Application Research of Support Vector Regression in Coal Mine Ground-water-level Forecasting

Liu Taian[1], Xue Xin[2], Liu Xinying[1], Zhao Huiqi[1]
[1] Department of Information and Engineering,Shandong University of Science and Technology(SDUST),Taian 271019,China
[2] Department of Mathematics and System Science, Taishan University (TSU), Taian 271021, China
lta999@163.com

ABSTRACT: The forecast of the mine Ground-water-level is an issue with many influencing factors, highly non-linear and temporal series. SVR (Support Vector Regression) is applied to forecast Coal Mine Ground-water-level in this paper. Appropriate kernel function and parameters are chosen based on the analysis to SVR regression algorithm. This paper proposes the Forecasting Model of Coal Mine Ground-water-level basing on SVR regression algorithm and determines the forecast of the input factor and the output factor according to the physical geography and the hydrology geology situation of the chosen mining area. The numerical test results show that the forecast results have compatibility with the actual measurement result. We verify that the forecast model of Coal Mine Ground-water-level has effect, and provide a new effective method to the Forecasting of Coal Mine Ground-water-level.

KEYWORDS: support vector regression algorithm; kernel function; cross validation methods; coal mine ground-water-level; forecasting model

I. INTRODUCTION

In the energy industry, coal exploitation is a high risk industry, which often has some mine flooding accidents and brings heavy loss to the life and property security of our country and people. So the research on the forecast of the coal mine underground water level has important theoretical and practical significance, which is an issue with many influencing factors, highly non-linear and temporal[1,2,3] series.

Support Vector Regression Algorithm(SVRA) is a method to regression prediction and function approximation. SVRA has not only strict theory base[4,5,6], but also can well resolve such practical problem as non-linearity, high dimension and local minima. So, it is a good method to forecast Coal Mine Ground-water-level with SVRA.

II. THE THEORY OF PREDICTION WITH SUPPORT VECTOR REGRESSION ALGORITHM

SVM is originally proposed for classification. It is a new Machine Learning method based on Statistical Learning Theory. Because of the use of kernel function and Structural Risk Minimization, SVM has good performance to classification with limited samples. SVM is formulated as a convex quadratic programming, so the minima which are found are global optimal solutions. SVRA is a method to regression prediction and function approximation.

A. Support Vector Regression Algorithm

SVM is good at solving regression prediction problem by utilizing a appropriate loss function[7]. The ε -insensitive loss function is used usually. When $|y-f(x)|\le\varepsilon$, there is no loss. The ε -insensitive loss function is that

$$L_\varepsilon(y)=\max\{0,|y-f(x)|-\varepsilon\} \tag{1}$$

We consider a given train data set $D=\{(x_1,y_1)\ldots,\quad(x_l,y_l)\},x\in R^n,y\in R$ .For linear regression problem, SVR would like to find regression prediction function:

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

Where, $(\omega\cdot x)$ is inner product, $\omega\in R^n,b\in R$.

By utilizing largest margin principle, ε -insensitive loss function and Lagrange function, we get the optimization problem[6]:

$$\min\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)(x_i\cdot x_j)-$$

$$\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)y_i+\sum_{i=1}^{l}(\alpha_i+\alpha_i^*)\varepsilon \tag{3}$$

$$s.t.\ 0\le\alpha_i,\alpha_i^*\le C\ \ ,i=1,2,\ldots,l \tag{4}$$

$$\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)=0 \tag{5}$$

Where, C is penalty parameter， α,α^* are Lagrange multipliers.

By solving convex quadratic programming (3)-(5), we get the optimal solution

$$\bar{\alpha}=(\bar{\alpha}_1,\bar{\alpha}_1{}^*,\bar{\alpha}_2,\bar{\alpha}_2{}^*,\ldots,\bar{\alpha}_l,\bar{\alpha}_l{}^*)^T\ ,$$

So, we get the regression prediction function

$$f(x)=\sum_{i=1}^{l}(\bar{\alpha}_i{}^*-\bar{\alpha}_i)(x_i\cdot x)+b \tag{6}$$

For nonlinear regression problem, by utilizing an appropriate kernel function, we can get nonlinear regression function. The samples can be project to high dimension space by inner product where the linear approximation can be done. Kernel function

$$K(x_i,x_j)=\varphi(x_i)\varphi(x_j) \tag{7}$$

978-0-7695-3600-2/09 $25.00 © 2009 Crown Copyright
DOI 10.1109/IFITA.2009.61

So, we get the regression prediction function for nonlinear regression problem

$$f(x)=\sum_{i=1}^{l}(\overline{\alpha}_i{}^{*}-\overline{\alpha}_i)K(x_i,x_j)+b \tag{8}$$

The advantages of utilizing ε-insensitive loss function in support vector regression machine is that only the Lagrange multipliers $\alpha_i-\alpha_i^*$ of the samples outside the Strip region are not to equal zero. The vectors are named as support vector that their Lagrange multipliers are not equal to zero.

