Scholar Publishing Group
Machine Learning Theory and Practice
https://doi.org/10.38007/ML.2022.030407
ISSN 2790-0983 Vol. 3, Issue 4: 53-60

Machine Learning in the Prediction of Commodity Sales

Yuxi Zhou*
Philippine Christian University, Philippine
2945942870@qq.com
**corresponding author*

Keywords: Machine Learning, Commodity Sales, Data Processing, Sales Forecast

Abstract: With the development of national economy, people's consumption level has also increased. In order to improve the utilization rate of products, reduce waste, and meet the requirements of environmental protection, it is necessary to improve the commodity sales forecast to reduce the waste of resources. Therefore, this paper proposes the role of machine learning algorithm in commodity sales forecasting. This paper mainly uses the cluster analysis method and the comparison method to carry on the experiment and the data analysis to the commodity sales volume forecast system. The experimental results show that the response time of the system designed in this paper is less than 2s, which can well meet the system requirements. Therefore, machine learning has played a greater role in the prediction of commodity sales.

1. Introduction

In the commodity sales forecast, the model is used more and more widely. Machine learning is also an important and indispensable method. It can process and analyze data and knowledge through computer technology. Machine learning is an effective and direct method in forecasting the sales volume of goods. It combines data sets, knowledge sets and algorithms. It includes a lot of repeated input and output information and implicit training distribution function. It can combine data mining technology, decision tree and neural network.

There are many theoretical achievements on the application of machine learning in commodity sales forecasting. For example, some experts pointed out that commodity sales is one of the indicators that enterprises attach great importance to. It can help merchants formulate appropriate strategies for current sales to maximize profits [1-2]. Some scholars have proposed a prediction model based on k-means clustering and machine learning regression algorithm for the prediction of multi commodity sales in the retail industry [3-4]. In addition, some experts believe that accurate prediction of commodity sales can improve warehouse efficiency, reduce raw material consumption, reduce inventory occupation, and better meet market demand [5-6]. Therefore, the research on commodity sales in this paper is a platitude topic, and the use of machine learning algorithm is a practical application of existing technical means.

This paper first analyzes the application of data mining in commodity sales. Secondly, machine learning algorithm and related knowledge are discussed, and clustering algorithm is proposed. Then the sales forecast system is designed. Finally, the experimental environment is set up, and the algorithm comparison and system response time related research are carried out to draw conclusions.

2. Application of Machine Learning in Commodity Sales Forecast

2.1. Application of Data Mining in Commodity Sales

There are many factors that affect the sales of goods, including consumers' purchasing power, personal preferences and other factors. In different periods of time, there will be differences in sales volume due to commodity sales prices and income levels. Commodity sales forecast is a purposeful and systematic information collection process. There are many methods to predict sales volume, among which regression analysis is commonly used. It is based on statistical principles to establish mathematical models for modeling and calculation. In practical application, grey system theory and analytic hierarchy process can be used. When forecasting the sales volume, mathematical analysis can be used to establish a forecasting model for scientific forecasting. The method of time series prediction is to organize data into time series and predict the future value of data with mathematical methods. Through machine learning, model prediction can be made by constructing feature engineering. After establishing the prediction model, we must evaluate the accuracy of the model through certain evaluation indicators. Common methods for evaluating the accuracy of prediction regression models include average absolute error, percentage average error, mean square error and percentage average error [7-8].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{x}(b_i - \hat{b})^2}{x}} \tag{1}$$

Where, b_i refers to the actual value of the ith sample, but $\hat{b}_i$ predicts value of the ith sample, and x refers to the actual number of samples. In this study, we cannot carry out an unlimited number of experiments, so we can only find the best results.

Data mining is an important step of data mining, and also a necessary part of building prediction models. In the face of massive data, we need to start from the data, analyze and extract the data. In order to accurately predict the relationship between commodity sales, it is necessary to analyze and discuss the different dimensions of commodity sales, so as to find the potential connotation relationship hidden in the commodity sales data [9-10].

