International Journal of Innovative Computing, Information and Control ICIC International ©2017 ISSN 1349-4198
Volume 13, Number 2, April 2017 pp. 521–533

AN INTEGRATED RS-SVM MODEL USING GENETIC ALGORITHM AND CROSS VALIDATION FOR PREDICTING THE POTENTIAL DEMAND OF AGRICULTURAL SUPPLIES

Yumeng Wang* and Xiangzhi Kong

School of Agricultural Economics and Rural Development
Renmin University of China
No. 59, Zhongguancun Street, Haidian District, Beijing 100872, P. R. China
*Corresponding author: wymmyw@ruc.edu.cn

Received October 2016; revised February 2017

Abstract. *In the distribution process of agricultural supplies, there are some factors impacting the potential demand, so precise supply orders are often not known in advance. Meanwhile, the dimensions of these impact factors are different and measured with various criteria. Thus, it is difficult to determine the indexes to predict the potential demand of agricultural supplies. In order to cope with the problem, this paper identifies and presents 16 specific indexes of the potential demand prediction depending on four aspects, i.e., customer, enterprise, society and logistics. Then, a hybrid algorithm, that is, RS-SVM, is proposed based on rough set (RS) and support vector machine (SVM). In the proposed algorithm, genetic algorithm (GA) as one of attribute reduction approaches in RS is chosen and used to achieve the attribution reduction, by which the number of indexes is reduced from 16 to 6. Then, SVM is designed and performed to predict the potential demand of agricultural supplies in the agricultural distribution. Finally, a numerical example is given and the comparison results demonstrate the effectiveness and applicability of the proposed algorithm.*
Keywords: Agricultural supplies, Demand prediction, Genetic algorithm, Rough set, Support vector machine

1. **Introduction.** Generally, agricultural supplies involve seeds, fertilizers, pesticides, production and processing equipment, raw supplies, fuels, and so on [1]. Agricultural logistics aims at timely delivering agricultural supplies to ensure the agricultural production and the daily lives of farmers. Thus, the distribution of agricultural supplies is to deliver products from the production line to farmers.

Numerous papers have paid attention to the modeling for logistic distribution of agricultural supplies [2-4], but most of them considered the given and known customers in advance. Few studies consider the potential customers in the delivery, which often results in extra delivery cost and time. Customers especially farmers in one region have the herd behavior. For example, if one farmer buys some agricultural supply that seems to be with a good cost performance, other farmers may follow the purchase. If no extra supplies are carried, delivery providers have to arrange extra deliveries. Thus, the consideration of potential demands can effectively reduce delivery costs. The demand prediction of potential customers for agricultural supplies is defined as predicting and analyzing the demand of customers, who have never bought the products but in the future they might be the actual customers [5,6]. Identifying the potential demand of customers is of great benefit to increase the profit and improve the customers' satisfaction.

The focus on the potential demand of customers has been studied in previous works [5-11], but few works proposed the predicting model of potential demands in the distribution process of agricultural supplies. Kim et al. predicted the purchase behavior of customers using the multiple classifiers approach based on genetic algorithm to reveal more potential customers [7]. Liu and Yu used the self-organizing feature map network to predict the potential customers [8]. Guo and Fang divided the hidden customers' behavior mode into different target customer groups by using support vector machine to mine the Web log files and find the hidden customer resource [9]. Egri considered the demand prediction in a multiple newsvendor-type purchasing problem [10]. Guan et al. used the link prediction approach to estimate the potential trade links [11]. To the best of our knowledge, the study on the potential demand prediction in the agricultural logistics distribution is not still explored in the literature. However, it is very vital to research on which factors affect the potential demand and then to predict the potential demand in the distribution process for agricultural supplies.

Rough set and support vector machine have been extensively used in many works, and the results show that they are more accurate and effective in comparison with some other methods [12-16]. However, the prediction of potential demands often involves various factors, and these factors may be dependent. That is to say, before the prediction, we need to determine which specific factors play the most useful to the prediction accuracy. This reduction on dependent factors can improve the prediction efficiency. Thus, a joint of rough set and support vector machine is able to get better accuracy since the hybrid algorithm can eliminate redundant data by using rough set, and support vector machine has fault tolerance and generalization ability to perform the classification [17]. Therefore, this paper designs a hybrid algorithm called as RS-SVM, and adopts it to predict the demand of potential customers in the logistics distribution for agricultural supplies.

The contributions of this paper are as follows: (1) an index system is constructed to predict the potential demand of agricultural supplies; (2) the proposed algorithm RS-SVM is designed to predict the potential demand based on the historical data and the error between the prediction results by RS-SVM and the original values are compared and considered to be acceptable. The remaining of this paper is outlined as follows. Section 2 introduces the index system for predicting potential demand of agricultural supplies. The proposed algorithm RS-SVM is presented in Section 3. A numerical example is given to illustrate the advantage of the proposed algorithm in Section 4. Finally, Section 5 concludes this paper and some research topics in the future are given.

2. **Establishment of the Index System of Predicting Potential Demand.**

The following four kinds of factors are considered to establish the index system of predicting potential demand.

(1) **Customer** factors

Customer factors mainly include the purchasing power and purchasing desire, the consumption level, crop planting area and agricultural customer number, etc. [18]. Purchasing power is mainly based on the level of customer revenue, which implies the determination for the ability of customers to buy which agricultural products. Purchase desire is the desire of customers to buy which kind of products, which is often dependent on the customer demand and the previous purchase behavior. The consumption behavior of customers changes vastly with the development of economy and the improvement of the level of per capita income. The area of crops relies on the customers' subjective decision, which determines the demand for agricultural supplies. The number of customers has a direct link to the quantity of products to be sold by companies.

