Iranica Journal of Energy & Environment 5 (4): 345-353, 2014
ISSN 2079-2115
IJEE an Official Peer Reviewed Journal of Babol Noshirvani University of Technology
DOI: 10.5829/idosi.ijee.2014.05.04.02

BUT

Prediction of Degree of Soil Contamination Based on Support Vector Machine and K-Nearest Neighbor Methods: A Case Study in Arak, Iran

Faridon Ghadimi

Arak University of Technology, Arak, Iran

Received: August 4, 2014; Accepted in Revised form: October 5, 2014

Abstract: The degree of soil contamination in an urban region can be changed by heavy metals. This might result in endangering safety of an urban region. This paper presents an approach to build a prediction model for the assessment of degree of contamination index, based upon heavy metals changes. The heavy metal concentration of Pb, Cu, Ni, Zn, As, Cr and Ni as input was used to build a prediction model for the assessment of degree of contamination. Two prediction models were implemented such as support vector regression (SVR) and k-nearest neighbor regression method (KNNR). A comparison was made between these two models and the results showed the superiority of the SVR model. Furthermore, a case study in Arak, Iran was conducted to illustrate the capability of the support vector machines (SVM) model.

Key words: Degree of contamination · Heavy metals · Support vector machines · K-Nearest Neighbor · Arak

INTRODUCTION

With rapid development in industrialization, soil contamination has become a serious problem in many countries. Contamination and negative impact on the quality of air, water and soil by population growth, rapid urbanization and industrial activities have been discussed in literature [1-3]. Among the most significant soil contaminants resulted from both natural and manmade sources, heavy metals are of prime importance due to their long-term toxicity effect [4-7]. Increase in metal content in soils is generally observed in areas of intense industrial activities. Metal accumulation in these areas is a few times higher than uncontaminated sites [8]. The most important impact of soil pollution on environmental health is that contaminants in soil may be transmitted into the food chain through plant vegetation and direct use or animal consumption of feedstock [9-11].

Arak is one of the industrialized cities in Iran where the impact of rapid population growth and industrialization on limited natural sources is progressively high. Due to expanding industrialization and urbanization in Arak and the unrestrained disposal of factory wastes to soil or waters and their transport through the air, it is thought that heavy metal contents of soils in this region are high [10, 12]. Therefore, monitoring of these changes and prediction of contamination in soils has gained special attention.

Moreover, over the years, the application of support vector machines (SVM) in environmental sciences has been growing [13, 14]. In recent years, there is a growing interest of using SVM to assist building a reasonable model structure for nonlinear systems [15]. SVM have a special capacity to approximate the dynamics of nonlinear systems in many applications. Given sufficient input-output data, SVM is able to approximate any continuous function to arbitrary accuracy [16].

In this work we focused on the development of heavy metals for estimation of the soil contamination using SVM and on some of its methodological aspects. An investigation was conducted to identify whether SVM method was applicable for this purpose. For the purpose of automated models' parameters search genetic algorithms were used as an optimization framework. A new form of the aim function used for models' parameters search is proposed, which allows for the suitable selection of models parameters.

Corresponding Author: Faridon Ghadimi, Arak University of Technology, Arak, Iran.
Tel:+988633663562, Fax: +988633670020.

Please cite this article as: Faridon Ghadimi, 2014. Prediction of Degree of Soil Contamination Based on Support Vector Machine and K-Nearest Neighbor Methods: A Case Study in Arak, Iran . Iranica Journal of Energy and Environment, 5 (4): 345-353.

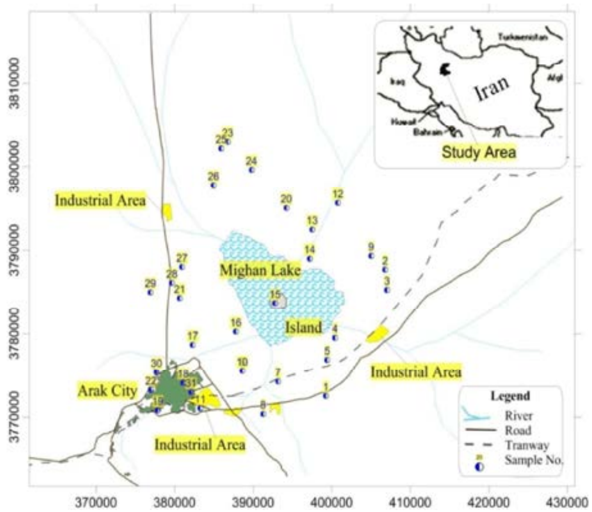

Fig. 1: Location of some of the collected samples in Arak.

