Prediction Model of Nonlinear Combination Based on Support Vector Machine

Yuping Yuan, Zenglong An and Yanting Sun

***Abstract*—**Grain yield prediction is a kind of randomness and complexity and has a strong nonlinear prediction problem. Adopting an intelligent optimization algorithm for support vector machines and combined forecasting technology. First, using the SOM method of self -organizing neural network to discretization attribute in order to establish information systems and decision table. Second, Transform determining weight coefficient into the evaluation of attribute significance among standard rough set theory, work out weight coefficient of single model amid combination prediction model. Use constructed combination prediction model, predict the historical data of grain gross output. At last, in order to reduce the risk of declining grain production, increase punish to the risk of grain production. The prediction model of support vector regression machine based on Ramp controlled asymmetric loss function is established. It shows the high accuracy of constructed combination prediction model in predicting the grain output.

Keywords—combination forecasting, rough set, support vector machines (SVMs)

I. INTRODUCTION

Fire Grain yield prediction was complex nonlinear relationship, with randomness and mutation. it is difficult to describe by using traditional linear model, and this model has low prediction accuracy. Combination forecasting method has proved a linear combination of the various models can effectively improve the model fit capability and enhance the prediction accuracy under certain conditions. Currently, the combination forecasting method has two main directions: First, learn to combine optimization algorithm with the traditional single forecasting model [1]-[3]. The second, make multiple single model combined by linear combinations [4]-[5], construct combination forecasting model by an appropriate form of weighted average, in which the reasonable weighting factor will greatly improve the prediction accuracy.

The key and the difficulty of combination forecasting model lie in determination of the weight coefficient. In the literature [6]-[7] the weight method of combination forecasting model has been studied using rough set theory , gives the determining step for the weight coefficient of the combined predictive model , but the discretization of attribute data has not been elaborated in detail ,resulting in the lack of applicability of the method . Meanwhile, in order to overcome the shortcomings of the traditional prediction method in the model selection, the prediction accuracy and stability of grain output are improved.

Employ new data mining method - support vector machine which present the unique advantage in solving small sample size, nonlinear and high dimensional pattern recognition problem, and this method also shows a good ability to solve problems in many larger data sets. In order to overcome deficiency of the food production, support vector machine model, a linear regression model, exponential model, Gaussian model and logarithmic functions model are selected to constitute combination forecasting model . Using self-organizing neural network SOM method to discrete attribute characteristics, to establish a knowledge representation systems and decision tables. Transforming the problem of determining the weights coefficient into evaluation questions about attribute importance in standard rough set theory, and gives Matlab application program which make the attribute data discretization, calculates the dependence and importance of prediction method on prediction model as well as weight coefficient of each single model in combination forecasting model; Standard support vector machines are usually used in time series prediction, while standard support vector machines are mainly used to predict with $\varepsilon -$ insensitive loss functions [8]-[9]. The $\varepsilon -$ insensitive loss function has the nature of the same penalty for the positive and negative parts of the true value deviating from the predicted value. In the process of minimizing the empirical risk and regularization item in the support vector machine, the regression function will be deviated and the generalization ability of the support vector regression machine is affected. By setting up $\varepsilon_1 < \varepsilon_2$, in order to increase the punishment for falling risk to reduce the risk of downfall, the penalty for the real value deviating from the negative part of the prediction value is greater than that of the deviation from the positive part. The prediction model of the support vector regression machine based on the Ramp controlled asymmetric loss function is used to predict the grain yield. Using constructed combination prediction model to carry out combination prediction of the historical data of grain output in Heilongjiang reclamation area. The analysis shows that established combined model is highly accurate for predicting

This work was supported by the project of Daqing's science and technology instruction (Grant: zd-2016-078); The project of Heilongjiang August first land reclamation university fostering fund (Grant: XZR2017-14).

Yuping Yuan is with the College of Sciences, Heilongjiang Bayi Agricultural University, Daqing 163319, China.

Zenglong An is with the College of Economics & Management, Heilongjiang Bayi Agricultural University, Daqing 163319, China (corresponding author; e-mail: bynd_yyps@sina.com).

Yanting Sun is with the School of Civil Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China.

grain output. It is highly consistent with the actual value , the average absolute error of the forecasting results is significantly lower than traditional model methods.

