Perspectives in Science (2016) 8, 629—631

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/pisc

Support vector regression model for predicting the sorption capacity of lead (II)☆

Nusrat Parveen, Sadaf Zaidi*, Mohammad Danish

Department of Chemical Engineering, Aligarh Muslim University, India

Received 14 February 2016; received in revised form 19 April 2016; accepted 11 June 2016
Available online 4 July 2016

KEYWORDS
Heavy metals;
Low cost biosorbent;
Biosorption;
Support vector regression (SVR)

Summary Biosorption is supposed to be an economical process for the treatment of wastewater containing heavy metals like lead (II). In this research paper, the support vector regression (SVR) has been used to predict the sorption capacity of lead (II) ions with the independent input parameters being: initial lead ion concentration, pH, temperature and contact time. Tree fern, an agricultural by-product, has been employed as a low cost biosorbent. Comparison between multiple linear regression (MLR) and SVR-based models has been made using statistical parameters. It has been found that the SVR model is more accurate and generalized for prediction of the sorption capacity of lead (II) ions.

Introduction

The toxic and hazardous nature of lead (II) has become a major concern across the world. Most commonly used techniques for the abstraction of lead like chemical precipitation, ion exchange, filtration, membrane processes, and coagulation. But, nowadays biosorption process is found to be an ideal alternative to these existing methods as it offers many benefits such as environmental friendliness, low cost, high efficiency of metal removal and minimum generation of sludge that can be easily treated (Abbas et al., 2014).

In the recent past, support vector machines (SVMs), a soft computing technique, have been adapted as a regression (SVR) predicting tool which has several advantages over the traditional neural networks. These include few model parameters to be chosen, avoidance of over-fit to the data and unique, optimal and global solution. Recently, SVR has been applied to many actual issues like predicting the performance of compact heat exchangers (Peng and Ling, 2015), modelling of heat transfer in a thermosyphon reboiler (Zaidi, 2015), etc. In the current study, support vector regression has been used to predict the sorption capacity (mg/g) of lead (II) ions.

Theory and basic principles of support vector regression

The detailed description of SVM is therein several excellent works (Vapnik, 1995; Gunn, 1997). A regression consists of a

☆ This article belongs to the special issue on Engineering and Material Sciences.

* Corresponding author. Tel.: +91 9411212432.
E-mail addresses: nusratalig@gmail.com (N. Parveen), sadaf63in@yahoo.com (S. Zaidi), mdanish@rediffmail.com (M. Danish).

http://dx.doi.org/10.1016/j.pisc.2016.06.040

Table 1 Parameters of the SVR model for the sorption capacity of lead (II) ion.

Model	C	$\gamma = 1/2\sigma^2$	ε	Kernel type	Loss function	Quantity of support vectors	Number of training data
Sorption capacity (mg/g)	32,768	0.058	0.09	Radial basis function	ε-insensitive	70	116

training data as given: $P = \{(a_1, b_1), (a_2, b_2), \ldots, (a_N, b_N)\}$, so that a_i is a vector of real independent variables and b_i the corresponding scalar real dependent variable. The regression equation in the feature space can be approximated by:

$$z(a, \mathbf{w}) = (\mathbf{w} \cdot \phi(a) + c) \tag{1}$$

where, $\mathbf{w}$ defines the weight vector, c is a constant, $\phi(a)$ is the feature function and $(\mathbf{w} \cdot \phi(a))$ the dot product therein.

