

Prediction of toxic metals concentration using artificial intelligence techniques

R. Gholami · A. Kamkar-Rouhani ·
F. Doulati Ardejani · Sh. Maleki

Received: 17 May 2011 / Accepted: 21 September 2011 / Published online: 1 November 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Groundwater and soil pollution are noted to be the worst environmental problem related to the mining industry because of the pyrite oxidation, and hence acid mine drainage generation, release and transport of the toxic metals. The aim of this paper is to predict the concentration of Ni and Fe using a robust algorithm named support vector machine (SVM). Comparison of the obtained results of SVM with those of the back-propagation neural network (BPNN) indicates that the SVM can be regarded as a proper algorithm for the prediction of toxic metals concentration due to its relative high correlation coefficient and the associated running time. As a matter of fact, the SVM method has provided a better prediction of the toxic metals Fe and Ni and resulted the running time faster compared with that of the BPNN.

Keywords Prediction · Toxic metals · Support vector machine · Sarcheshmeh cooper mine · Back-propagation neural network

Introduction

Copper exploitation causes a major water quality problem due to acid mine drainage (AMD) generation in Sarcheshmeh mine, Kerman Province, southeast Iran. The oxidation of sulphide minerals particularly pyrite exposed to atmospheric oxygen during or after mining activities generates acidic waters with high concentrations of dissolved iron (Fe), sulphate (SO_4) and both of the heavy and toxic metals (Williams 1975; Moncur et al. 2005). The low pH of AMD may cause further dissolution and the leaching of additional metals (Mn, Zn, Cu, Cd, and Pb) into aqueous system (Zhao et al. 2007). AMD containing heavy and toxic metals has detrimental impact on aquatic life and the surrounding environment. Shur River in the Sarcheshmeh copper mine area is polluted by AMD with pH values ranging between 2 and 4.5 and high concentrations of heavy and toxic metals. The prediction of toxic metals in Shur River is useful in developing proper remediation and monitoring methods. Environmental problems due to the oxidation of sulphide minerals and hence AMD generation in the Sarcheshmeh copper mine and its impact on the Shur River have been investigated in the past (Marandi et al. 2007; Shahabpour and Doorandish 2008; Doulati Ardejani et al. 2008; Bani Assadi et al. 2008; Derakhshandeh and Alipour 2010). In addition, several investigations have been carried out using artificial neural networks (ANN) multiple linear regression (MLR) in different fields of environmental engineering in the past few decades (Karunanithi et al. 1994; Lek and Guegan 1999; Govindaraju 2000; Karul et al. 2000; Bowers and Shedrow 2000; Kemper and Sommer 2002; Dedecker et al. 2004; Kuo et al. 2004, 2007; Khandelwal and Singh 2005 Almasri and Kaluarachchi 2005; Kurunc et al. 2005; Sengorur et al. 2006; Messikh et al. 2007; Palani et al. 2008; Hanbay et al. 2008; Chenard and Caissie 2008; Dogan et al. 2009; Singh et al. 2009).

R. Gholami (✉) · A. Kamkar-Rouhani · F. Doulati Ardejani ·
Sh. Maleki
Faculty of Mining, Petroleum and Geophysics, Shahrood
University of Technology, Shahrood, Iran
e-mail: Raof.Gholami@Gmail.com

A. Kamkar-Rouhani
e-mail: Kamkarr@yahoo.com

F. Doulati Ardejani
e-mail: fdoulati@shahroodut.ac.ir

Sh. Maleki
e-mail: Sh.Maleki.Ch@Gmail.com

Recently, a novel machine learning technique, called support vector machine (SVM), has drawn much attention in the fields of pattern classification and regression forecasting. SVM was first introduced by Vapnik (1995). SVM is a kind of classification methods on statistic study theory. This algorithm derives from linear classifier, and can solve the problem of two-kind classifier, later this algorithm applies in non-linear fields, i.e. to say, we can find the optimal hyperplane (large margin) to classify the samples set. It is an approximate implementation to the structure risk minimization (SRM) principle in statistical learning theory (SLT), rather than the empirical risk minimization (ERM) method (Kwok 1999). Compared with traditional neural networks, SVM can use the theory of minimizing the structure risk to avoid the problems of excessive study, calamity data, local minimal value and etc. For the small samples set, this algorithm can be generalized well. SVM has been successfully used for machine learning with large and high-dimensional datasets. These attractive properties make SVM become a promising technique. This is due to the fact that the generalization property of a SVM does not depend on the complete training data but only a subset, the so-called support vectors. Now, SVM has been applied in many fields as follows: handwriting recognition, three-dimension objects recognition, faces recognition, text images recognition, voice recognition, regression analysis, and so on (Carbonneau et al. 2008; Chen and Hsieh 2006; Huang 2008; Seo 2007; Trontl et al. 2007; Wohlberg et al. 2006). The aim of this paper is to predict the concentration of two toxic metals namely Fe and Ni using SVM. For making a good comparison, the obtained results will be compared with those given by a back-propagation neural network (BPNN).

Study area

Sarcheshmeh copper mine is located 160 km to southwest of Kerman and 50 km to southwest of Rafsanjan in Kerman province, Iran. The main access road to the study area is Kerman-Rafsanjan-Shahr Babak road. This mine belongs to Band Mamazar-Pariz Mountains. The average elevation of the mine is 1,600 m. The mean annual precipitation of the mine area varies from 300 to 550 mm. The temperature varies from +35°C in summer to −20°C in winter. The area is covered with snow about 3–4 months per year. The wind speed sometimes exceeds to 100 km/h. A rough topography is predominant at the mining area. Figure 1 shows the geographical position of the Sarcheshmeh copper mine.

