Research Article

Prediction of Merchandise Sales on E-Commerce Platforms Based on Data Mining and Deep Learning

Xiaoting Yin [ID] **and Xiaosha Tao**

School of Digital Trade, Jiangxi University of Engineering, Xinyu 338000, China

Correspondence should be addressed to Xiaoting Yin; 8078yinxiaoting@yzpc.edu.cn

Online business has grown exponentially during the last decade, and the industries are focusing on online business more than before. However, just setting up an online store and starting selling might not work. Different machine learning and data mining techniques are needed to know the users' preferences and know what would be best for business. According to the decision-making needs of online product sales, combined with the influencing factors of online product sales in various industries and the advantages of deep learning algorithm, this paper constructs a sales prediction model suitable for online products and focuses on evaluating the adaptability of the model in different types of online products. In the research process, the full connection model is compared with the training results of CNN, which proves the accuracy and generalization ability of CNN model. By selecting the non-deep learning model as the comparison baseline, the performance advantages of CNN model under different categories of products are proved. In addition, the experiment concludes that the unsupervised pretrained CNN model is more effective and adaptable in sales forecasting.

1. Introduction

At present, the research results on product sales forecasting are relatively rich, and the research methods are different. Bi and Wei improved BP neural network from the two aspects of sample quality and initial weight by using principal component analysis method and particle swarm optimization algorithm [1]; Qu et al. established a neural network prediction model for cigarette sales by using the improved BP neural network Levenberg–Marquardt algorithm [2]. Research results using time series prediction method are also common. For example, Peng and Yu use RBF neural network to predict product sales based on time series analysis and optimize the prediction model [3]; Wang extracted product clusters according to product sales commonness and established a product reclassification time series sales prediction model based on sales data [4]. Some scholars have also adopted the support vector machine prediction method, such as Wu and Lin. Taking the cigarette sales of specific tobacco enterprises as the research object, they have

proposed a hybrid method for cigarette sales prediction based on support vector machine [5]. The research methods in the above literature have their own advantages, but there are also many disadvantages: first, online product data samples often have diversified characteristics, while most models do not have diversified data processing ability; second, with the increasing scale of online product sales, the resulting massive sales data not only is the basic basis for sales forecasting, but also reflects the deficiency of traditional forecasting methods in dealing with large-scale data. For example, Bi et al. used shallow neural network, which has advantages in big data processing, but the prediction accuracy needs to be improved. Liu et al. established crown model based on deep learning algorithm, on the basis of fully considering the characteristics of agricultural e-commerce sales data, and used this model to realize the classified prediction of online agricultural product sales [6]. Deep learning algorithm has its unique advantages in online product sales forecasting. Firstly, deep learning improves the training algorithm based on BP neural network, and the

gradient disappearance problem is effectively solved, so that the effective time of training is longer. Secondly, online product sales forecasting needs high generalization model support. The deep learning model with high complexity capacity has good generalization in the big data environment. Thirdly, compared with the general model, deep learning can extract more and more effective information from massive data. Finally, deep learning has the feature of building layer by layer, which can extract higher-level features from the existing data, decompose the influencing factors of interaction into independent and more effective factors, and improve the prediction accuracy of the model. Based on the above advantages, this paper aims to establish a relatively perfect index system of influencing factors of online product sales and use deep learning algorithm to build a sales prediction model of all kinds of online products. Because the online product sales forecasting model based on deep learning algorithm usually classifies products and designs the model according to the characteristics of a certain kind of products, such a model has poor adaptability. Once the product type changes, the influencing factor index and model must be redesigned. Therefore, this paper not only evaluates the prediction accuracy and generalization ability of the model, but also focuses on the adaptability of the model.

Convolutional neural network is an effective structure in deep learning. Deep learning is a machine learning structure containing multiple hidden layers. It is very good at extracting and calculating the eigenvalues of objects or problems with complex structure and looking for potential complex rules without destroying the useful structural information of objects or problems. Considering the limitations of neural network in the representation ability of shallow structure function, the deep structure of multilayer nonlinear mapping of deep network can not only realize the effective approximation of complex function, but also obtain the main driving variables of input data through layer by layer learning algorithm. Following are the advantages of deep learning theory:

(1) Distributed representation is a basic concept in machine learning and neural network research. Its appearance is very helpful to solve the defects of "dimensional disaster" and local generalization. It is also an important reason why deep learning theory has advantages over traditional machine learning algorithms. Distributed representation is a compact coding method, which is of great significance for machine learning. It can not only reduce the amount of computation, but also effectively use the sample data, so as to avoid the phenomenon of overfitting. In addition, for the same model structure, the compactness of distributed representation is exponentially proportional to other local representation methods. Single decision tree can linearly divide the input space, and the number of divided regions is shown in Figure 1.

