," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, pp. 9–16, 2021.

[28] C. Yan, G. Pang, X. Bai et al., "Beyond triplet loss: person re-identification with fine-grained difference-aware pairwise loss," *IEEE Transactions on Multimedia*, 2021.

[29] Y. Tong, L. Yu, S. Li, J. Liu, H. Qin, and W. Li, "Polynomial fitting algorithm based on neural network," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 1, pp. 32–39, 2021.

[30] G. Suchacka and S. Stemplewski, "Application of neural network to predict purchases in online store," in *Proceedings of the 37th International Conference on Information Systems Architecture and Technology–ISAT 2016–Part IV*, pp. 221–231, Cham, NY, USA, September 2017.

[31] C. Giri, S. Thomassey, J. Balkow, and X. Zeng, "Forecasting new apparel sales using deep learning and nonlinear neural network regression," in *Proceedings of the 2019 International Conference on Engineering, Science, and Industrial Applications (ICESI)*, pp. 1–6, Tokyo, Japan, 22-24 Aug. 2019.

[32] Y.-S. Shih and M.-H. Lin, "A LSTM approach for sales forecasting of goods with short-term demands in E-commerce," in *Proceedings of the Asian Conference on Intelligent Information and Database Systems*, pp. 244–256, Phuket, Thailand, April 2019.