CHEMICAL ENGINEERING TRANSACTIONS
VOL. 51, 2016
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet
Guest Editors: Tichun Wang, Hongyang Zhang, Lei Tian

ISBN 978-88-95608-43-3; **ISSN** 2283-9216
DOI: 10.3303/CET1651154

The Research on the Electronic Commerce Sales Prediction based on the Improved LSSVM Algorithm

Xiuhong Dong

Dongying Vocational Institute; Dongying, Shandong 257091 China
Dongxiuhong_1976@126.com

The emergence of the electronic commerce provides the consumers with more choices and provides the enterprises with another way of life. In recent years, the development of the electronic commerce is very rapid while it brings the huge competition. Like the offline enterprises, the electronic commerce needs to face with the problems of the inventory. It can better solve the problems of the high inventory and the stock to predict the electronic commerce sales. The enterprises can also take the relevant measures to increase the profits according to the results of the prediction. In this paper, we propose an improved LSSVM algorithm to improve the accuracy of e-commerce sales prediction and use the method to predict the sales. The experiment has achieved the good results.

1. Introduction

The prediction of the electronic commerce was a very important question in the management of the electronic commerce enterprises (Dakshata Argade, Hariram Chavan, (2015)). The accurate sales prediction can not only reduce the inventory costs, but also can bring more profits for the enterprise. The development of the electronic commerce in China was later. However, the development was very fast (Neil Towers, Kiki Xu (2016)). From the beginning of 1997, the Internet users in China increased very quickly. It had provided the broad basis for the development of the e-commerce activities in China (Hefu Liu et al. (2016)). In recent years, with the emergence of the Internet of things, the cloud computing (Lackermair Georg (2011)) and other technologies, the e-commerce has the new energy.

According to the different standards, the electronic commerce can be divided into the different types. From the scope of the definition of the electronic commerce, it can be divided into the broad electronic commerce and the narrow electronic commerce. From the perspective of the development of the e-commerce, it can be divided into the traditional e-commerce and the modern e-commerce. Traditional e-commerce is to carry out the business activities by using the electronic tools of the non-Internet forms. Aiming at the participating subject, the online support platform, the contents of the transaction, the nature of the transaction and the classification of the geographical scope of the transaction of the modern electronic commerce, we make the summarize and manage. The specific categories are shown in the following table.

Table 1: Electronic commerce classification standard

Classification standard	Classification
Participating subject	B2C, B2B and C2C,
Online support platform	Enterprise internal network, enterprise external network and Internet
Contents of the transaction	Indirect e-commerce and direct E-commerce
Nature of the transaction	International, ordinary e-commerce and finance e-commerce
Geographical scope of transaction	Local, regional and international e-commerce

Lssvm algorithm is a kind of machine learning algorithm (Niyaz Mohammad Mahmoodi et al. (2014)). Compared with the SVM algorithm, LSSVM algorithm is simple, the less operation time and the higher accuracy of dealing with the large-scale sample set (Safari et al. Hossein (2014)). Therefore, LSSVM

Please cite this article as: Dong X.H., 2016, The research on the electronic commerce sales prediction based on the improved lssvm algorithm, Chemical Engineering Transactions, 51, 919-924 DOI:10.3303/CET1651154

algorithm has been widely used (Zheng et al. Pan-Pan (2014)). At the same time, the LSSVM algorithm also has appeared many derivatives in the process of the discovery.
In this paper, in order to be able to better predict the e-commerce sales, we propose an improved LSSVM and use the algorithm to predict the e-commerce sales. Compared with the traditional LSSVM algorithm, the results which are obtained in this paper are more accurate.

2. LSSVM algorithm

We assume that the hypothetical training sample set is $T=\{(x_k, y_k)|k=1,2,3,\cdots,n\}$, $x_k \in R^n$ and $y_k \in R$. x_k is the input data. y_k is the output data. In the original space, the optimization problem can be described as,

$$\min C\sum_{i=1}^{n}(\zeta_i+\zeta_i^*)+\frac{1}{2}\omega_i^2 \tag{1}$$

Subject to

$$\begin{cases} f(x)=(w,x)+b \\ y_i - f(x_i) - e \le \zeta_i \\ f(x_i) - y_i - e \le \zeta_i \\ \zeta_i^* \ge 0 \end{cases} \tag{2}$$

We use the sum of the error square to be instead of the slack variables, transform the inequality constraints to equality constraints and obtain the regressed optimization problems.