The orebody in Sarcheshmeh is oval shaped with a long dimension of about 2,300 m and a width of about 1,200 m. This deposit is associated with the late Tertiary Sarcheshmeh granodiorite porphyry stock (Waterman and

Hamilton 1975). The porphyry is a member of a complex series of magmatically related intrusives emplaced in the Tertiary volcanics at a short distance from the edge of an older near-batholith-sized granodiorite mass. Open pit mining method is used to extract copper ore in the Sarcheshmeh mine. A total of 40,000 tons of ore (average grades 0.9% Cu and 0.03% molybdenum) are approximately extracted per day from the Sarcheshmeh mine (Banisi and Finch 2001).

Sampling and field methods

Sampling of water in the Shur River downstream from the Sarcheshmeh mine was carried out in February 2006. Water samples consist of water from the Shur River (Fig. 1) originating from the Sarcheshmeh mine, acidic leachates of heap structure, run-off of leaching solution into the River and tailings along the Shur River. The water samples were immediately acidified by adding HNO₃ (10 cc acid to 1,000 cc sample) and stored under cool conditions. The equipments used in this study consisted of sample container, GPS, oven, autoclave, pH meter, atomic adsorption and ICP analysers. The pH of the water samples was measured using a portable pH meter in the field. Other field measured quantities were total dissolved solids (TDS), electric conductivity (EC) and temperature. Analyses for dissolved metals were performed using atomic adsorption spectrometer (AA220) in the Water Laboratory of the National Iranian Copper Industries Company (NICIC). The ICP (model 6000) analysis method was also used to analyze the concentrations of those metals, usually detected in the range of ppb. Table 1 gives the minimum, maximum and the mean values of the some physical and chemical measured quantities.

Support vector machine

In pattern recognition, the SVM algorithm constructs nonlinear decision functions by training a classifier to perform a linear separation in some high dimensional space which is nonlinearly related to input space. To generalize the SVM algorithm for regression analysis, an analog of the margin is constructed in the space of the target values (y) using Vapnik's ε -insensitive loss function (Fig. 2) (Quang-Anh et al. 2005; Stefano and Giuseppe 2006).

$$|y - f(x)|_{\varepsilon} := \max\{0, |y - f(x) - \varepsilon|\} \quad (1)$$

To estimate a linear regression

$$f(x) = (w \cdot x) + b \quad (2)$$

where w is the weighted matrix, x is the input vector and b is the bias term. With precision, one minimizes

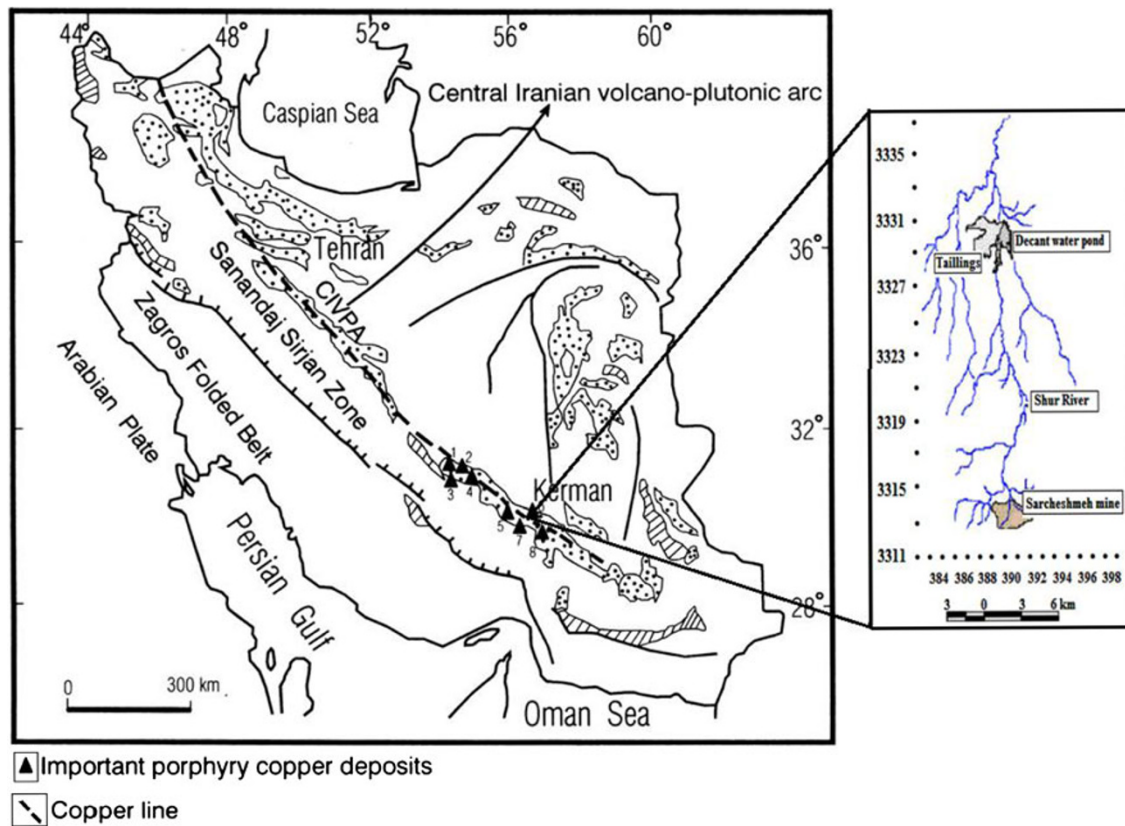


Fig. 1 Location of the Sarcheshmeh cooper mine and Shur River

Table 1 Maximum, minimum and mean physical and chemical constituents including toxic metals of the Shur River (concentrations of elements are given in ppm)

	pH	SO ₄ ²⁻	HCO ₃ ⁻	Ca ²⁺	Mg ²⁺	Ni	Fe	TDS	EC (μS/cm)
Min	3.3	27	0	92	13	0.02	0.01	446	870
Max	7.20	1,526	628	460	123	25	23	2,080.68	2,260
Mean	5.27	778.45	34.01	182.78	56.70	4.5	4.60	1,009.90	1,306.52

