# Demand Forecasting about Galvanized Pipe Based on SVM

Guotao WANG, Shaohua DONG, Shang LIU

*School of Mechanical Engineering,* University of Science and Technology, Beijing 100083, CHINA

***Abstract —*** **This paper creates a system for the demand forecasting of galvanized pipe with the algorithm of Support Vector Machine (SVM) in a pipe enterprise. Firstly, the paper analyzes the market demand about galvanized pipe of a pipe enterprise in Tianjin, one of China's top ten steel pipe production enterprise, with the algorithm of clustering analysis. Second, the paper studies the selection of prediction model parameters about SVM and pretreatment of demand data. Lastly, the demand of mature galvanized pipe is forecasted through SVM with Gaussian radial basis for kernel function. It shows that the forecast effect of SVM is better than traditional prediction methods.**

***Keywords - SVM; clustering analysis; demand forecasting***

## I. INTRODUCTION

Demand forecasting is to estimate the demand amount for the products within a certain time. The aim is to use the present and the past historical data, consider various factors combine with the actual situation of the enterprise, use appropriate scientific analysis method and put forward realistic demand targets so that it can customize order demand plans and guide the raw materials or goods orders. Forecasting the future using past data is an important tool to reduce uncertainty and support both individual and organization decision making [1]. In the very competitive and dynamic environment that most businesses face, forecasting is a beneficial tool and an indispensable strategy for business survival [2]. The field of Time Series Forecasting (TSF) deals with the prediction of a given phenomenon based on the past patterns of the same event. TSF has become increasingly used in distinct areas such as Agriculture, Finance, Production or Sales [1]. Reliable prediction of sales can improve the outcome of business, because it allows functional areas, such as production, sales and marketing, and finance, to effectively develop programs to grow the company. Examples of these programs include sales planning, budgeting, and promotion and advertising plan [2]. Numerous forecasting techniques are available to sales forecasting managers [3]. In comparison with the improvement of efficiency in industrial production in these decades, the accuracy of demand forecasting still remains at a low level. Mentzer and Bienstock pointed that forecast of sales influences various decisions at the organizational level. Agrawal and Schorling also pointed out that accurate demand forecasting plays an important role in profitable retail operations. Inaccurate forecast, therefore, often leads to severe business losses evident in both the cost of goods return and opportunity loss [4].

Several Computational Intelligence (CI) methods have been applied to TSF, such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM). CI models such as ANNs and SVMs are natural solutions for TSF, since they are more flexible when compared with classical TSF models, presenting nonlinear learning capabilities [1]. The ANN extracts the implicit non-linear relationship among input variables by learning from training data without making complex dependency assumptions among input variables. But learning algorithm of ANN lacks quantitative analysis and perfect results since it adopts empirical risk minimization (ERM) principle according to statistical learning theory (SLT), which tries only to minimize experience risk [5]. When compared with ANN, SVM presents theoretical advantages, such as the absence of local minima in the learning phase [1]. SVM is based on the structural risk minimization (SRM) principle to minimize the generalization error rather than the empirical error. According to SVM theory, regression problems can be converted into linear ones, and finally deduced to mathematical problems of quadratics programming [5].

Support vector machines (SVM) is based on the statistic theory raised by Vapnik [6]. At the early stage, SVM was only used for classification purposes. Later on, the regression computation of nonlinear function was added by solving a convex quadratic optimization problem [7]. It is a machine learning algorithm which began to be used in the middle of nineteenth century [6]. As a new algorithm on the foundation of structure risk minimum principle, the SVM algorithm has the superiority than other algorithm based on experience risk minimum principle [8]. It can lower the global error of model. It raises the generalization capability of the model, which is more prominent in the small-sample learning [6]. It solves the dimension problem ingeniously. The complexity of algorithm has nothing to do with the sample dimension. The SVM regression algorithm structures regression function according to the part of training sample, it doesn't need the prior information of the sample or the regression function structure [8]. Recently, the support vector machine (SVM) has gained popularity and is regarded as a state-of-the-art technique for regression and classification applications [9].

