Sales Prediction Based on State-of-art Machine Learning Scenarios

Yuanpu Hu*
*202042041@uibe.edu.cn

School of Finance University of International Business and Economics Shanghai China

Abstract-The way to predict sales volume accurately and efficiently is an important issue that enterprises have always paid attention to. Although the traditional time series prediction method plays a leading role in research and practice, it has some limitations. With the development of big data, e-commerce enterprises can obtain unprecedented data volume and data characteristics. It is difficult to accurately predict sales volume only based on past behaviors and trends. In this paper, we propose a combined forecasting model of cost aversion bias based on random forest, GBDT and XGboost algorithms, and uses the cost data of each commodity to realize the fine weighting of samples, so as to output the forecasting results. These results shed light on pointing that combined forecasting model can predict sales more accurately, which is of great significance to e-commerce enterprises to reduce commodity management costs.

Keywords-Sales forecast; Machine learning; Combination model; Feature construction; Sample weighting.

1 INTRODUCTION

Sales prediction is a kind of prediction of future sales by the sales implementation unit, which carries out law analysis and data mining according to the existing historical sales data, comprehensively considers a variety of influencing factors, and designs a scientific and reasonable prediction sales model for the future. The accuracy of sales prediction is a necessary condition for enterprises to carry out market demand planning. It is an important basis for market analysis and market research. It directly affects the project budget, capital return and business decision-making [1].

Generally, there are some traditional methods of sales prediction. The initial stage is oral prediction and report summary. Through the weekly sales meeting, the enterprise asks the salesperson about the current sales situation and the sales opportunities being followed up, and summarizes the salesperson's sales prediction through spreadsheet [2]. Such a sales forecasting process not only takes up the time for salespeople to communicate with customers, but also the accuracy of sales prediction is very low [3]. The intermediate stage is CRM sales management platform, which is the basic of modern middle office management of an enterprise. It can collect and store complete sales data as much as possible, and predict the sales trend through big data analysis. However, the operation experience of the United States in recent ten years has proved that CRM has great defects. Specifically, the sales insight report of CRM platform mainly predicts the sales trend through the general mathematical statistical model at present, and lacks the mathematical statistical model of business logic and market practice. The advanced stage is

FFIT 2022, October 28-30, Shenzhen, People's Republic of China

DOI 10.4108/eai.28-10-2022.2328458

Revenue Intelligence System. Companies use the business model with many years of sales analysis experience in the industry as the core, which is supplemented by advanced business intelligence technology, and the reports are through visual and automatic insight [4].

Therefore, machine learning, a new method of artificial intelligence, has attracted much attention because of its ability to enhance prediction performance and model nonlinear models. Some scholars have demonstrated the outstanding interpretability level, good accuracy and appropriate calculation time of Random Forest (RF). Random Forest is also considered a conventional tool for predictive analysis because it allows managers to understand the reasons behind the model and how it affects the final results. In addition, the gradient boosting decision tree (GBDT) algorithm using the idea of iteration and gradient lifting shows better performance and stability than the general model in the prediction of production and service demand [5]. Based on gradient lifting, extreme boosting (XGboost) algorithm has excellent accuracy in industrial practice, and its good accuracy has been verified in the research of sales prediction [6]. In the process of outputting prediction models, academic researchers and business practitioners often encounter an important problem: whether to choose appropriate modeling methods for prediction, or combine these different methods into a single prediction model? After a lot of research, it is found that the prediction effect of the combined model is generally better than that of the individual prediction model, and the prediction accuracy of the nonlinear combination is better than that of the linear combination model [7].

This paper will first introduce RF, GBDT and XGboost algorithms. In order to further optimize the prediction results, we will establish a combined prediction model of the three, focus on the characteristics of e-commerce commodities, analyze the characteristic factors affecting e-commerce commodities, and predict the sales volume of e-commerce commodities.

