M-GAN-XGBOOST model for sales prediction and precision marketing strategy making of each product in online stores

Song Wang and Yang Yang
College of Economics and Management, Shandong University of Science and Technology, Qingdao, China

Received 20 November 2020
Revised 23 March 2021
Accepted 20 April 2021

Abstract

Purpose – The rapid development of e-commerce has brought not only great convenience to people but a great challenge to online stores. Phenomenon such as out of stock and slow sales has been common in recent years. These issues can be managed only when the occurrence of the sales volume is predicted in advance, and sufficient warnings can be executed in time. Thus, keeping in mind the importance of the sales prediction system, the purpose of this paper is to propose an effective sales prediction model and make digital marketing strategies with the machine learning model.

Design/methodology/approach – Based on the consumer purchasing behavior decision theory, we discuss the factors affecting product sales, including external factors, consumer perception, consumer potential purchase behavior and consumer traffic. Then we propose a sales prediction model, M-GNA-XGBOOST, using the time-series prediction that ensures the effective prediction of sales about each product in a short time on online stores based on the sales data in the previous term or month or year. The proposed M-GNA-XGBOOST model serves as an adaptive prediction model, for which the instant factors and the sales data of the previous period are the input, and the optimal computation is based on the proposed methodology. The adaptive prediction using the proposed model is developed based on the LSTM (Long Short-Term Memory), GAN (Generative Adversarial Networks) and XGBOOST (eXtreme Gradient Boosting). The model inherits the advantages among the algorithms with better accuracy and forecasts the sales of each product in the store with instant data characteristics for the first time.

Findings – The analysis using Jingdong dataset proves the effectiveness of the proposed prediction method. The effectiveness of the proposed method is enhanced and the accuracy that instant data as input is found to be better compared with the model that lagged data as input. The root means squared error and mean absolute error of the proposed model are found to be around 11.9 and 8.23. According to the sales prediction of each product, the resource can be arranged in advance, and the marketing strategy of product positioning, product display optimization, inventory management and product promotion is designed for online stores.

Originality/value – The paper proposes and implements a new model, M-GNA-XGBOOST, to predict sales of each product for online stores. Our work provides reference and enlightenment for the establishment of accurate sales-based digital marketing strategies for online stores.

Keywords Neural network, Deep learning, Sales forecast, Precision marketing strategy, XGBOOST, Integration model

Paper type Research paper

1. Introduction

With the rapid development and widespread use of digital technology worldwide, digital marketing has become an important part of online store development strategies in the new era and the importance is underscored in several recent marketing trends (Herhausen *et al.*, 2020). Also, digital marketing creates, communicates and delivers value for the target market. A successful marketing strategy can precisely target consumers, satisfy the market demand and reduce the cost of communication between companies and consumers, so that the competitiveness and economic efficiency of online stores are improved (Kolkova and Andrea, 2020). However, the formulation of current marketing strategies relies more on

Data Technologies and Applications
Vol. 55 No. 5, 2021
pp. 749-770

2514-9288
DOI 10.1108/DTA-11-2020-0286

the experience and methodology of salespeople. In a dynamically changing market environment, there is a large imbalance in the level of marketing. There are many studies devoted to mine sales-based digital capabilities, such as customer relationship management in post-sales service (Agnihotri *et al.*, 2017) and digital sales resources deployment (Lopez-Fernandez and Perrigot, 2018) in the digital marketing. The core of sales forecasting is to estimate and judge the unknown market. Based on the forecasted product sales, merchants can accurately grasp the direction of the market, make timely response and adjust the strategies and then continue to expand product sales and seize market share. Low customer satisfaction, high inventory and high operating costs when forecasts vary significantly (Teresa *et al.*, 2006). The excessive sales can cause the short supply and the small sales can reduce revenue for the online stores, finally, it may result in loss of reputation and income substantially. In 2019, online sales of products and services were worth 10.6 trillion yuan in China (China Internet Network Information Center, 2020). And there are 23 million online stores for Taobao and 8 million stores for Jingdong. However, online stores have made marketing decisions through prediction of sales by sales teams, which has awful prediction accuracy. Moreover, 90% of these stores have different degrees of out-of-stock, slow sales and other problems (Cui *et al.*, 2020). To minimize or avoid these problems, online stores have an interest in developing a model for predicting sales and understanding the underlying factors associated with it. From an operational perspective, if stores know the sales volume of all kinds of products in the coming period, the uncertainty of markets will diminish in the future and the formulation of sales-based marketing strategies such as stock replenishment, optimal allocation of store resources and product promotion will become more scientific and reasonable. Furthermore, stores can achieve a comprehensive improvement in the level of management and management efficiency. Financially, sales volume helps provide an estimate of the cost due to the resources such as staff and stuff so that online stores can prepare in advance and offer high-quality customized service. Thus, accurate sales prediction plays a beneficial role in designing the marketing strategies, planning the inventory and making the logistics decisions, which is valuable in taking actions to make profits.

Modeling the sales prediction is a tedious and complex process as the sales are affected by price, promotions, evaluations and descriptions online, product life cycle, season, ranking online, etc. Moreover, sales volume is a dramatic fluctuation, in a specific period it shows a linear trend of increase or decrease, while certain phases may show the characteristics of nonlinear fluctuation because of various potential uncertainties. Besides, there is a large customer purchase behavior data information, which reflects the process of consumer purchase and the feedback of each product. Therefore, it is one of the critical research hotspots to mine, analyze and apply the data information to build a prediction method that is suitable for the mixed characteristics of both the linear and nonlinear changes and to employ these characteristics to predict sales (Pan and Zhou, 2020).

In general, sales prediction techniques are classified into two types, such as empirical methods and algorithms models (Razeef *et al.*, 2020). The empirical method depends on the analysis of past historical data of sales and the experience of salesmen, which has a low prediction accuracy. There are a lot of algorithms and models are employed to predict sales such as logistic regression, decision tree, random forest, gradient ascending decision tree, neural network and so on. Each model has its advantages in predicting sales, but it is tough to achieve accurate prediction of sales.

Accurate forecasting of sales has been one of the most important issues in marketing research because early warnings of severe sales can help prevent damages and make profits, if timely and accurately forecasted. This paper proposes a data-driven model to predict sales volume in the future and design sales-based marketing strategy for stores. The model is built using a detailed dataset from an online store in Jingdong. Our contribution can be summarized as follows. We put instantaneous data of factors as input and employ combined machine learning methods to build and test predictive models for sales using a real dataset that is comprehensive, which achieves excellent prediction accuracy. We find that the model utilizing instantaneous data as input is stronger compared to putting history data as input. It shows robust performance with similar prediction accuracy both in the training and test data. Based on the predictive results of consumer traffic and sales volume, we explore digital marketing for product positioning strategy, product display optimization strategy, zero inventory management strategy and product promotion strategy in our study.

1.1 The major contribution of the paper

M-GAN-XGBOOST model for adaptive prediction of the sales. We propose an M-GAN-XGBOOST model for effective sales prediction, which integrates the memory of time series data, the accuracy of fitting data and the robustness of prediction. Besides, the model we proposed forecasts the sales of each product in the store with instantaneous data characteristics for the first time. According to the sales prediction of each product, the marketing strategy of product positioning, product display optimization, zero inventory management and product promotion is designed for the online store, which provides reference and enlightenment for the establishment of accurate digital marketing strategies for online stores.

The remainder of the paper is organized as follows. In Section 2 we review the relevant literature. Section 3 discusses the selection of input variables that influence sales volume based on the consumer purchase behavior decision theory. In Section 4 we build M-GAN-XGBOOST models. Section 5 achieves the model by using real data. We explore and design the marketing strategy according to the predicting results in Section 6, and discuss key takeaways in our study and future research direction in Section 7.

2. Review of literature

In this section, existing work related to the predicting methods of the sales volume for online stores is discussed, which mainly include mathematical statistics, machine learning and combined method. Many existing research studies focused on the prediction of gross sales and the feature of lag phase as input, but none of the studies has considered the prediction of each product on the online store and the influence of consumer behavior on product sales.