The main principle of SVM is that for nonlinear problem, transforming it into a linear problem in high dimension space through nonlinear and seeking the optimal classification surface in transformation space. In other words, changing the input space into a high-dimensional space, which change the nonlinear problem into a linear problem in high-dimension space with the kernel function. Then obtaining the optimal classification surface or curve fitting, which minimizes structure risk, namely minimizes the difference between machine learning and real model [10].

Steel industry in China is facing great pressure. Manufacturing capacity is relative surplus, the market competition is intense, the downstream market demand falls, and cost is high and price is low. Decision makers in pipe enterprises forecast demand relying on historical data and subjective experience. In this paper, a model of demand forecasting about SVM is proposed to predict the demand of pipes base on a pipe enterprise in Tianjin. Then the paper forecasts with Matlab and proves that the SVM algorithm is reasonable and effective.

## II. SVM REGRESSION

This section presents the fundamental knowledge of SVM regression. Under the condition of linear inseparable, $\zeta_i^*$ and $\zeta_i$ called slack variable are introduced into a training sample (xi, yi). Punish coefficient is used to balance sample deviation and machine generalization ability. Assuming that nonlinear training set is $R = \{(x_i, y_i), i = 1,2,..., l\}$.

$R^m = \{(\Phi_{(x_i)}, y_i), i = 1,2,...,l\}$ is the training set which is mapped from input space to high-dimensional space. $f_{(x)} = |w^* * \Phi_{(x)}| + b^*$ is the decision function based on optimal hyper plane. Finding the optimal hyper plane is approximately equivalent with solving the following objective function.

$$\min \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{l}(\zeta_i + \zeta_i^*) \tag{1}$$

where

$$\begin{aligned} & w * \Phi_{(x)} + b - y_i \le \varepsilon + \zeta_i^* \\ & y_i - w * \Phi_{(x)} - b \le \varepsilon + \zeta_i \\ & \zeta_i, \zeta_i^* \ge 0, i = 1 \ldots l \end{aligned}$$

Changing it to Lagrange function of the dual problem as follows:

$$\begin{aligned} \max & -\frac{1}{2}\sum_{i,j=1}^{l}(a_i - a_i^*)(a_j - a_j^*)(\Phi_{(x_i)} * \Phi_{(x_j)}) \\ & + \sum_{i=1}^{l}(a_i - a_i^*)y_i - \varepsilon\sum_{i=1}^{l}(a_i + a_i^*) \end{aligned} \tag{2}$$

Dimension disaster will appear after the original data is mapped into a high dimensional space. When a kernel function K(x, x′) satisfies the Mercer condition, it corresponds to the inner product of a transform space according to the functional theory [11]. So avoiding dimension disaster after mapping, SVM uses kernel function meeting Mercer condition to express the inner product form of mapping function. Therefore, the form of Lagrange function can be expressed as follows:

$$\max - \frac{1}{2}\sum_{i,j=1}^{l}\left(a_i - a_i^*\right)\left(a_j - a_j^*\right)K\left(x, x_i\right) + \sum_{i=1}^{l}\left(a_i - a_i^*\right)y_i - \varepsilon\sum_{i=1}^{l}\left(a_i + a_i^*\right) \tag{3}$$

where

$$\sum_{i=1}^{l}\left(a_i - a_i^*\right) = 0$$

$$c \ge a_i \ge 0, c \ge a_i^* \ge 0$$

Solving the upper function to obtain the optimal solution, and the prediction regression function:

$$f(x) = \sum_{i=1}^{l}\left(a_i - a_i^*\right)K\left(x, x_i\right) + b \tag{4}$$

$W$ is the weight vector. c is the penalty parameter. $\varepsilon$ is the insensitive loss function. $\zeta_i^*$ and $\zeta_i$ are called slack variables. $\zeta_i^*$ is the upper and $\zeta_i$ is the lower training error subject to Vapnik' $\varepsilon$ . $a_i^* = \left(a_1, a_2, a_3 \ldots a_l\right)^T$ is the Lagrange optimal operator. $b^*$ is offset scalar. $\Phi(x)$ is the mapping from input vector to high-dimensional space. In this case, $X$ ,the training sample is called support vector. $X_i$ is unknown vector.

