Full Length Research Paper

Application of support vector regression to predict metallogenic favourability degree

Chunming Wu[1,2]*, Xinbiao Lv[1,2], Xiaofeng Cao[2], Yalong Mo[2] and Chao Chen[2]

[1]School of Geological Survey, China University of Geosciences, Wuhan, China.
[2]Faculty of Resources, China University of Geosciences, Wuhan, China.

Mineral resource prediction is becoming increasingly important as researchers attempt to resolve the prospect direction by mining geological data. In this paper, Support Vector Regression (SVR) is applied to predict iron deposit metallogenic favourability degree since SVR is a powerful tool to solve the problem characterized by smaller sample, nonlinearity, and high dimension with a good generalization performance based on structural risk minimization. The paper discusses the support vector regression algorithm in some detail, describes a SVR based-system that learns from examples to predict metallogenic favourability degree of iron deposit and contrasts this approach with Partial Least Squares (PLS). The experimental results show that SVR has high recognition rates and good generalization performance for small sample, especially good for treating the data of some nonlinearity in geology.

Key words: Support Vector Regression (SVR), metallogenic favourability degree, mineral resource, quantitative prediction.

INTRODUCTION

Support Vector Machine (SVM) is an excellent kernel-based tool for binary data classification and regression. The method to solve regression problems using SVM is called Support Vector Regression (SVR), which is one of the most important applications of function approximation. This learning strategy was introduced in the mid-1960s by Vapnik and co-workers (Vapnik and Chervonekis, 1991; 1999). It is developed within the area of statistical learning theory and structural risk minimization and hence has comparable or better results than neural networks and other traditional learning methods (Vapnik, 1999; 1999). Based on the structural risk minimization (SRM) principle, SVMs 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, such as in the pattern recognition, function estimation, signal processing, control and other fields (Scholkopf et al., 1997; Stitson et al., 1999).

In mineral resource prediction, some traditional methods, such as Artificial Neural Network (ANN), multiple statistical analysis and fuzzy sets theory, have been used to calculate the metallogenicl favourability degree. These methods often require a large number of sample data. Actually, large number of sample data is hard to collect in prospective mining area with less geological works. Therefore, it is difficult to predict metallogenic favourability degree by the traditional methods in the area.

Careful contrast of SVR and traditional methods shows up the advantages of the first one, especially in the condition of small sample and nonlinear data. In this article, SVR was adopted to predict the mineral potential of iron deposit. A brief summary of SVR was given. The adequate geological data was applied to the model and good prediction results were obtained.

SUPPORT VECTOR REGRESSION

SVR is a powerful machine learning method that is useful for constructing data-driven non-linear process models by a kernel function. It shares many features with artificial neural network (ANN) but possesses some additional desirable characteristics and is gaining widespread acceptance in data-driven non-linear modeling applications. SVR possesses good generalization ability of regression function, robustness of solution, addressing regression from sparse data and an automatic control of solution complexity. The method brings out the explicit data points from the input variables that are important for defining the regression function.

---

*Corresponding author. E-mail: wcmcug@cug.edu.cn.

This feature of SVR makes it interpretable in terms of the training data in comparison with the other black-box models including ANN, where the model parameters are difficult to interpret.

Given is a brief description of SVR. Vapnik has a more detailed description of SVR (Vapnik, 1999). Given dataset $D = \{(x_i, y_i)\}_{i=1}^{N}$ obtained from a latent function where x_i denotes the sample vector, y_i the corresponding response and N is the total number of samples. In the SVR, the original data is first nonlinearly mapped into a high dimensional feature space, and then a linear function is fitted to approximate the latent function between x and y.

