A Grey SVM based model for Patent Application Filings Forecasting

Sheng Xu, Huifang Zhao and Xuanli Lv

Abstract—**Tracking historical levels as well as estimating future levels of patent applications is an ongoing activity of considerable significance. The patent applications filings (PAF) are complex to conduct due to its nonlinearity of influenced factors. Support vector machines (SVM) have been successfully employed to solve nonlinear regression and time series problems. Grey theory is a truly multidisciplinary and generic theory that deals with systems that are characterized by poor information and/or for which information is lacking. Grey system theory successfully utilizes accumulated generating data instead of original data to build forecasting model, which makes raw data stochastic weak, or reduces noise influence in a certain extent. However, the application combining grey system theory and SVM for PAF forecasting is rare. In this study, a grey support vector machines with genetic algorithms (GSVMG) is proposed to forecast PAF. In addition, Grey system is used to add a grey layer before neural input layer and white layer after SVM layer. Genetic algorithms (GAs) are used to determine free parameters of support vector machines. Evaluation method has been used for comparing the performance of forecasting techniques. The experiments show that the GSVMG model is outperformed grey model and SVM with genetic algorithms (SVMG) model and PAF forecasting based on GSVMG is of validity and feasibility.**

I. Introduction

Accurate forecasting of numbers of patent application filings is crucial for resource planning at one country Patent Office. Tracking historical levels as well as estimating future levels of patent applications is an ongoing activity of considerable significance. The development of demand for patent rights is of interest as well to economists who seek to map patterns of technological development. There are many traditional regression methods of forecasting patent applications that could be used as linear regression model, Cobb-Douglas functional model [1] [2] [3]. Artificial intelligence techniques for load forecasting are superior to traditional forecasting approach. However, the training procedure of an artificial intelligence model is time consuming. Therefore, some approaches were proposed to accelerate the speed of converge [4].

Manuscript received November 28, 2007. This work was supported by the National Natural Science Foundation of China (70602017), National Social Science Foundation of China (07BJY034) and Social Science Foundation of China Ministry of Education (07JC630061).

S. Xu is with the School of Management, Hefei University of Technology, Hefei, CO 230009 China (phone: 86-0551-2901489-8203; fax: 86-0551-2904971; e-mail: xssx69@163.com).

H.F. Zhao is with the School of Management, Hefei University of Technology, Hefei, CO 230009 China (e-mail: hfang_zhao@yahoo.com).

X.L. Lv is with the School of management, Xi'an Jiaotong University, CO 710049 China, (e-mail: xuanlilv@mail.xjtu.edu.cn).

The SVM is grounded in the framework of statistical learning theory, which has been developed by Vapnik and co-works [5] [6], which is a novel learning machine based on statistical learning theory, and which adheres to the principle of structural risk minimization seeking to minimize an upper bound of the generalization error rather than minimize the training error (the principle followed by neural networks). This induction principle is based on the bounding of the generalization error by the sum of the training error and a confidence interval term depending on the Vapnik-Chervonenkis (VC) dimension. Based on this principle, SVM achieves an optimum network structure by striking a right balance between the empirical error and the VC-confidence interval. This balance eventually leads to better generalization performance than other neural network models. Additionally, the SVM training process is equivalent to solving linearly constrained quadratic programming problems, and the SVM embedded solution meaning is unique, optimal and unlikely to generate local minima. Originally, SVM has been developed to solve pattern recognition problems. However, with the introduction of Vapnik's 6-insensitive loss function, SVM has been extended to solve nonlinear regression estimation problems, such as new techniques known as Support Vector Regression (SVR). Compared with the Artificial Neural Networks (ANN), it provides high generalization ability and overcomes the over-fitting problem.

Grey theory, first derived by Prof. Deng Julong [7], is a novel mathematical approach, particularly designed for handling situations in which only limited data are available. The fields covered by grey theory include systems analysis, data processing, modeling, prediction, decision making and control. The grey theory mainly works on systems analysis with poor, incomplete or uncertain messages. Grey forecasting models have been extensively used in many applications. This theory describes random variables as a changeable interval number that varies with time factors and uses color to represent the degree of uncertainty in a dynamic system. It implies that a grey environment is a system that consists of partially known and partially unknown information. It is believed that the uncertainties existing in the whitening process mainly come from the insufficiency of understandable information. Therefore, by way of increasing system information, the degree of uncertainty could be changed or diminished over time. The output from a grey environment must have a certain degree of implication for

978-1-4244-1819-0/08/$25.00©2008 IEEE

behavior in this system. The measured database with discrete features can be formulated as a typical solution based on a so-called pseudo-differential equation. With a solution from the proposed pseudo-differential equation, the next output from a grey environment can be predicted, even based on a few observed data. Since the grey model can characterize such an unknown system and can make predictions based on a few data, it has been effectively applied in many fields such as agricultural, socioeconomic and environmental evaluations where limited samples are required. so, Grey system theory utilizes accumulated generating data instead of original data to build forecasting model, which makes raw data stochastic weak, or reduces noise influence in a certain extent, therefore, intrinsic regularity of data can be searched easily, and model can be built with relatively little data [8].