As a typical regression problem, logistic regression is a generalized linear regression model, which employs logical functions to predict classification problems widely based on linear regression (Yao and Wang, 2019). Since the logistic regression algorithm is good at explaining dichotomy problems and can better fit the functional relationship between independent variables and dependent variables (Vijaya and Sivasankar, 2019; Stripling *et al.*, 2018). Frank *et al.* (2004) utilized the multivariate fuzzy logistic regression model on sales forecasts so that the forecasting implementation is simple and easy. Fildes *et al.* (2016) put the promotion information and operational information through multi-stage LASSO regression (Least absolute shrinkage and selection operator) for model estimation, variable selection and prediction using a rolling scheme. Taylor (2004) employed exponential smoothing to predict short-term sales, with a high weight for instant data and decreasing in history data. Ramos *et al.* (2015) used the state-space model and ARIMA (Autoregressive Integrated Moving Average Model) to forecast market demand and found that the forecasting accuracy of the two methods was similar. Soopramanien *et al.* (2014) took competitive information into account in sales forecasting based on the autoregressive distributed lag model for prediction. Celia *et al.* (2003) developed three models for predicting total sales respectively, including exponential smoothing, ARIMA and multiple regression models, which showed that ARIMA had the best effect. Di Pillo *et al.* (2016) used SVM (Support Vector Machines) to predict e-commerce product sales. OEzden *et al.* (2009) found that regression trees were closer to the true value than linear regression models when there was a promotional impact. Sun *et al.* (2006) used Bayesian networks to make predictions on time series data. Prateek *et al.* (2013) used FTS (Fuzzy Time Series) for auto sales data to analyze performance and forecast sales. Bendato *et al.* (2015) used Monte Carlo simulation to predict the dynamics of information on sales volume, and the increase in the tolerance interval led to great progress in forecast accuracy. Fan *et al.* (2017) regarded consumer review information as one of the input features of the Bass model to forecast sales. Besides, the basic idea of the decision tree is that the classification rules of the representation form of the decision tree can be deduced from a bunch of random and unordered instances according to some criteria in a top-down recursive way (Kannadath *et al.*, 2018; Bell and Mgbemena, 2018; Sivasankar and Vijaya, 2019). Compared with other prediction methods, it can clearly and intuitively show logical classification. Random forest refers to the establishment of a forest with unrelated decision trees in a random way (Wager and Athey, 2018). After the forest is built, when a new sample is an input, all the decision trees in the forest are asked to judge the category of the sample, and which category is the most selected, the predicted result will be the same (Mau *et al.*, 2018; Mahdavinejad *et al.*, 2018).

Deep learning, as an emerging field of data mining in recent years, recognizes data by simulating the multilayer perceptual structure of the human brain. Neural networks embody excellent performance in processing unstructured data with a high fit. Choi *et al.* (2013) provided an overview of current analysis and prediction methods, pointing out that neural networks for prediction have become a hotspot for scholars with the development of artificial intelligence. For example, RNN (Recursive Neural Network) was specifically employed to process sequence-type data and has been widely used to predict time-series data (Manning *et al.*, 2011). As RNNs were unable to solve the long-term dependency problem, the Long Short-Term Memory (LSTM) model and a simplified version of the Gated Recurrent Unit (GRU) have been proposed (Hochreiter and Schmidhuber, 1997; Chung *et al.*, 2015). The convolutional neural network is an intelligent information processing technology that mimics the information processing process of the human brain with a grid-like structure to feature extraction in prediction problems (Krizhevsky *et al.*, 2017; Hanson *et al.*, 2018). It has the characteristics of self-organization, self-adaptation and self-learning (Liberis *et al.*, 2018; Pham and Le, 2018; Qiu *et al.*, 2019; Khaled *et al.*, 2019; Kuzovkin *et al.*, 2019). Generative Adversarial Networks (GAN) are based on a two-person zero-sum game and are used to estimate the potential distribution of data samples and generate new ones (Goodfellow *et al.*, 2014). Neural networks are highly adaptive and explain the influence of many factors on sales.

In addition to the use of neural networks alone, it has been found that the use of multiple learner integration methods can improve the validity of prediction results. Like the random forest, the gradient boosting decision tree (GBDT) is also a combinatorial model based on the decision tree (Rao *et al.*, 2019). Its idea is to build a decision tree each time in the direction where the loss function of the existing model decreases (Wang *et al.*, 2017). Guestrin and Chen (2016) described the principles and applications of XGBOOST (Extreme Gradient Boosting), a variant of GDBT (Gradient Boosting Decision Tree), which introduces L1 and L2 regularized logistic and linear regressions into loss function, making model overfit difficult (Friedman, 2001). Li *et al.* (2016) employed the BP neural network model to predict the total retail sales of consumer goods. Ryutaro *et al.* (2017) used the LSTM network to predict the sales of a Japanese supermarket chain. Guadagni *et al.* (2017) compared the algorithms of neural networks, SVM, random forest, XGBOOST, etc. The comparison shows that XGBOOST has a higher prediction accuracy (Taylor, 2004; Ramos *et al.*, 2015; Soopramanien *et al.*, 2014).

Both linear and nonlinear models mentioned above have their advantages. However, the effect of using a single model for forecasting is bound to have some drawbacks. Scholars gradually employed hybrid models for forecasting to further improve the accuracy of the models.

Choi *et al.* (2011) proposed a combined ARIMA and wavelet transform prediction model considering the seasonality and randomness of sales. Eskenazi and Fei (2013) used a combination of stable seasonal models and SVM to predict sales given the seasonal variation characteristics of e-commerce sales data. Trofimov *et al.* (2012) used multiple decision trees for prediction, and the prediction effect was greatly improved. Huang *et al.* (2014) proposed a model that first learned data features with a DBN (Deep Belief Network) and then predicted them through a regressor. Doganis *et al.* (2006) proposed a network structure that fused genetic algorithms and RBF (Radial Basis Function) for the marketing of fresh milk prediction. Wedding and Cios (1996) jointly predicted with RBF and ARIMA. Wang *et al.* (2016) propose the FNN (Factorization Machine Supported Neural Network) model, which first obtained the implicit vector of each feature dominated by a factorization machine and then employed it as input to a deep neural network to make predictions. Armano *et al.* (2005) constructed a model combining neural networks and genetic algorithms. Wu and Tan (2016) used CNN to find the spatial features of the data and then utilized two LSTMs to make predictions.

Together, these studies indicate that the prediction accuracy of models largely depends on the characteristics of input data. Many factors affect merchandise sales strongly. The consumer behavior data provided to merchants by platforms such as JD.com and Taobao are on the next day, which contribute to the unavailability of instantaneous data. On the other hand, it has not happened at the current moment and for a period of time in the future, so that there is no instantaneous consumer behavioral data such as sales and browsing number corresponding to the day generated, and it cannot be obtained from the existing data of the online stores. Therefore, lag period data were mostly employed as input variables of the model, which often ignored the importance of instant data in sales forecast so that caused the reduction of prediction accuracy. To solve this problem and improve the accuracy of short-term sales prediction effectively, this paper analyzes the influencing factors and characteristics of sales based on the online purchase decision-making processes and builds the M-GAN-XGBOOST short-term sales prediction model based on deep learning, with the instantaneous data of influencing factor as input variables to predict the short-term sales of all kinds of online store products at once. The strategy of product positioning, product display optimization, zero inventory management, product promotion and other precision marketing strategies are made according to sales prediction for online stores.

3. The selection of input variable

In this section, our focus is to identify the variables that can predict sales volume. According to the online purchase decision-making processes and Self-Determination Theory (SDT), we discuss each of these predictor variables and the justification for utilizing them in the prediction models (Karimi *et al.*, 2018; Kottke and Mellor, 1986).

After confirming a shopping demand, consumers make a purchase decision through relevant information collection, browsing and comparison, which is a process where consumers use subjective cognition to systematically reorganize relevant information. Besides, relevant behaviors are conducted after a series of mental activities such as weighing and perception that consumers carry out. SDT suggests that behavior is driven by both extrinsic and intrinsic motivation. Intrinsic motivation refers to the self-interest, needs and other factors that individuals rely on when making decisions, while extrinsic motivation is the way in which individuals are driven to behave in an external environment, triggered by the environmental factors in which the activity takes place, such as the evaluation of others. This paper argues that consumer purchase behavior is influenced by both intrinsic and extrinsic motivations. Four dimensions are divided about the factors influencing merchandise sales, including extrinsic factors, consumer traffic, consumer perception and potential consumer purchase behavior.

External factors are external stimuli that consumers receive, and they are not determined by the consumer or the online stores (Steinker *et al.*, 2017). They can influence the purchase behavior of external characteristics, mainly including weather and time factors, which have an impact on sales at the time of occurrence.

Consumer traffic is the changes in relevant variables caused by consumers visiting the relevant pages of the store over a period of time (Steinker *et al.*, 2017; Akram *et al.*, 2018), which is usually expressed in terms of the number of keywords searched, the number of views and other indicators. Besides, consumer traffic is distinctly time-sensitive, with significant differences in traffic on different dates, such as holidays. The larger the consumer traffic in the current period, the more possibility that consumers make a purchase it is in the current period. So the instantaneous consumer traffic data have a strong impact on product sales.

Consumer perception refers to the relevant information obtained by consumers from various approaches according to their demand (Yuan *et al.*, 2019), which measures include the number of reviews, positive ratings and merchant reputation of each product. When consumers refer to such information, the time frame is all moments prior to the current time point. So consumer perception has an influence on sales in the form of cumulative values.