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^m |y - f(x)|_\varepsilon \tag{3}$$

Where *C* is a trade-off parameter to ensure the margin ε is maximized and error of the classification ξ is minimized. Considering a set of constraints, one may write the following relations as a constrained optimization problem:

$$L(w, \xi, \xi') = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi'_i) \tag{4}$$

$$\text{Subject to } \begin{cases} y_i - w^T \cdot x - b \leq \xi_i + \varepsilon & (5) \\ y_i + w^T \cdot x + b \leq \xi'_i + \varepsilon & (6) \\ \xi_i, \xi'_i, x_i \geq 0 & (7) \end{cases}$$

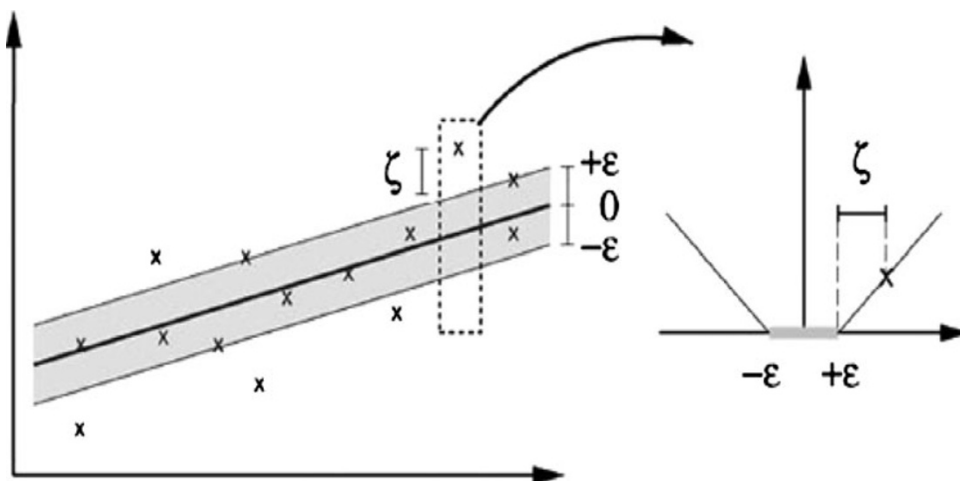
That according to relations (5) and (6), any error smaller than ε does not require a nonzero ξ_i or ξ'_i , and does not enter the objective function (4) (Lia et al. 2007; Hwei-Jen and Jih Pin 2009; Eryarsoy et al. 2009).

By introducing Lagrange multipliers (α and α') and allowing for $C > 0$, $\varepsilon > 0$ chosen a priori, the equation of an optimum hyper plane is achieved by maximizing of the following relations:

$$L(\alpha, \alpha') = \frac{1}{2} \sum_{i=1}^N \sum_{i=1}^N (\alpha_i - \alpha'_i) x'_i x_i (\alpha_i - \alpha'_i) + \sum_{i=1}^N ((\alpha_i - \alpha'_i) y_i - (\alpha_i + \alpha'_i) \varepsilon) \tag{8}$$

$$\text{Subject to } 0 \leq (\alpha_i - \alpha'_i) \leq C \tag{9}$$

Fig. 2 Concept of ϵ -insensitivity. Only the samples out of the $\pm\epsilon$ margin will have a non-zero slack variable, so they will be the only ones that will be part of the solution



Where, x_i only appears inside an inner product. To get a potentially better representation of the data in non-linearized case, the data points can be mapped into an alternative space, generally called feature space (a pre-Hilbert or inner product space) through a replacement:

$$x_i \cdot x_j \rightarrow \phi(x_i) \cdot \phi(x_j) \tag{10}$$

The functional form of the mapping $\phi(x_i)$ does not need to be known since it is implicitly defined by the choice of kernel: $k(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ or inner product in Hilbert space. With a suitable choice of kernel, the data can become separable in feature space while the original input space is still non-linear. Thus, whereas data for n-parity or the two spirals problem are non-separable by a hyper plane in input space, it can be separated in the feature space by the proper kernels (Scholkopf et al. 1998; Walczack and Massart 1996; Rosipal and Trejo 2004; Mika et al. 1999; Scholkopf and Smola 2002; Gunn 1997). Table 2 gives some of the common kernels.

Then, the nonlinear regression estimate takes the following form:

$$y_i = \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha'_i) \phi(x_i)^T \phi(x_j) + b = \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha'_i) K(x_i, x_j) + b \tag{11}$$

Table 2 Polynomial, normalized polynomial and radial basis function (Gaussian) kernels (Wang 2005)

Kernel function	Type of classifier
$K(x_i, x_j) = (x_i^T x_j + 1)^\rho$	Complete polynomial of degree ρ
$K(x_i, x_j) = \frac{(x_i^T x_j + 1)^\rho}{\sqrt{(x_i^T x_i)(x_j^T x_j)}}$	Normalized polynomial kernel of degree ρ
$K(x_i, x_j) = \exp\left[-\frac{\ x_i - x_j\ ^2}{2\sigma^2}\right]$	Gaussian (RBF) kernel with parameter σ which controls the half-width of the curve fitting peak

Where b is computed using the fact that equation (5) becomes an equality with $\xi_i = 0$ if $0 < \alpha_i < C$, and relation (6) becomes an equality with $\xi'_i = 0$ if $0 < \alpha'_i < C$ (Chih-Hung et al. 2009; Sanchez 2003).