B. *Choosing appropriate kernel function and the parameters*

Polynomial kernels, Gaussian radial basis function kernels and sigmoid kernels are often used. Suppose the train set is given, the first thing we should do is to choose appropriate kernel function and the parameters of support vector machine. Then regression prediction function could be found.

1) *Choice of kernel function:*

There are several reasons we choose the Gaussian radial basis kernels

$$K(x_i,x_j)=\exp(-\frac{\|x_i-x_j\|^2}{\sigma^2}) \tag{9}$$

They are

a) Number of parameters of hyperplane is related to the choice of model. The more of the number of parameters, the more complicated of computing of model is. The number of parameters of hyperplane with Gaussian radial basis function kernels is little than that with Polynomial kernels.

b) In network of Gaussian radial basis function kernels, the output of basis function from input layer to hidden layer is nonlinear. But the output is linear. Feature space nonlinear separeted is mapped another high-dimensional space by the network of Gaussian radial basis function kernels. The samples in the new feature space are linear separeted. Then we could get the regression prediction function of old nonlinear regression problem[8].

c) Gaussian radial basis function kernels avoid the dimensional disaster.

2) *Choice of parameters*

In the method of regression prediction, penalty parameter C, parameters of kernels and ε of insensitive loss function should be chosen. In this paper we utilize cross validation methods via parallel grid search to choose penalty parameter C, kernels parameter σ [9]. By numerical experiments and comparative analysis to the actual measurements of the mime chosen, the values of the parameters of model (10) are given as follows:

C=16, σ=0.326, ε=0.035.

III. Establishing Forecast Model of Coal Mine Ground-Water-Level

A. *Determining forecast factor and output factor*

Dynamic underground water level is a direct reflection of equilibrium status of the regional underground water level, which is an integrated externalization of the aquifer structure, properties, circulating conditions and the interaction between the supplies and the drainage factors. The actual measurement data of a digging in Shandong Yankuang Group are taken to do the forecast model of underground water level experiment. Based on the analysis on document [10], the main factors that affect the underground water level of digging are amount of precipitation, evaporation capacity, exploitation and riverway flux. Besides, with a comprehensive consideration of the hysteresis of all the factors and the self-correlation of the underground water level and regarding the amount of precipitation and measured underground water level in the former period as the impact factors, six impact factors are obtained as hefts of the input vectors supporting vector regression machine: amount of precipitation X_1, evaporating capacity X_2, exploitation X_3, river flux X_4, precipitation in the former period X_5 and measured underground water level in the former period X_6. Forecasting underground water level Y is taken as the output value of the model. So the forecasting issue of underground water level in the digging can be transformed to a support vector regression forecasting one with six characteristic variables and one output.

B. *The forecasting model of Coal Mine Ground-water-level based on SVR*

We are given the actual measurements of ground-water-level $\{(x_1,y_1),...,(x_l,y_l)\}$, where pattern $x_i\in R^6$ consists of the factors which effect the forecast of ground-water-level and $y_i\in R$ is the value relating to ground-water-level, $i=1,2,...,l$. We can train the given sets $\{(x_1,y_1),...,(x_l,y_l)\}$ by SVR. Then we can get the forecast model of ground-water-level,

$$y=\sum_{i=1}^{l}(\overline{\alpha}_i{}^{*}-\overline{\alpha}_i)K(x_i,x_j)+b \tag{10}$$

IV. Numerical Experiments

We proceed the numerical experiments under the conditions that the CPU is Pentium IV3.0G，Memory is 512MB，and operating system is Windows XP. The testing program of forecast model of ground-water-level utilize Matlab6.5. The numerical experiments are based on the history data of A digging in Shandong Yankuang Group.

All data is distributed into train set and test set firstly. We get the regression prediction function based on train set. The accuracy of regression prediction function is tested based on test set.

The data of numerical experiments consist of observation value of long-time observation hole O3-1. History data 1 to 30 are used as train samples. History data 31 to 66 are used as test samples. The comparative results of the forecast values of ground-water-level of hole O3-1 with factual values are showed in Table I.

TABLE I. COMPARATIVE RESULT OF THE FORECAST VALUES OF GROUND-WATER-LEVEL OF HOLE O3-1 WITH FACTUAL VALUES

Time	Factual values(m)	Forecast values(m)	Relative error(%)
2007.3	13.93	13.86	0.50
2007.4	15.46	15.60	0.91
2007.5	19.92	19.86	0.30
2007.6	21.51	21.43	0.37
2007.7	21.76	21.67	0.41
2007.8	26.65	26.58	0.26
2007.9	22.83	22.75	0.35
2007.10	19.97	19.86	0.55
……	……	……	……

From Table I, we conclude that forecast values are consistent well with factual values. Average relative error is 0.46%， which verifies that the forecast model (10) is effective.