Consumer potential purchase behavior refers to consumers' hesitant purchase behavior based on perceived relevant information (Park and Sang-June, 2016), which is often expressed by indicators such as favorites and add-to-cart. Products in shopping carts and favorites may be converted into purchase orders at some time in the future. So the cumulative value of the current potential purchase behavior for all consumers is an important reference for the prediction of sales in online stores.

After the analysis of each influencing factors on sales in consumer decision process, we find that the instant factors are important for predicting instant sales. So this paper selects instant external factors, instant consumer traffic, instant cumulative consumer perception and instant potential purchase behavior as the input characteristics, and sales as the output variables to build a model for online stores to predict sales. We list the predictor variables in Table 1 along with their definitions.

4. The algorithm of M-GAN-XGBOOST

This research work develops a sales prediction model that forecasts the sales volume of all products in online stores with the variables listed in Table 1. We would ideally like a parsimonious model that has good predictive power and can easily be implemented in practice.

Table 1.
The definitions of predictor variables

Name	Symbol	Meaning
External Factors	*ecause*	Consumers receive external stimuli that can affect purchasing behavior
Consumer Traffic	*Ctra*	Changes in related variables caused by consumers visiting pages of the store
Consumer Perception	*Cper*	Related product information obtained by consumers
Potential Purchase Behavior	*Cpot*	Consumer hesitation
Sales	*Y*	Merchandise sales

We assume the current time is t. External factors for known time characteristics can directly use the current date as the instantaneous data, i.e. $ecause_t$. Cumulative values are employed for consumer perception data and consumer potential purchase behavior data. With less new data generated each day on such characteristics, a lagged period of cumulative values can be applied as an approximate substitute for the immediate data in cases where the cumulative length of the data is large. Consumer perception data and consumer potential purchase behavior are expressed as $cp\widehat{er}_t$ and $cp\widehat{ot}_t$ respectively. The instantaneous consumer traffic has a large impact on consumers' immediate purchases. With highly time-sensitive, it varies widely from one period to the next. In addition, merchants do not have access to the instantaneous data of consumer traffic. Therefore, in this paper, a generator of M-GAN deep learning model is constructed with LSTM neural networks, considering that GAN (Generative Adversarial Networks) deep learning model has high fitting properties and unique sample generation capabilities, and LSTM (Long Short-Term Memory) neural network has the capable of learning long dependencies of time-series type data (Goodfellow *et al.*, 2014; Hochreiter and Schmidhuber, 1997). The instantaneous data of external factors $ecause_t$ and the historical data of consumer traffic $ctra_1$-$ctra_{t-1}$ are input for the M-GAN deep learning model. Then we can predict the imitation instantaneous data of consumer traffic $ct\widehat{ra}_t$ by training.

The inputs for sales forecasting contain multiple factors and online stores have multiple products. And each input has a different impact on sales. The XGBOOST as a set of tree-boosting extensible machine learning systems has high accuracy in the prediction problem performance (Guestrin and Chen, 2016). It can be used to satisfy the need for multiple inputs and outputs by constructing multiple tree models. Historical data of immediate external factors $ecause_t$, imitation immediate consumer traffic $ct\widehat{ra}_t$, imitation immediate consumer perception $cp\widehat{er}_t$, imitation immediate consumer potential purchase behavior $cp\widehat{ot}_t$ and product sales Y_1-Y_{t-1} are fed into the XGBOOST model to predict all product sales $\widehat{Y}_t$.

The above two step-by-step prediction models are tied together to form the M-GAN-XGBOOST deep learning model to predict the short-term sales for products in online stores, which operates based on the error found between the predicted output and the real data. Since the M-GAN-XGBOOST model makes full use of these deep learning models and integrates XGBOOST, GAN and LSTM neural networks according to the advantages of each model, the learning is more effective, generalize is better and long-time dependence is better than the other conventional neural networks. Based on the predicted products sales and predicted products consumer traffic values obtained from the distribution prediction model, an accurate marketing strategy for the online shop is developed.

4.1 The structure of M-GAN-XGBOOST

In the following, we define the relevant influencing factors as:

$$X_t = \{ecause_t, ctra_t, cper_{t-1}, cpot_{t-1} | t = 2, 3, \ldots, m\}. \tag{1}$$

We express the sales volume of products in online stores as:

$$Y = \{Y_t | t = 1, 2, 3, \ldots, m\}. \tag{2}$$

We represent the temporal length of sales data as m, and the number of products in the online store as n. Then we predict Y_t according to X_t. X_t contains multiple variables and online stores have multiple products. *step* denotes the predicted time step, $t = step+1, step+2, \ldots, m$. The mathematical model of the M-GAN-XGBOOST is given as:

$$\widehat{Y}_t = XGBOOST(ecause_t, M - GAN(ecause_t, ctra_{t-1}, ctra_{t-2}, \ldots, ctra_{t-step}), cper_{t-1}, cpot_{t-1}). \tag{3}$$

The structure of the model is shown in Figure 1.

The prediction of the sales volume is performed adaptively using the M-GAN-XGBOOST model. The weights are calculated by loss function between the predicted output and the real data. Initially, the data of consumer traffic is generated based on the time dependence of LSTM cells in *G* model. And a fully connected neural network is used to determine whether the generated consumer traffic data is consistent with the real consumer traffic data distribution in *D* model. Once the output of the *G* model is calculated, the error is estimated with respect to the real data. By minimizing the error, the instantaneous consumer traffic data is obtained by training. Finally, the predicted consumer traffic data is employed as input to predict the sales of each product for the online store in XGBOOST model. The training that assists with the minimum value of the error is employed for updating the weights of the M-GAN-XGBOOST model.

w_f, w_i, w_o, w_c denote the input connection weights of the oblivion gate, input gate, output gate and unit state respectively. b_f, b_i, b_o, b_c denote the offset of the oblivion gate, input gate, output gate and unit state respectively. $c_{t\text{-}step}$, $h_{t\text{-}step}$ denote the unit state transmitted from the previous moment to this moment, representing the long-term memory up to this moment. σ represent the sigmoid function. We assume that $ctra_{t\text{-}step+1}$ is the current input.

In M-GAN model, LSTM cells are employed to generate the instantaneous consumer traffic data, as shown as box *G* in Figure 1. Every blue square box in Figure 1 is represented a LSTM cell, which has three gates that can ensure the memory for the previous consumer traffic data in the model, and they are given as:

(1) Oblivion gate determines the number of United States for consumer traffic data from the previous moment to be preserved to the current moment. The oblivion gate of the model is given as:

$$f_{t-step+1} = \sigma(w_f \cdot [h_{t-step}, ctra_{t-step+1}] + b_f). \tag{4}$$

(2) The input gate determines the number of inputs saved to the current moment's unit state for consumer traffic data is given by:

$$i_{t-step+1} = \sigma(w_i \cdot [h_{t-step}, ctra_{t-step+1}] + b_i). \tag{5}$$

(3) The output gate controls the number of unit state outputs to the current output value for consumer traffic data at the current moment. We calculate the output gate as:

$$o_{t-step+1} = \sigma(w_o \cdot [h_{t-step}, ctra_{t-step+1}] + b_o). \tag{6}$$

Then the memory level of the current moment for consumer traffic data is represented as:

$$\tilde{c}_{t-step+1} = \sigma(w_c \cdot [h_{t-step}, ctra_{t-step+1}] + b_c). \tag{7}$$

Compared with the consumer traffic data for next moment, the long-term memory state on the current moment is given as:

$$c_{t-step+1} = f_{t-step+1} {}^* c_{t-step} + i_{t-step+1} {}^* \tilde{c}_{t-step+1}. \tag{8}$$