2.2. Machine Learning Algorithm and Related Knowledge

Pattern prediction is to simulate the automatic prediction system of human intelligence and use machine learning algorithm to process data intelligence to judge object attributes. The general process of pattern prediction is data acquisition, preprocessing, variable generation, variable selection, pattern classification, etc. Variable selection and pattern classification are completed through machine learning. The research object of pattern prediction is different depending on the field. It can be genes, proteins, images, signal waves or other measurable objects with attributes. Therefore, its core problem is classification. The pattern prediction system first obtains a certain number of data sets for the prediction of unknown samples. According to the problem area studied, the data types obtained are also different, which may be binary data, character data, entity data, etc. The model prediction algorithm is trained and used to predict the samples of unknown attribute classes and make appropriate decisions. Whether there are supervised learning methods or unsupervised learning methods, each method has its own advantages and disadvantages. It is necessary to select an appropriate pattern prediction algorithm for different problems and the characteristics of the feature data to be processed [11-12].

Linear regression can be divided into single linear regression and multiple linear regression analysis according to the number of independent variables. The dependent variable A of linear regression analysis is related to an independent variable B, which is the simplest form of regression. The functional relationship is expressed as a linear function of B, namely:

$$B = \mathrm{y} + \mu A \tag{2}$$

Multiple linear regression analysis is an extension of multiple independent variable linear regression analysis. Clustering analysis is one of the important methods to find the natural aggregation structure of data. It can improve the understanding of data categories, and can also be used as the premise of other data mining methods such as classification and prediction [13-14].

2.3. Design of Sales Forecast System

Sales forecast information system is a part of enterprise management information system. It should integrate with the surrounding systems, give full play to the advantages of each system, reduce the opportunities for manual data entry, improve the data interaction ability between systems, help improve the feasibility of the system, improve the internal operation efficiency of the company, and reduce management costs [15-16]. The peripheral system integration of the sales forecast system is shown in Figure 1:

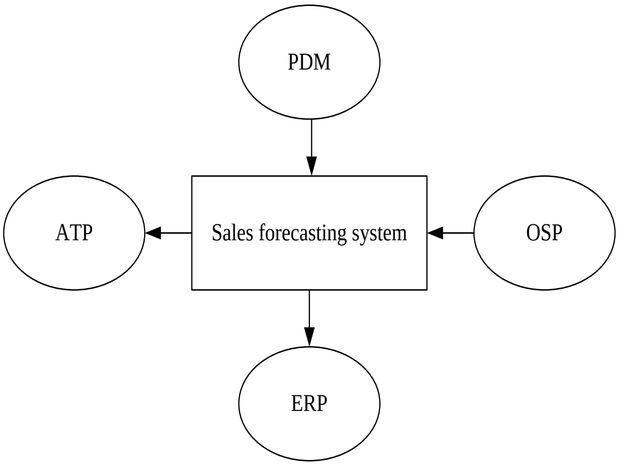

Figure 1. The peripheral system of the sales forecast system

The sales forecast system is based on the basic product data. When filling in the forecast, the product manager of the representative office can only select complete basic product data. The source of basic product data is PDM system. In addition to basic product data, PSO is also a system that needs to collect data. The actual contract/order processing is performed in PSO. From the sales forecast of the product manager to ATP and SOP, the plan BOM is a dimension. Sales forecast is one of the input data of ATP system. The ATP system shall match the demand information and quotation information in ERP. SOP data is an important part of ERP weekly plan. The MDS module of ERP takes the SOP data as the final total demand according to the demand forecast and contract quantity balance. In addition, the promotion and distribution system is integrated with the sales forecast system, but other systems are not the main functions and requirements of the sales forecast system [17-18].

The system will be divided into six menus: application management, basic data management, revenue forecast, SOP processing, ATP commitment management, and statistical report. The system defines roles to control the permissions of each user. Each role represents the user level and limits the operation permission of the role. After configuring the organization type, it need to configure the entity and company organization structure for the corresponding organization type in the system. After configuring the organization type and organization, it need to manage the relationship between organizational structures at all levels to transfer data among organizations at all levels according to rules. The authority management of the sales forecast information system is completed. First, control the action menu and configure the effectiveness of the action menu for each role. Second, it must configure product permissions. To have menu permissions, it must also have product permissions to perform appropriate actions on the product.

3. System Test

3.1. System Environment

In order to support ongoing business operations, the system will support these operations. Similar to the B/S architecture, the B/S architecture does not allow users to access and operate the external network, but allows them to access the LAN of any representative office. The system needs a complex report query interface. Use the report development tools provided by SQL Server 2017 to develop some interface elements using DEXPRESS controls in order to better realize complex data entry and user operability. The design shall fully reuse the existing and verified functional modules to ensure effective support for project activities and minimize the scope of secondary development. In order to enable the system to cover all business operations, it must also be able to monitor data throughout the process, conduct comprehensive user authority management, and exchange system data with surrounding systems to reduce data redundancy.