(2) **Enterprise** factors

The enterprise factors include the popularity, product quality, product price, service level, management level and market share, etc. [19]. The popularity means the size of the intensity of external publicity. Product quality directly affects the agricultural purchasing behavior of customers. Reasonable prices of agricultural products are more likely to promote the purchase behavior of customers. Service level refers to the level of technical services, after-sales service level, and the better the level of service enterprises is, the more strong the purchase behavior of the masses is. Management level is defined as the means of enterprise management and the management level of enterprises in a large extent affects the sales of agricultural products. Market share indicates that the proportion of enterprises shared in the market, which is also represented by the strength and competitiveness of enterprises.

(3) **Social** factors

Social factors mainly include cultural differences, economic level, government policy, competitive ability, etc. [20]. Cultural differences refer to different population and culture in various regions. The economic level mainly refers to the level and scale of economic development between different regions, which determines the agricultural products demand. Government policy refers to the government attitude towards enterprise development and agricultural customers demand, and it has shown that the government's system and policy guide the direction of the enterprises' development and agricultural customers' choice and plan development of agricultural products. Competition ability refers to the degree of competition between agricultural enterprises.

(4) **Logistics** factors

Logistics has a dominant position in the agricultural supplies distribution, mainly including logistics services, transportation, delivery cycle, distribution mode, vehicle scheduling and some other factors [18-20]. Logistics service is described as the logistics scale and professional advantages in the process of logistics distribution of enterprises, which helps to improve commodity turnover rate, and save the enterprise cost. Transportation is related to the choice of what kind of vehicle to deliver agricultural products, and the quality of transport condition directly affects the customer satisfaction. A reasonable delivery cycle can reduce the distribution cost, optimize the inventory, and improve the service level and customer satisfaction. Distribution mode means a kind of way or strategy adopted by agricultural enterprises, and an effective and efficient logistics distribution mode can provide customers with high-quality agricultural products and service quality. Reasonable vehicle scheduling can guarantee the completion of the transportation tasks on schedule, and promote the orderly conduct of transport with the minimal transport capacity.

According to the above analyses for the impact factors, the indexes used to measure the potential demand in the distribution process of agricultural supplies are shown in Table 1, in which the possible value range of all indicators is also given.

3. Model of Predicting the Potential Demand of Agricultural Supplies Using RS-SVM.

3.1. Design of the hybrid algorithm RS-SVM. Rough set theory, proposed by Pawlak in 1982, is mainly used to solve incomplete and uncertain problems as a mathematical tool which has been developed extensively [21]. It aims to reveal the hidden rules from the known knowledge. The fault tolerant ability and generalization ability of rough set theory have been proved to be weak, which can only deal with the discrete data, but it is the advantage of SVM method. The SVM cannot determine the redundant information in the data and fails to reduce the space dimension, so the training time of SVM will cost more when the data dimension is very large, which reduces the speed of machine learning and affects the demand predicting accuracy [22,23]. Additionally, rough set theory does

TABLE 1. Definitions of the indexes and their possible value

Definitions of the indexes (notation)	**Possible value**
The quality of agricultural products (F_1)	perfect, good, general, poor, bad
The price of agricultural products (F_2)	RMB
Transport condition (F_3)	perfect, good, general, poor
Transport speed (F_4)	very fast, fast, general, slow, very slow
Delivery cycle (F_5)	day
Crop planting area (F_6)	hectare
The number of farmers (F_7)	household
The household density of farmers (F_8)	very high, high, average, low, very low
Average rate of contact (F_9)	per time
Income per capita (F_{10})	RMB
GDP per capita (F_{11})	ten thousand
Market share (F_{12})	percent
Competitive advantage (F_{13})	very strong, strong, average, weak, very weak
Reputation (F_{14})	perfect, good, general, poor, very poor
Service level (F_{15})	perfect, good, general, poor, very poor
Enterprise scale (F_{16})	very large, large, general, poor, very poor
Objective (Y)	The demand increment of agricultural supplies

not require any prior knowledge and can remove the redundancy information by finding the relationship between data and meanwhile retaining the basic information. Thus, the attribute reduction of rough set exactly makes up for the shortcoming of SVM.

The construction process of predicting the potential demand in the logistics distribution of agricultural supplies using RS-SVM is as follows. Since there are not only the discrete attribute data, but also the continuous attribute data, the first step aims at the discretization of initial sample data. Then under the premise of without affecting the classification results, rough set theory is used for the high dimensional index reduction. Followed by it, SVM is trained to perform the classification of potential customers' demand, i.e., the prediction for the potential demand.

3.2. **The establishment of the potential demand model.**

3.2.1. *Data discretization.* Since rough set fails to deal with the continuous data, it is important to discretize data firstly in the practical application and then RS is applied subsequently. This paper adopted Boolean operation method and using the ROSETTA software is available to achieve the discretization of the data.

3.2.2. *Attribute reduction.* Many attribute reduction algorithms have been proposed in [24,25], in which genetic algorithm is common to perform the attribute reduction. GA is a type of adaptive random search method based on the attribute dependent degree to avoid the blind random generated population, reduce the search space, accelerate the convergence speed, and also improve the accuracy of attribute reduction [26]. The basic steps of reduction algorithm based on attribute importance are as follows.

Step 1: Initialize the data; i.e., discrete processing data.

Step 2: Calculate the dependence of the decision attribute on the conditional attribute. The attribute set is divided into the conditional attribute set $\boldsymbol{C}$ and the decision attribute set $\boldsymbol{D}$, such that $\boldsymbol{U/C}$ and $\boldsymbol{U/D}$ are obtained in which $\boldsymbol{U}$ is defined as the objective's set; in the following, *POSC(D)* is got and then the dependence $\boldsymbol{D}$ on $\boldsymbol{C}$ is calculated.

Step 3: Calculate the attribute importance. It means that $\boldsymbol{U/C_i}$ and $\boldsymbol{POSC_i(D)}$ are obtained by deleting the attribute a_i from $\boldsymbol{C}$, and finally the importance attribute σ_i of a_i is obtained.