II. DETERMINING WEIGHTS COEFFICIENT

Assuming there is m kinds of prediction model to predict a same object , the combination forecasting model constituted by m kinds of single prediction model be taken as .

$$\hat{y}_t = \sum_{i=1}^{m} k_i \cdot y_t^{(i)} \tag{1}$$

Where, $\hat{y}_t$ indicates the prediction value in t Moment of combination forecasting model $\hat{y}_t^{(i)}$ represents the predicted values $(i = 1, 2, \cdots, m)$ of No. i prediction model in t moment , k_i Represents weight coefficient of No. i predictive model in moment $(i = 1, 2, \cdots, m)$, $\sum_{i=1}^{m} k_i = 1$, and $k_i \geq 0$ .The fitted values of each single model in combination forecasting model be taken as conditional attributes $C = \{\hat{y}^{(1)}, y^{(2)}, \cdots, y^{(m)}\}$,The observed value of forecasting object be taken as the decision attribute $D = \{y\}$, $U = \{u_1, u_2, \cdots, u_n\}, u_t = (\hat{y}^{(1)}, y^{(2)}, \cdots, y^{(m)}, y_t)$, $t = 1, 2, \cdots, n$, Where in , $\hat{y}^{(1)}, y^{(2)}, \cdots, y^{(m)}$ and y_t are respectively the fitted value of the single predictor model and historical data of the predicted objects . Within the observation period, the two-dimensional data table constituted altogether by the fitted values of each single model and historical data is a information system relating to the combination forecasting model , in the table each row describes an object , and each column describes an attribute of an object .Use Rough set theory to analyse the degree of importance of each single model , based on the premise that to build knowledge representation system can be fulfilled after the discretization of condition attribute values.Therefore, the discretization of continuous attributes is an important part in the practical application of rough set theory .

Currently the typical methods on continuous data discretization are mainly based on hierarchical clustering method , genetic algorithms, conditional information entropy and self-organizing neural networks (SOM) and other discrete methods respectively [10] . In this paper, self-organizing feature map neural network SOM is used to process discretization of continuous attribute values .

Calculating the weight coefficient of each single model:

(1) Calculate the dependence of decision attribute D on condition attribute C :

$$k = \gamma_C(D) = \sum_{i=1}^{m} \left|POS_C(y_i)\right| \Big/ |U| \tag{2}$$

Among them, $C = \{\hat{y}^{(1)}, y^{(2)}, \cdots, y^{(m)}\}$, $D = \{y\}$, $|POS_C(y_i)|$ decision attribute D is about positive domain of condition attribute C . cardinal number of U in $|U|$ set , represents the number of elements contained in the collection for the finite collection.

(2) Delete No. i prediction model, calculate dependence of decision attribute on condition attribute .

$$\gamma_{C-\{c_i\}}(D) = \sum_{i=1}^{m} \left|POS_{C-\{c_i\}}(y_i)\right| \Big/ |U|, i = 1, 2, \cdots, m \tag{3}$$

(3) Calculate the degree of importance of No. i predictive models in all of predictive models:

$$\sigma_{CD}(c_i) = \gamma_C(D) - \gamma_{C-\{c_i\}}(D), i = 1, 2, \cdots, m \tag{4}$$

(4) Calculate the weight coefficient of No. i predictive model $k_i = \dfrac{\sigma_{CD}(c_i)}{\sum_{i=1}^{m} \sigma_{CD}(c_i)}, i = 1, 2, \cdots, m \cdot$

III. NONLINEAR COMBINATION FORECASTING MODEL

3.1 Support vector machine prediction model

SVM (Support Vector Machine, SVM) is a new technique in data mining which is also a learning algorithm proposed by structural risk minimization principle in statistical learning theory [11], a new tool to solve machine learning by means of optimization method. The mathematical formulation of regression problems : According to the given training set

$$T = \{(x_1, y_1), \cdots, (x_l, y_l)\} \in \{X \times Y\}^l,\ x_i \in X \subset R^n,$$

$y_i \in R,\ i = 1 \cdots l$, Among them, the training set $X \times Y$ is based on the assumption of a uniform probability distribution of independent and identically distributed sample points ,And suppose given loss function $c(x, y, f)$, looking for real-valued function $f(x) = \omega \cdot x + b$ from a input space X to the output space Y , so that the expected risk reaches a minimum .