Given the training data, the linear ε-SVR algorithm theoretically aims to solve the optimizing problem, which can be written in the following form with an ε-insensitive loss term: Minimize:

$$\frac{1}{2}\|w\|^2 + \frac{c}{N}\sum_{i=1}^{N} L(y_i - f(x_i), \varepsilon)$$

(1)

where C, penalty parameter is a predefined regularizing parameter. The above minimizing problem can further be expressed in the following form with the slack variable $\xi_i^{(*)}$ introduced: Minimize:

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} (\xi_i + \xi_i^*)$$

(2)

Subject to:

$$\begin{cases} y_i - \langle w, x_i \rangle - b \le \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \le \varepsilon + \xi_i^*, i = 1, 2 ..., N \\ \xi_i, \xi_i^* \ge 0 \end{cases}$$

Then, we can obtain a lagrange function from the above objective function (also referred as the primal objective function) and the corresponding constraints by introducing a dual set of variables. It can be shown that this function has a saddle point with respect to the primal and dual variables at the solution. The Lagrangian is as follows:

$$L := \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} (\xi_i + \xi_i^*) - \sum_{i=1}^{N} (\eta_i\xi_i + \eta_i^*\xi_i^*)$$
$$- \sum_{i=1}^{N} \alpha_i(\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b)$$
$$- \sum_{i=1}^{N} \alpha_i^*(\varepsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b)$$

where $\eta^{(*)}, \alpha^{(*)}$ are Lagrange multipliers.

It follows from the saddle point condition that the partial derivatives of L with respect to the primal variables (w, b, ξ_i, ξ_i^*) have to vanish for optimality. Thus, the following equations hold.

$$\partial_b L = \sum_{i=1}^{N} \alpha_i^* - \alpha_i = 0$$

$$\partial_w L = w + \sum_{i=1}^{N} (\alpha_i^* - \alpha_i)x_i = 0$$

$$\partial_{\xi_i^{(*)}} L = C - \alpha_i^{(*)} - \eta_i^{(*)} = 0$$

Substituting above equations into L functions yields a dual optimization problem: Maximize:

$$-\frac{1}{2}\sum_{i,j=1}^{N} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle x_i, x_j \rangle$$
$$-\varepsilon\sum_{i=1}^{N} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*)$$

Subject to:

$$\sum_{i=1}^{N} (\alpha_i - \alpha_i^*) = 0$$

and

$$\alpha_i, \alpha_i^* \in [0, C]$$

With the help of Lagrange multiplier method and QP algorithm, the regression function can be derived as

$$f(x) = \sum_{i=1}^{N} (\alpha_i^* - \alpha_i)(x_i^t x) + b$$

(3)

$$b = y_i - \sum_{i=1}^{N} (\alpha_i^* - \alpha_i)(x_i^t x) + \varepsilon$$

(4)

where α_i^* and α_i are the optimized Lagrange multipliers.

For nonlinearity, the original input was first nonlinearly mapped into the feature space by introducing the kernel function and the resulted ε-SVR becomes so flexible that it can be used to deal with the complicated nonlinear regression problem in geology.

As the deriving procedure of the final decision function is quite similar to that in the linear case, we here only give the ultimate mathematical form as,

$$f(x) = \sum_{i=1}^{N} (\alpha_i^* - \alpha_i)k(x_i, x) + b$$

(5)

$$b = y_i - \sum_{i=1}^{N} (\alpha_i^* - \alpha_i)(x_i x_j) + \varepsilon$$

(6)

where α_i^* and α_i are the optimized Lagrange multipliers, and $k(x_i, x)$ is the kernel function.

For the regression by SVR, the user has to select three parameters, namely insensitivity parameter ε, the penalty parameter C and the shape parameter of kernel function. The choice of these parameters is vital to good regression. If C is too small then insufficient stress will be placed on fitting the training data. If C value is too large then the algorithm will over fit the training data and over fit implies poor generalization. Maximum value that C can take is infinity.

The performance of SVR largely depends on the choice of the kernel type and the kernel parameters. However, there is no

Table 1. Input and output variables[a].

Key input variables		Output variable
Geology	**Geophysics**	
HSY1, HSY2, HSY3, DLJL, DLJH, QRY1, QRY2	DZX,YCJL, HCJZ, SNJL, DYYC	MFD

HDY1-C1x/C1a intermediate-acid volcanic rocks; HSY2-C2sh sedimentary -volcanic rocks; HSY3- C2m intermediate-basic volcanic rocks; DLJL-distance between unit and fracture; DLJH-crossing of different fractures in a unit; QRY1-$\gamma\delta_4^2$ intermediate intrusive rocks; QRY2-$\gamma\delta_4^2$ acid intrusive rocks; DZX- the shapes of bouguer gravity contour; YCJL-distance between unit and residual gravity anomaly; HCJZ-the average value of aeromagnetic survey in a unit; SNJL-distance between unit and positive magnetic anomaly in the range of SN 45°; DYYC-exist or adjacent to magnetic anomaly caused by iron deposit in a unit; MFD-metallogenicl favourability degree.