$$\min_{\omega,b,e} J(w,e)=\frac{1}{2}\omega^T\omega+\frac{\gamma}{2}\sum_{i=1}^{N}e_k^2 \tag{3}$$

Subject to,

$$y_k=\omega^T\varphi(x_k)+b+e_k \tag{4}$$

Where,

$$\varphi(\cdot): R^n \to R^m \tag{5}$$

$\varphi(\cdot)$ is the kernel space mapping function. $w \in R^m$ is the weight vector. $e_k \in R$ is the error vector. b is the deviation vector. The loss function J is the sum of the error and the regular quantity. γ is the adjustable function. The purpose of the kernel space mapping function is to extract the features from the original space. It maps the sample in the original space into a vector of the high dimensional space in order to solve the linear non-separable problem in the original space.
Lagrange function,

$$L(\omega,b,\xi_i,\beta)=\frac{1}{2}\omega^T\omega+\frac{1}{2}\gamma\sum_{i=1}^{N}\xi_i^2-\sum_{i=1}^{N}\beta_i\left[\omega^T\varphi(x_i)+b+\xi_i-y_i\right] \tag{6}$$

Where, Lagrange factor is belongs to $\alpha_k \in R$. The optimized formula is,

$$\begin{cases} \dfrac{\partial L}{\partial \omega}=0, \omega=\sum_{i=1}^{N}\alpha_k\varphi(x_k) \\ \dfrac{\partial L}{\partial b}=0, \sum_{i=1}^{N}\alpha_k=0 \\ \dfrac{\partial L}{\partial e_k}=0, \alpha_k=\gamma e_k \\ \dfrac{\partial L}{\partial \beta_i}=0, \omega^T\varphi(x_k)+b+e_k-y_k=0 \end{cases} \tag{7}$$

The matrix function is,

$$\begin{bmatrix} 0 & E^T \\ E & \phi \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{8}$$

Where,

$$\begin{cases} b = \dfrac{E^T \phi^{-1} y}{E^T \phi^{-1} E} \\ \alpha = \phi^{-1}(y - E \cdot \dfrac{E^T \phi^{-1} y}{E^T \phi^{-1} E}) \end{cases} \tag{9}$$

$$\phi = \begin{bmatrix} K(x_1, x_1) + \dfrac{1}{2\gamma} & K(x_1, x_2) & \cdots & K(x_1, x_N) \\ & K(x_2, x_2) + \dfrac{1}{2\gamma} & \cdots & K(x_2, x_N) \\ & & & \\ K(x_N, x_1) & K(x_N, x_2) & \cdots & K(x_N, x_N) + \dfrac{1}{2\gamma} \end{bmatrix} \tag{10}$$

The function of the LSSVM least squares support vector machine is estimated as,

$$y(x) = \sum_{k=1}^{N} \alpha_k K(x, x_k) + b \tag{11}$$

3. Improved LSSVM algorithm

If we want predict the monthly sales and get the better results, we need to find out the similar historical month with the month which is to be predicted as the training sample. The data of the training sample are as the data source to be modeled. There are many methods to determine the training samples, such as the clustering method, the correlation analysis method etc. This paper is to predict the sales of the e-commerce. Therefore, we adopt the grey incidence theory which has the smaller calculation to analyze. Grey relational analysis theory is an important part of the grey system theory and is a kind of multivariate statistical analysis. It is based on the sample data of the various factors. It uses the grey correlation degree to describe the strength, size and order among the factors. Grey correlation degree is relatively to compare with the close to the data curve geometry. The closer the geometry is, the closer the change trend is, the greater the correlation degree is.

We constitute the daily maximum volume to a vector $[x(1), x(2), \cdots, x(k)]^T$. We assume that x_0 and x_i are the vector that constitutes of the daily maximum volume for the month that to be predicted and the i month.

We can get,

$$x_0 = [x_0(1), x_0(2), \cdots, x_0(k)]^T \tag{12}$$

$$x_i = [x_i(1), x_i(2), \cdots, x_i(k)]^T \tag{13}$$

In addition, we can get the correlation coefficient,

$$\xi_i(k) = \frac{\min i \min k \left|\Delta i(k)\right| + \rho \max i \max k \left|\Delta i(k)\right|}{\left|\Delta i(k)\right| + \rho \max i \max k \left|\Delta i(k)\right|} \tag{14}$$

$\xi_i(k)$ is the correlation coefficient in k point of x_0 and x_i .