3.2. System Function Test

The user account management of the system is related to the security and stability of the system, and may also affect the maximization of commercial sales profits for sales enterprises. Therefore, detailed testing and analysis are required. As the basis of data analysis and prediction, data acquisition plays an important role. Test this part of functions. We use ID3 decision tree algorithm to build the system test model. The improved ID3 algorithm needs to process continuous numerical data. For the combination of the actual data and the establishment of the decision tree, the noise data of the actual data and the interference of other data with low availability must be considered. Sales forecast is the core function of this paper. We use BP algorithm to design and make appropriate improvements.

3.3. System Performance Test

In addition to the functions of the system, users also care about the performance of the system. That is, the system is not only required to be usable, but also easy to use. This section mainly tests the page response delay and data reading delay in system performance. We used several experiments to test the response time of each page jump. The experiment was conducted 10 times, mainly for the login interface, main interface and sales analysis interface. For most system development, it is basically to interact with data, so data reading in the database is a frequent operation. If the user is using the system, the waiting time for data processing results is too long. They used 10 experiments to test the response time of two data reading modules of cigarette sales analysis system and sales prediction system.

4. Analysis of Experimental Results

4.1. Comparison of Test Results between the Univariate Linear Regression Analysis Model and Bp Neural Network Model

This paper intends to verify the performance of BP algorithm. We use the univariate linear regression technique to conduct data comparison and analysis. The proportion of good prediction times of BP neural network model is better than that of single variable linear regression model. See Table 1 for details:

Table 1. Comparison of test results between unary linear regression analysis and BP neural network model

	Unitary linear regression analysis	BP neural network
MAPE	15.85%	14.05%
Maximum APE	52.61%	64.22%
Minimum APE	0.55%	2.12%
Times of good predictions	66.68%	80.15%
Unqualified predicted times	6.65%	6.65%

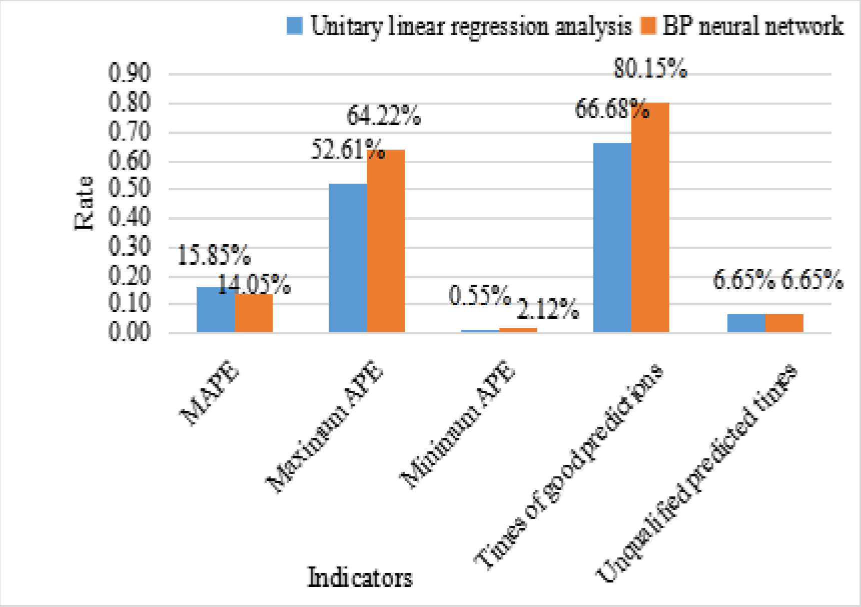

Figure 2. Comparison of test results between unary linear regression analysis and BP neural network model

As shown in Figure 2, we can see that the BP neural network model is slightly stronger than the unary linear regression model in terms of the most important indicator, the average absolute percentage error (MAPE). Therefore, relatively speaking, BP neural network model has better prediction accuracy and generalization ability.