Step 4: Reduce attribute. Based on the importance value of attribute obtained in Step 3, if $\sigma_i \leq \sigma_0$, a_i is considered to be insignificant. Repeat Step 4 until all conditional attributes are judged and end.

Step 5: Output results.

3.2.3. *Data preprocessing.* Since the dimension of different indicators is various and the possible value may be different from each other, the original data should be standardized before predicting the potential demand, which avoids the impact of values from different indicators on the predicting results. There are some common data standardization methods [27], for example, by dividing the maximum value, sigmoid function, trigonometric function sin or cos, ratio conversion method. Consider that ratio conversion method can be used to deal with the positive index and negative index, so it is adopted in this paper. The calculation of ratio conversion method is shown in Equation (1). Furthermore, we use the function "mapminmax" in MATLAB to achieve normalization.

$$T = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

where T denotes the objective data, X is the original data, and $X_{\max}$ and $X_{\min}$ are the maximum and minimum from the original data set.

3.2.4. *Determination of SVM kernel function and parameters.* This paper selects the radial basis function (RBF) as the kernel function due to its extensive use in [28]. Then, the selection of parameters is to find a set of the penalty factor c and kernel parameter g to improve the accuracy of the predicting model. The goodness of regression fit or classification is expressed by the penalty factor c, and the larger the value c is, the higher the goodness of regression fit or classification for the sample is. Meanwhile, the value of kernel parameter g is also related to the performance of SVM. If the value of g is very small, the number of the support vectors in SVM increases such that it is prone to lead to over-learning; however, a large enough g could cause under-learning. Therefore, the reasonable value for g is vital to achieve the good performance. Grid search algorithm is often used to optimize the value of the penalty factor c and kernel parameter g, and has been applied by many researchers [24].

$T = \{(x_1, y_1), \cdots, (x_l, y_l)\} \in (X, Y)^l$ is defined as the sample set, in which the input vector is $x_i \in X = R^n$, the output vector is $y_i \in Y = R$ and $i = 1, \cdots, l$, and the loss function can be described as $c(x, y, f)$. The objective is to conclude a function $y = f(x)$ from R^n, i.e., the determination of the dependence of y on the input variable x, and obtain the minimal expected risk $R(f) = \int c(x, y, f) dP(x, y)$. In order to calculate the loss function, the insensitive loss coefficient ε is introduced to describe the difference between the actual value y and the predicted value $f(x, \alpha)$. Firstly, the optimization function is defined as:

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

The constraints are given by

$$w \cdot \phi(x) + b - y_i \leq \xi_i + \varepsilon, \quad i = 1, 2, \cdots, l \tag{3}$$

$$y_i - w \cdot \phi(x) - b \leq \xi_i^* + \varepsilon, \quad i = 1, 2, \cdots, l \tag{4}$$

$$\xi_i, \xi_i^* \geq 0, \quad i = 1, 2, \cdots, l \tag{5}$$

where ξ_i, ξ_i^* are relaxation factors, $\phi(x)$ is the nonlinear mapping function, w denotes the weight vector and b is a threshold. To solve Equation (2), a Lagrangian parameter is applied into Equation (6) and obtained as:

$$\begin{aligned} L(w, b, \alpha, \alpha^*) = \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{l}(\xi_i + \xi_i^*) - \sum_{i=1}^{l}\alpha_i(\xi_i + \varepsilon - y_i + w \cdot \phi(x_i) + b) \\ - \sum_{i=1}^{l}\alpha_i^*(\xi_i^* + \varepsilon + y_i - w \cdot \phi(x_i) - b) - \sum_{i=1}^{l}\eta_i(\xi_i + \xi_i^*) \end{aligned} \tag{6}$$

where α_i, $\alpha_i^* \geq 0$, $i = 1, 2, \cdots, l$. So, the dual form of the original optimization problem can be written as:

$$\min_{\alpha_i, \alpha_i^*} \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i - \alpha_i^*)\left(\alpha_j - \alpha_j^*\right)K(x_i - x_j) - \sum_{i=1}^{l}y_i(\alpha_i - \alpha_i^*) + \varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) \tag{7}$$

$$\text{s.t.}, \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0, \quad 0 \leq \alpha_i, \quad \alpha_i^* \leq c, \quad i = 1, \cdots, l \tag{8}$$

The nonlinear regression function is obtained by solving Equation (7), which is shown as:

$$f(x) = \sum(\bar{\alpha}_i - \bar{\alpha}_i^*)K(x_i, x_j) + b \tag{9}$$

where the kernel function $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ is used to represent the kernel function, and $\sum$ is the output function.

3.2.5. *The process of RS-SVM model.* The following are the steps of RS-SVM model used to predict the potential demand of agricultural supplies.

(1) The discretization of the sample data. The Boolean operation method configured by the ROSETTA software is used to realize the discretization of the data in this paper.

(2) Genetic algorithm configured by the ROSETTA software performs the attribute reduction and saves the obtained data after the reduction.

(3) The normalization of data, which can be finished by the normalization function "mapminmax" in MATLAB as mentioned in the above.

(4) Dividing the data after the normalization into two parts randomly, i.e., the training data and the test data, in order to train and test the predicting model.

(5) Using the model's training function "svmtrain" in the libsvm-mat toolbox aims to train the training data, so the kernel function c and the kernel parameter g can be obtained and determined. It implies that the predicting model is available. The following is the usage of the function "svmtrain":

Model = svmtrain(training_label_vector, traing_instance_matrix, ['libsvm options'])

(6) Realize the prediction based on the obtained predicting model and the test data, and the accuracy of the predicting model can be obtained.

4. **Numerical Examples.** Based on the process of RS-SVM model mentioned in Section 3, this section gives the case study by taking area of Yongnian County in China as the research object.