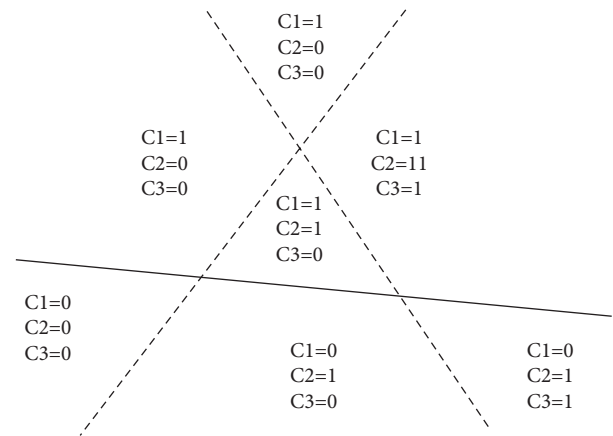Figure 1 illustrates that a single decision tree can linearly divide the input space, and the number of

Figure 1: Single decision tree can linearly divide the input space.

divided regions (hereinafter referred to as subregions) is equal to the number of parameters, that is, the number of decision tree leaves. The number of areas that can be divided by multiple tree combinations (i.e., random forest) and the number of "trees" refer to several times, that is, the exponential times of the total number of parameters in random forest. Then, each subregion corresponds to any leaf of any tree in the random forest. It can be seen that the number of parameters and samples required for the distributed representation constructed by the above algorithm is significantly less than the number of subregions, which is also the main reason for avoiding data generalization.

(2) The advantages of depth structure: The learning algorithm of depth structure can effectively express the function, and its learning process can learn some functions that other algorithms cannot effectively learn. This is not only a theoretical advantage of deep learning, but also a potential limitation of other shallow structures, such as SVM, random forest, and BP algorithm. The so-called effective expression of function means that the expression of function is compact; that is, when the degree of freedom required to learn parameters in the model is low, the expression of the model is compact. When the number of samples is limited and there is a lack of foreign prior knowledge, the compact expression of the objective function will produce better generalization. More precisely, for a function that can be expressed by a structure with a depth of k, the number of computing units required at k-1 depth increases exponentially. This is because the number of computational elements that a learning structure can afford depends on the number of samples that can be used to learn the structural parameters. When using a shallow structure to describe a function, it will inevitably lead to poor generalization. It should be noted here that the depth structure has three advantages for the compact representation of complex functions: first, it can make effective use of sample data; second, the number of computing units

is very small; third, little foreign prior knowledge is needed [7, 8].

(3) Unsupervised pretraining: The existing standard training mode of deep learning tends to put the parameters in a parameter space with poor generalization. This method is often used in the process of experimental training. Based on this idea, deep belief network (DBN) and stack self-encoder (SAE) methods appeared in 2006, which are an effective breakthrough in deep learning training strategy. Both DBN and SAE use a similar algorithm strategy; that is, after greedy layer by layer unsupervised pretraining, supervised fine-tuning of the depth structure based on gradient optimization algorithm is conducted This is because each layer of unsupervised training can learn the nonlinear mapping relationship of the main factors in the input characteristics, and unsupervised pretraining is equivalent to setting an initial stage for the fine-tuning of depth structure under the condition of supervised training. In essence, unsupervised pretraining is an uncommon regular form. It can minimize variance and introduce bias, so as to promote the deep learning process into a parameter space useful for unsupervised training. In addition, in the highly nonconvex deep learning structure, a special initialization point can also be defined, which can enhance the parameter constraints. This is because the initialization point can specify which minimum value points (outside a large number of possible minimum value points) are acceptable and allowed in the cost function [9].

Data mining, machine learning, and deep learning have widely been used in sales forecasting and predictions in research and business. Peng [10] constructed a sales prediction model for retail stores using deep learning approach. Sales data was used across three years from a store, and a model that predicted sales on a day by analyzing the sales on the previous day was constructed. Accuracy above 93% was achieved with deep learning model, while less than 86% was achieved using traditional learning models. Yin et al. [11] used fuzzy clustering and deep learning to build a model for forecasting products sale. They used the weight of product similarity attributes and fuzzy clustering rough set method which provides basis for collecting historical data of similar products sale. The prediction error is adjusted through an LSTM based deep learning model. Wang et al. [12] used different historical sales ranking and other data of books from different online stores to build a model for predicting book sales. The data is preprocessed, feature selection is carried out, and models are built and trained on prediction results. They used the Generative Adversarial Network (GAN) to build their deep learning model.