4.2. System Page Response Test

The login interface has the smallest jump delay, mainly because this interface and the main interface jump. The main reason is that the class package import takes a certain amount of time without too much participation of the database, so the response is fast. Other performance conditions are shown in Table 2:

Table 2. System page response test

	Login page	Main interface	Sales analysis interface
1	0.5	1.2	1.6
2	0.4	1.3	1.7
3	0.3	1	1.4
4	0.4	1	1.3
5	0.5	1.1	1.4

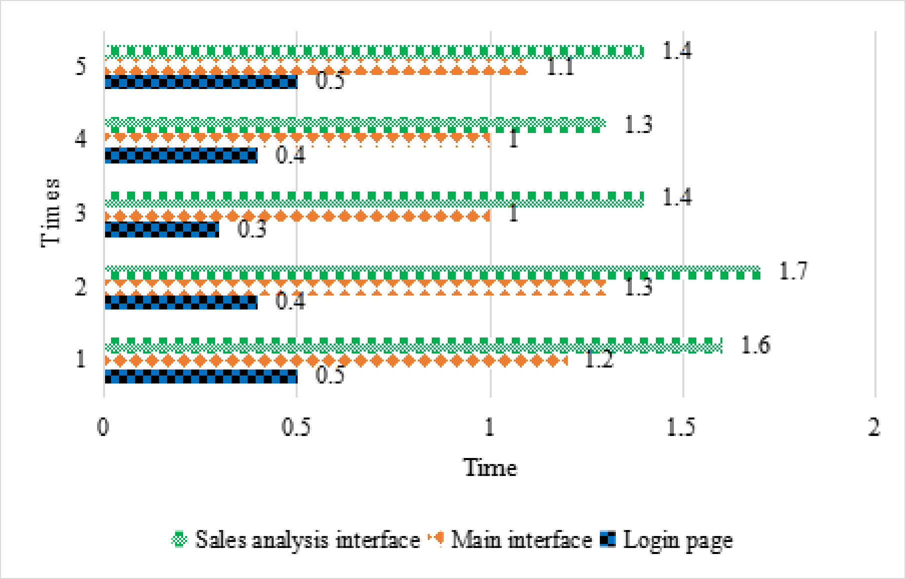

Figure 3. System page response test

As shown in Figure 3, we can see that the main interface and many other modules need to interact, so the jump response time is relatively more expensive. For the sales analysis interface, it is mainly responsible for reading data and submitting it to the page for display after complex calculation. Therefore, its response time is obviously much less.

5. Conclusion

This paper studies the factors that affect the sales volume of goods, uses machine learning theory and methods, and combines case analysis to verify the problems in the application of the model. In mathematical modeling. Bayesian classification algorithm is used for data mining, and the predicted value is significantly different from the actual value. When the correlation model established by the time series method is used to predict the sales volume, it does not take into account the change of variables affected by random interference, so it cannot accurately reflect the trend and law of commodity sales volume, error rate and other characteristic indicators. Therefore, it is necessary to study the univariate linear regression analysis model and BP neural network model. The number of experimental data in this paper is not much, and relevant data needs to be added to achieve the experimental purpose.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] *Seyed Ali Hasheminejad, Masoud Shabaab, Nahid Javadinarab: Developing Cluster-Based Adaptive Network Fuzzy Inference System Tuned by Particle Swarm Optimization to Forecast Annual Automotive Sales: A Case Study in Iran Market. Int. J. Fuzzy Syst. 24(6): 2719-2728 (2022). https://doi.org/10.1007/s40815-022-01263-6*

[2] *Xuan Bi, Gediminas Adomavicius, William Li, Annie Qu: Improving Sales Forecasting Accuracy: A Tensor Factorization Approach with Demand Awareness. INFORMS J. Comput. 34(3): 1644-1660 (2022). https://doi.org/10.1287/ijoc.2021.1147*

[3] *Marlene A. Smith, Murray J. Côté: Predictive Analytics Improves Sales Forecasts for a Pop-up Retailer. INFORMS J. Appl. Anal. 52(4): 379-389 (2022). https://doi.org/10.1287/inte.2022.1119*

[4] *Kohei Takahashi, Yusuke Goto: Embedding-Based Potential Sales Forecasting of Bread Product. J. Adv. Comput. Intell. Intell. Informatics 26(2): 236-246 (2022). https://doi.org/10.20965/jaciii.2022.p0236*