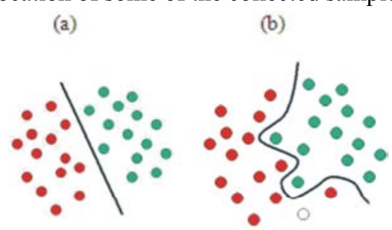

Fig. 2: Separating line (a) and curve (b) in boundary all objects

MATERIALS AND METHODS

Study Area: Arak is located in the center of Iran. Economy of the district is mainly based on industrial activities and it is one of the rapidly growing and developing regions in central province (Fig. 1). The rapid expansion of industrializations, particularly, after the 1990s has given unexpected rise in population of the city as well hosting several plants belong to various industrial sectors. There are several organized industrial small towns in Arak. Industrial facilities including paint, plastic, electric, metal, automotive supply industry, food, cosmetics, packing, machinery and chemical sectors are currently in operation in these organized industrial area. Normally, Arak has cold winter and rainy while summer is hot and very less rain. Arak annual precipitation is in the range of 250 and 400 mm. Soils in the region are well-developed, dark-colored and rich in organic-material and are included in brown soil group.

Sampling and Analysis: A total of 60 soil samples were collected from the outer surface (5– 10 cm) after removing surface contamination. Fig. 1 shows the location of the some of the soil samples collected from the area. Plastic spatula was used for sample collection. Soil samples were dried at room temperature and ground before analysis. The materials under 80-mesh sieve were sent to laboratory (Department of Mining Engineering, Arak University of Technology) for analyses. During the analysis, 1 g of soil sample was left in 2 ml HNO_3, 2M solution for 1 h. The samples were then added to 6 ml of 2:2:2 $HCl–HNO_3–H_2O$ solutions, dissolved at 95°C for 1 h and analyzed with ICP-MS.

Assessment of Degree of Contamination: A significant number of indicators designed to approximate the quality of soils can be found in literature [17, 18]. In our case, assessment of soil contamination level is performed by the quantification of the Pollution Index (C_f) known as contamination factor (C_f) and by the Contamination Degree (C_d) [17]. For each soil sample and each heavy metal the C_f has been calculated as the ratio between the metal concentrations with its background values as established for the study area by Guillén *et al*. [19]:

$$C_f = C_{heavy\ metal} / C_{background} \quad (1)$$

$$C_d = \Sigma C_f \quad (2)$$

where C_f (Contamination Factor) is the ratio between the concentrations of each metal in the soils and the reference background value (Table 1); and C_d is the contamination degree calculated as the sum of the C_f of each of considered metals. According to the literature [17] the variation in C_d can be defined as:

* $C_d < n$: low degree of contamination
* $n < C_d < 2n$: moderate degree of contamination
* $2n < C_d < 3n$: high degree of contamination
* $C_d > 3n$: very high degree of contamination

where n is the number of contaminants involved in the C_d determination.

Support Vector Machines: Support vector machines (SVM) is based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships [20, 21]. A schematic pattern is shown in Fig. 2. The separating line defines a boundary on the all objects. Most classification tasks, however, are not that simple and often more complex structures are needed in

Table 1: Concentrations of heavy metals in some soil samples around the Arak (All concentration in mg/kg)

Sample No.	Sampling satiation	Pb	Cu	Ni	Zn	As	Cr	Co
1	Shaveh	8.77	15.44	47.9	5.5	2.01	55.1	11.2
2	Salabad	7.9	14.95	61.7	8.5	1.41	68	9.05
3	Mazreh	3.9	28.35	63.1	8.41	1.54	25	10.1
4	Ghorogh	11.69	136.3	48.9	6.55	2.02	49	11.2
5	Moradabad	9.34	15.5	40.8	5.93	2.4	36	12.1
	Valid N	31	31	31	31	31	31	31
	Mean	8.99	37.60	39.70	12.16	2.48	58	12.47
	Median	8.05	20.39	43.67	7.20	2.34	52	11.23
	Minimum	3.01	9.60	121	0.85	1.41	25	7.54
	Maximum	17.36	195.55	63.12	68.91	4.50	130	24
	Std.Dev.	3.65	44.16	11.81	13.65	0.85	27.82	4.60
	Skewness	0.62	2.58	0.54	2.90	0.95	1.22	1.58
	Kurtosis	1.05	6.15	0.33	9.76	0.50	0.88	1.58
	Earth crust averages [29]	17	13	18	47	5	92	17