Original space robust support vector regression model. For the random training sample regression problem $T = \{(x_i, y_i)\}_{i=1}^{N}$, the $x_i \in R^n$ is the input index and the y_i is the corresponding function value. The support vector machine algorithm based on the ε- insensitive loss function is:

$$\begin{aligned} &\min_{w,b} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \\ &s.t \quad y_i - (w \cdot x_i + b) \leq \varepsilon + \xi_i \\ &\qquad w \cdot x_i + b - y_i \leq \varepsilon + \xi_i^* \\ &\qquad \xi_i, \xi_i^* \geq 0, i = 1, 2, \cdots, N \end{aligned} \tag{5}$$

Here C is a compromise that regulates the complexity of the model and the error of training. In (5), the relaxation variable

$\xi_i, \xi_i^*, i = 1, \cdots, N$ is eliminated, and the unconstrained optimization problem is obtained:

$$\min_{w,b} L_\varepsilon(w,b) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} H_\theta(z_i) \tag{6}$$

Here we have established two asymmetric control type Ramp loss function asymmetric forms:

$$H_\theta(z_i) = \min\left(\theta^2, H_A(z_i)\right)$$

Among

$$H_A(z_i) = \begin{cases} 0, & -\varepsilon_1 \le z_i \le \varepsilon_2 \\ (z_i - \varepsilon_2)^2, & z_i > \varepsilon_2 \\ (-z_i - \varepsilon_1)^2, & z_i < -\varepsilon_1 \end{cases}$$

For the simplicity, the following derivation does not consider the influence of bias b without affecting the generalization performance of the model. At the same time, the kernel function strategy is introduced and transformed to Eyre Bert space H to get the optimization problem:

$$\min_f L(f) = \frac{1}{2}\|f\|_H^2 + C\sum_{i=1}^{N} H_\theta(z_i) \tag{7}$$

By the expression theorem [12], (6) the optimization function $f(x)$ of the type can be expressed as:

$$f(x) = \sum_{i=1}^{N} \beta_i k(x, x_i) \tag{8}$$

Replace (8) into (7) and get:

$$\min_\beta L(\beta) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \beta_i\beta_j k(x_i, x_j) + C\sum_{i=1}^{N} H\left(\sum_{j=1}^{N}\beta_j k(x_i, x_j) - y_i\right) \tag{9}$$

Order $\beta = [\beta_1, \beta_2, \cdots, \beta_N]^T$, $z_i = \sum_{j=1}^{N}\beta_j k(x_i, x_j) - y_i$

K is a kernel matrix, among $K_{ij} = k(x_i, x_j)$,

$i, j = 1, 2, \cdots, N$, The (9) form can be transformed into:

$$\min_\beta L(\beta) = \frac{1}{2}\beta^T K\beta + C\sum_{i=1}^{N} H(z_i) \tag{10}$$

among $z_i = K_i^T\beta - y_i$.(Detailed process submitted for publication [13]).

Algorithm 3.1(The prediction model algorithm of support vector regression machine based on Ramp controlled asymmetric loss function modeling):

Step 1: Input training sample $\chi = \{(x_i, y_i)\}_{i=1}^{n}$; selecting any small positive real number $\rho, \varepsilon_1, \varepsilon_2, \theta, h$, $k = 0$;

Step 2: First, the initial $f^0(x)$ is trained on a subset of the set χ;

Step 3: The training samples are divided into seven categories according to the size of the

$$|z^0| = |f^0(x) - y|$$

region: $SV_1, SV_2, SV_3, SV_4, ESV_1$, ESV_2, NSV;

Step 4: Calculate the gradient $\nabla^n L(\beta)$, if $\|\nabla^n L(\beta)\| \le \rho$, then stop the calculation, otherwise turn one step;

Step 5: Let $\lambda^k = \delta w^i, \delta > 0$ is the initial step, and i is the smallest nonnegative integer to set up , $b_1 \in (0,1)$

$$L(\beta^k) - L(\beta^k + \delta w^i d_k) \ge -b_1\delta w^i d_k^T \nabla L(\beta^k)$$,;

Step 6: According to the (10) formula, the function is obtained by (8), then step 2.