Network training: the over-fitting problem

One of the most common problems in the training process is the over fitting phenomenon. This happens when the error on the training set is driven to a very small value, but when new data are presented to the network the error is large. This problem occurs mostly in case of large networks with only few available data. Demuth and Beale (2002) have shown that there are a number of ways to avoid over-fitting problem. Early stopping and automated Bayesian regularization methods are the most common. However, with immediate fixing the error and the number of epochs to an adequate level (not too low/not too high) and dividing the data into two sets: training and testing, one can avoid such problem by making several realizations and selecting the best of them. In this paper, the necessary coding was added through MATLAB multi-purpose commercial software to implement the automated Bayesian regularization for training both the SVM and BPNN. In this technique, the available data are divided into two subsets. The first subset is the training set, which is used for computing the gradient and updating the network weights and biases. The second subset is the test set. This method works by modifying the performance function, which is normally chosen to be the sum of squares of the network errors on the training set. The typical performance function that is used for training neural networks is the mean sum of squares of the network errors according to the following equation:

$$mse = \frac{1}{N} \sum_{i=1}^N (e_i)^2 = \frac{1}{N} \sum_{i=1}^N (t_i - a_i)^2 \tag{12}$$

where N represents the number of samples, a is the predicted value, t denotes the measured value and e is the error.

It is possible to improve generalization if we modify the performance function by adding a term that consists of the mean of the sum of squares of the network weights and biases which is given by:

$$\text{msereg} = \gamma \text{mse} + (1 - \gamma) \text{msw} \quad (13)$$

Where, msereg is the modified error, γ is the performance ratio, and msw can be written as:

$$\text{msw} = \frac{1}{N} \sum_{i=1}^N w_i \quad (14)$$

Performance function will cause the network to have smaller weights and biases, and this will force the network response to be smoother and less likely to over fit (Demuth and Beale 2002).

Data set

One of the main objectives of this study is to predict the concentrations of Ni and Fe in the samples collected from the Shur River nearby the Sarcheshmeh cooper mine using SVM and BPNN methods. In this regard, physical and chemical constitutions (given in Table 1) are considered as the inputs, whereas Ni and Fe concentrations are taken as the output of the network in the both methods. In view of the requirements of the SVM and back-propagation neural computation algorithms, the data of both the input and output variables were normalized to an interval by a transformation process. In this study, normalization of data (inputs and outputs) was carried out that the normalized results were transformed to the range of $(-1, 1)$ using equation (15) and the number of train data (40) and test data (16) were then selected randomly.

$$p_n = 2 \frac{p - p_{\min}}{p_{\max} - p_{\min}} - 1 \quad (15)$$

where, p_n is the normalized parameter, p denotes the actual parameter, p_{\min} represents the minimum of the actual parameters and p_{\max} stands for the maximum of the actual parameters. In addition, the leave-one-out (LOO) cross-validation of the whole training set was used for adjusting the associated parameters of the networks (Liu et al. 2006).

Prediction of toxic metals concentration by SVM

Similar to other multivariate statistical models, the performance of SVM for regression depend on the

combination of several parameters. They are capacity parameter C , ε of ε -insensitive loss function, the kernel type K and its corresponding parameters. C is a regularization parameter that controls the trade-off between maximizing the margin and minimizing the training error. If C is too small then insufficient stress will be placed on fitting the training data. If C is too large then the algorithm will overfit the training data. However, Wang et al. (2003) indicated that prediction error was scarcely influenced by C . To make the learning process stable, a large value should be set up for C (e.g., $C = 100$).

The optimal value for ε depends on the type of noise present in the data, which is usually unknown. Even if enough knowledge of the noise is available to select an optimal value for ε , there is the practical consideration of the number of resulting support vectors. ε -insensitivity prevents the entire training set meeting boundary conditions, and so allows for the possibility of sparsity in the dual formulations solution. So, choosing the appropriate value of ε is critical from theory.

Since in this study the nonlinear SVM is applied, it would be necessary to select a suitable kernel function. The obtained results of previous published researches (e.g. Dibike et al. 2001; Han and Cluckie 2004) indicate the Gaussian radial basis function has superior efficiency than other kernel functions. As seen in the Table 1, the form of the Gaussian kernel is as follow:

$$K(x_i, x_j) = e^{-|x_i - x_j|^2 / 2\sigma^2} \quad (16)$$

In addition, where σ is a constant parameter of the kernel and can either control the amplitude of the Gaussian function and the generalization ability of SVM. We have to optimize σ and find the optimal one.

To find the optimum values of two parameters (σ and ε) and prohibit the overfitting of the model, the dataset was separated into a training set of 40 compounds and a test set of 16 compounds randomly and the LOO cross-validation of the whole training set was performed. The LOO procedure consists of removing one example from the training set, constructing the decision function on the basis only of the remaining training data and then testing on the removed example (Liu et al. 2006). In this fashion one tests all examples of the training data and measures the fraction of errors over the total number of training examples. The root mean square (RMS) error was used as an error function to evaluate the quality of model.

Detailed process of selecting the parameters and the effects of every parameter on generalization performance of the corresponding model are shown in Fig. 3. To obtain the optimal value of σ , the SVM with different σ were trained, the σ varying from 0.01 to 0.3, every 0.01. We calculated the RMS errors for different σ , according to the