If combine grey system theory with SVM to build Grey support vector machines (GSVM), we can exploit sufficiently the characteristic of grey system model requiring less data and feature of nonlinear map of SVM, and develop both advantages, thus raise predicting precision much more. Grey correlation coefficient analysis is a method to find the key input factors and GM (1, N) from grey system theory proposed by DENG [7] is used to add a grey layer before neural input layer and white layer after SVM layer. SVM with genetic algorithms are used to determine the weights between nodes. Finally, the proposed Grey support vector machines with genetic algorithms model (GSVMG model) is applied to forecast China patent applications filings. Evaluation methods are used for comparing the performance of forecasting techniques, which shows that the GSVMG model is outperformed the GM (1, N) model and support vector machines with genetic algorithms model (SVMG model). The experiment shows that this kind information manipulation and forecasting method based on GSVMG is of validity and feasibility.

This paper is organized as follows: Section 2 describes Support vector machines regression with GAs parameter selection. Section 3 introduces the proposed GSVMG model for forecasting. Section 4 the proposed GSVMG model is applied to the prediction patent application filings forecasting. In Section 5, experimental results are shown. Finally, conclusions are described in Section 6.

II. SVM Regression with GAs

Support vector machines (SVM) and kernel methods (KM) have become, in the last few years, one of the most popular approaches to learning from examples with many potential applications in science and engineering. As a learning method, it is often used to train and design radial basis function (RBF) networks. Given a set of examples $\{(x_i,y_i), x\in R^n, y\in R, i=1,\cdots,N\}$,the SVM learning method in its basic form creates an approximation function $f(x)=b+\sum_{j=1}^{m} y_j\cdot\alpha_j\cdot K(x_j,x)$ with $y\approx f(x)$ for regression. For that purpose, a subset of support vectors $\{x_j, j=1,...,m\}\subset\{x_i, i=1,...,N\}$ is determined, the kernel function K is chosen, and the parameters $b,\alpha_j, j=1,...,m$ are estimated.

KM is method that use kernels of the form $K(x_1,x_2)=\phi(x_1)\cdot\phi(x_2)$, is an inner product and Φ is in general a nonlinear mapping from input space X onto feature space Z. The symmetry of the inner product determines the symmetric of the kernel. The necessary and sufficient conditions for a symmetric function to be a kernel is to be positive definite, thus statistically seen, kernels are covariances. In practice, the kernel function K is directly defined. Φ and the feature space Z are implicitly derived from its definition. Kernel substitution of the inner product can be applied for generating SVMs for classification based on margin maximization.

In the ε-SVM regression [9], [10], the goal is to find a function $f(x)$ that has at most ε deviation from the actually obtained targets y_i for all the training data, and at the same time, is as flat as possible. The ε-insensitive loss function reads as follows:

$$e(f(x)-y)=\begin{cases}0, |f(x)-y|\le\varepsilon\\ |f(x)-y|-\varepsilon, otherwise\end{cases} \tag{1}$$

To make the SVM regression nonlinear, this could be achieved by simply mapping the training patterns by Φ: $R^n\to F$ into some high dimensional feature space F. Suppose $f(x)$ takes the following form:

$$f(x)=w\cdot x+b \tag{2}$$

A best fitting function:

$$f(x)=(w^{'}\cdot\phi(x))+b \tag{3}$$

is estimated in feature space F, where "." denotes the dot product in the feature space F. Flatness in the case of (3) means that one can seek small $w^{'}$. Formally this problem can be written as a convex optimization. By constructing the Lagrangian function, the dual problem can be given as follows:

$$\begin{aligned}\max imize \quad & -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)(x_i\cdot x_j)\\ & +\sum_{i=1}^{l}y_i(\alpha_i-\alpha_i^*)-\varepsilon\sum_{i=1}^{l}(\alpha_i+\alpha_i^*)\\ subject\ to \quad & \sum_{i=1}^{l}(\alpha_i-\alpha_i^*)=0\\ & 0\le\alpha_i,\alpha_i^*\le C\end{aligned} \tag{4}$$

Where α_i and α_i^* are the Lagrange multiplier coefficients for the ith training example of regression, and obtained by solving the dual optimization problem in support vector learning [5], [6]. The non-negative coefficients α and α^* are bounded by a user-specified constant C. The training example for which $\alpha \neq \alpha^*$ is corresponded to the support vectors.