3.2 Establish a single model in combination forecasting model

Using cross validation method to get the ideal parameters, selection of crop sowing area, agricultural fertilizer, agricultural machinery power, disaster area, effective irrigation area, engaged in agricultural production and the total number of main affecting factors (see Table 1) as input indicators. The grain output of Heilongjiang reclamation area in Heilongjiang province is divided into two groups, 1990-2002 year data as training set and 2003-2012 year as a forecast set. Select the radial basis kernel function:

$K(x, y) = \exp(-\|x - y\|^2 / \sigma^2)$, Using the grid search method, the optimal parameter of μ and σ is selected from $\{2^{-10}, 2^{-9}, \cdots, 2^{10}\}$, and the robust support vector regression model of the original space is obtained.

According to food production trends a certain area under 1990-2006 , the agricultural chemical fertilizer factor regard as factor affecting food production, to determine the support vector machine regression model , a linear regression model, exponential model , Gaussian model and logarithmic function model as 5 single prediction models in combination forecast model.

Model 1: $\hat{y}_t^{(1)} = \sum_{i=1}^{1}(\alpha_i^* - \alpha_i)K(x_i - x) + b^*$,Where the parameter values $\varepsilon = 0.002, C = 500, \sigma = 4, \upsilon = 0.02$;

Model 2: $\hat{y}_t^{(2)} = 22.66 * x - 385.9$,Goodness of fit $R^2 = 0.9710$;

Model 3: $\hat{y}_t^{(3)} = 176.1 \times e^{0.0281 * x}$, Goodness of fit $R^2 = 0.9653$;

Model 4: $\hat{y}_t^{(4)} = 1083 * e^{(-((x-70.65)/35.34)^2)}$, Goodness of fit $R^2 = 0.9531$;

Model 5: $\hat{y}_t^{(3)} = 1114 * \ln x - 3587$, Goodness of fit $R^2 = 0.9481$.

Table 1 Statistics about all the influencing factors of total grain output

Year	grain output (million tons)	agricultural fertilizer (tons) X_1	agricultural machinery power (ten thousand kilowatts) X_2	grain sown area (1000 HA) X_3	agricultural labor force (thousands of people) X_4	effective irrigation area (ten thousand hectares) X_5	application of pesticides (tons) X_6
1990	460.3	174152	274.2	1636	44.2	16.66	5234
1991	366.6	181392	277.2	1647	45.2	18.36	5250
1992	374.9	189939	277.2	1463	46.2	20.35	5288
1993	402.0	178787	281.3	1635	40.3	23.35	6052
1994	414.4	179317	276.9	1621	36.97	22.38	4865
1995	514.6	198211	245.0	1609	38.96	28.60	5295
1996	715.6	229103	266.4	1732	40.83	43.82	5946
1997	852.0	243228	290.7	1812	41.75	61.22	6742
1998	868.8	257048	305.6	1861	42.19	71.98	7405
1999	905.3	256937	323.6	1846	40.92	76.33	7020
2000	814.1	259703	331.6	1819	42.17	77.41	6721
2001	860.8	254820	340.3	1832	45.20	78.99	6831
2002	810.6	260909	351.0	1793	43.78	82.98	6682
2003	755.3	261618	370.9	1639.9	44.0	74.9	7030
2004	937.5	300601	401.3	1875.9	46.6	89.2	8412
2005	1026.5	310988	433.6	1904.8	46.5	94.54	8364
2006	1132.2	339183	472.3	2083.7	49.7	105.97	9233
2007	1246.4	379139	519.3	2151.5	59.5	121.48	11022
2008	1420.6	394732	564.3	2296.4	61.1	125.39	11159
2009	1652.6	437102	604.5	2544.2	60.3	134.31	12103
2010	1818.0	483542	671.6	2702.9	60.2	155.74	13686
2011	2037.0	527655	745.6	2744.07	59.95	173.25	14846
2012	2105.0	579642	818.6	2797.784	59.87	184.09	14793

Among, Goodness of fit of the model 1-5, respectively 0.9133 、0.8910、0.8353、0.9031、0.9081 .The historical data about food production and the fitted values of each model are shown in Table 2 .