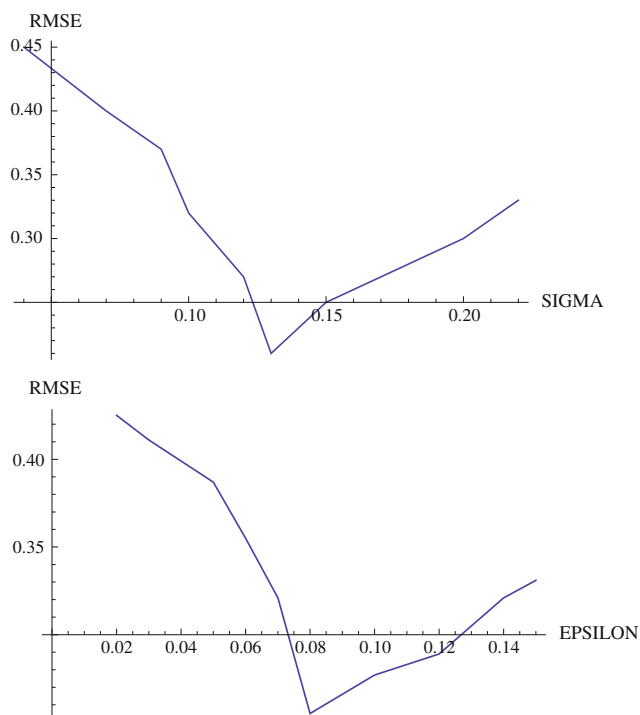
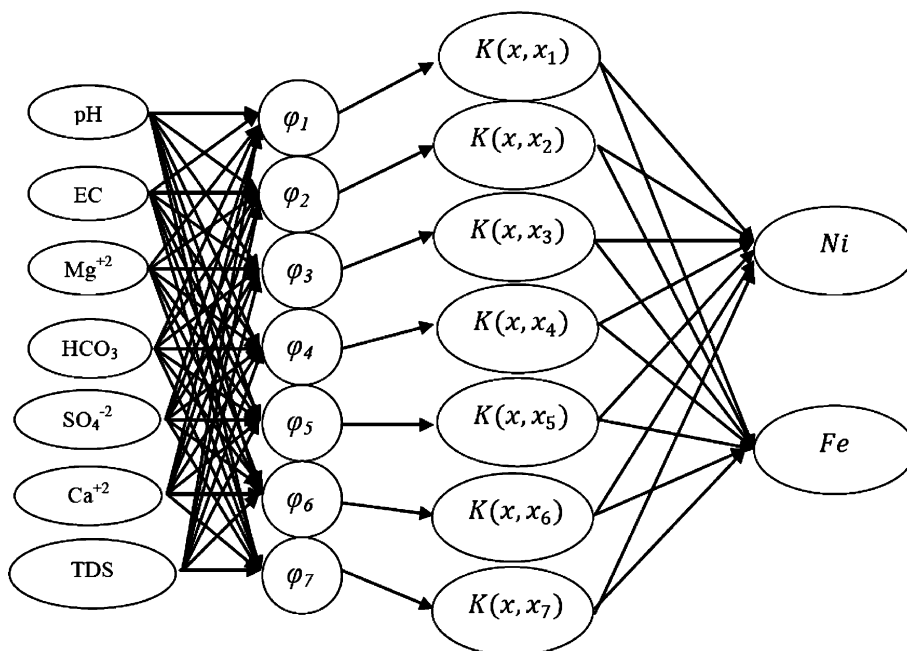


Fig. 3 RMS error versus σ (left) and RMS error versus ε (right) in LOO cross-validation stage

generalization ability of the model based on the LOO cross-validation for the training set to determine the optimal one. The curve of RMS error versus the sigma was shown in Fig. 3. The optimal σ was found as 0.13. To find an optimal ε , the RMS errors for different ε s, were calculated. The curve of the RMS error versus ε was shown in Fig. 3. From Fig. 3, the optimal ε was found as 0.08.

Fig. 4 Schematic diagram of optimum SVM for prediction of Ni and Fe concentration



From the above discussion, the σ , ε and C were fixed to 0.13, 0.08 and 100, respectively, when the support vector number of the SVM model was 48. Figure 4 is a schematic diagram showing the construction of the SVM.

Afterward, the most relevant input variables for predicting the concentration of Ni and Fe among many combinations of attributes (different physical and chemical parameters provided in the Table 1), the best input variables were selected by the trial and error method (Table 3). Two criteria were used to evaluate the effectiveness of each network and its ability to make accurate predictions. The RMS error can be calculated as follows:

$$\text{RMS} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (17)$$

where, y_i is the measured value, \hat{y}_i denotes the predicted value, and n stands for the number of samples. RMS error indicates the discrepancy between the measured and predicted values. The lowest the RMS, the more accurate the prediction is. Furthermore, the efficiency criterion, R , is given by:

$$R = \sqrt{1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n \hat{y}_i)^2}{n}}} \quad (18)$$

Where, R efficiency criterion represents the percentage of the initial uncertainty explained by the model. The best fitting between measured and predicted values, which is unlikely to occur, would have $\text{RMS} = 0$ and $R = 1$. It was found that combination of seven parameters (pH, SO_4 , HCO_3 , TDS, EC, Mg, and Ca) is the most suitable input

Table 3 Correlation coefficient (R) and RMS error of the prediction based on the different input variables

Input variables	Fe				Ni			
	R (train)	R (test)	RMSE (train)	RMSE (test)	R (train)	R (test)	RMSE (train)	RMSE (test)
pH, SO ₄ , HCO ₃	0.910	0.78	0.34	0.46	0.923	0.79	0.28	0.43
pH, SO ₄ , HCO ₃ , TDS, EC	0.935	0.81	0.27	0.36	0.944	0.83	0.25	0.33
pH, SO ₄ , HCO ₃ , TDS, EC, Mg	0.955	0.87	0.22	0.29	0.961	0.89	0.19	0.23
pH, SO ₄ , HCO ₃ , TDS, EC, Mg, Ca	0.98	0.92	0.12	0.21	0.98	0.95	0.13	0.2

variable. Table 3 gives the correlation coefficient (R) and RMS of the prediction based on the different input variables.

In the next stage, the performance of SVM was evaluated using the measured and predicted concentrations. Figure 5 can provide a good insight into the process of prediction.

As it is quite observed in Fig. 5, there is an acceptable agreement (correlation coefficient of 0.92 for the Fe and correlation coefficient of 0.95 for Ni) between the predicted and measured dataset. Based on the Fig. 5, the SVM is a proper method for the prediction of toxic metal concentration. Nonetheless, the performance of this method should be compared with another suitable method for highlighting the highly performance of the SVM.

Back-propagation neural network

Back-propagation neural networks are usually recognized for their prediction capabilities and ability to generalize well on a wide variety of problems. These models are a supervised type of networks, in other words, trained with both inputs and target outputs. During training, the network tries to match the outputs with the desired target values. Learning starts with the assignment of random weights. The output is then calculated and the error is estimated. This error is used to update the weights until the stopping

criterion is reached. It should be noted that the stopping criteria are usually the average error of epoch.