4.1. **Index reduction based on GA.** According to the index system in Table 1, we investigated 20 farmers to analyze the effect degree of each indicator on the potential demand, which is subject to four levels. They are very important, important, general, and unimportant, which are represented by 4, 3, 2 and 1, respectively. Also, there are three kinds of decisions according to the occurrence frequency of the potential demand, i.e., more frequently, occasionally and never. Similarly, numbers 3, 2 and 1 are used to express them. Table 2 shows farmers' opinion on the effect of indicators on the potential demand. The following step is to perform the attribute reduction using GA reduction rules in ROSETTA, and the reduction results are $\{F_1, F_2, F_5, F_8, F_9, F_{14}\}$.

TABLE 2. Decision table of the original indicators for predicting the potential demand of agricultural supplies

No.	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	F_{13}	F_{14}	F_{15}	F_{16}	D
1	4	3	3	2	2	2	3	4	3	4	4	2	3	1	3	2	3
2	3	3	4	2	3	2	2	3	3	3	3	1	3	2	1	3	2
3	3	4	2	3	4	3	2	3	2	2	3	2	4	3	2	2	2
4	4	3	2	3	3	3	2	2	3	3	3	2	2	4	3	1	3
5	2	3	3	4	2	2	3	4	1	3	3	4	3	3	3	3	2
6	3	2	3	2	2	3	4	4	1	4	2	3	2	3	3	4	1
7	4	3	2	3	2	4	3	3	2	3	4	3	3	4	3	3	2
8	3	3	2	4	2	3	4	3	3	2	4	4	3	3	3	4	3
9	4	3	4	2	3	4	4	3	4	2	4	4	3	3	3	2	2
10	3	4	3	3	3	4	4	4	3	3	3	3	2	2	2	1	1
11	2	3	3	3	4	3	4	3	2	3	3	3	3	3	3	2	3
12	3	4	3	3	2	3	2	3	4	3	4	2	3	3	3	2	2
13	3	2	3	4	3	3	2	3	3	3	2	2	3	4	3	2	2
14	4	3	2	3	3	2	3	3	2	3	4	2	3	3	3	2	2
15	3	4	4	3	4	3	2	2	2	3	4	3	3	2	3	3	3
16	3	2	3	3	4	2	2	3	2	3	3	2	3	3	2	3	3
17	4	3	3	3	4	3	4	3	3	2	4	4	3	3	2	3	2
18	3	4	3	3	3	4	3	4	2	2	3	3	4	2	4	2	1
19	3	4	3	3	4	2	3	1	2	3	4	4	3	4	4	4	1
20	3	2	3	4	4	4	2	3	4	3	4	3	3	2	3	2	3

4.2. **Forecasting based on RS-SVM potential agricultural supplies demand.** 16 samples with the occurrence of the potential demand are selected as the training data and 6 test samples are selected to predict their potential demand. The data collected are shown in Table 3, in which the values of the quality of agricultural products are 5 (perfect), 4 (good), 3 (general), 2 (poor), 1 (bad); the unit of the price of agricultural products is RMB/kg; the delivery cycle is calculated by per day; the household densities of farmers are denoted by 5 (very high), 4 (high), 3 (average), 2 (low), 1 (very low); average rate of contact is measured by times per day; reputation is expressed by 5 (perfect), 4 (good), 3 (general), 2 (poor), 1 (very poor); and the unit of the potential demand is kg.

For the original data in Table 3, the function "mapminmax" is run to achieve the normalization. After the normalization, we use the Libsvm toolbox on the normalized training samples for training. As mentioned in the above, an important issue of SVM is the selection of kernel function (g) and penalty parameter (c). In this paper, we first give the training results of the default parameters provided by the Libsvm toolbox, and then select some parameters to analyze the influence of these parameters on the prediction results. Finally, we propose a cross validation method to find the optimal results. In this paper, the performance of SVM is measured by the following two criteria: the root mean

TABLE 3. The original data of the reduced indicators for predicting the potential demand of agricultural supplies

No.	F_1	F_2	F_5	F_8	F_9	F_{14}	D
1	5	32	1	4	15	5	75
2	4	24	1	4	10	5	15
3	4	40	2	3	5	4	25
4	5	32	1	5	20	3	25
5	5	32	1	4	15	4	50
6	4	24	2	4	10	5	40
7	3	18	2	3	5	3	10
8	4	32	1	4	5	4	25
9	5	40	1	5	20	5	100
10	5	40	1	4	10	3	25
11	4	26	2	4	10	5	50
12	4	26	2	3	15	5	40
13	5	40	1	3	5	5	50
14	5	40	1	5	15	5	75
15	4	32	2	4	10	3	30
16	4	32	2	3	5	4	25
17	4	40	2	4	10	2	0
18	3	32	2	3	5	3	0
19	4	26	3	2	5	4	0
20	4	26	3	3	5	2	0
21	3	24	2	4	10	3	0
22	3	24	1	3	10	2	0