Online business has grown abundantly during the last 10–15 years, and the industries are now focusing on online business more than on offline business. However, online business has become more and more complex with the advance in data science and machine learning. Different machine learning and data mining techniques are needed to know the users' preferences and perform better decision making. Keeping in mind the importance of ML and data mining in online business, this paper contributes to the existing research in the field to develop a machine learning and data mining based method for predicting sales on the online platforms.

The key contributions of this paper include the following.

Firstly, we determine the key influencing factors that have an impact on online business. We develop a Python crawler that captures the required products data from Taobao and Tmall, and samples are collected based on the influencing factors system. Then, we develop the deep convolutional network based deep learning product sales forecasting model and introduce denoising autoencoder to pretrain the network. Finally, we carry out a number of experiments with different models for different categories of products. The experimental results show that the model achieves a high accuracy reaching up to 97%.

The rest of the paper is organized as follows. Section 2 discusses the data collection and processing performed in this paper, Section 3 presents the detailed sales forecasting model, Section 4 discusses the experiments and experimental results, and Section 5 is the conclusion of our work.

2. The Data Processing

First of all, the evaluation index system needs to be constructed. Because the online product trading mode is very different from the traditional trading mode, the influencing factors of online product sales are more complex than offline products. In addition, this paper aims to build a sales prediction model of all kinds of online products, so the characteristics of a certain type of products should not be too prominent in the selection of influencing factors and indicators, and the common characteristics of most products should be comprehensively considered to ensure the reliability and adaptability of the model. Therefore, this paper selects the influencing factors of sales from five aspects: product attributes, merchant attributes, buyer attributes, competitor attributes in the same industry, and main marketing channels of products.

2.1. Product Attributes. The characteristics of the product itself are the main factors affecting the sales volume. The product attribute indexes finally determined in this paper are as follows: price (I1), praise rate (I2), quality grade (I3), collection volume (I4), and cumulative comment volume (I5). Among them, indicators other than "quality grade" can directly obtain sample data from the platform to which the product belongs.

2.2. Merchant Attributes. This paper describes the characteristics of businesses through the attributes of business time (I6), store level (I7), sales volume (I8), and score (I9). The "scoring" index includes the scoring results of "product description consistency score," "logistics service level score,"

and "comprehensive service score." Therefore, this paper uses vectors to represent the input values of this index. The sample data of merchant attribute indicators are directly obtained from the merchant store page.

2.3. Buyer Attributes.
Because the basic information of customers, such as gender, age, education, information of products purchased, and online transaction amount, involves personal privacy issues, there is no channel for obtaining it on major e-commerce platforms. Therefore, this paper selects buyer loyalty (I10) to reflect customers' satisfaction with products and their personal consumption preferences.

2.4. Attributes of Competitors in the Same Industry.
This paper selects industry average quality index (I11) and product average price (I12) to represent the characteristic attributes of product industry competitors.

2.5. Main Marketing Channels of Products.
The marketing channel suitable for product characteristics and market positioning is the key factor for the increase of product sales. This paper classifies the conventional marketing channels of online products into five categories, search engine, social network, e-mail and information, e-commerce platform, and traditional offline channels, and assigns values according to the promotion effect of each marketing channel on product sales, which is used as the sample data of marketing channel (I13) index. Based on the above influencing factor indicators, the input eigenvector of the online product sales prediction model constructed in this paper is expressed as Ia = (I1, I2, I3..., I13), and the vector dimension is I3.

Then, there are data sources and technical means. Alibaba is the largest e-commerce platform in China, and its C2C e-commerce platform Taobao (including Tmall) covers a wide range of products, involving the leading domain. In addition, the platform was established earlier, the transaction data of each product category is relatively complete and comprehensive in the domestic C2C e-commerce platform, and its transaction data continuity is obvious. Therefore, this paper takes Taobao (including Tmall) e-commerce platform as the main object for data capture, in which the web crawler program is written in Python. Due to the complex technical environment of Taobao platform and considering the feasibility and stability of data capture, selenium framework is adopted for data capture technology. The basic idea is as follows: take the keyword of Taobao product search as the entry, and traverse all products displayed in the search results under the keyword page by page. Finally, samples are collected based on the influencing factor index system, and the specific crawling steps are shown in Figure 2.

In this paper, the six fields of agriculture and animal husbandry, clothing, personal consumer goods, furniture, second-hand cars, and food are selected as the primary category, and several keywords are designed as the secondary category under each field, as shown in Table 1. It

should be noted that some products and keywords can be subdivided into multiple keywords, which are not considered one by one in this paper. The division method used in this paper is only for the purpose of this study, and other scholars can make corresponding adjustments according to the actual needs.