Table 2 Statistic of grain yield and forecast values of each model

Year	Grain output（million tons）	Model 1	Model 2	Model 3	Model 4	Model 5	Combination model
1990	460.3	472.1	416.3	476.2	400.4	386.3	440.457
1991	366.6	374.54	504.6	531.3	493	502.7	464.519
1992	374.9	368.81	482	516.6	468.5	474	447.234
1993	402	420.33	468.4	508	454	456.4	456.820
1994	414.4	433.64	434.4	487	418.7	411.2	438.739
1995	514.6	504.02	475.2	512.3	461.2	465.3	491.841
1996	715.6	690.42	611.1	606.3	613.3	628.6	650.434
1997	852	797.86	706.3	682.3	723.4	730.1	752.131
1998	868.5	863.02	797	763.5	824.6	819	834.257
1999	905.3	890.2	803.8	769.9	831.9	825.3	848.423
2000	814.1	840.92	797	763.5	824.6	819	826.221
2001	860.8	854.67	808.3	774.3	836.7	829.6	838.398
2002	810.6	834.26	842.3	807.6	872	860.8	852.316
2003	755.3	804.6	835.5	800.8	865.1	854.7	837.308
2004	937.5	936.99	1019	1005.5	1020	1010.6	994.919
2005	1027	1153.76	1032.6	1022.6	1028.2	1021.4	1081.4
2006	1132	1158.22	1179.9	1227.5	1080.9	1131.4	1160.4

For each value of the property is divided into several feature values, then substitute the property value for feature values. Use SOM network to discrete the data in Table into three level so that knowledge representation systems on the combination forecasting will be established(see Table 3).

Table 3 Forecast values of each model and forecast object(kW)

Domain of discourseU	Model 1	Model 2	Model 3	Model 4	Model 5	Grain output y
u_1	1	1	1	1	1	1
u_2	1	1	1	1	1	1
u_3	1	1	1	1	1	1
u_4	1	1	1	1	1	1
u_5	1	1	1	2	1	1
u_6	1	1	1	2	1	1
u_7	2	1	1	2	2	1
u_8	2	2	2	2	1	2
u_9	2	2	2	3	1	2
u_{10}	2	2	2	3	2	3
u_{11}	2	2	2	3	2	3
u_{12}	3	2	2	3	2	3
u_{13}	3	2	2	3	2	3
u_{14}	3	2	2	3	2	3
u_{15}	3	3	3	3	3	2
u_{16}	3	3	3	3	3	3
u_{17}	3	3	3	3	3	3

3.3 Establish Combination Forecasting Model

According to equation (2) calculating , get the dependence degree of the total grain output on five kinds of model $\gamma_C(D)$=15/17=0.8824 , According to equation (3) calculating, after deleting a predictive model , dependence degree of Total grain output on the remaining 4 models was got $\gamma_{C-\{c_i\}}(D)$, $i=1,2,\cdots,5$, According to formula (4) calculating, get the degree of importance of various forecasting models and weighting coefficients are shown in Table 4 .

Table 4 The dependence、important and weights of various prediction model

Model	Model 1	Model2	Model 3	Model 4	Model 5
$\gamma_{C-\{c_i\}}(D)$	0.647	0.882	0.765	0.765	0.706
$\sigma_{CD}(c_i)$	0.235	0	0.118	0.118	0.177
k_i	0.364	0	0.182	0.182	0.273

According to Weight coefficients in Table 4, combination forecasting model of total grain output in Heilongjiang Province can be established based on rough set theory of support vector machine:

$$\widetilde{\hat{y}_t} = 0.3636y_t^{(1)} + 0.1819y_t^{(3)} + 0.1819y_t^{(4)} + 0.2727y_t^{(5)}$$

Using the established combination forecasting model, predicting values on grain yield can be obtained under1999-2006 in Heilongjiang Reclamation area, see the last row shown in Table 2, 0.9133、0.8910、0.8353、0.9031、0.9081 .

The evaluation index of the performance of the algorithm is:

Root mean square error： $RMSE=\sqrt{\frac{1}{M}\sum_{t=1}^{M}(y_t-\hat{y}_t)^2}$.