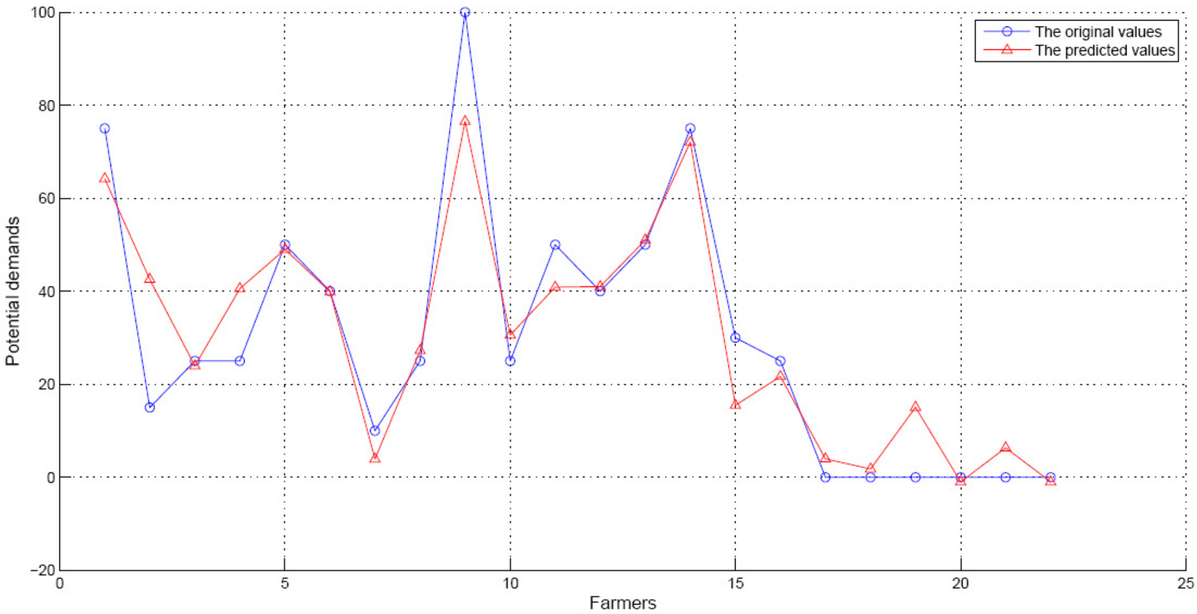

FIGURE 1. The prediction results with default SVM parameters, where the blue and red lines denote the original and predicting values, respectively

square error (RMSE) and the squared correlation coefficient (R) [29,30]. For RMSE, the smaller value represents the better performance of the SVM prediction. However, the larger R shows the better performance of the predicting model. The train runs and ends until 1.036336 seconds, and we obtained RMSE = 0.0106828, R = 86.5512%. The predicted regression results are shown in Figure 1, and the absolute error and relative error are shown in Figures 2 and 3, respectively.

From the prediction results with default SVM parameters, it can be seen that the reduced indicators can achieve a fitting result of 86.5512%. This verifies the effectiveness of the reduced indicators to some extent. However, there are still some absolute errors and relative errors, as shown in Figures 2 and 3. In order to analyze whether it is caused