*Caritate *et al.*, 2012

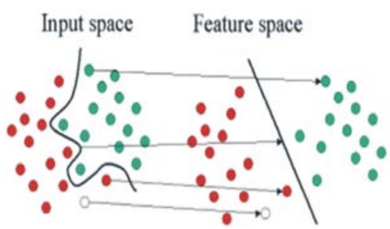

Fig. 3: Illustration basic idea about support vector machines

order to make an optimal separation, i.e., correctly classify new objects (test cases) that are available (train cases) [22]. Classification tasks based on drawing separating lines to distinguish between objects of different class memberships are known as hyper-plane classifiers. Support vector machines are particularly suited to handle such tasks [23]. The illustration in Fig. 3 shows the basic idea behind support vector machines. Here we observe the original objects (left side of the schematic) mapped, i.e., rearranged, using a set of mathematical functions, known as kernels. The process of rearranging the objects is known as mapping (transformation). Note that in this new setting, the mapped object (right side of the schematic) is linearly separable and, thus, instead of constructing the complex curve (left schematic), all we have to do is to find an optimal line that can separate the objects.

To construct an optimal hyperplane, SVM employs an iterative training algorithm, which is used to minimize an error function. According to the form of the error function, SVM models can be classified into two distinct groups known as: Classification SVM type 1 and Regression SVM. For classification SVM Type 1, training involves the minimization of the error function:

$$\frac{1}{2}w^T w + C\sum_{i=1}^{N}\zeta_i$$

$$y_i(w^T\phi(x_i)+b)\ge 1-\zeta_i \text{ and } \zeta_i \ge 0, i=1,...,N \quad (3)$$

where C is the capacity constant, w is the vector of coefficients, b is a constant and ξ_i represents parameters for handling non separable data (inputs). The index i label the N training cases. Note that $y \in \pm 1$ represents the class labels and x_i represents the independent variables. The kernel ϕ is used to transform data from the input (independent) to the feature space [24]. It should be noted that the larger the C, the more the error is penalized. Thus, C should be chosen with care to avoid over fitting. The task of regression SVM type1 is then to find a functional form for f(x) that can correctly predict new cases that the SVM has not been presented with before [13]. This can be achieved by training the SVM model on a sample set, i.e., training set, a process that involves, like classification, the sequential optimization of an error function. For this type of SVM the error function is:

$$y = f(x) + \text{noise} \quad (4)$$

$$\frac{1}{2}w^T w + C\sum_{i=1}^{N}\zeta_i + C\sum_{i=1}^{N}\zeta_i^{*} \quad (5)$$

$$w^T\phi(x_i)+b-y_i \le \varepsilon+\zeta_i^{*}$$

$$y_i - w^T\phi(x_i)-b_i \le \varepsilon+\zeta_i$$

$$\zeta_i, \zeta_i^{*} \ge 0, i=1,...,N$$

There are number of kernels that can be used in support vector machines models. These include linear, polynomial, radial basis function (RBF) and sigmoid:

$$K(X_i, X_j) = \begin{cases} X_i \bullet X_j & Linear \\ (\gamma X_i \bullet X_j + C)^d & Polynomial \\ \exp(-\gamma |X_i - X_j|^2) & RBF \\ \tanh(\gamma X_i \bullet X_j + C) & Sigmoid \end{cases} \quad (6)$$

where $K(X_i, X_j) = \phi(X_i) \bullet \phi(X_j)$
that is, the kernel function, represents a dot product of input data points mapped into the higher dimensional feature space by transformation ϕ. The RBF is by far the most popular choice of kernel types used in support vector machines. This is mainly because of their localized and finite responses across the entire range of the real x-axis. The SVM is a learning method with a theoretical root in statistical learning theory. The SVM was originally developed for classification and was later generalized to solve regression problems. This method is called support vector regression (SVR). The model produced by support vector classification only depends on a subset of the training data, because the cost function for building the model does not care about training points that lie beyond the margin [25, 26]. Analogously, the model produced by SVR only depends on a subset of the training data, because the cost function for building the model ignores any training data that are close to the model prediction.