2 BASIC DESCRIPTIONS OF SALES PREDICTION

For a long time, the sales forecast of commodities has been an important topic in the retail industry. Accurately predicting the sales volume of single products can improve the stock efficiency of stores, so as to reduce commodity loss, reduce inventory occupation, and better meet the market demand. Due to the complexity of real-life environment and the scarcity of data, accurate sales prediction is a very difficult problem. Most of the time, sales prediction stays in the total amount prediction analysis, while the fine-grained prediction of single products is difficult to achieve. With the development of information technology and the emergence of various data sensors, we can obtain various factors affecting commodity sales from a fine granularity. Meanwhile, with the significant improvement of computer computing power, machine learning, especially deep learning, is gradually rising. The algorithm model has achieved great success in image recognition, audio understanding, natural speech processing etc. In this case, it is possible for us to realize the fine-grained analysis and prediction of commodity sales through multi-directional data analysis and efficient model algorithm. In recent years, e-commerce and e-commerce logistics have flourished in the context of the Internet era, shortening the delivery cycle and increasing customer expectations. In order to win sustainable competitive advantage, e-commerce enterprises, with limited resources, sales forecast becomes more important [8]. At the same time, e-commerce enterprises can also obtain a large amount of data from consumer behavior. Data has become the core competitiveness of enterprises in the future development. Massive data is their unique advantages and resources for enterprises [9]. The way to effectively use data and accurately predict sales has become the focus of e-commerce enterprises.

The accuracy of sales prediction is a necessary condition for enterprises to carry out market demand planning. It is an important basis for market analysis and market research, which directly affects the project budget, capital return and business decision-making.

The current situation of sales predictions is that the sales modernization level of B2B enterprises in the United States is in the middle to advanced transformation, while the sales management of B2B enterprises in China started relatively late [10]. With understanding of middle platform management and the recognition of artificial intelligence big data analysis, it is rapidly updating from the primary stage to the intermediate stage.

3 MACHINE LEARNING APPROACH

3.1 Random Forest

RF is a regression model learning algorithm based on decision tree [11]. Firstly, we should know the configurations and principles of decision tree. The decision tree uses the tree structure to specify the sequence of decisions and results. For a given *n* independent variables X = {x_1, x_2, ... x_n}, the dependent variable *y* is predicted. The prediction is realized through the decision tree constructed by nodes and branches. On each node of the decision tree, select a characteristic branch and traverse the tree downward, and finally reach an end point, and then make a decision. We can sample n samples again from the original data set to form a new data set. The probability of a sample being sampled is:

$$\lim_{n\to\infty} 1 - \left(1 - \frac{1}{n}\right)^n \approx 63\% \tag{1}$$

Assuming that the dimension of each sample is a, K features are randomly selected to train a decision tree; Repeat the above two steps m times to obtain the random forest model (m decision trees).

3.2 Gradient Boosting Decision Tree model

GBDT is a decision tree algorithm [12] derived from the idea of iteration. It predicts the results by integrating the base learner, cart regression tree which is used to form a strong learner. For a dataset containing n samples $D = \{(x_i, y_i)\}(|D| = n, x_i \in R, y_i \in R$. The specific steps of the algorithm are as follows:

(1) Initialize learner

$$f_0(x) = argmin\Sigma_i^n L(y_i, C) \tag{2}$$

Among them, the initial constant C is generally set as the mean value of the real value of the sample, and L() is the required value [13].

(2) Iterative model, where the number of iterations M = 1,2…M

1) For each sample I = 1,2… N, calculate the negative gradient, which is residual r_{im}*:*

$$r_{im} = -\left[\frac{\partial L\left(y_i, f(x_i)\right)}{\partial f(x_i)}\right]_{f(x)=f_{m-1}(x)} \tag{3}$$

2) The obtained residual is updated to the true value of the sample, and the data (x_i, r_{im}) is used as the training data of the m-th tree, and its corresponding leaf node is R_{jm}, j = 1,2 …J, J is the number of leaf nodes of the regression tree.

3) Calculate the best fitting value for the leaf area j = 1,2..., J

$$Y_{jm=arg\min_{Y} \sum_{x_{i\in R_{jm}}} L(y_i, f_{m-1}(x_i)+Y)} \tag{4}$$

4) Update strong learner

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J} Y_{jm} I\ , x \in R_{jm} \tag{5}$$

where *I* is the indicating function, and the value is 1 when x∈ R_{jm}, otherwise it is 0.