## III. CLUSTER ANALYSIS FOR PRODUCT DEMAND

The sale data of galvanized pipe from January 2011 to December 2013 is extracted from the ERP of a pipe enterprise in Tianjin. The paper uses SPSS statistical analysis software to divide galvanized pipes into three levels, respectively, best-selling products, general products and the unbest-selling products. And the corresponding levels are A, B and C respectively. The number of cases in each cluster is shown in table 1. The result of corresponding single factor variance is shown in table 2.

From table1, we can see that there are ten kinds of products of class A, twenty-one kinds of products of class B and thirteen kinds of products of class C. The significance level of each class is less than 0.05, which indicates that the difference among each class is significant. So we can think that the result of classification is reliable. Cluster analysis classifies different specifications of pipe, which can more clearly analyze the needs of each type of specification pipe. Specific clustering result is shown in table 3.

TABLE 1. THE NUMBER OF CASE IN EACH CLUSTER

| **Classification** | **Number** |
|---|---|
| **A** | 10 |
| **B** | 21 |
| **C** | 13 |
| **effective** | 44 |
| **Invalid** | 0 |

TABLE 2. ANALYSIS TABLE OF VARIANCE WITH SINGLE FACTOR

| | **Clustering** | | **Error** | | **Significance** |
|---|---|---|---|---|---|
| | Sum of Squared Residuals | Degrees of Freedom | Mean Square Error | Degrees of Freedom | |
| **Average Monthly Demand** | 17.154 | 2 | 0.212 | 41 | 0.000 |
| **Coefficient of Variation** | 17.773 | 2 | 0.182 | 41 | 0.000 |
| **Sales Revenue** | 17.138 | 2 | 0.213 | 41 | 0.000 |

TABLE 3. THE RESULTS OF PRODUCT CLASSIFICATION

| **Classification** | **Specification** |
|---|---|
| **A** | 26.3X2.5X6.0, 25.8X2.5X6.0, 32.5X2.5X6.0, 33.0X2.75X6.0, 33.0X3.0X6.0, 42.0X2.75X6.0, 42.0X3.0X6.0, 47.5X3.0X6.0, 59.5X3.25X6.0, 59.0X3.0X6.0 |
| **B** | 32.5X2.2X6.0, 21.3X2.75X6.0, 60.3X3.0X6.0, 26.5X2.75X6.0, 33.5X3.25X6.0, 20.5X1.8X6.0, 48.3X3.5X6.0, 42.3X3.25X6.0, 59.0X2.75X6.0, 25.5X1.8X6.0, 20.8X2.2X6.0, 20.5X2.0X6.0, 47.0X2.5X6.0, 59.0X2.5X6.0, 47.0X2.75X6.0, 47.5X3.25X6.0, 25.5X2.0X6.0, 41.5X2.5X6.0, 20.8X2.5X6.0, 26.0X2.2X6.0, 47.0X3.0X6.0 |
| **C** | 41.5X1.8X6.0, 48.0X3.5X6.0, 48.5X3.5X6.0, 88.5X 3.0X6.0, 88.5X3.5X6.0, 88.5X3.25X6.0, 60.3X2.75X6.0, 60.3X3.5X6.0, 60.3X3.75X6.0, 75.0X2.75X6.0, 75.0X3.0X6.0, 75.0X3.25X6.0, 75.0X3.5X6.0 |

According to the above of clustering results, the paper selects five kinds of specification products from the mature ones as the forecast test object. They are 32.5X2.5X6.0, 33X2.75X6.0, 33X2.0 X6.0, 42X2.75X6.0, 59X3.0X6.0. The previous data of those product demand is shown in table 4.