Mean absolute error:

$$M_{MAE}=\frac{1}{M}\sum_{t=1}^{M}\left|y_t-\widetilde{\hat{y}_t}\right|,\quad M_{MAEneg}=\frac{1}{M}\sum_{t=1,y_t<\hat{y}_t}^{M}\left|y_t-y_t\right|.$$

Where y_t is the real output of grain in the year t ; $\hat{y}_t$ is a predictive value, M is the number of samples, M_{MAE} is the mean absolute error, M_{MAEneg} is the error of the real and predicted value of grain output. M_{MAEneg} 1 is increase in output of grain , M_{MAEneg} 2 is reduction in output of grain.

Table 5 Comparison of forecasting results of the algorithm 3.1 and the SVR algorithm

Algorithm	M_{MAE}	$RMSE$	M_{MAEneg} 1	M_{MAEneg} 2
ε-SVR	5.052	12.906	5.060	7.846
Algorithm 3.1	3.515	12.156	2.784	9.372

IV. CONCLUSION

Because combination forecasting model collects all useful information contained in each single prediction model contains, and thus on the overall it has more advantage than the single model in the following aspects: the ability to adapt to future changes , stability, and to reflect the trends and accuracy of the forecasting results and so on. Table 4 has showed that, the average absolute error 5.8901% caused by combination forecasting model based on the rough set , is smaller than average absolute error in 5 single prediction model ; The prediction model of the support vector regression machine based on the Ramp controlled asymmetric loss function is used to predict the grain yield, and the small error is obtained. Further, due to the diversity of each prediction model and sample data discretization method, making the data discretization difficult to be completely rational, which issue needs further study in the field of using rough set theory to determine combined forecasting model weights .

REFERENCES

[1] M. Tamimi, R. Egbert, “Short term electric load forecasting via fuzzy neural collaboration”, *Electric Power Systems Research*, vol.56, no.3,pp. 243-248,2000.

[2] N.L. Tai, Z.J. Hou , T. Li, “New principle based on wavelet transform for power system short-term load Forecasting”, *Proceedings of the CSEE*, vol.23 no.1,pp. 45-5,2003.

[3] Z.Q. Qi, Y.J. Tian and Y.Shi, “Successive overrelaxation for laplacian support vector machine”, *IEEE Trans on Neural Networks and Learning System*, vol.26,no.4,pp. 674-683,2015.

[4] Y.H. Shao, W.J. Chen and Z. Wang, “Weighted linear loss twin support vector machine for Large-Scale classification”, *Knowledge-Based Systems*, vol.73,no.1,pp. 276-288,2015.

[5] C.Q. Kang, C.Xu and Q. Xia , “ Novel Approach Considering Load-relative Factors in Short-term Load Forecasting”, *Electric Power Systems Research*, vol. 70, no. 2, pp. 99-107,2004.

[6] C. Chen, B.He and Y. Ye, “The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent”, *Mathematical Programming*, vol. 155, no. 2, pp. 57-79,2016.

[7] G.Y. Gu, B.S.He and X. M.Yuan, “Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems :a unified approach”, *Computer Optim. Apply*, vol. 59, no. 1, pp. 135 -161,2014.

[8] D. L. Liu, Y. Shi and Y.J.Tian , “ Ramp loss nonparallel support vector machine for pattern classification”, *Knowledge-Based Systems*,vol.85,no.1,pp. 224-233,2015.

[9] H.D. Zhu, Y. Zhong, " Feature Selection Method Based on Parallel Immune Cloning Algorithm", *Journal of Shanghai Jiaotong University*, vol.43, no. 12,pp. 1847-1852 ,2009.
[10] X.Q. Zhao, X.Y. Cao and Z.Q. Lan, "Study of the method for determining weighting coefficient of coal ash slagging fuzzy combination forecast based on rough set theory", *Journal of China coal society*, vol.29, no.2, pp. 222-225, 2004.
[11] X.G. Zhang, "Introduction to statistical learning theory and support vector machines", *Acta automatica sinica*, vol.26,no. 1,pp. 32-42,2000.
[12]P.Shige,Y.S.Song and J.F.Zhang ,"A complete representation theorem for G-martingales", *Stochastics an international journal of probability and stochastic processes* , vol.86,no. 4, pp. 609-631,2014.
[13] Y.P. Yuan, Z.L. An, "Support vector machine in the primal space based on the ramp loss function", *Journal of East China Normal University(Natural Science*) , vol.186, no.2, 20-29,2016.