K-Nearest Neighbors: K-Nearest Neighbors (KNN) achieves this by finding K examples that are closest in distance to the query point, hence, the name k-Nearest Neighbors [27]. For regression problems, KNN predictions are based on averaging the outcomes of the K nearest neighbors; for classification problems, a majority of voting is used. After selecting the value of k, you can make predictions based on the KNN. For regression, KNN prediction is the average of the K-nearest neighbors outcome. Since KNN predictions are based on the intuitive assumption that objects close in distance are potentially similar, it makes good sense to discriminate between the K nearest neighbors when making predictions, i.e., let the closest points among the K- nearest neighbors have more say in affecting the outcome of the query point. This can be achieved by introducing a set of weights W, one for each nearest neighbor, defined by the relative closeness of each neighbor with respect to the query point. Thus:

$$W(x, p_i) = \frac{\exp(-D(x, p_i))}{\sum_{i=1}^{k} \exp(-D(x, p_i))} \quad (7)$$

where D(x, p_i) is the distance between the query point x and the ith case p_i. It is clear that the weights defined in this manner above will satisfy:

$$\sum_{i=1}^{k} W(x_\circ, x_i) = 1$$

$$y = \sum_{i=1}^{k} W(x_\circ, x_i) y_i \quad (8)$$

For classification problems, the maximum of the above equation is taken for each class variables.

RESULTS AND DISCUSSION

Prediction of Degree of Contamination Classes: In this paper, the SVM are applied to model the degree of contamination index in statistical environment. Concentration of heavy metals was introduced as input parameters into the SVM models and contamination degree as output. The proposed models were trained with 45 data sets obtained from a degree of contamination equation in Arak. The 10 samples of training data sets are shown in Table 2.

The main objective of this study was to predict the classes of contamination degree of heavy metals in soil in the study area. The radial basis function (RBF) kernel function is used to prepare the best discrepancy between the samples [24]. Applying this function requires the determination of the degree of capacity constant (C) [28]. After testing several degrees of capacity constant in train samples, the optimal C equal 71 is obtained. The correct classification was calculated by above degree of capacity constant (C) is shown in confusion matrix (Table 3). The confusion matrix was obtained by dividing the summation of diagonal elements by overall samples. This matrix function allowed the comparison of the groups of contamination degree (low, moderate, high and very high classes) (Table 3). After testing several degrees of capacity constant in test samples, the optimal C equal 71 is also obtained (Table 4) and difference is in low class of contamination degree for C equal 71 that is 75% of correct samples. Table 5 displays class of correct and incorrect samples for some of train and test samples. Therefore, there is the best classification for contamination degree based on training and test samples and it can apply to predicting contamination degree parameter. All of heavy metals display measured contamination degree classes are corresponding with predicted contamination

Table 2: Samples of the training data sets used for learning the SVM

No.	Input: Pb	Cu	Ni	Zn	As	Cr	Co	Output: Degree of Contamination
1	0.51	1.18	2.66	0.11	0.40	0.59	0.66	6.14
2	0.45	1.03	3.41	0.17	0.28	0.70	0.52	6.60
3	0.22	2.18	3.50	0.17	0.30	0.27	0.59	7.27
4	0.21	2.11	3.47	0.17	0.31	0.26	0.58	7.13
5	0.68	10.48	2.71	0.13	0.40	0.53	0.66	15.62
6	0.61	10.15	2.63	0.13	0.40	0.52	0.64	15.11
7	0.54	1.19	2.26	0.12	0.48	0.39	0.70	5.71
8	0.44	1.23	1.12	0.59	0.90	1.35	1.41	7.07
9	0.38	1.68	2.36	0.14	0.31	0.83	0.47	6.18
10	0.36	1.63	2.27	0.13	0.30	0.81	0.47	6.01

Table 3: Optimization of the capacity constant in the radial basis function kernel in the train sample (Optimum gamma=0.14)

Capacity Constant	Class	Total Samples	Correct Samples	Incorrect Samples	Correct (%) Samples	Incorrect (%) Samples
	High	4	2	2	50	50
C=5	Low	25	25	0	100	0.00
	Moderate	14	6	8	42.85	57
	Very High	1	0	1	0.00	100
	High	4	4	0	100	0.00
C=41	Low	25	25	0	100	0.00
	Moderate	14	12	2	85.71	14.28
	Very High	1	1	0	100	0.00
	High	4	4	0	100	0.00
C=71	Low	25	25	0	100	0.00
	Moderate	14	14	0	100	0.00
	Very High	1	1	0	100	0.00

Table 4: Optimization of the capacity constant in the radial basis function kernel in the test sample (Optimum gamma=0.14)