TABLE 4. HISTORICAL DATA FOR PRODUCT DEMAND

| **specifications （2011）** | **Jan.** | **Feb.** | **Mar.** | **Apr.** |
|---|---|---|---|---|
| 32.5X2.5X6.0 | 1289.038 | 735.977 | 586.402 | 488.261 |
| 33X2.75X6.0 | 3011.692 | 2620.872 | 1897.952 | 406.615 |
| 33X2.0X6.0 | 694.586 | 1280.090 | 951.545 | 755.201 |
| 42X2.75X6.0 | 1665.928 | 2713.556 | 1787.065 | 917.884 |
| 59X3.0X6.0 | 1775.160 | 1308.757 | 1267.202 | 2277.779 |
| **specifications （2011）** | **May.** | **Jun.** | **Jul.** | **Aug.** |
| 32.5X2.5X6.0 | 548.340 | 1560.065 | 1218.748 | 990.643 |
| 33X2.75X6.0 | 727.841 | 2663.126 | 3583.690 | 2698.830 |
| 33X2.0X6.0 | 721.595 | 1506.350 | 858.399 | 987.234 |
| 42X2.75X6.0 | 2036.234 | 1641.905 | 1207.832 | 185.110 |
| 59X3.0X6.0 | 1343.352 | 2270.503 | 1483.426 | 1886.242 |
| **specifications （2011）** | **Sep.** | **Oct.** | **Nov.** | **Dec.** |
| 32.5X2.5X6.0 | 1801.344 | 1156.583 | 1212.816 | 1620.756 |
| 33X2.75X6.0 | 3044.098 | 2908.092 | 2997.370 | 2387.311 |
| 33X2.0X6.0 | 879.681 | 693.690 | 970.088 | 732.008 |
| 42X2.75X6.0 | 2102.529 | 1406.248 | 172.066 | 2061.828 |
| 59X3.0X6.0 | 2086.168 | 2094.056 | 1518.195 | 1620.756 |
| **specifications (2012)** | **Jan.** | **Feb.** | **Mar.** | **Apr.** |
| 32.5X2.5X6.0 | 1191.541 | 583.505 | 1097.823 | 1326.590 |
| 33X2.75X6.0 | 3097.279 | 1870.227 | 3510.818 | 2818.396 |
| 33X2.0X6.0 | 1603.534 | 394.529 | 1436.826 | 988.549 |
| 42X2.75X6.0 | 1943.393 | 1984.182 | 1302.028 | 2206.775 |
| 59X3.0X6.0 | 2051.000 | 1659.219 | 1709.763 | 1663.442 |
| **specifications (2012)** | **May.** | **Jun.** | **Jul.** | **Aug.** |
| 32.5X2.5X6.0 | 542.405 | 1315.008 | 450.456 | 863.940 |
| 33X2.75X6.0 | 387.801 | 3304.757 | 2387.997 | 2632.580 |
| 33X2.0X6.0 | 1417.097 | 1155.795 | 1330.483 | 1109.029 |
| 42X2.75X6.0 | 1986.333 | 753.793 | 2339.990 | 1247.646 |
| 59X3.0X6.0 | 1261.147 | 1820.068 | 1537.279 | 1998.673 |
| **specifications (2012)** | **Sep.** | **Oct.** | **Nov.** | **Dec.** |
| 32.5X2.5X6.0 | 1082.552 | 1308.562 | 856.133 | 862.967 |
| 33X2.75X6.0 | 2714.791 | 2426.564 | 2995.617 | 2868.110 |
| 33X2.0X6.0 | 2019.868 | 1021.257 | 1433.294 | 787.020 |
| 42X2.75X6.0 | 404.040 | 866.628 | 1357.155 | 1847.915 |
| 59X3.0X6.0 | 2157.939 | 1528.151 | 1283.213 | 2111.519 |
| **specifications (2013)** | **Jan.** | **Feb.** | **Mar.** | **Apr.** |
| 32.5X2.5X6.0 | 1134.524 | 1212.669 | 762.613 | 1147.036 |
| 33X2.75X6.0 | 2253.127 | 2576.315 | 3150.140 | 3857.655 |
| 33X2.0X6.0 | 1545.393 | 890.379 | 1079.929 | 992.059 |
| 42X2.75X6.0 | 2756.552 | 1235.178 | 2182.009 | 1620.977 |
| 59X3.0X6.0 | 1698.775 | 1382.187 | 1868.332 | 1570.931 |
| **specifications (2013)** | **May.** | **Jun.** | **Jul.** | **Aug.** |
| 32.5X2.5X6.0 | 979.799 | 1756.510 | 1115.174 | 1035.524 |
| 33X2.75X6.0 | 3021.325 | 2224.810 | 2060.372 | 2145.231 |
| 33X2.0X6.0 | 1936.873 | 851.000 | 1257.215 | 1252.691 |
| 42X2.75X6.0 | 1135.530 | 1889.779 | 1628.689 | 1768.315 |
| 59X3.0X6.0 | 1061.836 | 1361.608 | 1566.575 | 1935.802 |
| **specifications (2013)** | **Sep.** | **Oct.** | **Nov.** | **Dec.** |
| 32.5X2.5X6.0 | 1142.722 | 1294.016 | 792.549 | 1139.636 |
| 33X2.75X6.0 | 2822.701 | 3154.596 | 2696.655 | 2426.298 |
| 33X2.0X6.0 | 1240.039 | 962.240 | 1303.870 | 1003.457 |
| 42X2.75X6.0 | 2225.697 | 1573.542 | 2300.340 | 1780.891 |
| 59X3.0X6.0 | 2147.172 | 1752.356 | 1649.774 | 1967.262 |