The optimal network for this study is a multilayer perceptron (Cybenko 1989, Hornik et al. 1989, Haykin 1994, Noori et al. 2009, 2010), that has one input layer with seven inputs (i.e. pH, SO₄, HCO₃, TDS, EC, Mg, and Ca), one hidden layers with six neurons that each neuron has a bias and is fully connected to all inputs and utilizes sigmoid activation function. The output layer has two neurons (Fe and Ni) with linear activation function (purelin) without any bias. Figure 6 can properly show the performance of BPNN in the prediction process.

As seen in Fig. 6, BPNN provides a good prediction for the Ni and Fe concentrations, but it is not as good as the SVM prediction. Nonetheless, there is good agreement between the measured and predicted concentration provided by BPNN (correlation coefficient of 0.88 for the Fe and correlation coefficient of 0.901 for Ni). Hence, BPNN can be considered as an alternative approach after the SVM for the prediction of the toxic metal concentration.

Discussion

In this research work, we have demonstrated one of the applications of SVM in forecasting the concentration of

Fig. 5 Relationship between the measured and predicted concentration by SVM; Fe (left) and Ni (right)

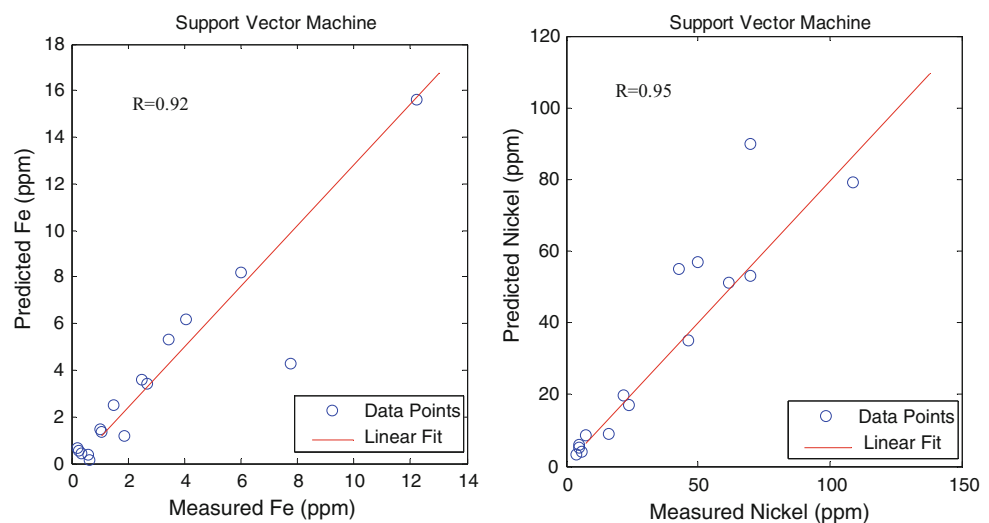


Fig. 6 Relationship between the measured and predicted concentration by BPNN; Fe (left) and Ni (right)

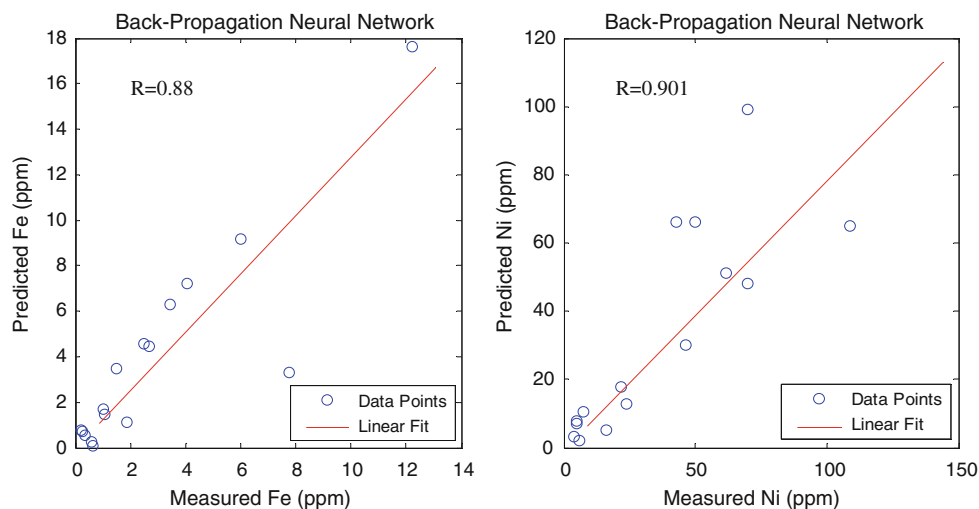


Table 4 Comparing the performance of SVM and BPNN methods in the training and testing steps

Type of network	Fe						Ni					
	<i>R</i> (train)	<i>R</i> (test)	RMSE (train)	RMSE (test)	Running time (train)	Running time (test)	<i>R</i> (train)	<i>R</i> (test)	RMSE (train)	RMSE (test)	Running time (Train)	Running time (Test)
SVM	0.98	0.92	0.12	0.21	4 s	3 s	0.98	0.95	0.13	0.2	4 s	3 s
BPNN	0.98	0.88	0.12	0.3	7 s	5 s	0.98	0.901	0.13	0.4	7 s	5 s

toxic metals. We interrogate the performance of this method by comparing the results with the best performed work of BPNN model. When compared SVM with this model (Table 4), we can see overall better performance of SVM over BPNN approach in terms of RMS error in both training and testing steps.

According to this table, the RMS error of SVM model is quite smaller than that of the BPNN. In terms of running time, In addition, the SVM consumes a considerably less time for prediction compared with that of the BPNN. For determining the relative running time of each network, Matlab multipurpose software has been used (i.e. relative codes of both networks have written in the Matlab software environment). As it is completely clear in the Table, the associated running time of SVM in training set is even less than that of the BPNN in the testing process. All of these expressions can introduce the SVM as a robust algorithm for the prediction process.