At the optimal solution from (4), the regression function takes the following form:

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) K(x_i \cdot x_j) + b \tag{5}$$

Where $K(.,.)$ is a kernel function, where b is found by the Karush-Kuhn-Tucker conditions at optimality.

Symmetric and positive semi-definite function, which satisfies Mercer's conditions, can be used as a kernel function in the SVM context. Mercer's conditions can be written as follows:

$$\iint K(x,z) g(x) g(z) dxdz > 0, \int g^2(x) dx < \infty$$

$$\text{where} \quad K(x,z) = \sum_{i=1}^{\infty} \alpha_i \psi(x) \psi(z) \tag{7}$$

$$\alpha_i \geq 0$$

In this paper, the Gaussian kernel function used in the SVM regression method is as follows:

$$K(x,z) = \exp\left(\frac{-\|x-z\|^2}{\sigma^2}\right) \tag{8}$$

The SVM regression can be expressed by the following chat. Where $x_i (i = 1,...,n)$ is the input vectors, $K(x_j, s)(j = 1,...s)$ is the kernel function corresponding to support vectors (amount to s entries).y is the output.

The selection of the three positive parameters, δ, ε and C, of a SVM model is important to the accuracy of the regression for forecasting. Therefore, genetic algorithms are used in the proposed SVM model to optimize the parameter selection. The framework of optimize the parameter selection is proposed by figure 1.

A negative mean absolute percentage error (MAPE) is used as the fitness function for evaluating fitness [11]. The MAPE is as follows:

$$MAPE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{a_i - f_i}{a_i} \right| \times 100\% \tag{6}$$

Where a_i and f_i represent the actual and forecast values and N is the number of regression forecasting periods. GAs are used to yield a smaller MAPE by searching for better combinations of three parameters in SVM, which is described below:

The first step is the creation of an initial population of chromosomes. The three free parameters, δ, ε and C, are encoded in a binary format; and represented by a chromosome. Secondly, the fitness of each chromosome is evaluated by the cross-validated predictive accuracy of the SVM model. Based on fitness functions, chromosomes with higher fitness values are more likely to yield offspring in the next generation. The roulette wheel selection principle is applied to choose chromosomes for reproduction [12].

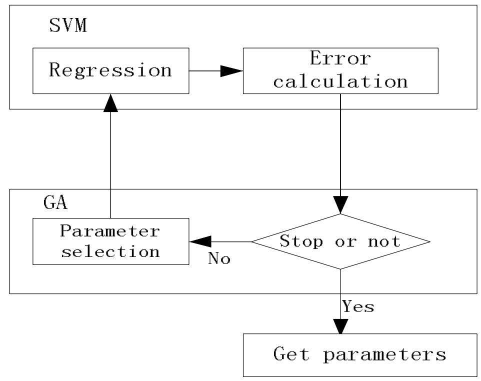

Fig 1. The framework of optimize the parameter selection

Thirdly, it is Crossover and mutation. Mutations are performed randomly by converting a "1" bit into a "0" bit or a "0" bit in to a "1" bit. The single-point-crossover principle is employed. Segments of paired chromosomes between two determined break points are swapped. The rates of crossover and mutation are probabilistically determined. In this study, the probabilities of crossover and mutation are set to 0.5 and 0.1, respectively. Fourthly, a new population is created for the next generation. Finally, if the number of generations equals a given scale, then the best chromosomes are presented as a solution; otherwise turn back to the second step. After these steps, the optimal parameters, δ, ε and C, of a SVM model are obtained.

III. GSVMG Model for Forecasting

Grey correlation coefficient analysis is a method to determine whether or not variables are correlated and to determine the degree of their correlation. By calculation of characteristic serial curves and the degree of geometrical similarity of these curves, key factors can be determined.