## IV. COMPUTATION AND ANALYSIS WITH SVM

The first step of SVM is to extract the training data and testing data through the historical data. Then normalizing the data. And getting the forecasting model with SVM training by processing the training data. Then using the test set for testing with the model and comparing the predictive value get through the model with actual value.

### *A. Preparation of training set and test set*

The paper treats the demand amount data of five kinds of specifications of galvanized pipe as the training and test data, the demand amount data for thirty-three months before

DOI 10.5013/IJSSST.a.17.43.19

ISSN: 1473-804x online, 1473-8031 print

as a training set of machine learning and the last of three months as basis of verifying the effect of analysis and prediction.

Taking the specification for 32.5x2.5x6.0 galvanized pipe as an example. Treat the demand amount of each month from January 2011 to August 2013 as training set and six months as a training unit. The SVM training matrix is expressed as:

$$\begin{bmatrix} 735.977 & 548.340 & 1212.816 & 1035.524 & 1115.174 & 586.402 \\ 548.340 & 1212.816 & 1035.524 & 1115.174 & 586.402 & 1289.038 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1560.065 & 583.505 & 856.133 & 1142.722 & 1308.562 & 1082.552 \\ 583.505 & 856.133 & 1142.722 & 1308.562 & 1082.552 & 1801.344 \end{bmatrix}$$

The corresponding test matrix is the month demand from July 2011 to September 2013. The regression prediction matrix is expressed as:

$$\begin{bmatrix} 548.340 \\ 1212.816 \\ \cdots \\ 1801.344 \\ 1294.016 \end{bmatrix}$$

*B. Data preprocessing*

Normalized operation is to transform all the data to the data from zero to one, which is to eliminate the difference among different dimensions and prevent the prediction deviation extending due to the obvious difference between input and output data. The paper takes maximum minimum method for data preprocessing. Functional form is as follows:

$$\tilde{x}_k = \frac{x_k - x_{\min}}{x_{\max} - x_{\min}}$$

Where $x_{\min}$ is the minimum value of set of original demand amount. $x_{\max}$ is the maximum value of set of original demand amount.

*C. The selection of c and g*

Kernel function is the key technology to SVM, choice of kernel function will affect learning ability and generalization ability of machine learning [12]. The different types of nonlinear decision surface of learning machine can be constructed with different kernel functions, which lead to different support vector algorithm. The common prime kernel functions contain linear kernel function, polynomial kernel function, Gaussian radial basis function (RBF) and two layer perception kernel function. Table 5 shows the expressions:

TABLE 5. KERNEL FUNCTIONS AND ITS EXPRESSIONS

| name | expression |
|---|---|
| **Linear kernel function** | $K(x, x_i) = x^T x_i$ |
| **Polynomial kernel function** | $K(x, x_i) = (gx^T x_i + g)$ |
| **Gaussian radial basis function** | $K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{g^2}\right)$ |
| **Two layer perception kernel function** | $K(x, x_i) = \tanh(gx^T x_i + r)$ |

The most common method of choice of kernel function is to find out the corresponding relationship between distribution characteristics of observation data and optimal hyperplane. Then choose the types and parameters according to some prior knowledge. The other way is to construct new types optimized gradually in the training process [13]. The selection of kernel function form and the parameters directly influences the type of classifier, and controls the performance of the classifier at the same time, which determines the complexity of the classifier. We should make the concrete analysis according to the practical problems in practice.

Based on the analysis of various kernel functions, this paper uses the optimal performance of RBF kernel function which expresses excellent learning ability in practice application [12]. If the Gaussian radial basis function is adopted, the coefficients in the model are c and g, they influence both the complexity and the generalization error of the classifier in which c is used to adjust the confidence interval's range of learning machine in certain data sub-spaces, and g alters the mapping function implicitly, in order to change the complexity of sample data sub-space, that is the linear hyperplane's VC dimension [14]. The pros and cons of RBF properties are directly affected by the error punishment parameters c and Gaussian width g, so reasonable parameters will be conducive to perform advantages of RBF-SVM in better way [12]. Taking the specification of 32.5 X2.5 X6.0 as an example, generating parameters in [-8,8] randomly with the method of selecting the parameters of kernel function .Then forecasting with them. Table6 shows the results of forecasting. The table shows that the mean square error and deviation between the predictive value and actual value is very big. Fitting parameters can be obtained possibly on the basis of experience. But it still cannot ensure that the current parameters are the most optimal. So this paper tries to use cross validation method， also known as the meshing method to select parameters. The basic idea is that the training set is divided into N subsets, and each subset is used to forecast serve as test set, while the remaining is used to train as training set. Assigning c and g in a certain range and finding out the best parameter in a certain sense. Also taking 32.5 X2.5 X6.0 as an example, the results with cross validation method are shown in Fig.1 and table 7. When c=0.3 and g=2.4, it achieves the best effect and the mean square error is 13.2199 which is far less than the results achieved by adopting the method of random select parameters.

TABLE 6. PREDICTION ACCURATE UNDER THE RANDOM PARAMETER SELECTION

| Run NO. | c | g | Regression prediction mean square error |
|---|---|---|---|
| **1** | 7.5853 | -3.0003 | 456.1771 |
| **2** | 1.3765 | -0.6122 | 597.2511 |
| **3** | 3.7843 | 4.9630 | 34.4555 |
| **4** | -6.7564 | 5.8857 | 782.1644 |

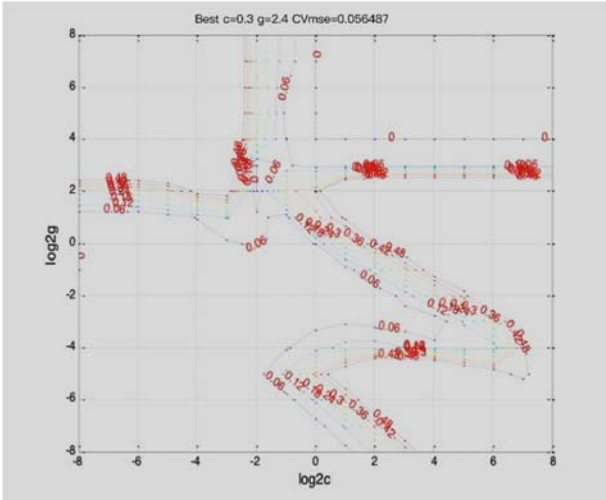


Fig1. Contour map of parameter selection results

TABLE 7. PREDICTION RESULTS OF CROSS VALIDATION METHOD

| Validation method | c | g | Regression prediction mean square error |
|---|---|---|---|
| **Cross validation method** | 0.3 | 2.4 | 13.2199 |

## D. *Experimental results*

Taking all of previous data of demand amount into the trained network and getting SVM prediction of the galvanized pipe about five specifications in October, November and December,2013. Comparing with Weighted moving average forecasting method whose mobile step is three and weighted coefficients are 0.5,1,1.5. Table8, Table9 and Table10 shows the comparison results of two kinds of algorithm. The most error rate of forecasting using SVM is less than the one using the weighted moving average forecasting for each specification. Besides, the gap is large. Three-month average error rate of SVM and the weighted moving average forecasting about five specifications are shown in Table 11. It can be seen that the average error produced by using SVM is lesser obviously. Which indicates that the forecasting results using SVM is more accurate.