Minimize the following equation:

$$\text{Minimize} : Q(f) = C\frac{1}{N}L_\varepsilon(b, z(a, \mathbf{w})) + \frac{1}{2}||W^2|| \tag{2}$$

and,

$$L_\varepsilon(b, z(a, \mathbf{w})) = \begin{cases} 0 & \text{if} \quad |b - z(a, \mathbf{w})| \le \varepsilon \\ |b - z(a, \mathbf{w})| - \varepsilon & \text{otherwise} \end{cases} \tag{3}$$

In Eq. (2), the LHS term represents the empirical error and the term C gives a measure of the optimization between the empirical error and the model complexity given by the second term of the said equation. Eq. (3) defines a loss function called ε-insensitive loss function (Vapnik et al., 1996). The optimization problem is converted into the dual problem by incorporating Lagrangian multiplier β and β^*. Only the non-zero coefficients, along with their input vectors, $\mathbf{a}_i$, are termed the support vectors. The final form comes out as follows:

$$z(a, \beta_i, \beta_i^*) = \sum_{i=1}^{N_{sv}} (\beta_i - \beta_i^*)(\phi(\mathbf{a}_i) \cdot \phi(a_j)) + c \tag{4}$$

By the help of kernel function $K(x_i, x_j)$, the SVR function can be obtained as given below:

$$z(a, \beta_i, \beta_i^*) = \sum_{i=1}^{N_{sv}} (\beta_i - \beta_i^*)K(a, a_i) + c \tag{5}$$

The term, c, is calculated by using the Karush–Kuhn–Tucker conditions.

Results and discussion

In this study, SVR has been applied to predict the sorption capacity (mg/g) of lead (II) ions with the input parameters being initial lead ion concentration, pH, temperature and contact time using agricultural by-product, tree fern as a biosorbent. Biosorption data for SVR have been taken from the available published literature (Ho et al., 2004; Ho, 2005). The whole dataset consisting of 145 runs are randomly divided into training dataset (116 data points; 80% of the total data) and the test dataset (29 data points; 20% of the data) (Zaidi, 2015). Afterward, model hyper-parameter (C, ε) and kernel parameter (RBF kernel, γ) are optimized using 10-fold cross-validation. Then, the support vector model is evaluated on the basis of statistical parameters and then compared with the MLR model.

Analysis of the support vector regression model

In this paper, the radial basis function (RBF) kernel has been employed as it has been found to give a good general performance and only one kernel parameter (γ) needs to be adjusted (Zaidi, 2015). Grid search methodology with 10-fold cross validation has been applied to get the optimal model parameters, C, ε, γ. These parameters have been varied in a wide range as: C [2^5,2^{15}], γ [2^{-15}, 2^2] and ε [2^{-15}, 2^4]. The model parameters for the support vector regression model are provided in Table 1. LIBSVM package developed by Chang and Lin (2001) was employed to obtain the support vector regression model for the sorption capacity (mg/g) of lead (II) ions.

Fig. 1 has been plotted between actual values of lead (II) sorption capacity (mg/g) and the values predicted by the support vector regression model for the sets of training data and the test data. The statistical evaluation parameters of the SVR-based model predicted values for the training data have been found as: average absolute relative error (AARE) of 3.99%, correlation coefficient (R) of 0.9901, root mean square error (RMSE) of 0.0041, standard deviation (SD) of 8.2529 and mean relative error (MRE) of 0.0398 and the corresponding values for the test data are: 3.38%, 0.9915, 0.0025, 7.8962 and 0.0338 respectively. It is clearly evident that the support vector regression model has given such values of the statistical evaluation parameters, which demonstrate its excellent prediction performance.

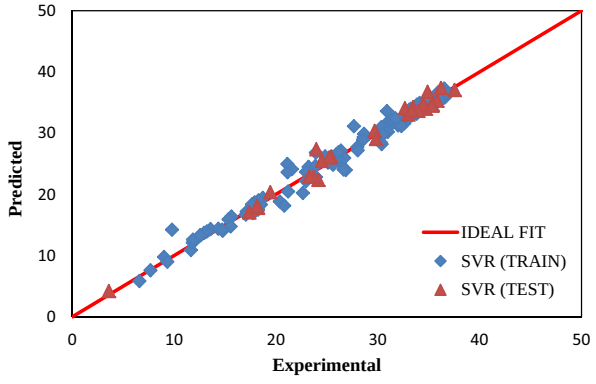

Figure 1 SVR simulation of sorption capacity (mg/g) of lead (II) for the training data and test data.