Due to the complex technical environment of Taobao, the repetition rate of captured data is high, and there are missing values of attributes. Therefore, it is necessary to clean the data, a process which is divided into two stages: The first is data deduplication stage. In this paper, the two attributes of product title and store name are used as the key fields to identify duplicate products. As long as samples with the same value appear on these two attributes in the data, only the first captured samples are retained according to the captured time sequence. The second is data missing value processing stage. Each missing value in the sample is supplemented according to the average value of the attribute; that is, the average value of the index of each classification is taken as the missing value.

After data cleaning, the number of samples is 13000. In order to further describe the data, this paper draws the influencing factor indicators into a straight square diagram. Since each index contains a certain number of extreme values, in order to avoid the impact of extreme values on the visualization of data distribution, each index is sorted according to the value, and the maximum (small) extreme value accounting for 5% of the total is removed. The histograms of some indicators are listed here to illustrate the data, as shown in Figure 3. As can be seen from the data distribution of the "sales volume" index in Figure 3, the sales volume of most products is not large, the sales volume of only a few products is very large, and the data presents a long tail distribution. The samples with long tail distribution will have an impact on the model training effect. The data distribution shape can be improved by increasing the number of samples.

3. Online Product Sales Forecasting Model

First of all, the full connection layer is the most initial structure in deep learning. Each layer of the network is a dense layer, and the neurons between layers are completely connected. Then, convolutional neural network is a deep learning structure inspired by biological vision system, which has excellent performance in the field of vision. Convolution neural network contains four different layer types: input layer, output layer, convolution layer, and pool layer. The overall structure is similar to the full connection layer and is composed of multiple layers. Convolution layer and pooling layer can be arbitrarily combined according to specific tasks and placed in the hidden layer to achieve the best model performance. The convolution layer can share the parameters of the convolution kernel, which greatly reduces the amount of model parameters. Pooling layer can further extract useful features and reduce model parameters again. Therefore, compared with the fully connected network, the convolutional neural network greatly simplifies the model parameters and makes the model training easier [13–15]. In addition, the full convolution network after
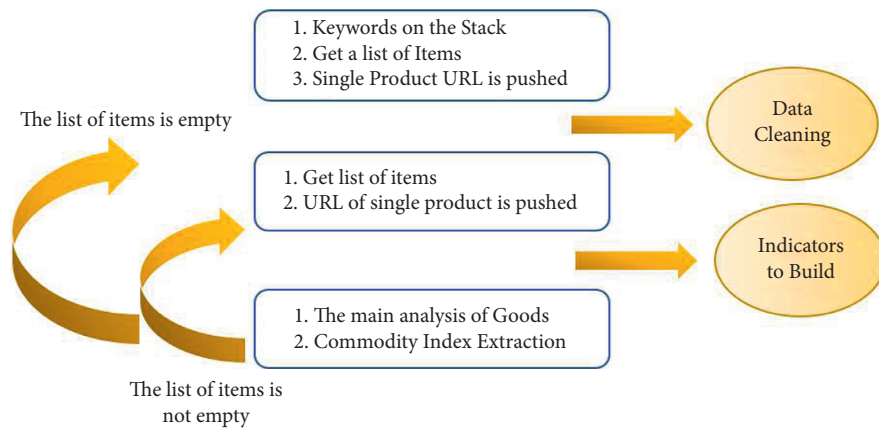
FIGURE 2: Data crawling and processing flow.

TABLE 1: Product classification.