(3) The final predicted value of the strong learner is

$$\hat{f}(x) = f_M(x) \tag{6}$$

3.3 Extreme gradient boosting (XGboost)

XGboost is an efficient implementation of GBDT algorithm. XGboost is an integrated lifting tree learning model proposed by Chen Tianqi and others It efficiently implements GBDT. The base learner in XGboost can be cart or linear classifier [14]. For a dataset containing n samples $D = \{(x_i, y_i)\}(|D| = n, x_i \in R, y_i \in R$, the specific steps of the algorithm are as follows:

(1) The objective function is defined, which is composed of loss function and regular term

$$L(\varphi) = \sum_{i}^{n} (\hat{y_i}, y_i) + \sum_{k}^{K} \Omega\ (f_k) \tag{7}$$

The regular term part is shown in the following formula:

$$\Omega\ (f_k) = \gamma T + 0.5\lambda\omega^2 \tag{8}$$

In the above formula: K represents a total of K trees, f_k represents the k-th tree model, T represents the number of leaf nodes of each tree, ω represents the weight value of leaf nodes of each tree, γ and λ are coefficients, which need to be adjusted during training.

(2) The model strategy is the same as that of GBDT. The solution of the objective function is also based on the idea of iteration. For the t-th iteration:

$$L^{(t)} = \sum_{i}^{n} l\left(y_i, \hat{y_i}^{(t-1)} + f_t(x_i)\right) + \Omega\ (f_k) \tag{9}$$

The above formula is expanded by Taylor formula to obtain:

$$L^{(t)} = \sum_{i}^{n} l\left(g_i f_t(x_i) + 0.5 h_i {f_t}^2(x_i)\right) + \Omega\ (f_k) \tag{10}$$

Here,

$$g_i = \partial_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}^{(t-1)}\right), h_i = \partial^2{}_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}^{(t-1)}\right) \tag{11}$$

(3) In addition to some differences between the algorithm and the traditional GBDT: XGboost also made a lot of optimization in engineering practice. In general, the differences and connections between the two can be summarized into the following aspects.

1) GBDT is a machine learning algorithm, and XGboost is the engineering implementation of the algorithm.

2) When using cart as the base classifier, it is displayed by XGboost. Regular term is added to control the complexity of the model, which is helpful to prevent over fitting and improve the generalization ability of the model

3) GBDT only uses the first derivative information of the cost function in model training. XGboost carries out the second-order Taylor expansion of the cost function, and the first-order and second-order derivatives can be used at the same time

4) The traditional GBDT uses card as the base classifier, and XGboost supports many types of base classifiers, such as linear classifiers

5) Traditional GBDT uses all data in each iteration. XGboost adopts a strategy similar to random forest to support data sampling

(6) The traditional GBDT is not designed to process missing values. XGboost can automatically learn the processing strategy of missing values [14].

4 APPLICATION IN SALES PREDICTION

Some scholars have studied candidate models about selection strategy and combination strategy. Found that the prediction performance of the combination model is generally better than the individual prediction model, the accuracy of the nonlinear combination is better than the linear [15]. Therefore, to optimize the prediction performances, the scholar constructs a new feature set, through the characteristics of e-commerce products and the characteristic factors that affect e-commerce products. A combined prediction model is established using RF, GBDT, and XGboost algorithms to establish to predict the sales of e-commerce products [16].

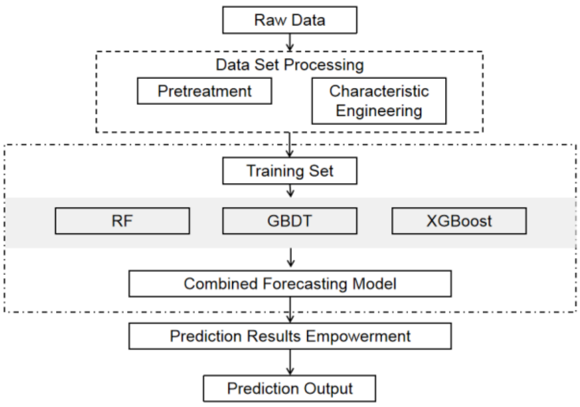

Figure 1. A sketch of forecast flow.