Conclusions

Support vector machine is a novel machine learning methodology based on SLT, which has considerable features including the fact that requirement on kernel and nature of the optimization problem results in a uniquely global optimum, high generalization performance, and

prevention from converging to a local optimal solution. In this research work, we have shown the application of SVM compared with BPNN model for prediction of the concentrations of two toxic metals, namely Fe and Ni, based on those chemicals and physical parameters obtained by conducting a sampling program nearby the Sarcheshmeh cooper mine, Iran. Although both methods are data-driven models, it has been found that SVM makes the running time considerably faster with the higher accuracy. In terms of accuracy, the SVM technique resulted in a less RMS error compared with that of the BPNN model (Table 4). Regarding the running time, SVM requires a small fraction of the computational time used by BPNN, which is an important factor to choose an appropriate and high-performance data-driven model.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

- Almasri MN, Kaluarachchi JJ (2005) Modular neural networks to predict the nitrate distribution in ground water using the on-ground nitrogen loading and recharge data. *Environ Model Softw* 20:851–871

- Bani Assadi A, Doulati Ardejani F, Karami GH, Dahr Azma B, Atash Dehghan R, Alipour M (2008) Heavy metal pollution problems in the vicinity of heap leaching of Sarcheshmeh porphyry copper mine. In: 10th International Mine Water Association Congress, June 2–5, 2008, Karlovy Vary, Czech Republic, pp 355–358
- Banisi S, Finch JA (2001) Testing a floatation column at the Sarcheshmeh copper mine. *Mineral Eng* 14(7):785–789
- Bowers JA, Shedrow CB (2000) Predicting stream water quality using artificial neural networks. *WSRC-MS-2000-00112*
- Carbonneau R, Laframbois K, Vahidov R (2008) Application of machine learning techniques for supply chain demand forecasting. *Eur J Oper Res* 184(3):1140–1154
- Chen RC, Hsieh CH (2006) Web page classification based on a support vector machine using a weighted vote schema. *Expert Syst Appl* 31(2):427–435
- Chenard JF, Caissie D (2008) Stream temperature modelling using neural networks: application on Catamaran Brook, New Brunswick. *Can Hydrol Process*. doi:1002/hyp.6928
- Chih-Hung W, Gwo-Hshung T, Rong-Ho L (2009) A Novel hybrid genetic algorithm for kernel function and parameter optimization in support vector regression. *Expert Syst Appl* 36:4725–4735
- Cybenko G (1989) Approximation by superposition of a sigmoidal function. *Math Control Signals Syst* 2:303–314
- Dedecker AP, Goethals PLM, Gabriels W, De Pauw N (2004) Optimization of artificial neural network (ANN) model design for prediction of macro invertebrates in the Awalm river basin (Flanders Belgium). *Ecol Model* 174:161–173
- Demuth H, Beale M (2002) Neural network toolbox for use with MATLAB, User's Guide Version 4
- Derakhshandeh R, Alipour M (2010) Remediation of Acid Mine Drainage by using Tailings Decant Water as a Neeutralization Agent in Sarcheshmeh Copper Mine. *Res J Environ Sci* 4(3):250–260
- Dibike YB, Velickov S, Solomatine D, Abbott MB (2001) Model Induction with support vector machines: Introduction and Application. *J Comput Civ Eng* 15(3):208–216
- Dogan E, Sengorur B, Koklu R (2009) Modeling biochemical oxygen demand of the Melen River in Turkey using an artificial neural network technique. *J Environ Manag* 90:1229–1235
- Doulati Ardejani F, Karami GH, Bani Assadi A, Atash Dehghan R (2008) Hydrogeochemical investigations of the Shour River and groundwater affected by acid mine drainage in Sarcheshmeh porphyry copper mine. In: 10th International mine water association congress, June 2–5, 2008, Karlovy Vary, Czech Republic, pp 235–238
- Eryarsoy E, Koehler Gary J, Aytug H (2009) Using domain-specific knowledge in generalization error bounds for support vector machine learning. *Decision Support Syst* 46:481–491
- Govindaraju RS (2000) Artificial neural network in hydrology. II: hydrologic application, ASCE task committee application of artificial neural networks in hydrology. *J Hydrol Eng* 5:124–137
- Gunn SR (1997) Support vector machines for classification and regression, Technical Report, Image Speech and Intelligent Systems Research Group, University of Southampton. Southampton, UK
- Han D, Cluckie I (2004) Support vector machines identification for runoff modeling. In: Liang SY, Phoon KK, Babovic V (eds) Proceedings of the sixth international conference on hydroinformatics, June, Singapore, pp 21–24
- Hanbay D, Turkoglu I, Demir Y (2008) Prediction of wastewater treatment plant performance based on wavelet packet decomposition and neural networks. *Expert Syst Appl* 34(2):1038–1043
- Haykin S (1994) Neural networks: a comprehensive foundation, 2nd edn. Macmillan, New York
- Hornik K, Stinchcombe M, White H (1989) Multilayer feed forward networks are universal approximators. *Neural Netw* 2:359–366
- Huang SC (2008) Online option price forecasting by using unscented Kalman filters and support vector machines. *Expert Syst Appl* 34(4):2819–2825
- Hwei-Jen L, Jih Pin Y (2009) Optimal reduction of solutions for support vector machines. *Appl Math Comput* 214:329–335
- Karul C, Soyupak S, Cilesiz AF, Akbay N, German E (2000) Case studies on the use of neural networks in eutrophication modeling. *Ecol Model* 134:145–152
- Karunanithi N, Grenney WJ, Whitley D, Bovee K (1994) Neural networks for river flow prediction. *ASCE J Comput Civ Eng* 8:210–220
- Kemper T, Sommer S (2002) Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy. *Environ Sci Technol* 36:2742–2747
- Khandelwal M, Singh TN (2005) Prediction of mine water quality by physical parameters. *J Sci Ind Res* 64:564–570
- Kuo Y, Liu C, Lin KH (2004) Evaluation of the ability of an artificial neural network model to assess the variation of groundwater quality in an area of blackfoot disease in Taiwan. *Water Res* 38:148–158
- Kuo J, Hsieh M, Lung W, She N (2007) Using artificial neural network for reservoir eutrophication prediction. *Ecol Model* 200:171–177
- Kurunc A, Yurekli K, Cevik O (2005) Performance of two stochastic approaches for forecasting water quality and stream flow data from Yesilirmak River, Turkey. *Environ Model Softw* 20:1195–1200
- Kwok JT (1999) Moderating the outputs of support vector machine classifiers. *IEEE Trans Neural Netw* 10(5):1018–1031
- Lek S, Guegan JF (1999) Artificial neural networks as a tool in ecological modelling, an introduction. *Ecol Model* 120:65–73
- Lia Q, Licheng J, Yingjuan H (2007) Adaptive simplification of solution for support vector machine. *Pattern Recogn* 40:972–980
- Liu H, Yao X, Zhang R, Liu M, Hu Z, Fan B (2006) The accurate QSPR models to predict the bioconcentration factors of nonionic organic compounds based on the heuristic method and support vector machine. *Chemosphere* 63:722–733
- Marandi R, Doulati Ardejani F, Marandi A (2007) Biotreatment of acid mine drainage using sequencing batch reactors (SBRs) in the Sarcheshmeh porphyry copper mine. Cidu R, Frau F (eds) IMWA symposium 2007: water in mining environments, 27–31 May 2007, Cagliari, Italy, pp 221–225
- Messikh N, Samar MH, Messikh L (2007) Neural network analysis of liquid–liquid extraction of phenol from wastewater using TBP solvent. *Desalination* 208:42–48
- Mika S, Ratsch G, Wetson J, Scholkopf B, Muller KR (1999) Fisher discriminant analysis with kernels. In: Proc. NNSP'99, pp 41–48
- Moncur MC, Ptacek CJ, Blowes DW, Jambor JL (2005) Release, transport and attenuation of metals from an old tailings impoundment. *Appl Geochem* 20:639–659
- Noori R, Abdoli MA, Jalili Ghazizade M, Samieifard R (2009) Comparison of neural network and principal component-regression analysis to predict the solid waste generation in Tehran. *Iran J Public Health* 38:74–84
- Noori R, Khakpour A, Omidvar B, Farokhnia A (2010) Comparison of ANN and principal component analysis-multivariate linear regression models for predicting the river flow based on developed discrepancy ratio statistic. *Expert Syst Appl* 37:5856–5862
- Palani S, Liang S, Tklich P (2008) An ANN application for water quality forecasting. *Marian Pollut Bull* 56:1586–1597
- Quang-Anh T, Xing L, Haixin D (2005) Efficient performance estimate for one-class support vector machine. *Pattern Recogn Lett* 26:1174–1182
- Rosipal R, Trejo LJ (2004) Kernel partial least squares regression in reproducing kernel Hilbert space. *J Mach Learn Res* 2:97–123