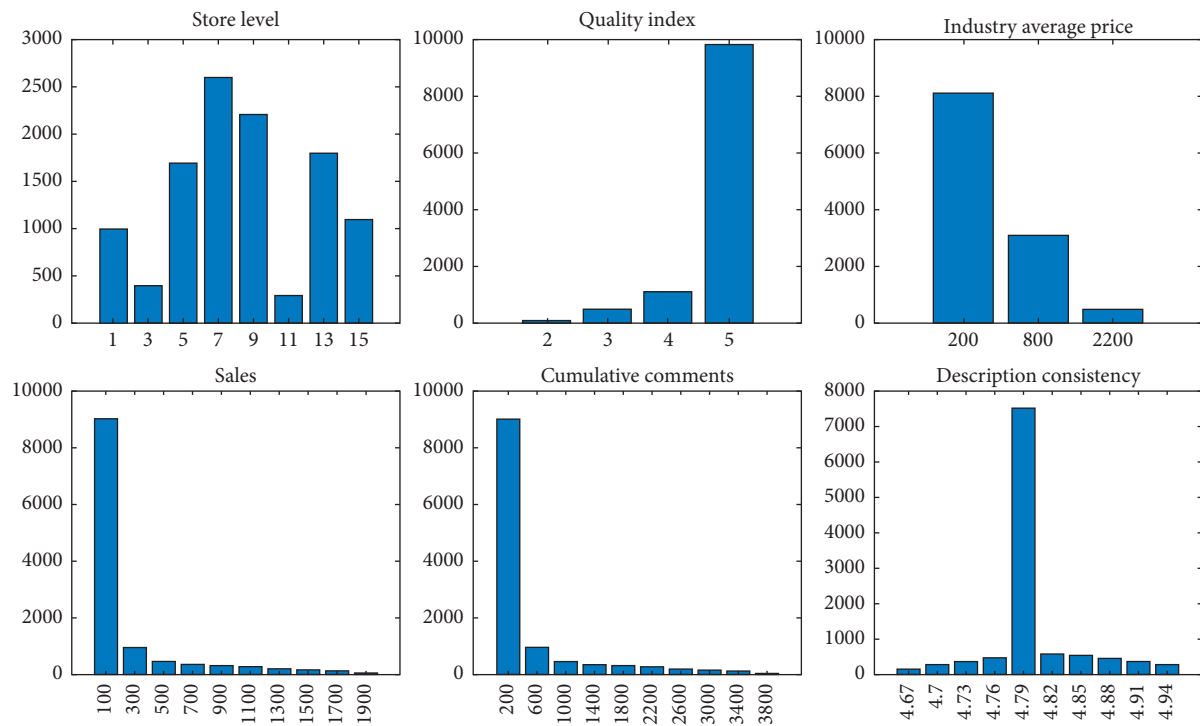| Primary classification	Secondary classification
Agricultural and animal husbandry products	Huaniu apple, yam, *Codonopsis pilosula*, *Angelica sinensis*, lily
Clothing	Children's wear, women's wear, underwear, shoes, boots
Personal consumer goods	Toiletries, watches, cosmetics, jewelry, department stores
Furniture	Furniture
Used car	Used car
Food	Meat, vegetables, fruits, milk, edible oil

FIGURE 3: Index distribution map.

removing the pooling layer still has a good performance in the visual field and can even reach the leading level in some image recognition tasks. The convolution layer with convolution kernel step size of 2 can act as the pooling layer [16]. Therefore, this paper decides to use the full convolution structure as the main structure of the convolution neural network.

Since unsupervised pretraining can make the deep learning model produce better training effect, this paper introduces denoising autoencoder (DAE) to pretrain the

network. The single-layer self-encoder (AE) adds a hidden layer with nonlinear mapping ability between the input layer and the output layer. The training goal is to make the vector input by the input layer completely equal to the vector output by the output layer to realize input reconstruction. Therefore, $X1 = F1(x)$ for the hidden layer and $y = F2(x1)$ for the output layer. Because the number of neurons in the hidden layer is not equal to the dimension of the input feature, AE does not simply make an identity mapping. Its middle hidden layer can extract the statistical features of samples and obtain the most effective influence factors in the input features. After the pretraining is completed, the output layer will be removed and a new output layer will be added according to the needs of supervised learning to carry out normal supervised learning [17]. However, when the number of neurons in the hidden layer is greater than the dimension of the input feature, it will lead to the over-completion of AE and the extraction of a lot of information irrelevant to the feature, and DAE can make the number of neurons in the hidden layer take any value. Therefore, DAE can corrupt the input samples; that is, one or more eigenvalues of the input values are zero [18] according to a certain probability. Other practices are completely similar to AE. This is also the reason why this paper uses DAE method for unsupervised pretraining. When the model needs to introduce multiple hidden layers, an output layer will be added behind each hidden layer. At the same time, the output of the previous hidden layer will be regarded as the input of this layer, and the output vector of the previous hidden layer will be reconstructed. Based on this, any number of hidden layers [18] can be introduced into the model. The basic training steps of the model constructed in this paper are shown in Figure 4.

As shown in Figure 4, after the unsupervised pretraining of the model is completed, the final classification model can be obtained by performing supervised training on the model parameters, that is, fine-tuning. When the model does not need unsupervised pretraining, it can directly enter the supervised training stage. In addition, in the model evaluation stage, because this paper is not continuous value prediction, it belongs to classification prediction, so the accuracy index is used to evaluate the training performance of the model.

4. Experimental Verification

Based on the index system of influencing factors of online product sales prediction and model training algorithm proposed above, this paper makes an empirical study on 13000 samples after cleaning. At the same time, in order to show the prediction accuracy and adaptability of the model in product sales in different industries, the model training results are deeply analyzed in this paper. Data preprocessing due to the inconsistent dimensions of the selected indicators can easily lead to the instability of model training. Therefore, this paper has standardized all indicator data.