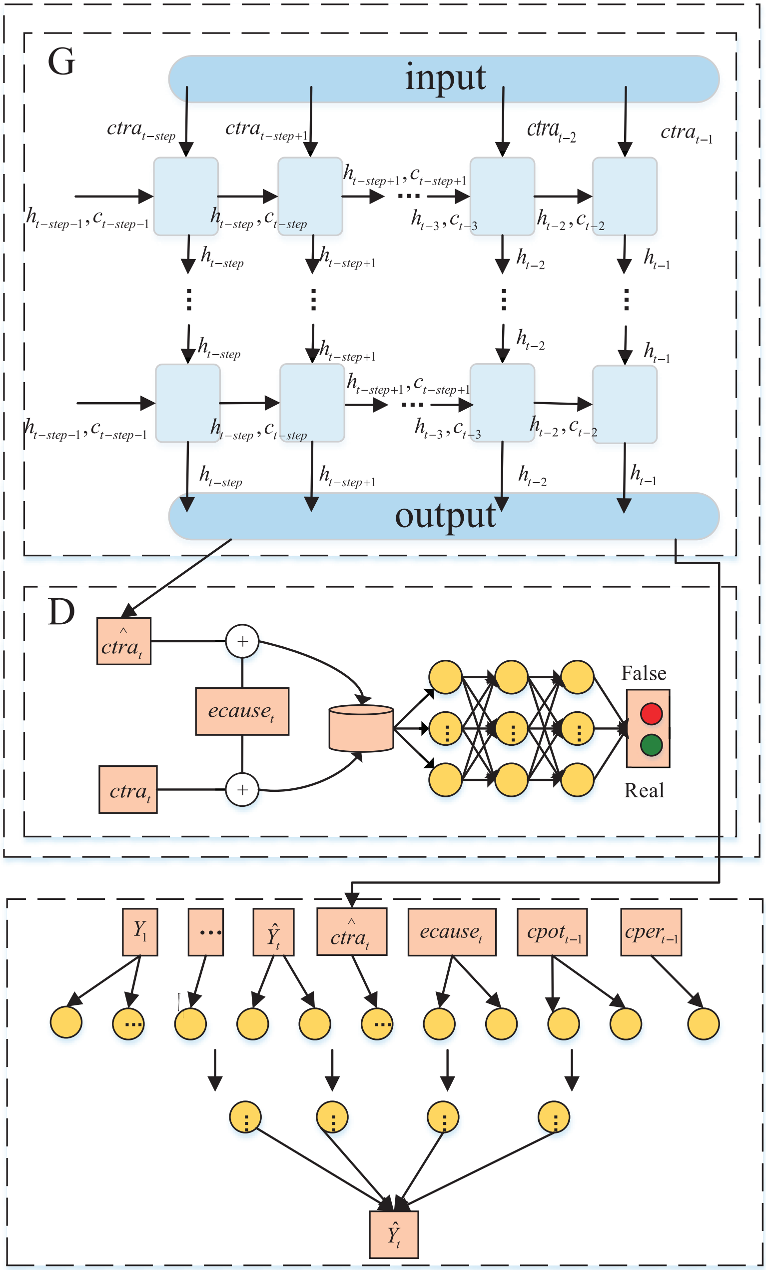

Figure 1.
The structure of M-GAN-XGBOOST model

The output of the consumer traffic data for current state to the next state is calculated as:

$$h_{t-step+1} = o_{t-step+1} \cdot \tanh(c_{t-step+1}). \tag{9}$$

$c_{t-step+1}$ and $h_{t-step+1}$ can be seen as right arrow in Figure 1. The G model is composed of many LSTM cells.

After the learning process described above, the G model can remember the *step* length of data and gives the value of the data at the next moment. The discriminant model D uses a fully connected neural network, which is shown as box D in Figure 1. The output is 1 if the input is real consumer traffic and 0 if the input is G-generated consumer traffic. The loss function is given as:

$$\min_G \max_D E[\log D(ecause_t, ct\widehat{ra}_t)] + E[\log(1 - D(G(ctra_{t-1\sim t-step})))]. \tag{10}$$

Through optimizing the loss function, $G(ctra_{t-1\sim t-step})$ is closer to $ctra_t$. The M-GAN model can be trained and the predicted value of instantaneous consumer traffic $ct\widehat{ra}_t$ can be obtained.

Then the instantaneous consumer traffic $ct\widehat{ra}_t$, instantaneous external factors $ecause_t$, cumulative consumer perception $cper_{t-1}$, cumulative potential purchase behavior $cpot_{t-1}$ and previous sales are input into XGBOOST model as feature data to predict the short-term sales of products in online stores, which is shown as bottom box in Figure 1. For XGBOOST model, we assume k prediction trees are constructed, the prediction function for sales is given as:

$$\widehat{Y}_t = \sum_{k=1}^{K} f_k(X_t), \tag{11}$$

f_k denotes a prediction function in function space, with a corresponding tree structure q and leaf weights ω. The final predicted product sales are the sum of the feature data on each prediction tree. Taken one prediction tree as example, the sales prediction value on this regression tree is the weight of the leaf node for the corresponding tree. The calculation formula is given as:

$$f_k(X_t) = \omega_{q(X_t)}, \tag{12}$$

where $q(X_t)$ represents the mapping between feature data and leaf nodes of this prediction tree.

The objective function of this prediction tree is calculated as given below:

$$l^{(k)} = \sum_{j=1}^{n} \sum_{i=1}^{m} l(Y_t, \widehat{Y}_t) + \Omega(f_k), \tag{13}$$

where the smoothing error is determined using the following formula:

$$l(Y_t, \widehat{Y}_t) = (Y_t - \widehat{Y}_t)^2, \tag{14}$$

and the regularized term is computed as:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2, \tag{15}$$

where T indicates the number of leaf nodes, γ denotes the number of control leaf nodes, λ denotes the fraction of control leaf nodes to prevent overfitting.

Starting from the root node of the input, the objective function is transformed into the structure of the prediction tree q and the leaf weights ω. As mentioned above, each prediction tree is trained through iterating according to the objective function. Then based on equations (11) and (12), we can get the sales of all items $\widehat{Y}_t$ on day t.

4.2 The evaluation criteria of model validity

To avoid possible assessment errors that may result from using one assessment method, this paper uses the root mean squared error and mean absolute error to evaluate the projections as:

$$RMSE = \sqrt{\frac{1}{n}(\widehat{Y}_t - Y_t)^2}, \tag{16}$$

$$MAE = \frac{1}{n}|\widehat{Y}_t - Y_t|, \tag{17}$$

According to the formula, the evaluation criterion is that the closer $\widehat{Y}_t$ and Y_t, the smaller the value of the corresponding sum is. It can indicate the higher the prediction accuracy of the model.

5. Numerical experiments and comparisons

In this section, we obtain the consumer data generated during the decision-making process of online shopping behavior on an online store of Jingdong Mall, including user number, product number, the time when the behavior occurs, browsing, ordering, concerning, commenting, adding the shopping cart and other behavioral data. There are 12068 dataset from 2018-2-1 to 2018-4-15. Based on the definition of the input variables for each dimension and the data obtained, the appropriate variables in the dataset are selected to represent the characteristics of each dimension. So, the day of the week as the time factor may imply the habits of consumer purchases, which can express external factors. The product's instantaneous browsing as the changes caused by consumers visiting the relevant pages can represent consumer traffic. The cumulative number of reviews are the important reference for consumer purchases, which can represent consumer perception. The cumulative number of add-carts and cumulative collections indicate the hesitation in making a purchase, which can represent potential purchase behavior. The dataset is used to implement the M-GAN-XGBOOST sales prediction model and verify the validity of the model. The model is implemented on the Pycharm and Tensorflow. The effect of different model parameters is tuned by comparing them.

5.1 Data preprocessing

For selected inputs and outputs, the data is pre-processed by eliminating outliers such as large promotions and supplementing the missing values through the insertion method. Statistics are presented in days, with sales volume as a label and time characteristics of weeks as numbers 1, 2, 3, 4, 5, 6 and 7. All input data are normalized by the normalization method Min–Max, which can eliminate order-of-magnitude differences in individual input data based on the following equation. The screenshots of the unprocessed and processed dataset can be seen in Figure 2.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}, \tag{18}$$

	user_id	sku_id	action_time	module_id	type
0	553413	0	2018-03-25 02:10:33	1323206	1
1	553413	0	2018-03-25 06:05:30	6310056	1
2	940229	0	2018-02-17 14:47:05	4896356	1
3	940229	0	2018-02-21 09:55:52	3724362	1
4	354207	0	2018-03-30 10:15:48	1800384	1
5	1593311	0	2018-03-30 14:10:15	6323238	1
6	1004652	0	2018-04-02 18:21:54	9132084	1
7	271980	0	2018-03-19 17:57:02	2532023	1
8	387091	0	2018-04-14 22:59:05	7301171	1
9	103676	0	2018-04-14 00:56:17	2422849	1

	date	time	week	weekend	0sku	1sku	2sku	3sku	4sku	5sku	6sku	7sku	8sku	9sku
0	2018-02-01	0.000000	0.500000	0	0.067227	0.336842	0.119303	0.120690	0.530973	0.226994	0.516746	0.268595	0.112554	0.261780
1	2018-02-02	0.013889	0.666667	0	0.058824	0.073684	0.120643	0.149425	0.265487	0.245399	0.669856	0.361570	0.158009	0.172775
2	2018-02-03	0.027778	0.833333	1	0.042017	0.242105	0.067024	0.201149	0.247788	0.061350	0.622010	0.431818	0.149351	0.303665
3	2018-02-04	0.041667	1.000000	1	0.109244	0.252632	1.000000	0.132184	0.265487	0.208589	0.153110	0.411157	0.162338	0.272251
4	2018-02-05	0.055556	0.000000	0	0.134454	0.494737	0.423592	0.155172	0.566372	0.656442	0.339713	0.417355	0.160173	0.282723
5	2018-02-06	0.069444	0.166667	0	0.176471	0.547368	0.193029	0.172414	0.566372	0.398773	0.354067	0.398760	0.203463	0.319372
6	2018-02-07	0.083333	0.333333	0	0.100840	0.410526	0.164879	0.103448	0.672566	0.220859	0.392344	0.427686	0.190476	0.387435
7	2018-02-08	0.097222	0.500000	0	0.075630	0.326316	0.136729	0.091954	0.539823	0.306748	0.282297	0.400826	0.207792	0.267016
8	2018-02-09	0.111111	0.666667	0	0.100840	0.357895	0.130027	0.155172	0.486726	0.165644	0.373206	0.311983	0.190476	0.298429
9	2018-02-10	0.125000	0.833333	1	0.092437	0.431579	0.360590	0.160920	0.256637	0.245399	0.339713	0.283058	0.335498	0.256545

Figure 2.
Unprocessed and processed dataset

5.2 Prediction results and discussion of M-GAN-XGBOOST

During the experiment, GAN was used to fit the consumer traffic data to the real sample. And then the traffic data was combined with the remaining variables as inputs to predict the sales of each variety of goods by using XGBOOST.

