**Journal of Novel Applied Sciences**
Available online at www.jnasci.org
©2014 JNAS Journal-2014-3-10/1095-1098
ISSN 2322-5149 ©2014 JNAS

# SVR parameters tuning by Imperialist Canonical Algorithm to predict pipes failure rates

**SeyedMahmood Tabatabaei[1], Mahdi Kalanki[2]* and Moosa Kalanaki[2]**

1- Department of Irrigation & Drainage Engineering, University of Zabol, Zabol, Iran
2- Department of electrical and computer engineering, Isfahan University of Technology,Isfahan,Iran

***Corresponding author***: Mehdi Kalanki

**ABSTRACT:** In this paper, Imperialist canonical algorithm (ICA) and Support Vector Regression (SVR) are combined to propose a model for predicting pipe failure rates. In this model, SVR used for simulation of the pipe failure rates and ICA algorithms tries to find the best SVR related parameters. Therefore, we can find the most appropriate relations of pipe failure rates and take necessary actions for decision-making that leads to resolve problems due to it.

***Keywords***: SVR, ICA, Failure rate prediction, intelligent method.

## INTRODUCTION

With regards to developing in science and technology and create Interdisciplinary fields we can apply it to develop industries and Improving level of living. One of the main objectives of efficient management and optimal operation of urban water distribution network is to present a model for predicting breakages of urban water distribution networks. These activities lead to demand water losses and costs of rehabilitations being less. In 1998, about one million accidents have occurred, attributing to themselves more than 20 percent of the total Water and Wastewater Companies budget used for repairs and rehabilitation. In most cases, accidents and pipe failures occur as a result several factors being measurable such as age, length, diameter, depth and pressure of the pipes (Tabesh, 2009). Hence, to achieve better results we need a comprehensive model to consider all these factors. Many studies have been done in this field with many types of methods more than two decades some of them are based on classical methods and others base on intelligent methods including ANN, ANFIS, fuzzy logic, and SVM in related field.

In another study comparing among NLR, ANN and ANFIS methods had been done. The results of the comparison between ANN and ANFIS showed that ANN model is more sensitive to pressure, diameter and age than ANFIS; So, ANN was more reliable (Tabesh, 2009).

SVM techniques used non-linear regression for environmental data and proposed a multi-objective strategy, MO-SVM, for automatic design of the support vector machines based on a genetic algorithm. MO-SVM showed more accurate in prediction performance of the groundwater levels than the single SVM (Giustolisi, 2006).

Rough set theory and support vector machine (SVM) was proposed to overcome the problem of false leak detection. For the computational training of SVM, used artificial bee colony (ABC) algorithm, the results are compared with those obtained by using particle swarm optimization (PSO).Finally; obtained high detection accuracy of 95.19% with ABC (Mandal, 2012).

A combined model (ANN -GA) presented to determine the effective parameters of pipe failure rates in water distribution system. In other words, ANN model was developed in order to relate parameters of breakage with pipe failure rates (kalanaki, 2013).

In this paper, using the SVR-ICA model and its abilities we can find effective parameters to predict pipes failure rate of water distribution networks. This research is aimed at optimizing parameters related to SVR and selecting the optimal SVR-ICA parameters to better pipes failure rate prediction.

## 2. Support Vector Machine

Support vector machine regression (SVR) is a method to estimate the mapping function from Input space to the feature space based on the training dataset (Vapnik., 2010). In the SVR model, the purpose is to estimate w and b parameters to get the best results. In SVR, differences between actual data sets and predicted results is displayed by ε. Slack variables are (ξ) considered to allow some errors that occurred by noise or other factors. If we do not use slack variables, some errors may occur, and then the algorithm cannot be estimate. Margin defines as $\mathrm{margin} = \frac{1}{\|w\|}$. Then, to maximize the margin, through minimizing$\|w\|^2$, the margin becomes maximize. These operations give in Equations (1 )and (2 )and these are the basis for SVR (Qi ., 2013).

$$\text{Minimize:} \frac{1}{2}\left|\left|w\right|\right|^2 + C\sum_{i=1}^{n} \xi_i \quad (1)$$

$$\text{subject to:}\ y_i\left(w^T x_i + b\right) \geq 1 - \xi_i\ ,\ \xi_i \geq 0 \quad (2)$$

xi is the input space and yi is the feature space. w is the weight vector and b is the bias, which will be computed by SVR in the training process. C is a parameter that determines the trade-off between the margin size and the amount of error in training.

A kernel function is a linear separator based on inner vector products and is defined as follows:

$$k\left(x_i, x_j\right) = x_i^T x_j \quad (3)$$

If data points are moved using $\varphi: x \rightarrow \varphi(x)$ to the feature space (higher dimensional space), their inner products turn into Equation (4) (Vapnik., 1992).

$$k\left(x_i, x_j\right) = \varphi(x_i)^T . \varphi(x_j) \quad (4)$$

xi is the support vectors and xj is the training data.