Then, the model structure and training parameters are selected. This paper selects the full connection model and CNN model in the deep learning model and uniformly adopts the early stop and learning rate decline strategy for training. Among them, the early stop strategy can stop the model training when the loss value on the verification set does not decrease for 30 consecutive echoes (epochs). The learning rate decay strategy can reduce the learning rate by a certain rate after each epoch. In this experiment, the reduction rate was 0.000001. The learning rate is 0.0001. For the fully connected model, it contains three hidden layers, each of which has 512 neurons. Each hidden layer and each input layer has a deactivation strategy (dropout). Except that the dropout probability of the input layer is set to 0.5, the others are set to 0.3. The model will "inactivate" neurons during training and will not participate in this training, so as to effectively prevent overfitting. For CNN model, full convolution structure is adopted. Therefore, the convolution kernel with step size of 2 is used to replace the original pooled layer effect, and the convolution layer steps at nonpooled positions are all 1. Thus, four alternating convolution hidden layers with step value of 1 and step value of 2 are generated, and their convolution cores are 60, 60, 120, and 120, respectively. In order to prevent overfitting, each hidden layer is still provided with dropout, and its setting rules are consistent with the full connection model. In the model pretraining stage, unsupervised pretraining is carried out for each layer of the two deep learning model structures, and fine-tuning is made for all layers in the pretraining stage, so that all layers can be trained under the supervision signal.

In this paper, CNN and full connection model are trained for product samples under primary classification, and unsupervised pretraining is introduced. Each primary classification product sample can get two model training results in each experiment, and each classification product has been tested for 5 times. In addition, the model classifies and forecasts the sales volume and adds a level every 500-sale volume; that is, when the sales volume is $< = 500$, the value is 1; the value of sales volume between 500 and 1000 is 2; and the value of sales volume between 1000 and 1500 is 3. By analogy, it is divided into 11 levels in total. When the sales volume is >5000, the maximum value of 11 is taken. In order to simplify the discussion, the training results of agricultural and animal husbandry products classification and personal consumer goods classification are selected for key analysis, as shown in Figure 5. The training results of other industry products will be given in the following model adaptability analysis. The reasons for selecting agricultural and animal husbandry products and personal consumer goods as the focus of the analysis are as follows: the product categories of the two industries are complex, the coverage is wide, and the product differences are obvious. Moreover, the merchants of the products of the two industries on Taobao and Tmall platform have a long operation time and have a large sales scale. Therefore, the samples obtained are less accidental and representative. In addition, if the model training results perform well in the products of these two industries, they will have a certain persuasion on the overall adaptability level of the model.

In the analysis of the model results, this paper introduces the accuracy index to explain the accuracy of the prediction results; it is the number of correctly classified samples
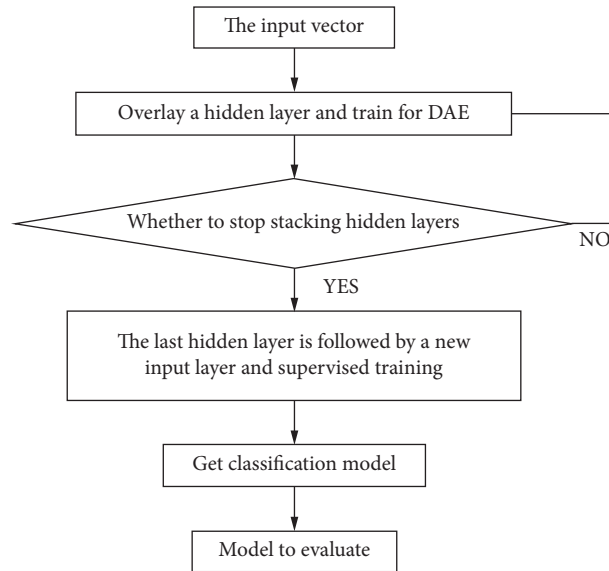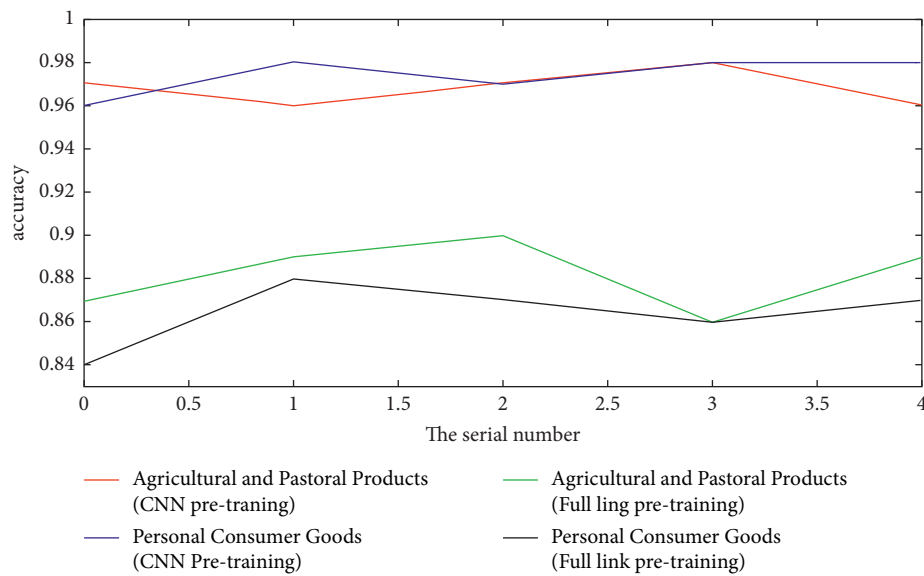
FIGURE 4: Model training process.