Take the time step of determining the *G* model as an example, we can determine the appropriate value of parameters in the generating and discriminating models by observing and comparing the changes in each training session. Through training, the prediction data is as close as possible to the sample training data of the model with the value of parameters. Figure 3 shows the mean-squared error of the GAN model when fitting consumer traffic data with a different value. It can be seen that as the number of iterations increases, the mean-squared error first increases and then decreases, and finally becomes stable. The training

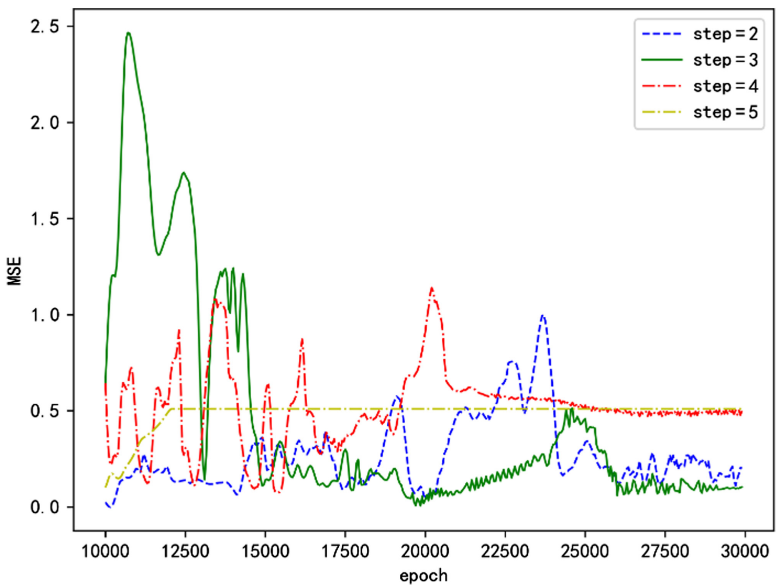

Figure 3.
The change of MSE on each *step*

process of GAN is unstable. From the final smooth results, when the *step* is 2, 3, 4 or 5, the final MSE is 0.20, 0.10, 0.50 and 0.51 respectively for training dataset. Therefore, the model has the least mean square error and the best performance when the *step* is 3. From an overall perspective, *step* is the length of memory in GAN that generates the model training is the past data, the larger it is, the greater the dependence of the current consumer traffic on the lagging consumer traffic. Considering *step* = *3*, it is also shown that consumer traffic varies with a period of 3 days. Therefore, 3 will be chosen for the *step* of the generative model while training the consumer traffic in this paper.

The instability of GAN increases the difficulty of training to some extent. Due to the limitation of hardware equipment, the training time of the model is long. In the existing studies, three improved models such as WGAN, LSGAN and BEGAN exist to alleviate this problem (Martin *et al.*, 2017; Mao *et al.*, 2017; David *et al.*, 2017). In this paper, different loss functions of these three models are used to train the data and make predictions. The one with the smallest error is employed as the GAN model structure for predicting sales in the online store. The variation on the different loss functions is shown in Figure 4. It can be seen that the mean square error of BEGAN is 0.02, when the WGAN is 0.20 and the LSGAN is 0.15, so that the BEGAN is the smallest among the three models. It is the smoothest throughout the training process. The BEGAN declines linearly between 10,000 and 12,500 repetitions and then stabilizes, whereas the WGAN and LSGAN stabilize at 30,000 repetitions. It fluctuates tremendously during the training process. As a result, BEGAN outperforms WGAN and LSGAN on the training set.

To avoid overfitting to the training dataset for the model, it is also necessary to compare the prediction results of the models on the test dataset using the same criterion in the same dataset case. Thus we can compare the suitability of the sample data to the WGAN, LSGAN and BEGAN data better. The comparison of the prediction results of each model on the test set is shown in Table 2. The error of BEGAN on the test set is much smaller than that of WGAN and LSGAN, which is 50% better than both models. Combined with the error of each model in fitting the data distribution on the training dataset, BEGAN is stronger in terms of

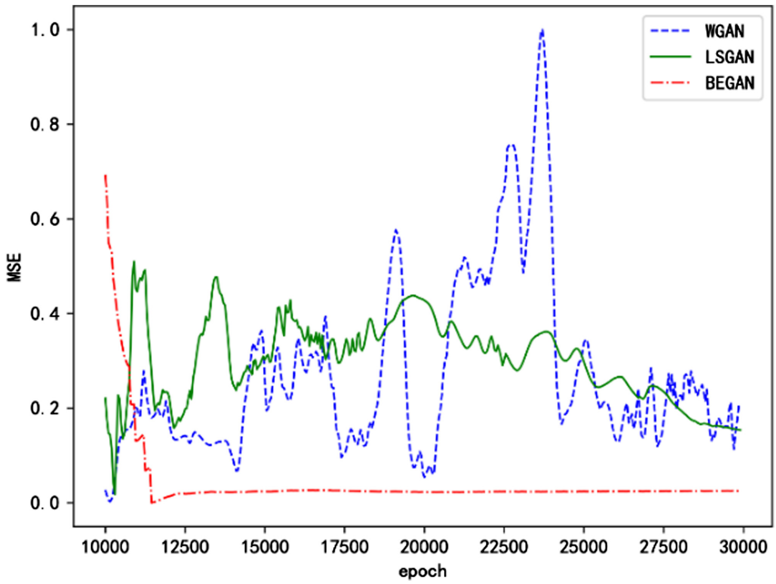

Figure 4.
The change of MSE on each model

training times, prediction accuracy, smoothness and robustness compared to WGAN and LSGAN. According to Figure 4, BEGAN has stabilized at 15,000 training times. Therefore, the BEGAN structure is used as the GAN model in this paper and training 15,000 times.

The fit of the BEGAN model to the consumer traffic data is shown in Figure 5, with the blue line being the real product sales and the red line being the product sales predicted using the model. It is seen that although the model has a poor fit to the maximum and minimum points in the data, it can effectively track and fit the changes in consumer traffic on the whole, which verifies the reliability and validity of the BEGAN model.

According to prediction results of consumer traffic on BEGAN, the remaining data features are combined as input for sales prediction. The choice among the parameters of the XGBOOST model is measured and adjusted by the equal error criterion of sales. The training was performed with the XGBOOST package in Pycharm, and the parameters and meanings used for the final XGBOOST model are shown in Table 3.

5.3 The comparison between instant data and lagged data

We use instant input features in the sales forecast model, while some studies have used lagged data. To further verify the influence of the input features of the instantaneous and lag period on the prediction accuracy of the model, and to test the prediction performance of the