The standard pretreatment process of source series is divided by the average value of that series, the correlation coefficient between y and x series can be described as below:

$$\gamma(y_1(k), x_i(k)) = \frac{\xi \max_i \max_k \Delta_{1i}(k)}{\Delta_{1i}(k) + \xi \max_i \max_k \Delta_{1i}(k)} \tag{7}$$

$$\gamma(y_1, x_i) = \sum_{k=1}^{n} \gamma(y_1(k), y_i(k)) / n \tag{8}$$

Where $\Delta_{li}(k)$ is the difference information of x_i, the Kth point on X, with regard to y_1, $r(y1(k),xi(k))$ is the grey-related modulus. $r(y_1,x_i)$ the degree of grey relation and ξ is the differentiation modulus. Usually this work takes ξ =0.5.

After grey correlation coefficient analysis, it finds the key input factors to GM (1, N) model. The GM (1, N) model is the grey compatibility model, and it is one that describes the relationship between one main factor as output, and all the other N-1 factors as input in a system. As well as their evaluation and logical, it is a model which quite suitable for the dynamics relative analysis of all variables and forecasting the output by input.

In our research, the mathematics model is GM (1, N) model. Sequences $y_1^{(0)}(k)$ is called the main factor in system, and $x_i^{(0)}(k)$. Where i =1, 2, 3... N-1 is called the influencing factors in system. The analysis steps of GM (1, N) model:

step1, Build up original sequences:

$$
\begin{aligned}
&y_1^{(0)}=\left\{y_1^{(0)}(1),y_1^{(0)}(2),y_1^{(0)}(3),.....y_1^{(0)}(k)\right\}\\
&x_1^{(0)}=\left\{x_1^{(0)}(1),x_1^{(0)}(2),x_1^{(0)}(3),.....x_1^{(0)}(k)\right\}\\
&x_2^{(0)}=\left\{x_2^{(0)}(1),x_2^{(0)}(2),x_2^{(0)}(3),.....x_2^{(0)}(k)\right\}\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&x_{N-1}^{(0)}=\left\{x_{N-1}^{(0)}(1),x_{N-1}^{(0)}(2),.......x_{N-1}^{(0)}(k)\right\}
\end{aligned}
\tag{9}
$$

Step2, build up accumulated generating operation (AGO) sequences from original sequences.

AGO

$$y^{(0)}=y^{(1)}=\left(\sum_{k=1}^{1}y^{(0)}(k),\sum_{k=1}^{2}y^{(0)}(k),...,...,\sum_{k=1}^{n}y^{(0)}(k)\right)$$

AGO

$$x^{(0)}=x^{(1)}=\left(\sum_{k=1}^{1}x^{(0)}(k),\sum_{k=1}^{2}x^{(0)}(k),...,...,\sum_{k=1}^{n}x^{(0)}(k)\right)$$

Step3, The GM (1, N) model is defined as

$$y_1^{(0)}(k)+az_1^{(1)}(k)=\sum_{i=1}^{N-1}b_i x_i^{(1)}(k) \tag{10}$$

Where, $z_1^{(1)}(k)=0.5y_1^{(1)}(k)+0.5y_1^{(1)}(k-1) \quad k\geq 2$

step4, Translate the above equations into matrix form:

Given $B=\begin{bmatrix}-z_1^{(1)}(2) & x_1^{(1)}(2) & \cdots & x_{N-1}^{(1)}(2)\\ -z_1^{(1)}(3) & x_1^{(1)}(3) & \cdots & x_{N-1}^{(1)}(3)\\ \vdots & \vdots & \ddots & \vdots\\ -z_1^{(1)}(n) & x_1^{(1)}(n) & \cdots & x_{N-1}^{(1)}(n)\end{bmatrix}$,

$y_N=\begin{bmatrix}y_1^{(0)}(2)\\ y_1^{(0)}(3)\\ \vdots\\ y_1^{(0)}(n)\end{bmatrix}$, $P_N=\begin{bmatrix}a\\ b_1\\ \vdots\\ b_{N-1}\end{bmatrix}$ then

$$y_N=BP_N \tag{11}$$

Where, $P_N=(B^TB)^{-1}B^Ty_N$

Step6, Obtain the white equation of Eq. (12)

$$\frac{dy_1^{(1)}}{dt}+ay_1^{(1)}=\sum_{i=1}^{N-1}b_i x_i^{(1)}(k) \tag{12}$$

Step7, Inverse and forecasting

According to Eq. (10), (11), (12), it gets Eq. (13)

$$
\begin{aligned}
\hat{y}_1^{(1)}(k+1)=&\left(y_1^{(1)}(0)-\frac{1}{a}\sum_{i=1}^{N-1}b_i x_i^{(1)}(k+1)\right)e^{-ak}\\
&+\frac{1}{a}\sum_{i=1}^{N-1}b_i x_i^{(1)}(k+1)
\end{aligned}
\tag{13}
$$