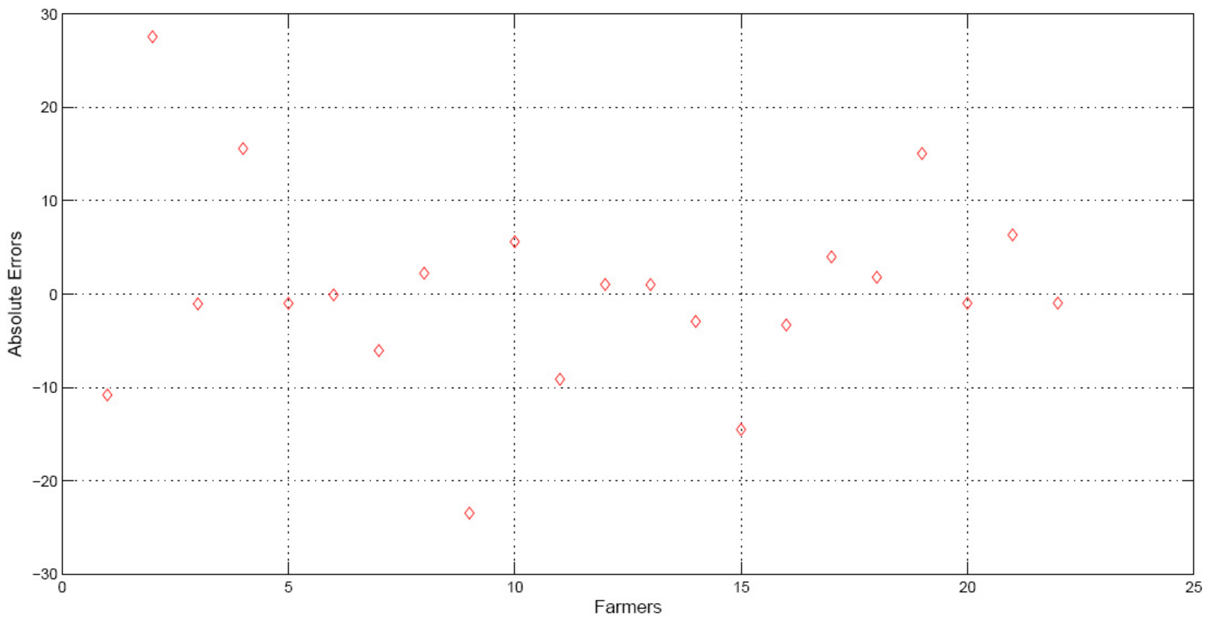

Figure 2. The prediction absolute errors with default SVM parameters

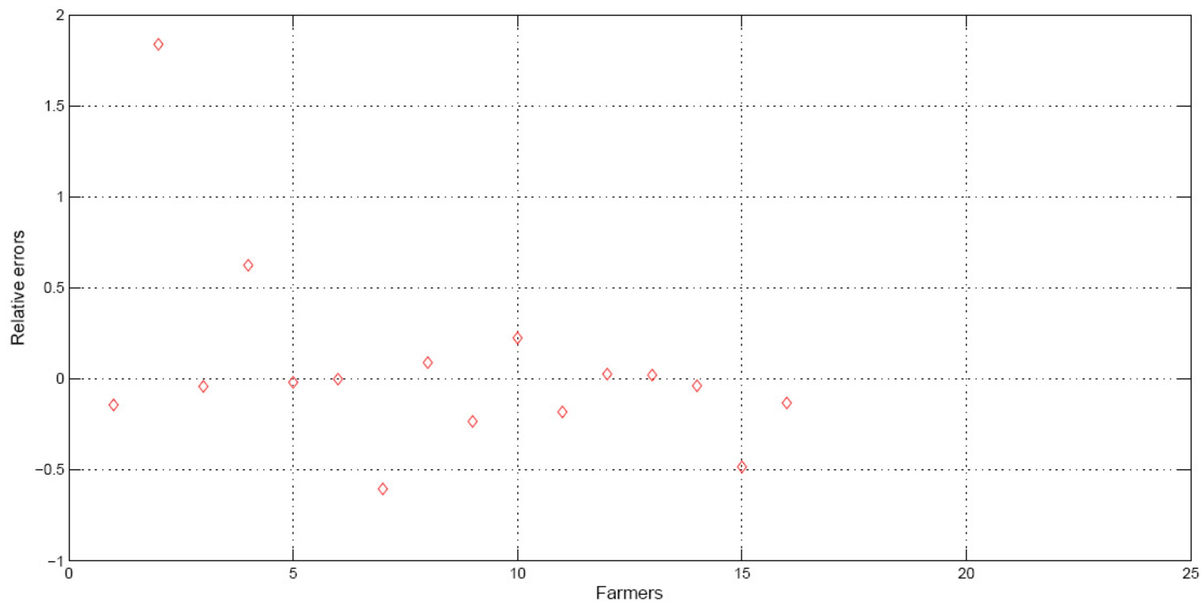

Figure 3. The prediction relative errors with default SVM parameters

by the parameters (g) and kernel function (c), this paper uses cross validation method to find the optimal SVM parameters. The pseudo code of this method is as follows:

```
GS_cg
Input training data
Start
%Initialization parameters
bestmse = Inf;
bestc = 0;
bestg = 0;
% Construct cg searching space
for c = 2^(cmin) : 2^(cmax)
    for g = 2^(gmin) : 2^(gmax)
        Use the train and train_label to train SVMs
        Record the mse into currentmse
        if (currentmse < bestmse)
            bestmse = currentmse; bestc = c; bestg = g;
        end
    end
end
Over
```

where $cmin$, $cmax$, $gmin$ and $gmax$ are the minimum, maximum values of the penalty parameter (c) and kernel function parameter (g), which determine the search space of cross validation. Set $cmin = 2\hat{}(-15)$, $cmax = 2\hat{}(15)$, $gmin = 2\hat{}(-15)$, and $gmax = 2\hat{}(15)$, and then the search space for cross validation is determined. According to the above cross validation method, the optimal penalty parameter (c) and kernel function parameters (g) are obtained after 14.588219 seconds, i.e., 1 and 181.019.

Using the optimal penalty parameter (c) and kernel function parameters (g), we train the training data and obtain RMSE = 0.0001376, R = 93.2908% after 0.981537 seconds.

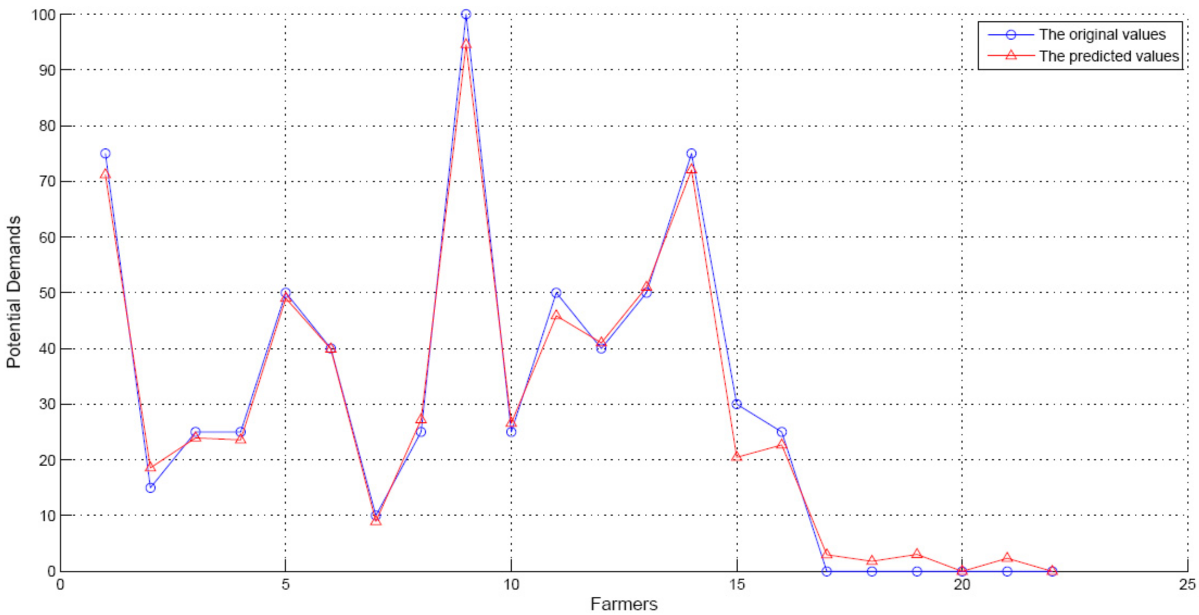

Figure 4. The prediction results with optimized SVM parameters, where the blue and red lines denote the original and predicting values, respectively