TABLE 8. ACTUAL VALUE, FORECASTING VALUE AND ERROR RATE OF EACH IN OCT.

| Specification | Actual value | SVM | Error rate | The weighted moving average forecasting | Error rate |
|---|---|---|---|---|---|
| 32.5X2.5X6.0 | 1294.016 | 1127.98 | 12.83% | 1102.398 | 14.81% |
| 59.0X3.0X6.0 | 1752.356 | 1703.888 | 2.77% | 1979.949 | 12.99% |
| 33.0X3.0X6.0 | 962.24 | 1005.969 | 4.55% | 1247.119 | 29.61% |
| 33.0X2.75X6.0 | 3154.596 | 3352.301 | 6.27% | 2469.823 | 21.71% |
| 42.0X2.75X6.0 | 1573.542 | 1731.155 | 10.02% | 1969.081 | 25.14% |

TABLE 9. ACTUAL VALUE, FORECASTING VALUE AND ERROR RATE OF EACH IN NOV.

| Specification | Actual value | SVM | Error rate | The weighted moving average forecastin g | Error rate |
|---|---|---|---|---|---|
| 32.5X2.5 X6.0 | 792.549 | 932.329 | 17.64% | 1200.503 | 51.47% |
| 59.0X3.0 X6.0 | 1649.774 | 1591.483 | 3.53% | 1914.536 | 16.05% |
| 33.0X3.0 X6.0 | 1303.87 | 1328.305 | 1.87% | 1103.248 | 15.39% |
| 33.0X2.75X6.0 | 2696.655 | 2904.439 | 7.70% | 2875.737 | 6.64% |
| 42.0X2.75X6.0 | 2300.34 | 2114.132 | 8.09% | 1808.143 | 21.40% |

TABLE 10. ACTUAL VALUE, FORECASTING VALUE AND ERROR RATE OF EACH IN DEC.

| Specification | Actual value | SVM | Error rate | The weighted moving average forecasting | Error rate |
|---|---|---|---|---|---|
| 32.5X2.5 X6.0 | 1139.636 | 1108.39 | 2.74% | 1018.067 | 10.67% |
| 59.0X3.0 X6.0 | 1967.262 | 1729.609 | 12.08% | 1766.868 | 10.19% |
| 33.0X3.0 X6.0 | 1003.457 | 1031.051 | 2.75% | 1179.355 | 17.53% |
| 33.0X2.75X6.0 | 2426.298 | 2535.807 | 4.51% | 2870.31 | 18.3% |
| 42.0X2.75X6.0 | 1780.891 | 1963.767 | 10.27% | 2067.372 | 16.09% |

TABLE 11. THE STATISTICS OF ERROR BETWEEN SVM AND THE WEIGHTED MOVING AVERAGE FORECASTING

| Specification | Average error rate using SVM | Average error rate using the weighted moving average forecasting |
|---|---|---|
| 32.5X2.5 X6.0 | 11．07% | 25．65% |
| 59.0X3.0 X6.0 | 6．13% | 13．08% |
| 33.0X3.0 X6.0 | 3．06% | 20．84% |
| 33.0X2.75X6.0 | 6．16% | 15．55% |
| 42.0X2.75X6.0 | 9．46% | 20．88% |

## V. CONCLUSION

This paper presents the use of SVM on the demand forecasting of galvanized pipe in a pipe enterprise in Tianjin. Firstly, classing steel pipes through cluster analysis according to the previous demand data. Secondly, support vector machine(SVM) algorithm is applied to the enterprise's demand forecasting. When selecting parameters, the paper uses the cross validation method which is better than the method of random parameter selection. Thirdly, comparing with weighted moving average algorithm, it is proved that the forecasting result of SVM is more effective and is feasible on the demand forecasting in pipe enterprise. Which can be used to make production plan and provides auxiliary support to reduce inventory.