Capacity Constant	Classes	Total Samples	Correct Samples	Incorrect Samples	Correct (%) Samples	Incorrect (%) Samples
	High	2	2	0	100	0.00
C=31	Low	8	7	1	87.50	12.50
	Moderate	5	4	1	80	20
	Very High	1	1	0	100	0.00
	High	2	2	0	100	0.00
C=41	Low	8	7	1	87.50	12.50
	Moderate	5	5	0	100	0.00
	Very High	1	1	0	100	0.00
	High	2	2	0	100	0.00
C=71	Low	8	6	2	75	25
	Moderate	5	5	0	100	0.00
	Very High	1	1	0	100	0.00

Table 5: Class of correct and incorrect classification in some of the dependent and predicted train and test samples

Type of sample	Sample No.	Pb	Cu	Ni	Zn	As	Cr	Co	Class Dependent	Class Predicted	Class Accuracy
	1	0.51	1.18	2.66	0.11	0.40	0.59	0.66	Low	Low	Correct
	2	0.45	1.03	3.41	0.17	0.28	0.70	0.52	Low	Low	Correct
Train	3	0.22	2.18	3.50	0.17	0.30	0.27	0.59	Moderate	Moderate	Correct
	4	0.21	2.11	3.47	0.17	0.31	0.26	0.58	Moderate	Moderate	Correct
	5	0.68	10.48	2.71	0.13	0.40	0.53	0.66	High	High	Correct
	1	0.61	1.36	1.50	0.15	0.62	0.52	0.70	Low	Low	Correct
	2	0.58	1.60	2.04	0.18	0.59	0.56	0.56	Low	Low	Correct
Test	3	0.54	1.50	1.94	0.17	0.59	0.55	0.52	Low	Low	Correct
	4	1.02	15.04	1.83	0.18	0.85	1.19	1.24	Very High	Very High	Correct
	5	0.61	3.53	1.38	0.11	0.46	0.55	0.76	Moderate	Low	Incorrect

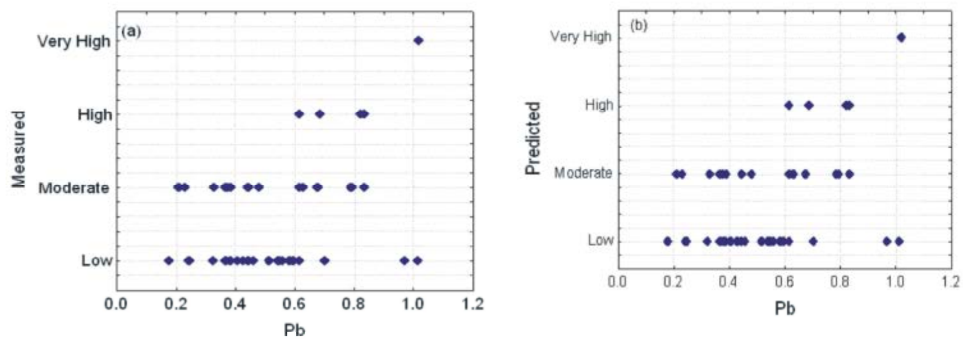

Fig. 4: Pb against classes of contamination degree in (a) measured and (b) predicted train samples

Table 6: Samples for testing the prediction models

	Input							Output
No.	Pb	Cu	Ni	Zn	As	Cr	Co	Contamination Degree
1	0.46	1.15	3.42	0.18	0.28	0.74	0.53	6.77
2	0.50	1.15	2.30	0.12	0.46	0.38	0.70	5.62
3	0.40	2.31	2.52	0.13	0.31	0.59	0.77	7.04
4	0.30	0.66	0.91	0.15	0.44	0.58	0.70	3.76
5	0.97	14.84	1.83	0.18	0.85	1.18	1.23	21.10

Table 7: A comparison between the results of two models.

Prediction model	RMSE	VAF	R^2
SVR	0.68	97.63	0.99
KNNR	1.84	86.74	0.92

degree classes. Fig. 4 shows this relationship for Pb and it has identified that other heavy metals have also similar relationship. Support vector regression (SVR) and K-nearest neighbor regression (KNNR) are two models to predicting contamination degree parameter.