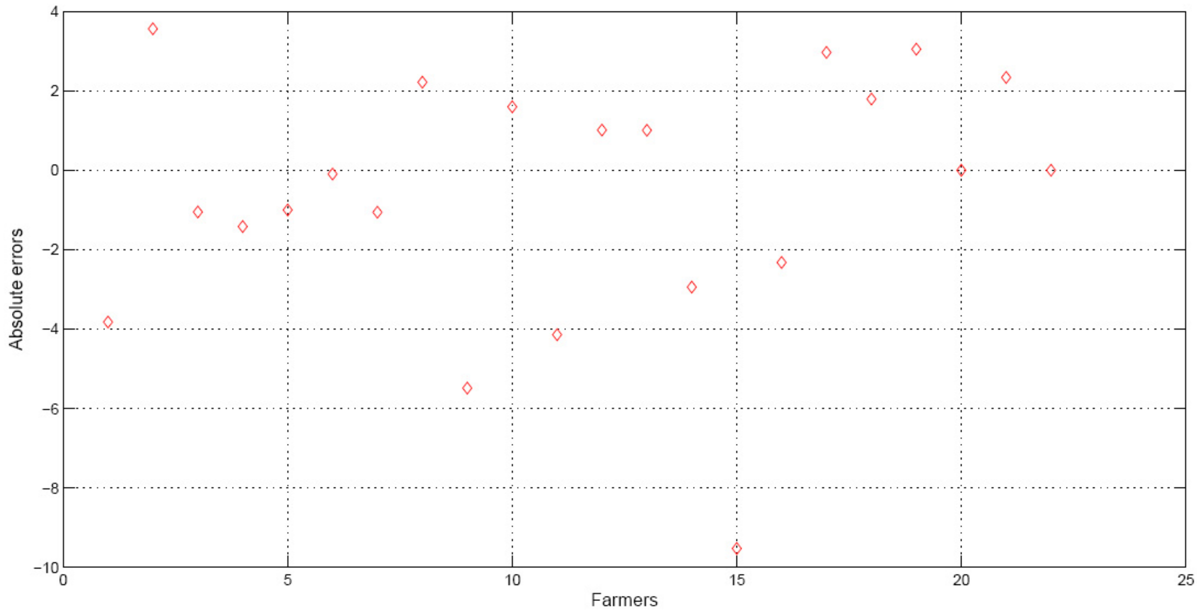

FIGURE 5. The prediction absolute errors with optimized SVM parameters

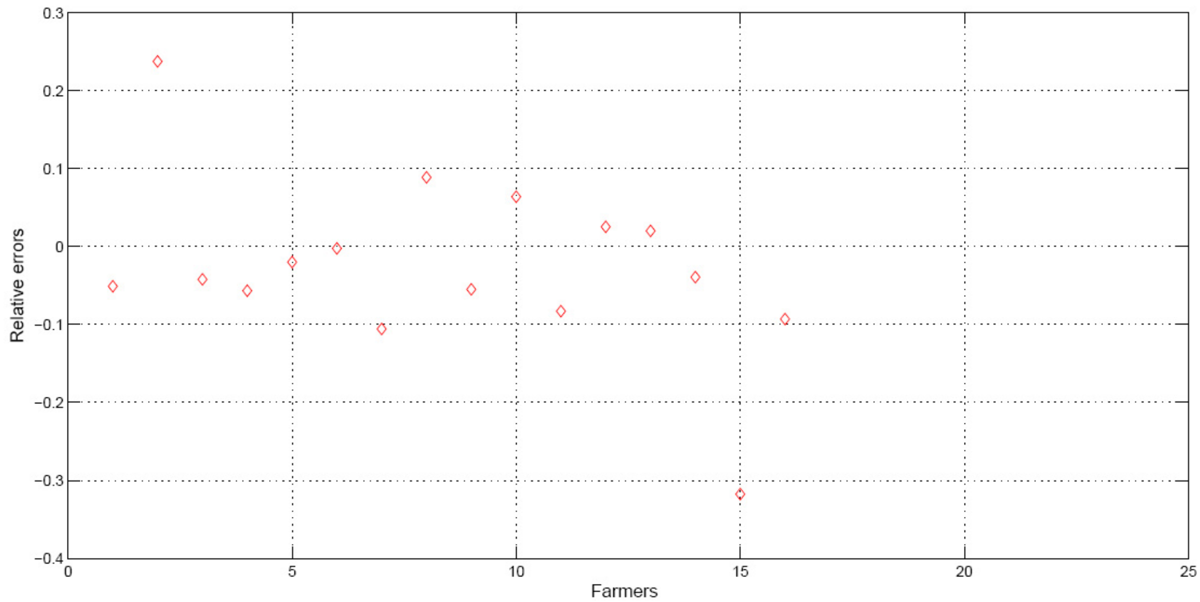

FIGURE 6. The prediction relative errors with optimized SVM parameters

The predicted regression results are shown in Figure 4, and the absolute error and relative error are shown in Figures 5 and 6. From results in Figures 4-6, the prediction accuracy after optimizing the SVM parameters is higher than the results of the default SVM parameters, which supports the effectiveness of the proposed predicting model.

5. **Conclusions.** There are many uncertain situations in the logistics distribution of agricultural supplies, especially the possible potential demand of customers. It is a key issue because it directly relates to a class of resources optimization and scheduling, for example, the vehicle scheduling. Thus, this paper focused on the prediction of the potential demand of customers of agricultural supplies. The index system for predicting the potential demand of customers including 16 indicators is firstly established, and then a hybrid predicting method RS-SVM is designed by taking the attribution reduction of RS

and the applicability of SVM prediction model into account. In the proposed method, the continuous sample data are firstly discretized by using Boolean operation, and then GA is used to reduce attributes from 16 indicators to 6 indicators. Consequently, SVM is trained to find the optimal kernel function and penalty parameter through MATLAB software, after which the optimal prediction model is obtained. The results obtained from numerical examples illustrated the applicability of the hybrid method RS-SVM and the comparison results show the proposed prediction model has a higher prediction accuracy compared to the traditional method. Our further studies will focus on the vehicle scheduling with consideration of potential demand, and more applications of the proposed prediction model.

Acknowledgment. This research is partially supported by the National Natural Science Foundation of China (Nos. 71273267 and 71303243), Education Ministry Social Science Foundation (No. 13YJC630175), and Specialized Research Fund for the Doctoral Program of Higher Education (No. 20130004120002).