Table I only includes forecast values of one long-time observation hole. When we forecast the observation values of other long-time observation hole, the input factor and the long-time observation data of observation hole which is affected geographically should be considered.

The comparative result of the forecast values of ground-water-level of 5 holes with factual values are showed in Table II.

TABLE II. COMPARATIVE RESULT OF THE FORECAST VALUES OF GROUND-WATER-LEVEL OF 5 HOLES WITH FACTUAL VALUES

Hole number	Factual values(m)	Forecast values(m)	Relative error(%)
O1-2	22.57	22.41	0.71
O2-1	25.65	25.73	0.31
O2-2	27.57	27.45	0.44
O3-2	24.56	24.41	0.61
O3-3	26.75	26.82	0.26

From Table II, we conclude that the forecast model (10) is effective for different observation hole. The forecast values are consistent well with factual values. Average relative error is 0.47%.

The real-time forecast values of observation hole O2-3 are listed in Table Ⅲ.

TABLE III. COMPARATIVE RESULT OF THE REAL-TIME FORECAST VALUES OF GROUND-WATER-LEVEL OF HOLE O2-3 WITH FACTUAL VALUES

Time	Factual values(m)	Forecast values(m)	Relative error(%)
2008.1	11.43	11.51	0.70
2008.2	12.16	12.23	0.58
2008.3	12.48	12.59	0.88
2008.4	13.26	——	——
2008.5	13.57	——	——
……	……	……	……

From Table Ⅲ, we conclude that he forecast model (10) is effective.

V. CONCLUSION AND INNOVATION POINT

There are two innovation points in this paper. The Forecasting Model of Coal Mine Ground-water-level is proposed based on SVR algorithm. Choosing the appropriate kernel function and the parameters of Forecasting Model in this paper based on cross validation methods via parallel grid search in SVR algorithm.

By factual testing and analysis to the Ground-water-level of a digging in Shandong Yankuang Group, we conclude that:

(1)Cross validation methods via parallel grid search is used in choosing the parameters of regression model, which avoids aimlessness and randomness of choice, and raises the forecast accuracy.

(2)The numerical experiments show that The Forecasting Model of Coal Mine Ground-water-level proposed t basing on SVR is effective. Experiments and forecast are stable. We propose a new effective method to the Forecasting Coal Mine Ground -water-level in this paper.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China(10571109), National Natural Science Foundation of Shandong(2007ZRB019FK) and Program of Shandong Tai'an Science and Technology(20082025).

REFERENCES

[1] X. H. Wu, H. L. Shi, and Z. L. Li, "The Forecasting Method of Coal Mine Ground-water-level Based on BP Neural Network", Industry and Mine Automation, Changzhou City, 2006,10, pp. 21-23.

[2] B. R. Chen, Groundwater Level trends and Prediction, Science Publication, Beijing City, 1998.

[3] J. Q. Yang, X. X. Luo, "A RBF Artificial Neural Network Model for the Prediction of Regional Groundwater Level", Journal of China Hydrology, Beijing City, 2001,8, pp. 1-3.

[4] X. F. Song, D. J. Chen, and H. J. Yu, "A Survey of Optimization Algorithms in Support Vector Machine", Computer Science, Chongqing City, 2003, Vol.30, No.3, pp. 12-20.

[5] S. W. Ji, W. T. Ji, "A Tutorial Survey of Support Vec-tor Machine Training Algorithms", Computer Technology and Development, Xian City,2004,1, pp.18-20.

[6] N. Y. Deng and Y. J. Tian , New Method in Data Mining –Support Vector Machines, Science Publication, Beijing City, 2004.

[7] V. Vapnik, S. Golowich and A. Smola, "Support vector method for function approximation, regression estimation and signal processing", In M. Jordan and T. Petsche, editors, Adances in Neural Information Processing System 9, MIT Press, Cambridge, MA, 1997, pp. 281-287.

[8] Z. Q. Bian, X. G. Zhang, Pattern Recognition (Edtion 2), Tsinghua University Press, Beijing City, 2000,1.

[9] M. Y. Ma, Seepage Monitoring Model and Safety Alarm Analysis of Ash Storage Dam. Dalian University of Technology, Dalian City, 2005.

[10] Southern Shandong Institute of Geological Exploration, A Report on Reconnaissance of Water supply and Hydrogeology. Southern Shandong Institute of Geological Exploration, Yanzhou City, 2002.