Prediction of Contamination Degree: The goal of support vector regression (SVR) and K-nearest neighbor regression (KNNR) method is to predict contamination degree based on measured samples of heavy metals. In the process of deciding the SVR parameters and choosing the type of kernel function, the 45 data selected as training sets, while the 15 data be regarded as the testing sets of SVR model (Table 6). SVR with radial basis function was considered as the destination according to generalization ability [24]. To evaluate the performances of the prediction models, the variance account for (VAF) and the root mean square error (RMSE) indices of data were used. A few samples of data sets for testing are presented in Table 7.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{1}^{n}(\mathrm{Y}-\mathrm{Y'})^2}{\mathrm{n}}} \tag{9}$$

$$\mathrm{VAF} = \left(1 - \frac{\mathrm{Var}(\mathrm{Y}-\mathrm{Y'})}{\mathrm{Var}(\mathrm{Y})}\right).100\% \tag{10}$$

where, var denotes the variance, y and y' are the measured and predicted values, respectively and N is the number of samples. The higher the VAF, the better is the model performance. For instance, a VAF of 100% means that the measured output has been predicted exactly (perfect model). VAF=0 means that the model performs as poorly as a predictor using simply the mean value of the data. Also, the lower RMSE indicates the better performance of the model. In addition, the determination coefficient (R^2) is calculated. Fig. 5 illustrates the correlation between measured and predicted values of the deformation modulus for two models.

A comparison between the results of two models is shown in Table 7. As it can be observed from this table, the SVM model with R^2= 0.99, VAF= 97.63 and RMSE = 0.68 performs better than the other two models for the modeling of contamination degree. SVM with the best performance was selected to predict contamination degree in the study area. In addition, only Cu displays the best relationship with contamination degree and prediction of contamination degree by use of other heavy metals is weak (Fig. 6). Correlation matrix shows correlation coefficient of Cu against contamination degree for measured and predicted samples is 0.97, but for other heavy metals is medium to weak (Tables 8 and 9). Therefore, we can use Cu for prediction of contamination degree in the study area.

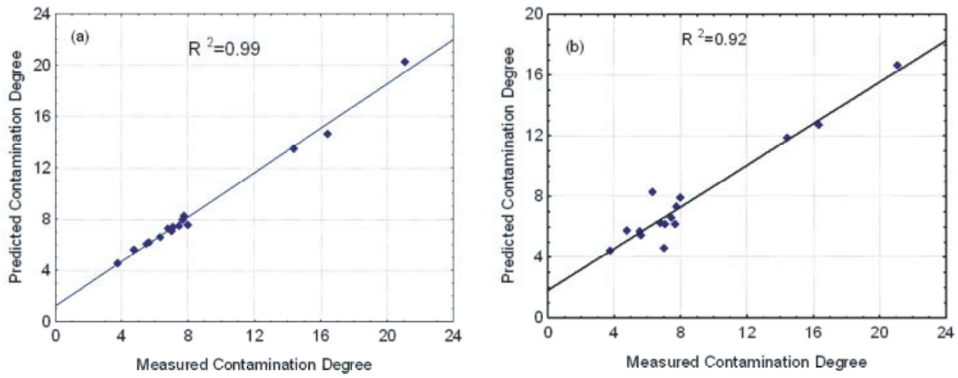

Fig. 5: Correlation between measured and predicted values of contamination degree: (a) SVM regression, (b) K-nearest neighbor regression

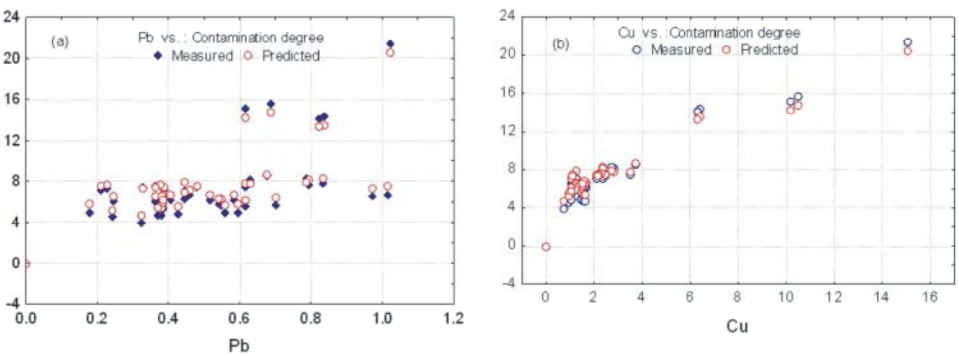

Fig. 6: Scatter plots of (a) Pb and (b) Cu against contamination degree

Table 8: Correlation matrix of measured contamination degree

	Pb	Cu	Ni	Zn	As	Cr	Co	Measured contamination degree
Pb	1.00							
Cu	0.43	1.00						
Ni	-0.09	0.00	1.00					
Zn	0.35	0.23	-0.04	1.00				
As	0.56	0.50	-0.26	0.56	1.00			
Cr	0.36	0.37	-0.14	0.53	0.65	1.00		
Co	0.33	0.42	-0.44	0.42	0.69	0.67	1.00	
Measured contamination degree	0.51	*0.97*	0.11	0.39	0.60	0.51	0.49	1.00