REFERENCES

[1] C. R. Baines, Thermal properties of foods and agricultural materials, *Journal of Thermal Biology*, vol.7, no.10, pp.2139-2156, 1981.
[2] J. X. Zang, W. C. Jiang and Y. G. Shen, Application and forecasting of agricultural materials logistics demand, *Hubei Agricultural Sciences*, no.8, pp.1985-1987, 2015.
[3] T. Zhang, An empirical analysis on agricultural materials logistics control and agricultural products safety: A case study of bi-chains management model for veterinary drugs in Pinggu District, *Asian Agricultural Research*, vol.5, no.2, pp.79-81, 2013.
[4] Q. Ali and D. Van Oudheusden, Logistics planning for agricultural vehicles, *IEEE International Conference on Industrial Engineering & Engineering Management*, pp.311-314, 2009.
[5] W. Yang, Forecasting and analysis of agricultural product logistics demand in Tibet based on combination forecasting model, *Asian Agricultural Research*, vol.7, no.9, pp.17-22,27, 2015.
[6] Y. L. Xu, R. Y. Pan, J. Zhang et al., Forecasting regional logistics demand of agricultural products logistics park in Yu Hang county based on artificial neural network, *International Conference of Logistics Engineering and Management*, pp.55-60, 2010.
[7] E. Kim, W. Kim and Y. Lee, Combination of multiple classifiers for the customer's purchase behavior prediction, *Decision Support Systems*, vol.34, no.2, pp.167-175, 2003.
[8] L. Liu and G. Yu, Application of discovering the potential customers based on self-organizing feature map (SOM) neural network, *Journal of Nanchang University*, vol.30, no.5, pp.507-510, 2006.
[9] B. B. Guo and Z. B. Fang, Application of SVM in mining potential customers from web log, *Journal of Industrial Engineering & Engineering Management*, vol.24, no.1, pp.129-133, 2010.
[10] P. Egri, Information elicitation for aggregate demand prediction with costly forecasting, *Autonomous Agents and Multi-Agent Systems*, vol.30, no.4, pp.681-696, 2016.
[11] Q. Guan, H. An, X. Gao, S. Huang and H. Li, Estimating potential trade links in the international crude oil trade: A link prediction approach, *Energy*, vol.102, pp.406-415, 2016.
[12] M. Kryszkiewicz, Rough set approach to incomplete information systems, *Information Sciences*, vol.112, nos.1-4, pp.39-49, 1998.
[13] Y. Zhao, S. K. M. Wong and Y. Y. Yao, Attribute reduction in decision-theoretic rough set models, *Information Sciences*, vol.178, no.17, pp.3356-3373, 2013.
[14] Z. Pawlak and R. So-Winski, Rough set approach to multi-attribute decision analysis, *European Journal of Operational Research*, vol.72, no.3, pp.443-459, 1994.
[15] S. Tong and D. Koller, Support vector machine active learning with applications to text classification, *Journal of Machine Learning Research*, vol.2, no.1, pp.45-66, 2012.
[16] D. R. Chen, Q. Wu, Y. Ying et al., Support vector machine soft margin classifiers: Error analysis, *Journal of Machine Learning Research*, vol.5, no.3, pp.1143-1175, 2004.
[17] P. Lingras and C. Butz, Rough set based 1-v-1 and 1-v-r approaches to support vector machine multi-classification, *Information Sciences*, vol.177, no.18, pp.3782-3798, 2007.
[18] E. Hämäläinena, O.-P. Hilmola and L. Hetemäki, Fluctuating demand and its impacts to a paper producer: Customer analysis, *Expert Systems with Applications*, vol.42, no.13, pp.5779-5788, 2015.

[19] L. Li, T. Chi, T. Hao and T. Yu, Customer demand analysis of the electronic commerce supply chain using big data, *Annals of Operations Research*, DOI: 10.1007/s10479-016-2342-x, 2016.
[20] M. L. Vonada, B. S. Bidner, K. E. Belk, F. K. McKeith, W. R. Lloyd, M. E. O'Connor and G. C. Smith, Factors influencing consumer demand for U.S. pork exported to the Republic of Korea (South Korea), *Journal of Animal Science*, vol.79, no.4, pp.907-911, 2001.
[21] Z. Pawlak, Rough set, *International Journal of Computer & Information Sciences*, vol.11, no.5, pp.341-356, 1982.
[22] W. Huang, Y. Nakamori and S. Y. Wang, Forecasting stock market movement direction with support vector machine, *Computers & Operations Research*, vol.32, no.10, pp.2513-2522, 2005.
[23] H. L. Tian, H. Li and B. Xu, Support vector machine prediction based on improved artificial fish swarm algorithm, *Computer Engineering*, vol.39, no.4, pp.222-225, 2013.
[24] A. R. Hedar, J. Wang and M. Fukushima, Tabu search for attribute reduction in rough set theory, *Soft Computing*, vol.12, no.9, pp.909-918, 2008.
[25] F. Wang, J. Liang and C. Dang, Attribute reduction for dynamic data sets, *Applied Soft Computing*, vol.13, no.1, pp.676-689, 2013.
[26] S. Chebrolu and S. G. Sanjeevi, Attribute reduction in decision-theoretic rough set models using genetic algorithm, *Swarm, Evolutionary, and Memetic Computing*, Springer, Berlin Heidelberg, pp.307-314, 2011.
[27] K. Tanioka and H. Yadohisa, Effect of data standardization on the result of k-means clustering, *Challenges at the Interface of Data Analysis, Computer Science, and Optimization – Proc. of the Conference of the Gesellschaft Für Klassifikation E. V. Karlsruhe*, pp.59-67, 2010.
[28] J. Nong, Radial basis function neural networks optimization algorithm based on SVM, *Communications and Information Processing*, pp.291-298, 2012.
[29] Y. Yao, L. Zhang, Y. Liu et al., An improved grid search algorithm and its application in PCA and SVM based face recognition, *Journal of Computational Information Systems*, vol.10, no.3, pp.1219-1229, 2014.
[30] F. F. Chen, Sensitivity of goodness of fit indexes to lack of measurement invariance, *Structural Equation Modeling*, vol.14, no.3, pp.464-504, 2007.