Where, $y_1^{(1)}(0)$ is $y_1^{(1)}(1)$

It will obtain the value of forecasting of GM (1, N) by inversing-accumulated generation, which is

$$\hat{y}_1^{(0)}(k+1)=\hat{y}_1^{(1)}(k+1)-\hat{y}_1^{(1)}(k) \tag{14}$$

After constructed the GM (1, N) model, it is need to combine SVM with GAs to GM (1, N) for forecasting. The GSVMG model has three basic parts: a grey layer, a gentler SVM with GAs, and a white layer. The grey layer before input nodes has accumulated generating operation (AGO) to initial input data, which is $x_i^{(1)}, i=1,2,..,N-1$ in GM (1, N) model. At last, the white layer after neural output nodes inverses accumulated generation to the output data of SVM with GAs, which is $\hat{y}^{(0)}(k+1)$ in GM (1, N) model. Therefore, the prediction value we need is obtained. The construction of GSVMG model is shown in figure 2.

In figure 2, where the pair [x, y] constitutes one train sample for SVM propagation model, y is input data and x is output data. Get a vector with m elements at one time from initial time series data in turn. if the length of initial time series data is K, we can obtain K-m train samples to train GSVMG. When the GSVMG is successfully trained, it can be used to forecasting. The forecast is estimated through one operation of the inverse of the accumulated generating operation.

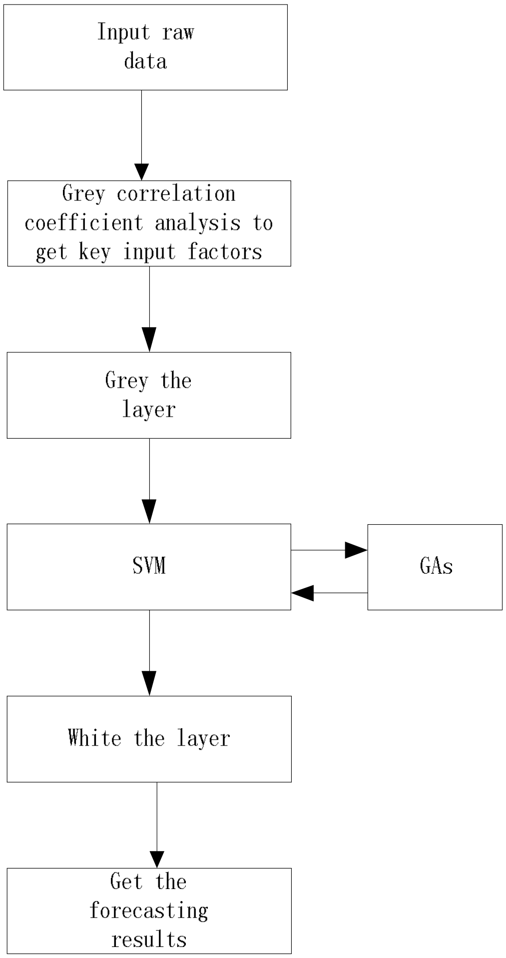

Fig 2. The framework of GSVMG model

IV. Application to Patent Application Filings Forecasting

Patents are intrinsically linked to the larger economic picture— to the process of innovation and to technological and scientific change. In the empirical literature on patenting activity, several economic variables also have been shown to influence the number of patent filings, Such as research and development (R&D), Gross National Product (GNP) and Gross Private Domestic Investment (GPDI) et al [1] [2].

To study the determinants of patenting activity and forecast the patent application filings in china, this paper should estimate the following equations:

$$y = f(x_i), i = 1, 2, ..., 8 \tag{15}$$

A number of crucial parameters, which mainly determine the patenting activity, are considered. Where:

x_1 is the variable of Gross Domestic Product;
x_2 is the Research and Development Expenditures;
x_3 is the Scientists and Engineers engaged in R&D;
x_4 is the Graduates in Science, Math, and Engineering;
x_5 is the Graduates in Science, Math, and Engineering (x_4), 10-Year China Government Bond Interest Rates;
x_6 is the Business Formations;
x_7 is the the full-time equivalent of R&D activities;
x_8 is the the China three expenses.