FIGURE 5: Model training results of agricultural and animal husbandry products and personal consumer goods.

divided by the total number of samples. As can be seen from Figure 5, the accuracy of the training results of each experiment of the two models of agricultural and animal husbandry products and personal consumer goods is more than 80%, and the average value is greater than 0.85. The accuracy of agricultural and animal husbandry products has reached more than 95%, which is a very satisfactory result. This result is related to the rich samples and various types of agricultural and animal husbandry products, and the probability distribution of price and sales volume is relatively concentrated. It can be seen that the accuracy of the model is acceptable and shows good adaptability in the classification of personal consumer goods and agricultural and animal husbandry products. In addition, the accuracy

index is calculated on the verification set. The model on the validation set shows good accuracy, which shows that the performance of the model on unknown data is also guaranteed and the phenomenon of overfitting is avoided.

Through the individual training of each industry data and the training of all data of the whole industry, the accuracy evaluation results of each model with and without pretraining can be obtained. The results are shown in Table 2. In order to make the evaluation more objective and the experiment more rigorous. Except for randomly selecting 10% of the samples as the validation set, the rest of the samples do not need to be trained. Each model was sampled and trained five times. Finally, the average accuracy of the five experiments is used as the basis for model evaluation, as

TABLE 2: Summary of model accuracy.

	AdaBoost	Full link	Full connection pretraining	CNN	CNN pretraining
Personal consumer	0.829117	0.861035	0.86267	0.877929	0.881199
Used car	0.890335	0.938462	0.938462	0.938462	0.938462
Agricultural products	0.953391	0.965934	0.967033	0.967582	0.97033
Furniture	0.919688	0.943137	0.937255	0.953431	0.957516
Clothing	0.616384	0.681448	0.681448	0.680543	0.684163
Food	0.658135	0.726351	0.725676	0.722973	0.723649
Whole industry	0.793311	0.841728	0.825175	0.834743	0.835662

shown in Table 2. In order to better reflect the accuracy and generalization ability of the model, this paper also selects AdaBoost, a widely used non-deep learning model, as the comparison model, and selects the decision tree as the basic classifier to optimize the performance of AdaBoost model. After repeated experiments on samples, the results are compared with the results of deep learning model, so as to investigate the prediction effect of deep learning CNN and full connection model.

Table 2 lists the accuracy of AdaBoost, full connection, and CNN models in six primary classification products, that is, their sales volume prediction accuracy in products of different industries, in which "whole industry" represents the mean value of the accuracy of different models. By analyzing the accuracy index in Table 2, the following results are obtained: first, the performance of the deep learning model is better than that of the AdaBoost model in all industries; second, the average prediction accuracy of the whole industry has reached more than 0.75, of which the accuracy index of deep learning model (full connection and CNN) is more than 0.8; third, the advantages of unsupervised pretraining are more obvious in CNN model. From the above analysis results, we can draw conclusions: first, as a comparative model, AdaBoost model reflects the advantages of the full connection and CNN deep learning model constructed in this paper in predicting product sales in different industries; second, supported by a relatively perfect index system, the full connection and CNN models are generally applicable to the sales volume prediction of products in different industries and have absolute advantages in prediction accuracy; third, in most forecasts, the pretraining full convolution model (CNN) is more effective than the full connection model, which shows that the unsupervised pretrained full convolution neural network (CNN) is more effective in sales forecasting and can capture the nonlinear mapping relationship between input value and sales output value. In addition, it is not difficult to increase the complexity of the deep learning model, so in theory, the deep learning model can adapt to a large data level, which many prediction models do not have. In the experimental training process of this paper, only about 10000 training samples are used. Under such a relatively small sample, the model shows better performance. Then, its performance advantage will be more obvious in the real large amount of data. It can be seen that the index system and model constructed in this paper will have better performance in practical application, which is suitable for the actual operation of product sales forecast in different industries. Based on the advantages of in-depth learning, combined with the product sales characteristics in the online trading mode, this paper constructs the influencing factor index system of online product sales, and the in-depth learning model to predict the sales of online products in different industries. According to the selected influencing factors and indicators, a large number of samples were taken from Taobao (including Tmall), an e-commerce platform. The automatic encoder (AE) method was used to mine the deep characteristics of online products in different industries, and a CNN model was constructed to predict the product sales.