Table 9: Correlation matrix of predicted contamination degree

	Pb	Cu	Ni	Zn	As	Cr	Co	Predicted contamination degree
Pb	1.00							
Cu	0.43	1.00						
Ni	-0.09	0.00	1.00					
Zn	0.35	0.23	-0.04	1.00				
As	0.56	0.50	-0.26	0.56	1.00			
Cr	0.36	0.37	-0.14	0.53	0.65	1.00		
Co	0.33	0.42	-0.44	0.42	0.69	0.67	1.00	
Predicted contamination degree	0.53	*0.97*	0.10	0.35	0.61	0.52	0.51	1.00

CONCLUSION

In this paper contamination degree index approach for the assessment of pollution of soils was proposed and the following remarks were concluded:

- Among the 7 effective heavy metals on the contamination degree, support vector machines was selected as the best model for the assessment of classes of contamination degree index.

- Based upon the results of support vector machines modeling, the class of the contamination degree such as low, medium, high and very high around Arak soils is significant and some of the region has high and very high contamination degree.
- A comparison was made between two statistical models, SVR and KNNR and based upon the performance indices; R^2, RMSE and VAF with R^2= 0.99, RMSE= 0.68 and VAF= 97.63 was selected as the best predictive model.
- The SVR modeling as a good tool can predict the pollution occurred due to industrial activity around the study area.
- Copper is the best predictor to determining of contamination degree around the study area.

ACKNOWLEDGEMENTS

Chemical analyses were conducted in the laboratory of the Department of Mining Exploration Engineering, Arak University of Technology. This study was financially supported by the Scientific Research Project Commission of Department of Mining Exploration Engineering of Arak University of Technology.

REFERENCES

1. Zamani, A.A., M.R. Yaftian and A. Parizanganeh, 2012. Multivariate statistical assessment of heavy metal pollution sources of groundwater around a lead and zinc plant, Iran. Journal of Environment Health Sciences Engineering, 9: 29-38.
2. Ren, H. and E. Yang, 2014. Distribution and controlling factors of selenium in weathered soil in Kaiyang County, Southwest China. Chinese Journal of Geochemistry, 33(3): 300-309.
3. Li, X., S. Zhang and M. Yang, 2014. Accumulation and risk assessment of heavy metals in Dust in main living areas of Guiyang City, Southwest China. Chinese Journal of Geochemistry, 33(3): 272-276.
4. Sharma, R.K., M. Agrawal and F. Marshall, 2007. Heavy metal contamination of soil and vegetables in suburban areas of Varanasi, India. Ecotoxicology and Environment Safety, 66(2): 258-266.
5. Sun, Y., Q. Zhou, X. Xie and R. Liu, 2010. Spatial, sources and risk assessment of heavy metal contamination of urban soils in typical regions of Shenyang, China. Journal Hazard. Material, 174: 455-462.
6. Kim, M.J. and T. Kim, 2011. Extraction of Arsenic and Heavy Metals from Contaminated Mine Tailings by Soil Washing. Soil and Sediment Contamination, 20(6): 631-648.
7. Ghomi, M., F. Ghadimi and A. Hajati, 2013. Evolution of soil contamination in the abandoned Lakan Lead and Zinc mine by heavy metals, Iran. Journal of Tethys, 1: 12-28.
8. Ghadimi, F., M. Ghomi, M. Ranjbar and A. Hajati., 2013. Statistical Analysis of Heavy Metal Contamination in Urban Dusts of Arak, Iran, Iranica Journal of Energy & Environment, 4(4): 406-418.
9. Chang, S.H., K.S. Wang., H.F. Chang., W.W. Ni, B.J. Wu., R.H. Wong and H.S. Lee, 2009. Comparison of Source Identification of Metals in Road-Dust and Soil, Soil and Sediment Contamination, 18(5): 669-683.
10. Ghadimi, F. and M.Ghomi, 2013. Assessment of the effects of municipal wastewater on the heavy metal pollution of water and sediment in Arak Mighan Lake, Iran, Journal of Tethys, 1(3): 205-214.
11. Manzano, R., J. Peñalosa and E. Esteban, 2014. Amendment Application in Multi contaminated Mine Soil: Effects on Trace Element Mobility. Water, Air, & Soil Pollution, 225: 1874-8.
12. Crnković, D., M.Ristić and D. Antonović, 2006. Distribution of Heavy Metals and Arsenic in Soils of Belgrade (Serbia and Montenegro). Soil and Sediment Contamination, 15(6): 581-589.
13. Ballabio, C., 2009. Spatial prediction of soil properties in temperate mountain regions using support vector regression. Geoderma, 151(3-4): 338-350.
14. Pasolli, L. and L. Bruzzone, 2011. Estimating Soil Moisture with the Support Vector Regression Technique, Dept. of Inf. Eng. & Comput. Sci., Univ. of Trento, Trento, Italy; Full Text Sign-In or Purchase, Geoscience and Remote Sensing Letters, 8(60): 1080-1084.
15. Sexton, R.S., R.E. Dorsey and N.A. Sikander, 2004. Simultaneous optimization of neural network function and architecture algorithm, Decis Support Syst, 36(3): 283-96.
16. Lamorski, K., C. Sbawinski, F. Moreno, G. Barna, W. Skierucha and J.L. Arrue, 2014. Modeling Soil Water Retention Using Support Vector Machines with Genetic Algorithm Optimization, The Scientific World Journal, Article ID 740521, 10 pages. http://dx.doi.org/10.1155/2014/740521
17. Hakanson, L., 1980. An ecological risk index for aquatic pollution control, a sedimentological approach, Water Research, 14: 975-1001.