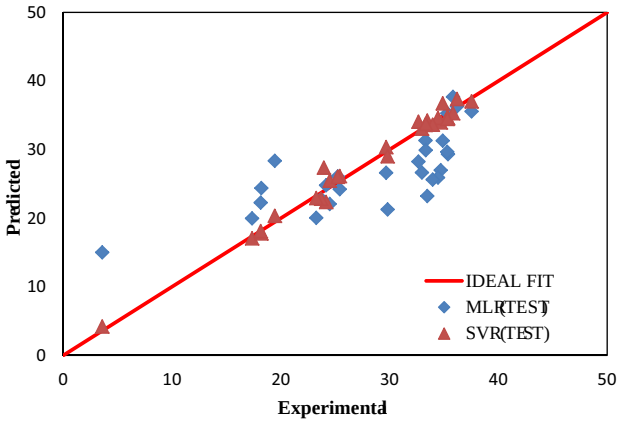

Figure 2 Prediction performance of the SVR model versus the MLR model for the sorption capacity of lead (II) ions using the test data.

Table 2 Evaluation of SVR model versus the MLR model for test data.

Evaluation index	Support vector regression model	MLR model
Average absolute relative error (AARE, %)	3.38	25.01
Correlation coefficient (R)	0.9915	0.7613
Root mean square error (RMSE)	0.0025	0.0587
Standard deviation (SD)	7.8962	7.9204
Mean relative error (MRE)	0.0338	0.1799

SVR-based model versus the MLR model

A comparison has been made of the performance of the MLR model with the support vector regression model. Fig. 2 clearly shows the superior performance of the support vector regression model. Table 2 exhibits the relevant statistical evaluation parameters for the two modelling methods utilizing the test data. It is thus seen that the support vector regression model outsmarts the MLR model.

Conclusion

Based on the statistical evaluation parameters and the comparison with the MLR model, the results prove that the support vector regression soft computing model is a superior model than the MLR model in terms of prediction performance, accuracy and generalization ability. Thus, SVR can be applied in chemical engineering and its allied fields.

Acknowledgment

We would like to profoundly acknowledge the authors whose research has been consulted in this work, especially those whose data has been used to develop the models.

References

Abbas, S.H., Ismail, I.M., Mostafa, T.M., Sulaymon, A.H., 2014. Biosorption of heavy metals: a review. J. Chem. Sci. Technol. 3, 74—102.

Chang, C.C., Lin, C.J., 2001. LIBSVM: A Library for Support Vector Machines. Department of Computer Science, National Taiwan University, Taipei, Taiwan.

Gunn, S., 1997. Support Vector Machines for Classification and Regression. ISIS Technical Report. University of Southampton, pp. 1—42.

Ho, Y.S., Chiu, W.T., Hsu, C.S., Huang, C.T., 2004. Sorption of lead ions from aqueous solution using tree fern as a sorbent. Hydrometallurgy 73, 55—61.

Ho, Y.S., 2005. Effect of pH on lead removal from water using tree fern as the sorbent. Bioresour. Technol. 96, 1292—1296.

Peng, H., Ling, X., 2015. Predicting thermal—hydraulic performances in compact heat exchangers by support vector regression. Int. J. Heat Mass Transfer 84, 203—213.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer, New York.

Vapnik, V.N., Golowich, S., Smola, A.J., 1996. Support vector method for function approximation, regression estimation and signal processing. Adv. Neural Inform. Process. Syst. 9, 281—287.

Zaidi, S., 2015. Novel application of support vector machines to model the two-phase boiling heat transfer coefficient in a vertical tube thermosiphon reboiler. Chem. Eng. Res. Des. 98, 44—58.