Table 2.
The error of each model

Index	WGAN	LSGAN	BEGAN
MAPE	32.96%	34.07%	*12.07%*
MAE	28.53	27.57	*11.03*
MSE	35.3	35.98	*15.33*

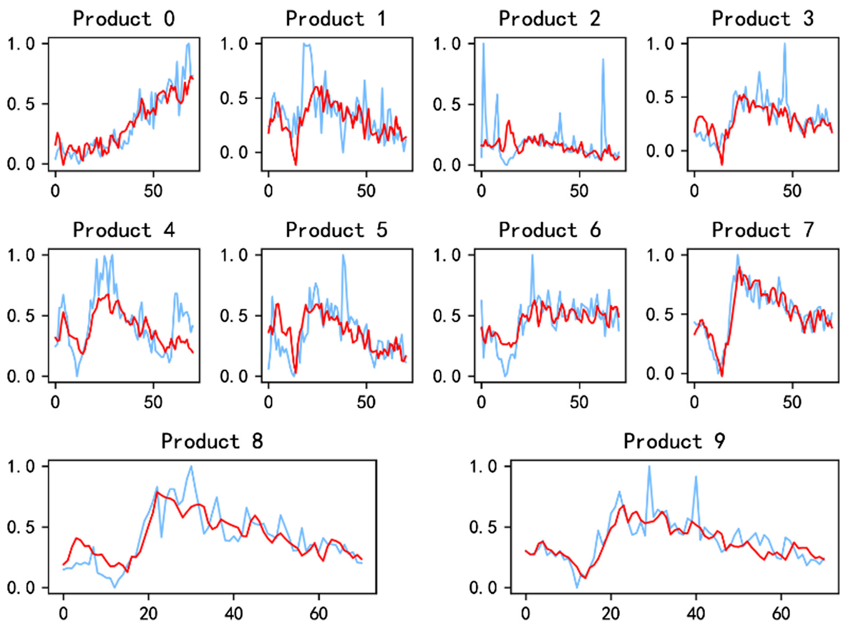

Figure 5.
The data-fitting of BEGAN

Name	Setting	Meaning
learning_rate	0.0001	Learning rate, which can control the step size of each weight when updating
n_estimators	50000	Training times
max_depth	9	The maximum depth of the tree
min_child_weight	1	The smallest sum of sample weights in child nodes
gamma	0.0001	The value of the minimum loss function
subsample	0.7	The proportion of the sub-sample of the training model to the entire sample
colsample_bytree	0.5	The proportion of random sampling of features when building the tree
eta	0.1	Shrink the step size, reduce the feature weight in the promotion calculation
silent	1	Print runtime information
eval_metric	*rmse*	The evaluation index of the verification data, the root mean square error
objective	*reg:linear*	Objective function, linear regression problem

Table 3. The parameters and setting of XGBOOST model

model, a comparison is made with the model using lag period features. In the same model structure, we use the lag period feature data as input to predict the sales volume of the product. The predictions using the instant feature data as input variables and those using the lag feature data as input variables are compared with the true values, to obtain their respective errors. The comparison results are shown in Table 4. Y_t denotes the true value, $\widehat{Y}_t$ denotes the predicted value of this model and $Y_t^{'}$ denotes the predicted value of the comparison model. The empirical comparison study shows that on the test set, the model using the prediction of instant data is smaller than the prediction model using the lag data for *MAE* and *RMSE*. For the lagged products, the model cannot accurately predict the sales of product 0. In summary, the M-GAN-XGBOOST model can predict the sales of each product in the short term of the online store. It can improve the prediction accuracy and enhance the robustness of instant sales.

6. Results and discussion of sales-based precision marketing strategy

It is the key point of precision marketing that how to effectively analyze data and use the results to help them make marketing decisions and guide their fine operations for businesses. Through the M-GAN-XGBOOST model proposed in this paper, businesses can know the number of user views and sales volume of each product in advance. The *MSE* of the forecast results is 11.03 and 8.2 for the comparison model with lagged period data and M-GAN-

Date		2018-4-13			2018-4-14			2018-4-15	
	Y_t	$\widehat{Y}_t$	$Y_t^{'}$	Y_t	$\widehat{Y}_t$	$Y_t^{'}$	Y_t	$\widehat{Y}_t$	$Y_t^{'}$
0	24	27	33	27	24	27	0	25	24
1	0	4	5	0	0	1	0	2	3
2	23	18	31	28	22	36	27	21	22
3	0	1	1	0	1	1	0	1	1
4	22	14	20	27	13	13	25	12	19
5	0	6	12	0	5	4	0	3	4
6	27	15	11	21	15	18	20	20	17
7	71	58	67	102	63	53	58	63	80
8	33	49	49	53	39	41	47	53	54
9	23	18	20	0	11	13	0	14	17
RMSE		8.6	9.3		14.6	17.3		10.5	12.2
MAE		7.3	7.6		9.9	10.5		7.5	9.2

Table 4. The comparison of predicting results

XGBOOST model with instantaneous data respectively. Besides, the accuracy of the M-GAN-XGBOOST model with instantaneous data as input is 74.66%. For different products, the accuracy rate is different, the accuracy rate of product 1 is up to 88.19%, but the accuracy of product 5 is only 53.26%. We checked the historical data of each product and found that the sales volume of product 5 fluctuated significantly, with many sharp increases and decreases, which resulted in low forecast accuracy. From the perspective of marketing scenarios and products, this paper combines products, data and services to optimize the business development of online stores. And it can provide effective support for operational activities such as inventory management and precision marketing for stores based on historical product traffic and sales volume as well as prediction results. Firstly, it provides a basis for the scientific replenishment of orders and reduces capital occupation. Secondly, it uses the forecast results to stimulate the enthusiasm of employees. Thirdly, the sales targets, pricing strategies and promotion strategies for each product are formulated in a targeted manner.

6.1 The strategy of product positioning

In the product dimension, the merchant can understand the purchasing power and attention information of the online store products based on the sales forecast and consumer traffic results. They can classify the products into primary and secondary sales categories based on the sales forecast $\widehat{Y}_i$. Through studying the life cycle of each product, they can adapt their promotion strategies for different product categories timely. In this numerical experiment, if we look at the past sales of the products, we will find that product 1 only has sales in 2 days. Product 1 has a certain degree of unsold characteristics and a short life cycle. Product 2 and product 5 have sales of 0 after some time, so they all have a cycle. Online stores on the shelves of goods should pay attention to its periodicity. According to Pareto's law, in general, profits are from best-selling products for online stores. In the application of the model product positioning, stores should confirm the characteristics of commodity sales and adjust product categories timely, so that they can distinguish between the sales of main products and supplementary products.

6.2 The optimizing strategy of product display

For online merchants, it is an important resource that the display of products on the home page for their websites and pushes information. From the perspective of marketing scenarios, it is crucial to adjust the order of products on the website according to the predicted consumer traffic data $ct\widehat{ra}_t$ in the formulation of marketing strategies, which can optimize shopping scenarios and enhance user experience. Products with high consumer traffic meet the needs of most people and play an important role in improving the conversion rate. They are the core selling point of the shop and should be placed in the position with the greatest weight of attractiveness on the page to improve the efficiency of visual information delivery and facilitate user browsing as a way to boost product sales. Through high traffic products to attract users to browse the shop, drive shop popularity and improve overall sales.

Set the flow of the ith commodity as $ctra_i$, the number of commodities as n, the number of commodity display positions as θ. $y_{i\theta}$ is whether the ith commodity is displayed in the θth display position, and the attraction weight of each display position is c_θ, then We can design the display order of the products according to the following model:

$$\begin{aligned} \text{Max} \quad & U(y) = \sum_i \sum_\theta c_\theta y_{i\theta} ctra_i \\ s.t. \quad & \sum_i y_{i\theta} = 1 \\ & \sum_\theta y_{i\theta} \leq 1 \\ & y_{i\theta} \in \{0,1\}, \forall i, \theta \end{aligned} \tag{19}$$

Merchants transfer the maximum degree of information to consumers in a limited space to increase sales volume. According to the above model, an appropriate adjustment of the page layout can be occurred based on the size of consumer traffic. The use of page resources is reasonable for different products with the proper position. Products with the largest consumer traffic allocate to the largest weight of the position, which has the biggest attractiveness.

6.3 The strategy of zero inventory management

The daily replenishment quantity is calculated according to the relationship between the predicted sales volume, historical sales volume and inventory quantity. The inventory manager dynamically adjusts the replenishment period to save inventory management costs and realize replenishment on demand. Set the zero point inventory as Q, the redundancy value of replenishment as C, the number of lanes occupied by goods as K, the number of goods in each lane as p and the ratio α. We can build a model to adjust the replenishment amount of products under special circumstances.

Based on the historical sales of each product, the average sales of each product $\overline{F}$ as a threshold value, we can develop the corresponding replenishment strategy. If the predicted value is greater than the average, the replenishment amount of the product per day is given as:

$$R_t = \min(\alpha(\widehat{Y}_t + C - Q), K^*p - Q). \tag{20}$$

If the predicted value is less than the average, the replenishment amount of the product per day is based on the following equation:

$$R_t = \widehat{Y}_t. \tag{21}$$

The redundancy value of replenishment C is initially 0. It is corrected according to the number of out-of-stock occurrences in actual operations. If out-of-stock occurs, the value of replenishment is given as:

$$C_1' = C + |\overline{Y}_t - R|. \tag{22}$$

We can calculate the accurate replenishment value by using the predicted sales of goods and the actual inventory of the online store, which can improve the turnover rate of goods, effective inventory management, save replenishment costs and approach zero inventory.

6.4 The strategy of product promotion

According to the predicted sales volume for each product, there is a large gap with the real sales volume. Merchants should take advantage of the herd effect in the consumer shopping process, to adopt different promotional strategies for different sales of products according to the product positioning.