Where,

$$\Delta i(k) = x_0(k) - x_i(k) \tag{15}$$

ρ is the resolution factor and $\rho \in [0,1]$.

In general, $\rho = 0.5$.

Comprehensive the correlation coefficient of each point, we can get the whole Grey correlation degree of x_0 and x_i.

$$\gamma_i = \sum_{k=1}^{n} \xi_i(k)\lambda_k \tag{16}$$

We select the similar months according to the size of the Grey correlation degree. We make the corresponding day of the month as the training sample of the largest sale and input into the LSSVM algorithm to predict.

At the same time, we improve the LSSVM algorithm. The influence of the Lagrange factor multiplier is the largest. Therefore, the bigger the value is, the bigger the effect is. We should pay more attention to the larger value. We define the vector degree as follows.

$$s_i = f(\alpha_i) = (1-\delta)\left(\frac{|\alpha_i| - |\alpha_{\min}|}{|\alpha_{\max}| - |\alpha_{\min}|}\right) \tag{17}$$

The steps of the improved LSSVM algorithm are as follows.

1. According to the Grey correlation degree, we select the similar months as the input of the LSSVM.
2. Training the inputted data set information and obtain the Lagrange multiplier $\{\alpha_i\}_{i=1}^{N}$.
3. Selecting the appropriate $0 \le \delta \le 1$ and using the Lagrange multiplier to determine the support vector degree.
4. Constructing new training data set, training the improved LSSVM and getting the model parameter $\{\alpha_i\}_{i=1}^{N}$ and b.
5. According to $\{\alpha_i\}_{i=1}^{N}$, we ascend the training data and minus a small part of the minimum value of the data points.
6. By the remainder of the Lagrange operator, we re-calculate and build the new training set. Then, we make the LSSVM training and get the new Lagrange multiplier.

If the fitting performance drops, the training ends. Otherwise, it goes to step 4.

The flow chart is shown as below.

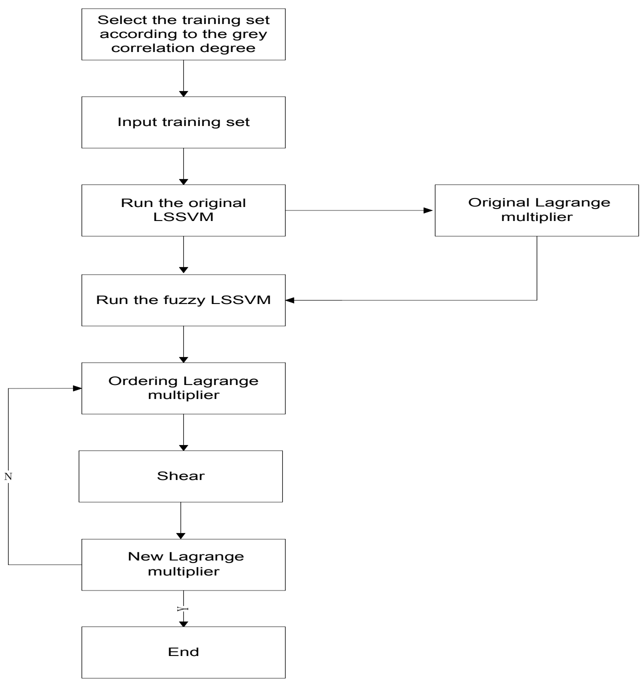

Figure 1: Flow chart of improved LSSVM algorithm

4. Experiment

In this paper, we use the improved LSSVM algorithm to study and predict the sales of the electronic commerce. According to the process of the proposed LSSVM algorithm, we need to select the month which has the biggest correlation degree with the predicted month and take these months as the training samples firstly. Then, we use the improved LSSVM algorithm to predict the target. Finally, we get the final predicted value.

In order to predict the sales volume in August 2015, we pretreat the sales data from July 2012 to July 2015. We selected 10 months which has the greater correlation degree as the source of the training sample. The calculation results of the grey correlation degree are shown in the following table.