Table 2. Partial variable values in training set.

Unit	HSY1	HSY2	HSY3	DLJL	DLJH	QRY1	QRY2	DZX	YCJL	HCJZ	SNJL	DYYC	MFD
001	0	0	0	1	0	0	0	0	0	0	0	0	1
002	0	0	0	0	0	0	0	1	0	0	1	1	1
003	0	0	0	0	0	0	1	1	0	0	1	0	1
-	-	-	-	-	-	-	-	-	-	-	-	-	1
040	0	1	1	0	0	0	1	0	0	0	1	0	1
041	0	0	0	1	0	0	0	0	0	0	1	1	0
042	0	0	0	0	0	0	1	0	0	0	1	1	0
043	0	0	1	0	0	0	1	1	1	0	1	0	0
-	-	-	-	-	-	-	-	-	-	-	-	-	0
080	1	1	0	0	0	0	0	1	1	0	1	0	0

theoretical guidance on how to choose a kernel function. The best choice of a kernel for a given problem is still an open research issue. In the absence of any known guidelines, kernel has to be chosen in a data dependent way.

CASE STUDY

In this section, SVR method was applied to predict the mineral potential of iron deposit, which was employed to explore the performance of SVR. This dataset collected geological data and geophysical data of 126 geological units in Eastern Tianshan area. Data in 126 geological units were randomly divided into the training set and test set for the construction and evaluation of SVR model, 80 dataset for training and 46 dataset for test. The input variables and the output variables are listed in Table 1.

Partial variable values for training are presented in Table 2. For input variable, when a geological unit exists an input variable, the variable value is 1, conversely, variable value is 0. For output variable, if there is iron deposit in a geological unit, MFD value is 1, conversely, MFD value is 0.

The libsvm package (Chang and Lin, 2001) is employed in our study to construct SVR models. The genetic algorithm is also applied for the parametric optimizing. In addition, all the other programs needed in our studies are coded in Matlab 7.1.

RESULTS AND DISCUSSION

Partial least squares (PLS) are a basic and powerful tool for modeling the linear relationship between x and y in geomathematics. It can also resist the nonlinearity to some extent. In this study, PLS is used as the reference method and the comparative analysis between PLS and SVR is conducted to give some insight into these two methods.

ε-SVR is applied here to construct the SVR model. The Radial Basis Function (RBF) is used as the kernel. The γ of the RBF kernel is set to the default value in the libsvm software. There are two parameters to be pre-defined before training. One is the regularizing factor C, the other is the sparsity parameter ε. So far the most commonly applied technique for determining the parameters model is cross-validation (CV). But it can not find the best parameter or combination of different parameters in a global and comprehensive manner. Here, the genetic algorithm (GA) is used to optimize these two parameters to predict the model because GA has theability to globally locate the optimized solution. In GA, the objection is to minimize the error rate of 10-fold cross-validation calculated by SVR. The optimized values of C and ε by GA are 59.69 and 0.3292, respectively. The SVR model is built with the optimized parameters using the training set.

The Root Mean Squared Error Of Prediction (RMSEP) value was used to evaluate the performance of the two

Table 3. RMSEP of PLS and SVR[a]

	Train	Test	Train[b]	R^2
PLS	2.4679	2.8941	3.3524	0.9633
SVR	1.9348	2.5836	3.3496	0.9739

The optimized parameters of SVR: C=59.69, ε =0.3292,. Notice that the original y-value are scaled into the region [0, 1] when building PLS and SVR model. But the results in this table are reported after transforming the scaled y-value into the original unit. RMSEP based on the 10-fold cross-validation.

methods. The RMSEP of the test sample set is calculated as follows:

$$RMSEP = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2} \tag{7}$$

where N represents the total number of the data points in the test set, and $\hat{y}_i$ represents the predicted value.