18. Caeiro, S., M.H. Costa, T.B. Ramos, F. Fernandes, N. Silveira, A. Coimbra, G. Medeiros and M. Painho, 2005. Assessing heavy metal contamination in Sado Estuary sediment: An index analysis approach. Ecological Indicators, 5(2): 151-169.
19. Guillén, M.T., J. Delgado, S. Albanese, J.M. Nieto, A. Lima and B. De Vivo, 2011. Environmental geochemical mapping of Huelva municipality soils (SW Spain) as a tool to determine background and baseline values. Journal of Geochemical Exploration 109: 59-69.
20. Schölkopf, B. and A.J. Smola, 2002. Learning with kernels Support vector machines, regularization, optimization and beyond. Cambridge, Massachusetts: MIT Press.
21. Vapnik, V., 1995. The nature of statistical learning theory. New York: Springer.
22. Waske, B., S. Van Der Linden, A. Benediktsson Rabe and P. Hostert, 2010. Sensitivity of support vectormachines to randomfeature selection in classification of Hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing, 48: 2880- 2889.
23. Abe, Sh. 2005. Support Vector Machines for Pattern Classification, Springer-Verlag London.
24. Shawe-Taylor, J. and N. Cristianini, 2004. Kernel Methods for Pattern Analysis, Cambridge University Press.
25. Ma, H., W. Chang and G. Cui, 2012. Ecological footprint model using the support vector machine technique. PLoS ONE 7, e30396.
26. Wan, X.B., Y. Zhao, X.J. Fan, H.M. Cai, Y. Zhang, M.Y. Chen, J. Xu, X.Y. Wu, H.B. Li, Y.X. Zeng, M.H. Hong and Q. Liu, 2012. Molecular prognostic prediction for Locally advanced nasopharyngeal carcinoma by support vector machine integrated approach. PLoSONE, 7: 31989.
27. Bishop, C., 1995. Neural networks for pattern recognition. Oxford: Clarendon Press.
28. Karatzoglou, A., D. Meyer and K. Hornik, 2006. Support vector machines in R. J. Stat. Softw, 15: 28.
29. Caritat, P.D. and C. Reimann, 2012. Comparing results from two continental geochemical surveys to world soil composition and deriving Predicted Empirical Global Soil (PEGS2) reference values, Earth and Planetary Science Letters, 319-320: 269-276.

Persian Abstract

چکیده

میزان آلودگی خاک منطقه شهری به فلزات سنگین دستخوش تغییر می شود. در نتیجه ممکن است سلامت ساکنین شهری به خطر افتد. این مقاله ارائه طریق مدلی است که شاخص آلودگی خاک را به تغییرات فلزات سنگین پیشگویی می کند غلظت فلزات سنگین Pb, Cu, Ni, Zn, As, Cr و Ni بعنوان داده ها استفاده گردیده تا میزان آلودگی سنجش گردد دو مدل SVR و ریگراسیون KNNR مورد استفاده قرار گرفت. نتایج مقایسه بین دو مدل نشان داده است که مدل SVR بر تر بوده است. کارایی مدل با استفاده از پشتیبانی ماشین بردار مدل SVM برای شهر اراک مورد آزمایش قرار گرفت.