## ACKNOWLEDGEMENTS

The author wishes to provide his thanks to teachers in University of Science and Technology Beijing for their helpful comments and suggestion, and to a pipe enterprise in Tianjin for providing a mass of data for this paper.

## REFERENCES

[1] Cortez,Paulo, Donate,Juan Peralta, "Evolutionary support vector machines for time series forecasting," 22nd International Conference on Artificial Neural Networks, ICANN 2012, Lausanne, Switzerland, 2012, pp523-530.

[2] Esmaeil Hadavandi, Hassan Shavandi, Arash Ghanbari. "An improved sales forecasting approach by the integration of genetic fuzzy systems and data clustering: Case study of printed circuit board. Expert Systems with Applications, Volume 38,pp9392-9399, August 2011.

[3] Chern Ching-Chin, Ao Ieong Ka Ieng, Wu Ling-Ling, Kung Ling-Chieh. "Designing a decision-support system for new product sales forecasting". Expert Systems with Applications, Volume 37, Issue 2, pp 1654-1665, March 2010.

[4] Kenji Tanaka. “A sales forecasting model for new-released and nonlinear sales trend products”. Expert Systems with Applications, Volume 37,pp 7387-7393, November 2010.

[5] Zheng, Yongkang, Zhu, Li, Zou, Xing, “Short-term load forecasting based on Gaussian wavelet SVM”, 1st International Conference on Smart Grid and Clean Energy Technologies, ICSGCE 2011, Chengdu, China,2011,pp 387-393.

[6] Dongxiao Niu, Yongli Wang, Desheng Dash Wu. “Power load forecasting using support vector machine and ant colony optimization”. Expert Systems with Applications, Volume 37, pp 2531-2539, 15 March 2010.

[7] Xing Yan, Nurul A. Chowdhury. “Mid-term electricity market clearing price forecasting: A multiple SVM approach”. International Journal of Electrical Power & Energy Systems, Volume 58, pp 206-214, June 2014.

[8] Yan, Ke-Wu, “Study on the forecast of air passenger flow based on SVM regression algorithm”, 2009 1st International Workshop on Database Technology and Applications, DBTA 2009, Wuhan, Hubei, China,2009, pp325-328.

[9] Shian-Chang Huang, Tung-Kuang Wu. “Integrating GA-based time-scale feature extractions with SVMs for stock index forecasting”. Expert Systems with Applications, Volume 35, pp2080-2088, November 2008.

[10] Wuneng Ling, Naishan Hang, Ruqi Li, “Short-term wind power prediction based on cloud support vector machine”, Electric Power Automation Equipment, Volume 33, pp 34-38, July 2013.

[11] Xibin Wang, Junhao Wen, Yihao Zhang, Yubiao Wang. “Real estate price forecasting based on SVM optimized by PSO”. Optik - International Journal for Light and Electron Optics, Volume 125, pp 1439-1443, February 2014.

[12] Gao, Hui-Sheng, Guo, Ai-Ling, Yu, Xiao-Dong , Li, Cong-Cong, “RBF-SVM and its Application on Network Security Risk Evaluation”, 2008 International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM 2008, Dalian, China, 2008, pp 1969-2016.

[13] Wanzhao Cui, Changchun Zhu, Wenxing Bao, Junhua Liu, “Chaotic time series prediction based on fuzzy support vector machine”, Journal of Physics, Volume 54, pp3009-3018, July 2005.

[14] Song, Huazhu, Ding, Zichun, Guo, Cuicui, Li, Zhe, Xia, Hongxia, “Research on combination kernel function of support vector machine”, International Conference on Computer Science and Software Engineering, CSSE 2008, Wuhan, Hubei, China, 2008, pp838-841.