5. Conclusion

The analysis results show that CNN model has good prediction accuracy and generalization ability and is suitable for sales prediction of online products in different industries. For the problem that the samples show a long tail distribution, although it does not significantly affect the training results of the model, in practical application, enterprise resources can be used to expand data sources. At the same time, the crawler algorithm is adjusted to avoid similar problems. By obtaining samples of a larger order of magnitude, the sales characteristics of products in the "long tail" can be effectively refined. This is of great help to improve the performance of the model and make it more in line with the practical requirements of product sales forecast. In addition, in the unsupervised pretraining method, other unsupervised pretraining algorithms such as restricted Boltzmann machine (RBM) and noise reduction self-encoder (DAE) can be selected to compare the pretraining effect between them, so as to achieve the best model training effect. This is also a potential breakthrough point to further optimize the model in future research.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] J. Bi and H. Wei, "Improved B-P neural network and its application in sales forecasting," *Journal of Shandong University of Technology*, vol. 125, no. 6, pp. 29–33, 201.

[2] Q. Qu, C. Liu, and X. Bao, "E-commerce enterprise supply chain financing risk assessment based on linked data mining

and edge computing," *Mobile Information Systems*, vol. 2021, Article ID 9938325, 9 pages, 2021.

[3] H. Peng and S. Yu, "Sales forecast of vending machines based on time series analysis," *Computer Science*, vol. 42, no. 6, pp. 122–124, 2015.

[4] J. Wang, "E-commerce sales forecast based on commodity clustering," *Computer System Application*, vol. 25, no. 10, pp. 162–168, 2016.

[5] M. Wu, H. Lin, and S. Li, "A cigarette sales prediction method based on support vector machine," *Tobacco Science and Technology*, vol. 49, no. 2, pp. 87–91, 2016.

[6] J. Liu, S. He, and Q. Zhu, "Research on online agricultural product sales forecasting model based on deep learning," *Computer Application Research*, vol. 34, no. 8, pp. 2291–2293, 2017.

[7] Z. Sun, X. Lei, and Y. Xu, "Review of deep learning research," *Computer Application Research*, vol. 29, no. 8, pp. 2806–2810, 2012.

[8] Y. Bengiy, "Learning for architects for AI," *Foundation & Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[9] D. Erhan, Y. Bengiy, A. Courville, P. A. Manzagol, P. Vincent, and S. Bengio, "Why is unsupervised pre training help deep learning?" *Journal of Machine Learning Research*, vol. 11, no. 3, pp. 625–660, 2010.

[10] H. Peng, "Research on credit evaluation of financial enterprises based on the genetic backpropagation neural network," *Scientific Programming*, vol. 2021, Article ID 7745920, 12 pages, 2021.

[11] P. Yin, G. Dou, X. Lin, and L. Liu, "A hybrid method for forecasting new product sales based on fuzzy clustering and deep learning," *Kybernetes*, vol. 5, 2020.

[12] L. Wang, M. Gao, and Z. Liang, "Application of data envelopment and internet of things technology for asset value evaluation," *Mobile Information Systems*, vol. 2021, Article ID 9934090, 8 pages, 2021.

[13] A. Krizhevsky, L. Sutskever, and G. E. Hinton, "Image Net classification with deep convolutional neural networks," in *Proceedings of the International Conference on Neural Information Processing Systems*, pp. 1097–1105, Curran Associates Inc, Lake Tahoe, NV, USA, December 2012.

[14] C. Liou, J. C. Huang, and W. C. Yang, "Modeling word usage using the Elman network," *Neurocomputing*, vol. 71, no. 16, pp. 3150–3157, 2008.

[15] Y. Lecun, L. bottou, Y. bengio, and P. Haffner, "Gradient based learning applied to document recognition," *Proceedings of the IEEE*, vol. 11, pp. 2278–2324, 1998.

[16] T. Springenbergj, A. Dosovitskiy, T. Brox, and M. Reidmillar, "Striving for simplicity: the all convolutional net," 2014, https://arxiv.org/abs/1412.6806.

[17] C.-Y. Liou, W.-C. Cheng, J.-W. Liou, and D.-R. Liou, "Autoencoder for words," *Neurocomputing*, vol. 139, no. 139, pp. 84–96, 2014.

[18] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. A. Manzagol, "Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, no. 12, pp. 3371–3408, 2010.