Table 2: *Calculation results of grey relational degree*

Month	Grey relational degree
January 2015	0.7934
February 2015	0.8121
March 2015	0.7453
April 2015	0.7890
May 2015	0.7481
June 2015	0.9317
July 2015	0.8976
September 2014	0.8213
August 2013	0.7783
September 2013	0.7269

We select RBF as the kernel function of LSSVM. According to the experiment of the different kernel function parameter selection, we select the experimental parameters. According to the kernel function and parameters, we model and train to get the LSSVM model. Then, we input the training samples which select according to the Grey correlation selection into LSSVM model to be trained. We can predict and get the sales prediction results of August 2015. In addition, in order to verify the accuracy of the method, we compare the prediction results with the traditional LSSVM method. Due to space limitations, we only give ten former data, the specific data are shown in the following table

Table 3: *Experimental data*

Data	Actual value	Method in this paper	LSSVM method
2015.8.1	35	37	39
2015.8.2	47	48	53
2015.8.3	43	44	52
2015.8.4	31	35	47
2015.8.5	25	28	38
2015.8.6	35	41	36
2015.8.7	41	42	43
2015.8.8	44	46	50
2015.8.9	52	54	58
2015.8.10	46	50	54

The predicted results and the actual results are shown in the following figure.

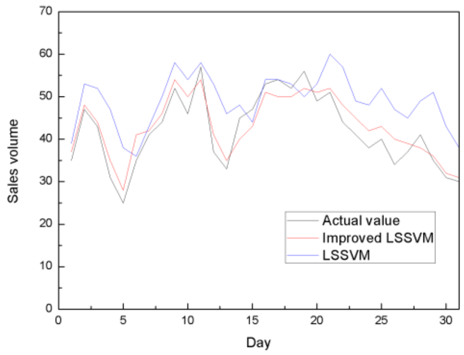

Figure. 2: Comparison of experimental results

From the above figure, we can see that the experimental results of the method which is proposed by this paper are more accurate. Compared with the actual value, the curve of the predicted result is more close to the predicted value. From the prediction accuracy, the results obtained from the LSSVM algorithm have some gaps compared with the improved LSSVM algorithm.

5. Conclusion

In recent years, the electronic commerce has been developing rapidly in our country. For the representative of "Taobao" and "Jingdong", these e-commerce enterprises occupy a significant market share. E-commerce has been integrated into people's daily life. At the same time, the competition of the electronic commerce enterprises is also very intense. To predict the e-commerce sales can get the future sales. At the same time, the electronic commerce enterprise can reduce or increase the inventory, reduce the inventory cost, and increase the flow of funds according to the prediction results. In this paper, firstly, we introduce the background of the research. After that, we introduce the LSSVM algorithm. In order to get the better prediction results, we propose an improved LSSVM algorithm and use this algorithm to predict the sales volume of the electronic commerce.

References

Argade D., Chavan H., 2015, Improve Accuracy of Prediction of User's Future M-commerce Behaviour, Procedia Computer Science, 49: 111-117, DOI: 10.1016/j.procs.2015.04.234

Lackermair G., 2011, Hybrid cloud architectures for the online commerce, Procedia Computer Science,3: 550-555, DOI: 10.1016/j.procs.2010.12.091

Liu H.F., Chu H.L., Huang Q., Chen X.Y., 2016, Enhancing the flow experience of consumers in China through interpersonal interaction in social commerce, Computers in Human Behavior,58: 306-314, DOI: 10.1016/j.chb.2016.01.012

Mahmoodi N.M., Arabloo M., Abdi J., 2014, Laccase immobilized manganese ferrite nanoparticle: Synthesis and LSSVM intelligent modeling of decolorization, Water Research, 67: 216-226, DOI: 10.1016/j.watres.2014.09.011

Safari H., Shokrollahi A., Jamialahmadi M., Ghazanfari M.H., Bahadori A., Zendehboudi S., 2014, Prediction of the aqueous solubility of BaSO4 using pitzer ion interaction model and LSSVM algorithm, Fluid Phase Equilibria,374: 48-62, DOI: 10.1016/j.fluid.2014.04.010

Towers N., Xu K., 2016, The influence of guanxi on physical distribution service quality availability in e-commerce sourcing fashion garments from China, Journal of Retailing and Consumer Services, 28: 126-136, DOI: 10.1016/j.jretconser.2015.09.003

Yarveicy H., Moghaddam A.K., Ghiasi M.M., 2014, Practical use of statistical learning theory for modeling freezing point depression of electrolyte solutions: LSSVM model, Journal of Natural Gas Science and Engineering, 20: 414-421, DOI: 10.1016/j.jngse.2014.06.020

Zheng P.P., Feng J., Li Z., Zhou M.Q. , 2014, A novel SVD and LS-SVM combination algorithm for blind watermarking, Neurocomputing, 142: 520-528, DOI: 10.1016/j.neucom.2014.04.005