The purpose is to establish a sales forecasting model, to predict the sales of e-commerce over a period of time. To this end, a combined machine learning model will be used to predict the future sales through the data of e-commerce goods. The overall research framework is shown in Figure 1. It mainly includes two Parts: dataset processing and combinatorial predictive model construction.

The data set studied consists of national warehouse data, regional warehouse data and commodity cost data. As for dataset processing, models and algorithms cannot determine the upper bounds of predicted outcomes. Therefore, when it is difficult for the algorithm to break through the bottleneck, good prediction results will be achieved with excellent combined features. Thus, the scholars deal with the missing values of the original data, preprocess the data of the outliers, and organize the re-construction of the sample features. Among them, the forecast period of commodity sales mostly takes weekly sales as the minimum forecast unit. Considering the characteristics of large demand for e-commerce goods, as well as it's complex logistics process, this paper uses two weeks as a forecast unit to resample the original data. The time sliding window method can eliminate the data noise and expand the training set [17].

Regarding to combinatorial predictive model construction, there are many factors affecting actual sales, the scholars expand the feature set on the basis of the original data, i.e., the new data set obtains more information. It can be learned by the algorithm. If a single model is used to predict, it may face a risk of decrease in accuracy or over-fitting. Dietterich pointed, the benefits may be brought from the model combination from three aspects: statistics, calculation and representation [18]. Therefore, the scholars build a combined sales forecasting model using the structural differences of different algorithms, it improves the accuracy and reduces the risk of model over-fitting. The specific steps are as follows: First, Base model training. The training set reconstructing the samples. Secondly, trained the features with random forest, GBDT, and XGboost models, respectively. Then, one inputs the test set feature set for prediction, and obtains the predicted value of each basic model, as shown in Equation (12):

$$\hat{Y}^{(m)} = \left[\hat{y}_1{}^{(m)}, \hat{y}_2{}^{(m)} \dots, \hat{y}_n{}^{(m)}\right]^T \ (m = 1,2,3) \tag{12}$$

Second, we need to combined models for forecasting. For e-commerce platforms, the difference between the forecast results and the actual sales directly determines the method, whether to use supplementary or supplementary cost to calculate the cost. Additionally, the cost of supplementary for different products is different. On account of that, the scholars will use the cost into ales forecast model modeling. Third, Predictive Empowerment. On the basis of combined forecasting, in order to improve the cost aversion bias of the forecasting results, each sample will be weighted. As shown in Figure 2, the sigmoid function is often used as a neural network, as a result of its good monotonically increasing properties.

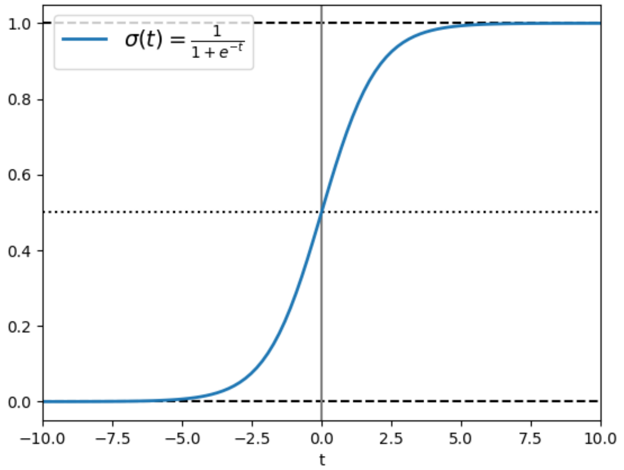

Figure 2. Sigmoid model

Table 1 Forecast Total Cost

Model	Total Cost(yuan)
RF	1,773,500
GBDT	1,915,300
XGboost	2,002,500
Characteristic Engineering + RF	1,707,800
Characteristic Engineering + GDBT	1,745,600
Characteristic Engineering + XGboost	1,803,400
Characteristic Engineering + Combined Model	1,490,400
Characteristic Engineering + Combined Model + Prediction Empowerment	1,274,300

Fourth, after calculating the weight value of each sample, assign a value to the combined prediction result of cost aversion bias, and the final prediction result is shown in Equation (13)

$$\widehat{Y}_i' = \begin{cases} \widehat{Y}_i \div w_i, a_i \leq b_i \\ \widehat{Y}_i \times w_i, a_i \geq b_i \end{cases} \tag{13}$$

After training the processed data sets b RF, GBDT, and XGboost models, the prediction results of the basic models are combined. On this basis, the combined prediction results are weighted by the supplementary cost, which can be used to evaluate the predictive ability of the combined model.