The primary sales products should get focused attention, which has large sales, such as products 0, 2, 4, 6, 7, 8, 9 in this numerical experiment. Before the sale, merchants can make personalized recommendations, which is categorizing users by interests and according to the results of user classification providing users in the category with primary sales products in correspondence with their interests. At the time of the sale, merchants can give away gifts to induce purchase motivation and guide consumers' exploratory behavior, which can enhance users' desire to buy. After the sale, rewarding recommendations can be made, such as a value-redemption mechanism to recommend users to buy complementary or similar products, which have appropriate discounts.

For supplementary products with low sales volume, such as products 1, 3, 5, merchants can use the traffic of primary sales products to drive theirs with low sales. Merchants can achieve the goal of enhancing sales of primary sales products and increasing sales of supplementary products by setting up product groups for bundling, which have discounts. If there is still no significant increase in sales of supplementary products, the demand plan for them should be promptly reduced and phased out of the market.

Merchants can determine the discount rate for the promotional product based on predicted sales $\widehat{Y}_i$, stock availability Q_i and replenishment R_i by minimizing the loss of profit from the promotion.

We assume the number of products to be promoted is given as:

$$L = \beta(Q + R_t - \widehat{Y}_t), \tag{23}$$

where β is a constant, which represents the proportion of the present goods that need to be promoted, $0 \le \beta \le 1$. We set the discount rate for the products is d and the price of the goods is h. And we suppose that there is a relationship between the price of the goods and sales volume according to the history data:

$$Y_i = f(h_i). \tag{24}$$

Then we can construct a model based on minimizing the loss of profit due to promotion, which is given as:

$$\begin{aligned} \min \quad & S = L \times h - f(hd) \times hd \\ s.t. \quad & f(hd) < L \\ & 0 \le d \le 1. \end{aligned} \tag{25}$$

We can obtain the relationship among sales, inventory and promotional factors based on the predicted consumer traffic, sales of products and the above model. Then the promotion strategy of products can be formulated scientifically and reasonably to achieve the expected effect of increasing sales and guaranteeing minimal loss of profit through the predicted results and the stores' known information.

7. Conclusion and future scope

Given the importance of short-term sales prediction for online stores in making business decisions, after the analysis of the influencing factors and characteristics of product sales based on the online purchase decision-making processes, the M-GAN-XGBOOST short-term sales prediction model is constructed to predict the sales of all products in online stores at once using the instantaneous data of the influencing factors. Because of the lack of adaptation of GAN to the dependence of time series data, the generative model is changed to an LSTM neural network. We implement the model using Tensorflow. Compared with using lag period influencing factors as input features, the model can improve the prediction accuracy and achieve a one-time prediction of all product sales for a multi-category online store. This paper explores the positive significance of using short-term product sales and consumer traffic prediction results in marketing decisions for online stores, which provides new ideas for short-term sales prediction of online store products. When merchants apply the model in practice, they can do subdivisions according to different colors and categories to predict the sales volume for each category of products respectively. This model can provide a basis for accurate digital marketing. The model can be improved and refined in the future by further optimizing the selection and measurement of immediate and lagged variables.

References

Agnihotri, R., Trainor Kevin, J., Itani, O.S. and Rodriguez, M. (2017), "Examining the role of sales-based CRM technology and social media use on post-sale service behaviors in India", *Journal of Business Research*, Vol. 81, pp. 144-154.

Akram, U., Khan, M.K., Peng, H., Tanveer, Y. and Akram, Z. (2018), "Development of E-commerce: factors influencing online impulse shopping in China", *Journal of Electronic Commerce in Organizations*, Vol. 16 No. 2, pp. 29-47.

Armano, G., Marchesi, M. and Murru, A. (2005), "A hybrid genetic-neural architecture for stock indexes forecasting", *Information Sciences*, Vol. 170 No. 1, pp. 3-33.

Bell, D. and Mgbemena, C. (2018), "Data-driven agent-based exploration of customer behavior", *Simulation*, Vol. 94 No. 3, pp. 195-212.

Bendato, I., Cassettari, L., Mosca, M., Mosca, R. and Rolando, F. (2015), "New markets forecast and dynamic production redesign through stochastic simulation", *International Journal of Simulation Modelling*, Vol. 14 No. 3, pp. 485-498.

Celia, F., Garg, A., Sztandera, L. and Raheja, A. (2003), "Forecasting women's apparel sales using mathematical modeling", *International Journal of Clothing Science and Technology*, Vol. 15 No. 2, pp. 107-125.

China Internet Network Information Center (2020), "The 45th China statistical report on internet development", available at: http://www.cnnic.cn/hlwfzyj/hlwxzbg/hlwtjbg/202004/P020210205505603631479.pdf.

Choi, T.-M., Yu, Y. and Au, K.-F. (2011), "A hybrid SARIMA wavelet transform method for sales forecasting", *Decision Support Systems*, Vol. 51 No. 1, pp. 130-140.

Choi, T.M., Liu, N., Ren, S.Y., Hui, C.L. and Ng, S.F. (2013), "Sales forecasting for fashion retailing service industry: a review", *Mathematical Problems in Engineering*, Vol. 2013, pp. 256-264.

Chung, J., Gulcehre, C., Cho, K. and Bengio, Y. (2015), "Gated feedback recurrent neural networks", *Computer Science*, pp. 2067-2075.

Cui, H., Rajagopalan, S. and Ward, A.R. (2020), "Predicting product return volume using machine learning methods", *European Journal of Operational Research*, Vol. 281 No. 3, pp. 612-627.

David, B., Thomas, S. and Luke, M. (2017), "BEGAN: boundary equilibrium generative adversarial networks", *Statistics*, Vol. 2, p. 10.

Di Pillo, G., Latorre, V., Lucidi, S. and Procacci, E. (2016), "An application of support vector machines to sales forecasting under promotions", *4OR*, Vol. 14 No. 3, pp. 309-325.

Doganis, P., Alexandridis, A., Patrinos, P. and Sarimveis, H. (2006), "Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing", *Journal of Food Engineering*, Vol. 75 No. 2, pp. 196-204.

Eskenazi, J. and Fei, Y. (2013), "Sales forecast using a hybrid learning method based on stable seasonal pattern and support vector regression", *Emerging Technologies for Information Systems, Computing, and Management*, Vol. 13 No. 10, pp. 1251-1259.

Fan, Z.P., Che, Y.J. and Chen, Z.Y. (2017), "Product sales forecasting using online reviews and historical sales data: a method combining the Bass model and sentiment analysis", *Journal of Business Research*, Vol. 74, pp. 90-100.

Fildes, R., Ma, S.H. and Huang, T. (2016), "Demand forecasting with high dimensional data: the case of SKU retail sales forecasting with intra- and inter-category promotional information", *European Journal of Operational Research*, Vol. 249 No. 1, pp. 245-257.

Friedman, J.H. (2001), "Greedy function approximation: a gradient boosting machine", *The Annals of Statistics*, Vol. 29 No. 5, pp. 1189-1232.

Frank, C., Sztandera, L., Vemulapali, B., Garg, A. and Raheja, A. (2004), *A Fuzzy Forecasting Model for Women's Casual Sales*, National Textile Center Annual Report, Vol. 11, pp. 46-47.

Goodfellow, I., Pougetabadie, J., Mirza, M., Xu, B. and Wardefarley, D. (2014), "Generative adversarial nets", *Advances in Neural Information Processing Systems* Vol. 27, pp. 2672-2680.

Guadagni, G., Hanrahan, K., Pomykalski, A. and Baker, J. (2017), "Application of machine learning methodologies to multiyear forecasts of video subscribers", *Systems and Information Engineering Design Symposium*, Vol. 17 No. 4, p. 17.

Guestrin, C. and Chen, T. (2016), "XGBoost: a scalable tree boosting system", *Acm Sigkdd International Conference on Knowledge*, pp. 785-794.

Hanson, J., Paliwal, K., Litfn, T., Yang, Y. and Zhou, Y. (2018), "Accurate prediction of protein contact maps by coupling residual two-dimensional bidirectional long short-term memory with convolutional neural networks", *Bioinformatics*, Vol. 34 No. 23, pp. 4039-4045.

Herhausen, D., Miocevic, D., Morgan, R.E. and Kleijnen Mirella, H.P. (2020), "The digital marketing capabilities gap", *Industrial Marketing Management*, Vol. 90, pp. 276-290.

Hochreiter, S. and Schmidhuber, J. (1997), "Long short-term memory", *Neural Computation*, Vol. 9 No. 8, pp. 1735-1780.

Huang, W., Song, G., Hong, H. and Xie, K. (2014), "Deep architecture for traffic flow prediction: deep belief networks with multitask learning", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 15 No. 5, pp. 2191-2201.

Kannadath, B.S., Cen, P. and Rowe, J. (2018), "Decision tree analysis of pancreatic cyst fluid data for the detection of mucinous cysts: 73", *American Journal of Gastroenterology*, Vol. 113, pp. S41-S42.