[5] *Mert Girayhan Türkbayragí, Elif Dogu, Y. Esra Albayrak: Artificial Intelligence Based Prediction Models: Sales Forecasting Application in Automotive Aftermarket. J. Intell. Fuzzy Syst. 42(1): 213-225 (2022). https://doi.org/10.3233/JIFS-219187*

[6] *Austin Schmidt, Md Wasi Ul Kabir, Md. Tamjidul Hoque: Machine Learning Based Restaurant Sales Forecasting. Mach. Learn. Knowl. Extr. 4(1): 105-130 (2022). https://doi.org/10.3390/make4010006*

[7] *Shaohui Ma, Robert Fildes: Retail Sales Forecasting with Meta-Learning. Eur. J. Oper. Res. 288(1): 111-128 (2021). https://doi.org/10.1016/j.ejor.2020.05.038*

[8] *Carlos Aguilar-Palacios, Sergio Muñoz-Romero, José Luis Rojo-Álvarez: Cold-Start Promotional Sales Forecasting Through Gradient Boosted-Based Contrastive Explanations. IEEE Access 8: 137574-137586 (2020). https://doi.org/10.1109/ACCESS.2020.3012032*

[9] *Nadide Caglayan, Sule Itir Satoglu, E. Nisa Kapukaya: Sales Forecasting by Artificial Neural Networks for the Apparel Retail Chain Stores-An Application. J. Intell. Fuzzy Syst. 39(5): 6517-6528 (2020).*

[10] *Shakti Goel, Rahul Bajpai: Impact of Uncertainty in the Input Variables and Model Parameters on Predictions of a Long Short Term Memory (LSTM) Based Sales Forecasting Model. Mach. Learn. Knowl. Extr. 2(3): 256-270 (2020). https://doi.org/10.3390/make2030014*

[11] *Aldina Correia, Isabel Cristina Lopes, Eliana Costa e Silva, Magda Monteiro, Rui Borges Lopes: A Multi-Model Methodology for Forecasting Sales and Returns of Liquefied Petroleum*

Gas Cylinders. Neural Comput. Appl. 32(16): 12643-12669 (2020). https://doi.org/10.1007/s00521-020-04713-0

[12] Carlos Aguilar-Palacios, Sergio Muñoz-Romero, José Luis Rojo-Álvarez: Forecasting Promotional Sales within the Neighbourhood. IEEE Access 7: 74759-74775 (2019). https://doi.org/10.1109/ACCESS.2019.2920380

[13] Giuseppe Craparotta, Sébastien Thomassey, Amedeo Biolatti: A Siamese Neural Network Application for Sales Forecasting of New Fashion Products Using Heterogeneous Data. Int. J. Comput. Intell. Syst. 12(2): 1537-1546 (2019). https://doi.org/10.2991/ijcis.d.191122.002

[14] Charu Gupta, Amita Jain, Nisheeth Joshi: DE-ForABSA: A Novel Approach to Forecast Automobiles Sales Using Aspect Based Sentiment Analysis and Differential Evolution. Int. J. Inf. Retr. Res. 9(1): 33-49 (2019). https://doi.org/10.4018/IJIRR.2019010103

[15] Abeer S. Desuky, Yomna M. Elbarawy, Samina Kausar, Asmaa Hekal Omar, Sadiq Hussain: Single-Point Crossover and Jellyfish Optimization for Handling Imbalanced Data Classification Problem. IEEE Access 10: 11730-11749 (2022). https://doi.org/10.1109/ACCESS.2022.3146424

[16] Marwan H. Hassan, Saad M. Darwish, Saleh M. El-Kaffas: An Efficient Deadlock Handling Model Based on Neutrosophic Logic: Case Study on Real Time Healthcare Database Systems. IEEE Access 10: 76607-76621 (2022). https://doi.org/10.1109/ACCESS.2022.3192414

[17] Kanu Goel, Shalini Batra: Dynamically Adaptive and Diverse Dual Ensemble Learning Approach for Handling Concept Drift in Data Streams. Comput. Intell. 38(2): 463-505 (2022). https://doi.org/10.1111/coin.12475

[18] Arjun Puri, Manoj Kumar Gupta: Improved Hybrid Bag-Boost Ensemble with K-Means-SMOTE-ENN Technique for Handling Noisy Class Imbalanced Data. Comput. J. 65(1): 124-138 (2022). https://doi.org/10.1093/comjnl/bxab039