V. Experimental Results

This paper focuses on the reality of the Intellectual property in China and collects data in a period as long as 15 years (statistical yearbooks on Intellectual property of china). The output y is the China domestic patent applications filings. The forecasting accuracy is measured by the root mean square error (RMSE), as given by Eq. (16). The time series data from year 1991 to year 2003 are used as known training data to forecast the last two years test data.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{16}$$

In the training stage, the training data set are fed into the GSVMG an SVMG model, and the structural risk minimization principle is employed to minimize the training error. While training errors improvement occurs, the three kernel parameters, δ, ε and C of GSVM, SVM model adjusted by GAs are employed to calculate the validation error. Then, the adjusted parameters with minimum validation error are selected as the most appropriate parameters. Note that the testing data sets are not used for modeling but for examining the accuracy of the forecasting model. Then, the kernel parameters, δ, ε and C, in the GSVM and SVM model with the smallest testing MAPE value is used as the most suitable model for this example. The MAPE of testing results and the suitable parameters for the GSVM, SVM models are found. The result of the GSVM, SVM model MAPE is 1.721% and 2.649%.

From the result, the percentage error indicates that the GSVMG is outperformed the SVMG model and the SVMG model is outperformed the GM (1, N). The RMSE of SVMG model is 0.2829, which is better than the GM (1, N) model at 0.3379, respectively, while the GSVMG model is the lowest at 0.1826. So, the test results showed that almost the GSVMG model yields improved forecast results and significantly outperform the other two forecasting models.

VI. Conclusions

This work introduced a novel forecasting technique, GSVMG, to investigate its feasibility in forecasting PAF in China. The experimental results indicate that the GSVMG model outperformed the SVM and GM (1, N) models in terms of forecasting accuracy. The superior performance of the GSVMG model has several causes. First, the GSVMG model exploits sufficiently the characteristic of the preprocessed data handled by the grey operation with stochastic reduced and regularity raised and the nonlinear map feature of SVM, which makes the convergent process of the SVM fast. Second, improper determining of these three parameters will cause either over-fitting or under-fitting of a SVM model. In this work, the GAs can determine suitable parameters to forecast PAF. Third, the GSVMG model performs structural risk minimization rather than minimizing the training errors. This investigation is the first to apply the Grey correlation coefficient analysis, GM (1, N) and SVM model with GAs to PAF forecasting. The empirical results obtained in this study demonstrate that the proposed model offer a valid alternative for application in the China PAF forecasting. In the future, other factors can be included in the GSVMG model for forecasting PAF. In addition, some other advanced searching techniques for suitable parameters selection can be combined with grey SVM to forecast the patent applications filings.

REFERENCES

[1] P. Hingley, M. Nicolas, "Methods for forecasting numbers of patent applications at the European Patent Office", *World Patent Information* no.26, vol.3, 2004, pp.191-204.

[2] K. Adams, "Modeling and forecasting U.S. Patent application filings", *Journal of Policy Modeling*, no.19, vol.5, 1997, pp.491-535.

[3] C. Bock, "Forecast of trademark applications in Switzerland" *World Patent Information*, no.26, vol.4, 2004, pp.275-282.

[4] P. F. Pai, W. C. Hong, "Software reliability forecasting by support vector machines with simulated annealing algorithms" *Journal of Systems and Software*, no.69, vol.6, 2006, pp.747-755.

[5] V. N. Vapnik, "The Nature of Statistical Learning Theory", *Springer*, New York, 1995.

[6] V. N. Vapnik, "An Overview of Statistical Learning Theory", *IEEE Trans. Neural Networks*, no.10, vol.5, 1999, pp.988-999.

[7] J. L. Deng, "Control problems of grey systems", *Syst. Control. Lett.*, no.5, vol.2, 1982, pp.288-294.

[8] P. Zhou, "A trigonometric grey prediction approach to forecasting electricity demand", *Energy* no.31, vol.14, 2006, pp.2839-2847.

[9] F. J. G. Castano, U. M. G. Palomares, R. R. Meyer, "Projection Support Vector Machine Generators", *Machine Learning*, no.540, vol.1, 2004, pp.33-44.

[10] G. M. Fung, O. L. Mangasarian, "Multicategory Proximal Support Vector Machine Classifiers", *Machine Learning*, no.59, vol.1, 2005, pp.77-97.

[11] P. F. Pai, W. C. Hong, "Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms", *Electric Power Systems Research*, no.74, vol.1, 2005, pp.417-425.

[12] J. Holland, "Adaptation in Natural and Artificial System", *University of Michigan Press*, AnnArbor, 1975.