The scholars use the three basic models of RF, GBDT and XGboost, adds the model of feature engineering, and the combined prediction model to predict the experimental data. Finally uses the total cost as the evaluation index. The results are shown in Table 1. According to the results, we can know the combined models improve the prediction accuracy. First, feature engineering can improve the prediction accuracy. Compared with the basic model, the total cost of the model is reduced after adding feature engineering. It shows that the feature construction process adds effective features from the original data, and is helpful to improved model accuracy. Second, the combined prediction model has better prediction accuracy. The combined model after adding feature engineering is predicted to be 1,274,300 yuan, which is decreased 420,000 to 520,000 yuan than the basic model after adding feature engineering. This figure shows that the superiority and the effectiveness of the combined model, compared with the basic model and the predictive weighting method

Finally, through proposing a combined forecasting model, which based on machine learning, obtains a more accurate forecast of the sales of e-commerce products. On the basis of the refined use of platform data, the overall storage cost is reduced, which is of great significance. As well as important for the e-commerce platform to understand the sales of goods in advance, and to reasonably formulate the inventory level.

5 GAPS AND FUTURE

Every model has some shortcomings. Firstly, random forest does not perform as well for regression as it does for classification, because it does not give a continuous output. When performing regression, random forest cannot make predictions beyond the range of the training set data, which can lead to over-fitting. As well as there may be many similar decision trees, masking the true results. For small data or low-dimensional data (data with few features), it may not produce good classification. Secondly, it is difficult to training data in the same time for GBDT due to dependencies between weak learners. However, partial parallelism can be achieved by self-sampling SGBT. If the data dimension is high, the computational complexity of the algorithm will increase. Thirdly, the space is too complexity to process presorting for XGboost. It not only needs to store the feature value, but also needs to store the index, which is equivalent to consuming twice. Finally, it should be pointed out that the time series prediction method does not consider the impact of external factors for the time being, so it has the defect of prediction error. When there is a big change in the outside world, there will be a large deviation. Short-term forecasts work better than long-term forecasts. On account of objective things, especially economic phenomena, are more likely to change in external factors in a relatively long period of time. It must have a significant impact on market economic phenomena. If when forecasting it happens, only the time factor is considered and the influence of external factors on the forecast object is not considered, and the forecast result will be seriously inconsistent with the actual situation.

In machine learning vision, natural language processing and other aspects, due to the good performance of deep neural network, scholars began to use it in sales forecasting, e.g., Wave Net [19], Long short-term Memory [20] and others. Compared with traditional prediction methods, deep neural network has better prediction performance, but the accuracy of its prediction results is poor. Therefore, the value of the predicted results is limited. In contrast, some scholars have shown in their studies that random forest (RF) has a better performance in good accuracy and appropriate computing time [21]. In addition, the gradient boosting decision tree (GBDT) algorithm using iterative and gradient boosting ideas shows better performance and stability, it's than general models in production and service demand forecasting [22]. On this basis, the extreme gradient boosting (XGBoost) algorithm has excellent accuracy in industrial practice, and its good accuracy has been verified in sales forecasting research. It is believed that we will be able to resolve distractions from the time and volume by combining and innovating models continually. Eventually, we will come up with more accurate sales forecasts models.

6 CONCLUSION

In summary, to increase sales of the company, this paper investigates sales prediction based on machine learning approaches. Specifically, the background material of machine learning was summarized and how it can be applied in the field of sales forecasting was examined. Then, some methods commonly used in machine learning and the importance of sales forecasting were summarized. Finally, this paper illustrate how machine learning can be applied to sales forecasting with an example, as well as illustrate its advantages and disadvantages. Owing to the limitations of traditional machine learning methods, one cannot accurately predict sales over long periods of time. By improving the method or combining multiple models, this problem will be resolved. In the future, we will have the ability to forecast sales over long periods of time, as well as sales with complex changes. Overall, these results offer a guideline for sales prediction in machine learning and pave a path for long-term sales forecasting.