Accordingly, with using kernel functions and determining derivatives of w and b, also using Lagrange multiplier the SVR function F(x) becomes the following function.

$$F(x) = \sum_{i=1} (\bar{\alpha}_i - \alpha_i) K(x_i, x) + b \quad (5)$$

$\alpha_i$ is the vector of Lagrange multipliers and represent support vectors. If these multipliers are not equal to zero, they are multipliers; otherwise, they represent support vectors (Vapnik & Chapelle, 2000).

Loss function determines how to penalize the data while estimating. A Loss function implies to ignore errors associated with points falling within a certain distance. If ε-insensitive loss function is used, errors between -εand +εare ignored. If C=Inf is set, regression curve will follow the training data inside the margin determined by ε (Smola& Scholkopf, 1998). The related equation is shown in Equation (6).

$$|\xi|_\varepsilon = \begin{cases} 0 & \text{if } |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{otherwise.} \end{cases} \quad (6)$$

### 2.1. Imperialist Canonical Algorithm

ICA algorithm is for optimization inspired by the imperialistic competition. Populations in this algorithm called country are in two types: colonies and imperialists that all together form some empires. Imperialistic competition among these empires forms the basis of the proposed evolutionary algorithm. During this competition, weak empires collapse and powerful ones take possession of their colonies (Atashpaz-Gargari & Lucas., 2007).

The colony moves toward the imperialist by x units. In this movement, x and θ are random numbers with uniform distribution and d is the distance between the colony and the imperialist.

Variable x ~U(0 , β × d ) where β is a number greater than 1 and d is the distance between colony and imperialist. β > 1 because the colonies to get closer to the imperialist state from both sides[10].To search different points around the imperialist we a random amount of deviation added to the direction of movement.

θ is a random number with uniform distribution. Then θ ~ U (-γ, γ) where γ is a parameter that adjusts the deviation from the original direction (Atashpaz-Gargari & Lucas., 2007).

## 3. Case Study

A part of a water distribution network of a city in Iran is considered as the study area. This city is one of the cities being frequently visited by travellers.The area of this district is 2,418 hectares, with a population of 93719 people, supplied with 579,860 meters of distribution pipes including steel pipes 800, 700 and 600 millimeters in diameter, asbestos cement and cast iron pipes 400, 300, 250, 200, 150, 100 and 80 millimeters in diameter.

The installation and execution of the network pipelines in this area were generally started in 1981. According to statistical records, this region has the highest failure rate especially on asbestos cement. In this study, due to incomplete data on steel and cast iron pipes, asbestos cement pipes are only used in the modeling process.

In order of modeling the failure rate of the asbestos cement pipes, the daily events have been recorded from 2005 to 2006 and analyzed as to 2438 record data including some information such as diameter, year of implementation

,installation depth, total accidents happened and the average of hydraulic pressure. These data have been collected from local water and water waste company.

## MATERIALS AND METHODS

In this study, an integrated (SVR-ICA) model is proposed to search the possible solutions.This research has been developed by MATLAB (version 7.12(R 2011a))and MATLAB SVM Toolbox and parameters were localized by ICA to solve these problems. Equation 10 is used for normalizing the Input values to the models.

$$x_n = 0.8\frac{(x - x_{min})}{(x_{max} - x_{min})} + 0.1 \quad (8)$$

x is the original value, x min is the minimum value and x max is the maximum value between input values, and x n shows normalized values. So that, input results are between [0.1, 0.9] [11]. Also, in this paper, the root of mean squared error (RMSE), normal root of mean squared error (NRMSE) and coefficient of determination($R^2$) are used as assessment criteria of the reliability of the model.

$$R^2 = \frac{\left(\sum_{i=1}^{n}(y_{actual} - \bar{y}_{actual})(y_{pred} - \bar{y}_{pred})\right)^2}{\sum_{i=1}^{n}(y_{actual} - \bar{y}_{actual})^2 \sum_{i=1}^{n}(y_{pred} - \bar{y}_{pred})^2} \quad (9)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{actual_i} - y_{pred_i}\right)^2} \quad (10)$$

$$NRMSE = \frac{RMSE}{var(y_{actual})} \quad (11)$$

Where $y_{actual}$ is the actual (observed) data, $y_{prediction}$ is the predicted data, $y_{average}$ is the average of data and n is the number of observations (Tokar & Johnson., 1999). Also, var($y_{actual}$) is the variance of actual data. Related parameters are chosen by author’s experiments. Number of iterations of all algorithms are set to 30 and initial population equalled to 25. Boundaries parameters that relate to SVR set as: $0 < \varepsilon \leq 1$ , $1 \leq \gamma \leq 10$ and $10 \leq C \leq 200$. Revolution rate in ICA sets to 0.3.

## RESULTS AND DISCUSSION

### *5. Results*

Table 1. indicates to results; this shows that RBF function offered the best results, because it has acted faster and showed better performance compared with other kernel functions as regards the data correlation and error parameters.