Karimi, S., Holland, C.P. and PapamichailNadia, K. (2018), "The impact of consumer archetypes on online purchase decision-making processes and outcomes: a behavioural process perspective", *Journal of Business Research*, Vol. 91, pp. 71-82.

Khaled, A., Ouchani, S. and Chohra, C. (2019), "Recommendations-based on semantic analysis of social networks in learning environments", *Computers in Human Behavior*, Vol. 101, pp. 435-449.

Kolkova and Andrea (2020), "The application of forecasting sales of services to increase business competitiveness", *Journal of Competitiveness*, Vol. 12 No. 2, pp. 90-105.

Kottke, J.L. and Mellor, S. (1986), "Intrinsic motivation and self-determination in human behavior", *Personnel Psychology*, Vol. 39 No. 3, pp. 672-675.

Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2017), "ImageNet classification with deep convolutional neural networks", *Communications of the ACM*, Vol. 60 No. 6, pp. 84-90.

Kuzovkin, D., Pouli, T., Meur, O.L., Cozot, R., Kervec, J. and Bouatouch, K. (2019), "Context in photo albums: understanding and modeling user behavior in clustering and selection", *ACM Transactions on Applied Perception (TAP)*, Vol. 16 No. 2, pp. 1-20.

Li, X., Xiang, S., Zhu, P. and Wu, M. (2016), "Establishing a dynamic self-adaptation learning algorithm of the BP neural network and its applications", *International Journal of Bifurcation and Chaos*, Vol. 25 No. 14, p. 1540030.

Liberis, E., Veličković, P., Sormanni, P., Vendruscilo, M. and Lio, P. (2018), "Parapred: antibody paratope prediction using convolutional and recurrent neural networks", *Bioinformatics*, Vol. 34 No. 17, pp. 2944-2950.

Lopez-Fernandez, B. and Perrigot, R. (2018), "Using websites to recruit franchisee candidates", *Journal of Interactive Marketing*, Vol. 42, pp. 80-94.

Mahdavinejad, M.S., Rezvan, M., Barekatain, M., Adibi, P., Baenaghi, P. and Sheth, A.P. (2018), "Machine learning for internet of things data analysis: a survey", *Digital Communications and Networks*, Vol. 4 No. 3, pp. 161-175.

Manning, C.D., Ng, A.Y., Lin, C.Y. and Socher, R. (2011), "Parsing natural scenes and natural language with recursive neural networks", *International Conference on Machine Learning*, pp. 129-136.

Mao, X.D., Li, Q., Xie, H.R., Lau, R.Y.K., Wang, Z., Smolley, S.P. and IEEE (2017), "Least squares generative adversarial networks", *2017 IEEE International Conference on Computer Vision*, pp. 2813-2821.

Martin, A., Soumith, C. and Léon, B. (2017), "Wasserstein GAN", *Statistics*, Vol. 3.

Mau, S., Pletikosa, I. and Wagner, J. (2018), "Forecasting the next likely purchase events of insurance customers: a case study on the value of data-rich multichannel environment", *International Journal of Bank Marketing*, Vol. 36 No. 6, pp. 1125-1144.

OEzden, G.A., Sayin, S., Tom van Woensel and Jan, F. (2009), "SKU demand forecasting in the presence of promotions", *Expert Systems with Application*, Vol. 36 No. 10, pp. 12340-12348.

Pan, H. and Zhou, H. (2020), "Study on convolutional neural network and its application in data mining and sales forecasting for E-commerce", *Electronic Commerce Research*, Vol. 54 No. 1, pp. 85-102.

Park and Sang-June (2016), "Potential solo consumers' consumption values", *The Academy of Customer Satisfaction Management*, Vol. 18 No. 4, pp. 97-107.

Pham, D.H. and Le, A.C. (2018), "Learning multiple layers of knowledge representation for aspect based sentiment analysis", *Data and Knowledge Engineering*, Vol. 114, pp. 26-39.

Prateek, P., Kumar, S. and Sandeep Shrivastav (2013), "Forecasting using fuzzy time series for diffusion of innovation: case of Tata Nano car in India", *National Academy Science Letters-India*, Vol. 36 No. 3, pp. 299-309.

Qiu, X., Suganthan, P.N. and Amaratunga, G.A.J. (2019), "Fusion of multiple indicators with ensemble incremental learning techniques for stock price forecasting", *Journal of Banking and Financial Technology*, Vol. 3 No. 1, pp. 33-42.

Ramos, P., Santos, N. and Rebelo, R. (2015), "Performance of state space and ARIMA models for consumer retail sales forecasting", *Robotics and Computer-Integrated Manufacturing*, Vol. 34, pp. 151-163.

Rao, H., Shi, X., Rodrigue, A.K., Feng, J., Xia, Y., Elhoseny, M., Yuan, X. and Gu, L. (2019), "Feature selection based on artificial bee colony and gradient boosting decision tree", *Applied Soft Computing*, Vol. 74, pp. 634-642.

Razeef, M., Muheet, A.B. and Majid, Z.B. (2020), "GWLM–NARX Grey Wolf Levenberg–Marquardt-based neural network for rainfall prediction", *Data Technologies and Applications*, Vol. 60 No. 2, pp. 235-249.

Ryutaro, I.C., Liu, Xin and Shi, Y. (2017), "Food sales prediction with meteorological data - a case study of a Japanese chain supermarket", *Data Mining and Big Data*, pp. 93-104.

Sivasankar, E. and Vijaya, J. (2019), "A study of feature selection techniques for predicting customer retention in telecommunication sector", *International Journal of Business Information Systems*, Vol. 31 No. 1, pp. 1-26.

Soopramanien, D., Huang, T. and Fildes, R. (2014), "The value of competitive information in forecasting FMCG retail product sales and the variable selection problem", *European Journal of Operational Research*, Vol. 237 No. 2, pp. 738-748.

Steinker, S., Hoberg, K. and Thonemann, U.W. (2017), "The value of weather information for E-commerce operations", *Production and Operations Management*, Vol. 26 No. 10, pp. 1854-1874.

Stripling, E., vanden Broucke, S., Antonio, K., Baesens, B. and Snoeck, M. (2018), "Profit maximizing logistic model for customer churn prediction using genetic algorithms", *Swarm and Evolutionary Computation*, Vol. 40, pp. 116-130.

Sun, S., Zhang, C. and Yu, G. (2006), "A Bayesian network approach to traffic flow forecasting", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 7 No. 1, pp. 124-132.

Taylor, J.W. (2004), "Volatility forecasting with smooth transition exponential smoothing", *International Journal of Forecasting*, Vol. 20 No. 2, pp. 273-286.

Teresa, M.M., Donna, F.D., Susan, L.G. and John, T.M. (2006), "The evolution of sales forecasting management: a 20-year longitudinal study of forecasting practices", *Journal of Forecasting*, Vol. 25 No. 5, pp. 303-324.

Trofimov, I., Kornetova, A. and Topinskiy, V. (2012), "Using boosted trees for click-through rate prediction for sponsored search", *International Workshop on Data Mining for Online*, p. 2.

Vijaya, J. and Sivasankar, E. (2019), "An efficient system for customer churn prediction through particle swarm optimization based feature selection model with simulated annealing", *Cluster Computing*, Vol. 22 No. 5, pp. 10757-10768.

Wager, S. and Athey, S. (2018), "Estimation and inference of heterogeneous treatment effects using random forests", *Journal of the American Statistical Association*, Vol. 113 No. 523, pp. 1228-1242.

Wang, J., Du, T. and Zhang, W. (2016), "Deep learning over multi-field categorical data: a case study on user response prediction", *arXiv*, p. 12.

Wang, J., Lin, L., Zhang, H., Tu, J.q. and Yu, P.h. (2017), "A novel confidence estimation method for heterogeneous implicit feedback", *Frontiers of Information Technology and Electronic Engineering*, Vol. 18 No. 11, pp. 1817-1827.

Wu, Y. and Tan, H. (2016), "Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework", *arXiv*, p. 14.

Yao, Y. and Wang, H.Y. (2019), "Optimal subsampling for softmax regression", *Statistical Papers*, Vol. 60 No. 2, pp. 235-249.

Yuan, G.H., Han, J.T., Wang, Y.Q., Liang, H.J. and Li, G.Y. (2019), "The product demand model driven by consumer's information perception and quality perception", *Physica a-Statistical Mechanics and its Applications*, Vol. 535, p. 122352.

Wedding, D.K. and Cios, K.J. (1996), "Time series forecasting by combining RBF networks, certainty factors, and the Box-Jenkins model", *NeuroComputing*, Vol. 10 No. 2, pp. 149-168.

Corresponding author
Yang Yang can be contacted at: 17863907712@163.com