Both the results obtained by PLS and SVR are shown in Table 3. From Table 3, the result on SVR model is promising in relative term to PLS method. The RMSEP of SVR for the training set and test set are lowered by 21.6% compared to that of PLS, which indicates an obvious improvement. The squared correlative coefficients of SVR and PLS on the whole data set are 0.9739 and 0.9633, respectively. This result can further prove the better prediction ability of SVR. It may be concluded that SVR has the ability to grasp the nonlinearity problem, especially in the condition of small sample.

The results show that SVR can predict iron resource more accurately. The reason may be that some nonlinearity component had been manually added to the dataset in the condition of small sample, such that the linear PLS can not work well. But the kernel-based SVR have the ability to handle it. The study clearly shows that SVR is indeed a good alternative for regression analysis in the condition of small sample. Thus, we may conclude that SVR is more powerful in capturing the latent data structure and modeling the unknown nonlinearity in the condition of small sample. From this point of view, SVR seems to be a promising method for modeling the real world data (Table 3).

It should be pointed out that SVR has its limitation and disadvantages when dealing with geological dataset. First, SVR is nonlinear, and this makes it difficult for the researcher to explain the results, e.g. which region or the combination of different regions is really meaningful for the model. Second, the optimization of parameters of SVR is a relatively time-consuming task compared to PLS, which may limit their applications.

Conclusion

Support vector machines are becoming increasingly popular in dealing with regression problem. Firmly rooted in the VC theory in the field of machine learning, this method was originally developed for the classification problem by Vapnik and the coworkers. By introducing the ε -insensitive loss function, it had been extended to solve the regression problem. In this paper, the dimensional superiority and linearly separability are discussed first. Then, the kernel function is introduced to perform the nonlinear mapping into a high dimensional feature space and calculate the inner product in the condition of small sample. Finally, the basic theory and algorithm of SVR are discussed in detail. It can be concluded that SVR work in the following two steps: (1) Nonlinearly mapping of the original data into a feature space, and (2) Then constructing a linear OSH with the maximal margin.

For demonstrating the performance of SVR, a set of geology, geophysics data for iron deposit resource are analyzed by the method. The results also show that the overall performance of SVR is better than that of PLS. Therefore, it should be further concluded that the mystery that SVMs are usually more flexible and have better generalization performance compared with the traditional statistical or machine learning methods is due to the intrinsic nature in mathematics.

Currently, in the field of mineral resource prediction, most of the regression problem are done using linear methods, e.g. PLS, which need much model data. However, most of data is nonlinearity and sample data set is smaller in mineral resource prediction. In this context, the traditional methods are hard. In the research a new approach, support vector regress is applied in prediction and solves the above problem. It is a meaningful attempt in mineral resource prediction, which will help geological engineers quantitatively predict an ore prospecing area and resources using nonlinear geological data in prospective mining area with less geological works. It is significant for geologists to extend ore-search thinking, to innovate ore-search techniques and methods and to enhance ore-search success rate. At the same time, SVR also gain a wide variety of applications in reality.

Several factors may affect the prediction accuracy. One is the diversity of iron deposit data. It is likely that not all possible types of iron deposit data are adequately represented in the training sets. This can be improved along with the availability of more iron deposit data. SVR optimization procedure and feature vector selection may

also be improved.

ACKNOWLEDGMENT

REFERENCES

Chang CC, Lin CJ (2001). LIBSVM: a library for support vector machines, 2001 (Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm).

Scholkopf B, Sung K, Burges C, Girosi F, Niyogi P, Poggio T, Vapnik V (1997). IEEE Trans. Signal Proces., 45: 2758–2765.

Stitson M, Gammerman A, Vapnik V, Vovk V, Watkins C (1999). Weston, Advances in Kernel Methods—Support Vector Learning, MIT Press, Cambridge, MA, 285–292.

Vapnik VN, Chervonekis AY (1991). The necessary and sufficient conditions forthe consistency of the method of empirical minimization, Pattern Recognit. Image Anal. 1: 284–305.

Vapnik VN (1999). The Nature of Statistical Learning Theory, Second Edition, Springer, New York.

Vapnik VN (1999). An overview of statistical learning theory, IEEE Trans. Neural Networks, 10(5): 988.