Table 1. Kernel function that used in this paper

| Kernel Type | Formula | Related variables |
| --- | --- | --- |
| Gaussian RBF | $k = e^{-\frac{(u-v)(u-v)'}{2p_1^2}}$ | P1 defines RBF function width, like as.δ |
| Exponential RBF | $k = e^{-\sqrt{\frac{(u-v)(u-v)'}{2p_1^2}}}$ | P1 defines eRBF function width like as RBF. |

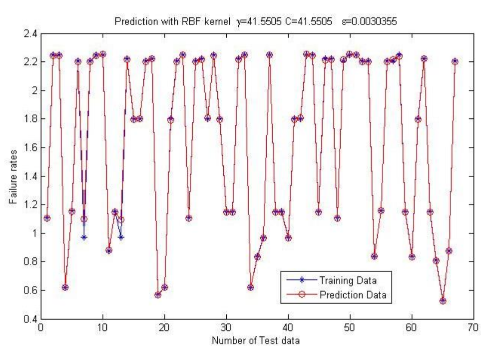


Figure 1. SVR-ICA result with RBF kernel function

In Figure 1, predicted results show higher accuracy because the predicted data are fitted to actual ones and according to Table 2. Results error computation formulas such as RMSE, NRMSE and $R^2$ are very good.

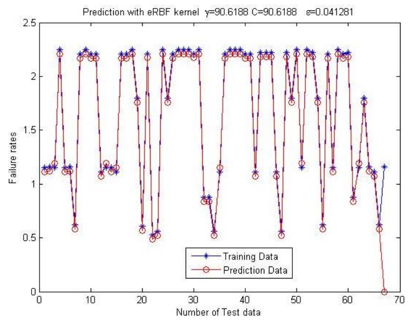

Figure 2. SVR-ICA result with eRBF kernel function

Figure 2. shows results base on eRBF kernel function and in comparing with RBF kernel function it shows less accurate.

Table 2. SVR-ICA algorithms with related kernel function

| Kernel Type | Time(s) | $R^2$ | RMSE | NRMSE | C | P1 | ε |
|---|---|---|---|---|---|---|---|
| RBF | 7144.2 | 0.9999836952884 | 0.003013534 | 0.008291992643 | 41.55052 | 1.986718 | 0.0030354991 |
| eRBF | 6308.8 | 0.9867128141217 | 0.084942491 | 0.231675449717 | 90.6188 | 4.60322 | 0.041280935 |

### *6. Conclusion*

In this paper, a hybrid SVR-ICA model presented to predict pipe failure rates in water distribution networks, in order of reducing the number of breaks. In the proposed model, it was tried to establish a relationship between the failure rate parameters in pipes with the number of events and failure of pipes considered as the main component of urban infrastructure, water supply and hygiene and health.

Also, using Imperialist canonical algorithm, the optimal kernel function type and SVR related parameters have been extracted. These parameters have more accuracy and results presented better performance between actual and predicted data. By comparing among achieved results RBF kernel, function showed better results.

## REFERENCES


Atashpaz-Gargari E and Lucas C. 2007. "Imperialist Competitive Algorithm: An algorithm for optimization inspired by imperialistic competition". IEEE Congress on Evolutionary Computation.

Giustolisi O. 2006. Using a multi-objective genetic algorithm for SVM construction, Journal of Hydroinf .8(2),125-139.

Kalanaki M, Soltani J and Tavassoli S. 2013. "The use of hybrid SVR-PSO model to predict pipes failure rates", International journal of science and engineering research ,Vol.4 , No.11, 2013, pp. 1022-1025.

Mandal SK, Tiwari MK & Chan FTS. 2012. Leak detection of pipeline: An integrated approach of rough.set theory and artificial bee colony trained SVM. Expert Systems with Applications, 39 (3), 3071–3080.

Qi Z, Tian Y and Shi Y. 2013. Robust twin support vector machine for pattern classification, Journal of Pattern Recognition. 46, 305-316.

Sajikumar, N. and Thandaveswara, B. S. 1999 A Non-Linear Rainfall-Runoff Modeling Using an Artificial Neural Network,J. Hydrology. 36 (4), 32-35.

Smola A and Scholkopf B. 1998. A tutorial on Support Vector Regression, J Statistics and Computing 14 (3) ,199 -22.

Tabesh M, Soltani J, Farmani R and Savic DA. 2009. Assessing Pipe failure Rate and Mechanical Reliability of water Distribution Networks Using Data Driven Modeling, J. Hydroinf 11,1-17.

Tokar AS and Johnson PA. 1999. Rainfall-Runoff Modelling Using Artificial Neural Networks,.J Hydrologic Eng 4232-239.

Vapnik VN. 2010. The Nature of Statistical Learning Theory. Springer Verlag, USA.

Vapnik VN. 1992. Principles of risk minimization for learning theory. Advances in Neural Information Processing Sys 4: 831-838.

Vapnik VN and Chapelle O. 2000. Bounds on error expectation for support vector machines. Neural Computation 12 (9): 2013–2036.