- Sanchez DV (2003) Advanced support vector machines and kernel methods. *Neurocomputing* 55:5–20
- Scholkopf B, Smola AJ (2002) *Learning with kernels*. MIT press, Cambridge
- Scholkopf B, Smola AJ, Muller KR (1998) Nonlinear component analysis as a kernel eigenvalues problem. *Neural Comput* 10:1299–1319
- Sengorur B, Dogan E, Koklu R, Samandar A (2006) Dissolved oxygen estimation using artificial neural network for water quality control. *Fresen Environ Bull* 15:1064–1067
- Seo KK (2007) An application of one-class support vector machines in contentbased image retrieval. *Expert Syst Appl* 33(2):491–498
- Shahabpour J, Doorandish M (2008) Mine drainage water from the Sarcheshmeh porphyry copper mine, Kerman, IR Iran. *Environ Monit Assess* 141:105–120
- Singh KP, Basant A, Malik A, Jain G (2009) Artificial neural network modeling of the river water quality—a case study. *Ecol Model* 220:888–895
- Stefano M, Giuseppe J (2006) Terminated ramp–support vector machines: a nonparametric data dependent kernel. *Neural Netw* 19:1597–1611
- Trontl K, Smuc T, Pevec D (2007) Support vector regression model for the estimation of c-ray buildup factors for multi-layer shields. *Ann Nucl Energy* 34(12):939–952
- Vapnik V (1995) *The nature of statistical learning theory*. Springer, New York, p 188
- Walczack B, Massart DL (1996) The radial basis functions—partial least squares approach as a flexible non-linear regression technique. *Anal Chim Acta* 331:177–185
- Wang L (2005) *Support vector machines: theory and applications*. Nanyang Technological University, School of Electrical & Electronic Engineering, Springer, Berlin
- Wang WJ, Xu ZB, Lu WZ, Zhang XY (2003) Determination of the spread parameter in the Gaussian kernel for classification and regression. *Neurocomputing* 55:643–663
- Waterman GC, Hamilton RL (1975) The Sar Cheshmeh porphyry copper deposit. *Econ Geol* 70:568–576
- Williams RE (1975) *Waste production and disposal in mining, milling, and metallurgical industries*. Miller-Freeman Publishing Company, San Francisco, p 489
- Wohlberg B, Tartakovsky DM, Guadagnini A (2006) Subsurface characterization with support vector machines. *IEEE Trans Geosci Remote Sens* 44(1):47–57
- Zhao F, Cong Z, Sun H, Ren D (2007) The Geochemistry of rare earth elements (REE) in acid mine drainage from the Sitai coal mine, Shanxi Province, North China. *Int J Coal Geol* 70:184–192