REFERENCES

[1] Y. Xue, et al. "A review on regional dynamical downscaling in intraseasonal to seasonal simulation/prediction and major factors that affect downscaling ability." Atmospheric research, vol. 147, 2014, pp. 68-85.

[2] P. A. Castillo, et al., "Applying computational intelligence methods for predicting the sales of newly published books in a real editorial business management environment," Knowledge-Based Systems, vol. 115, 2017, pp. 133-151.

[3] Dann, Christoph, et al, "Beyond value-function gaps: Improved instance-dependent regret bounds for episodic reinforcement learning." Advances in Neural Information Processing Systems, vol. 34, 2021.

[4] A. Touati, and Y. Ollivier., "Learning one representation to optimize all rewards," Advances in Neural Information Processing Systems, vol. 34, 2021.

[5] N. Whiteley, A. Gray, and P. R. Delanchy, "Matrix factorisation and the interpretation of geodesic distance," Advances in Neural Information Processing Systems, vol. 34, 2021.

[6] J. Gu, et al. "Unidoc: Unified pretraining framework for document understanding," Advances in Neural Information Processing Systems vol. 34, 2021, pp. 39-50.

[7] B. Xie, et al. "Finding Discriminative Filters for Specific Degradations in Blind Super-Resolution." Advances in Neural Information Processing Systems, vol. 34, 2021.

[8] D. Slack, et al, "Counterfactual explanations can be manipulated." Advances in Neural Information Processing Systems vol. 34, 2021.

[9] H. Zhang, et al. "From canonical correlation analysis to self-supervised graph neural networks." Advances in Neural Information Processing Systems vol. 34, 2021.

[10] Z. Luo, Z. Tang, H. Sang, and B. Mallick, "BAST: Bayesian Additive Regression Spanning Trees for Complex Constrained Domain." Advances in Neural Information Processing Systems vol. 34, 2021.

[11] R. M. Bell, and Y. Koren, "Lessons from the netflix prize challenge." Acm Sigkdd Explorations Newsletter vol. 9(2), 2007, pp. 75-79.

[12] A. Berg, J. Deng, and L. Fei-Fei. Large scale visual recognition challenge 2010. Retrieving from www.imagenet.org/challenges.

[13] L. Breiman, “Random forests,” Machine learning, vol. 45(1), 2001, pp. 5–32.

[14] D. Cire, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification." 2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012.

[15] Y. Lean, et al. "Time series forecasting with multiple candidate models: selecting or combining?," Journal of Systems Science and Complexity vol. 18(1), 2005, pp. 1-18.

[16] Y. Han, X. Gao, “E-commerce commodity sales forecast based on machine learning combination model.” Computer system applications, vol. 31(01), 2021, pp. 315-321.

[17] T. Y. Weng, W. Y. Liu, J. Xiao. “Supply chain sales forecasting based on light GBM and LSTM combination model.” Industrial Management & Data Systems, vol. 120(2), 2019, pp. 265–279.

[18] T. G. Dietterich, “Ensemble methods in machine learning.” Proceedings of 1st International Workshop on Multiple Classifier Systems. Cagliari: Springer, 2000, pp. 1–15.

[19] D. M. Duan. “Research on hotel online sales forecast model based on improved Wave Net,” Journal of Physics: Conference Series, vol. 1544, 2020, 012067.

[20] K. Bandara, P. B. Shi, C. Bergmeir, et al., “Sales demand forecast in E-commerce using a long short-term memory neural network methodology.” Proceedings of the 26th International Conference on Neural Information Processing. Sydney: Springer, 2019, pp. 462–474.

[21] L. Turgeman, J. H. May, R. Sciulli. “Insights from a machine learning model for predicting the hospital Length of Stay (LOS) at the time of admission.” Expert Systems with Applications, vol. 78, 2017, pp. 376–385.

[22] Z. Jiao, H. Jin, B. Liu, “Short-term demand forecast of shared bicycles driven by big data——Comparative analysis based on machine learning model.” Business Economics and Management, vol. 35(8), 2018, pp. 16–25.