ISSN 2229-5518

The use of hybrid SVR-PSO model to predict Pipes failure rates

Moosa Kalanaki[1], Jaber Soltani[2*], Sajjad Tavassoli[3]

Abstract—In this paper, by Particle Swarm Optimization algorithm (PSO) and Support Vector Regression (SVR) a hybrid model is proposed to predict pipe failure rates dataset. In this model, SVR used for simulation of the pipe failure rates and by PSO algorithms tries to find the best SVR related parameters. So, it 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. The results of this research indicate that the hybrid model is able to extract the optimal and effective parameters on the pipes failure rate among the factors affecting failure rates.

1 INTRODUCTION

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.

In most cases, accidents and pipe failures occur as a result of several factors some of which being measurable such as age, length, diameter, depth and pressure of the pipes [8]. 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.

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 [2].

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 [8].

A probabilistic measure of the failure rate was defined and formulated for cases where the pipe lifetimes follow parametric models. The resulting theoretical failure rates were time-invariant and the parametric models would be useful only if the failure rates of water distribution pipes were stationary random processes [1].

1,3. Department of Computer Engineering,Rouzbahan Higher Education Institute

2. Ph.D., in the field of Civil Engineering(Water and Environmental Eng.), Assistant Professor of Irrigation and Drainage Engineering Department, Aboureyhan Campus, University of Tehran, Email:jsoltani@ut.ac.ir

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 [7].

In this paper, effective parameters to predict pipes failure rate of water distribution networks are taken into SVR-PSO model then compared with own kernels, an analysis and comparison of various types of kernel is performed. This research is aimed at optimizing parameters related to SVR and selecting the optimal SVR-PSO parameters to better pipes failure rate prediction.

2 METHODOLOGIES

2.1 *Support Vector Regression*

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 [12]. 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 estimated. Margin is defined as" $margin = 1/\|w\|$". Then, to maximize the margin, through minimizing $\|w\|^2$, the margin becomes maximized. These operations give in (1) and (2) and these are the basis for SVR [10].

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

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

x_i is the input space and y_i 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

ISSN 2229-5518

in training data.

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

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

If data points are moved using $\varphi: x \rightarrow \varphi(x)$ to the feature space (higher dimensional space), their inner products turn into (4) [5].

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

x_i is the support vectors and x_j is the training data.

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

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

α_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 [11].

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 $-\varepsilon$ and $+\varepsilon$ are ignored. If C=Inf is set, regression curve will follow the training data inside the margin determined by ε [6]. The related equation is shown in (6).

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

2.2 Particle Swarm Optimization algorithm

The particle swarm optimization (PSO) was designed by [4]. This algorithm simulates the moving of social behaviour among individuals (particles) through a multi-dimensional search space, each particle represents a potential solution and has a position represented by a position vector [4].

A swarm of particles moves through the problem space, with the moving velocity of each particle represented by a velocity vector [3].PSO operates in three steps at first Define each particle as a potential solution to a problem and best positions have been selected. Each particle has a velocity and by selecting, the best of them these velocities will update.

$$v_i(t+1)=wv_i(t)+c_1r_1[\hat{x}_i(t)-x_i(t)]+c_2r_2[g(t)-x_i(t)] \quad (7)$$

$v_i(t)$ is the particle's velocity at time t and $x_i(t)$ is the particle's position at time t and $\hat{x}_i(t)$ is the particle's individual best solution in time t. g(t) is the swarm's best solution in time t and w is inertia weight.

.

2.4 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 (see Figure 1). 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 millimetres in diameter, asbestos cement and cast iron pipes 400, 300, 250, 200, 150, 100 and 80 millimetres 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 modelling process.

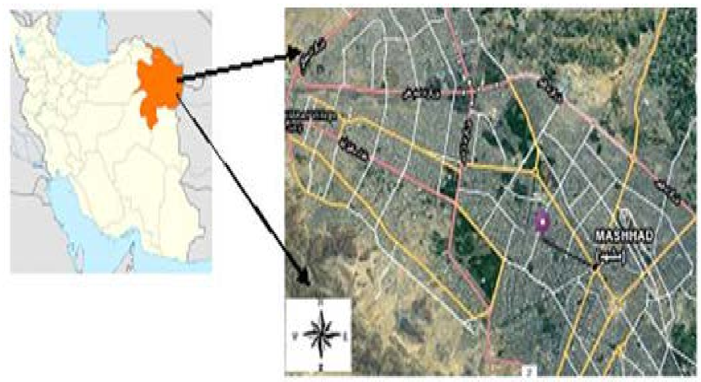

Figure1. Schematic of study area and pressure measurement points

In order of modelling the failure rate of the asbestos cement pipes, the daily events have been recorded from 2005 to 2006 and analysed 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.

3 Material and Methods

Generally, so far, previous researches have appropriated SVR parameters obtained by trial-and-error methods. In trial-and-error methods, each parameter is tested to approach the appropriated values. However, this method has been very time-consuming and is not sufficiently accurate. Therefore, in this study an integrated (SVR-PSO) model is proposed to search for the possible solutions.

This research has been developed by MATLAB (version 7.12(R 2011a)) and MATLAB SVM Toolbox and parameters were localized by PSO 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].

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 [9]. 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 $10 \leq C \leq 200$.

The inertia weight was 0.9; acceleration constants C_1 and C_2 were considered 0.9 and 1.7 respectively according to PSO parameters.

Table1
Kernel functions used in the SVR-PSO model

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.
Polynomial	$k = (UV + 1)^{p1}$	P1 determines Polynomial degree.
Spline	$Z=1+UV+\left(\frac{1}{2}\right)UV$ $min(u,v) - \left(\frac{1}{6}\right) min(u,v)^3$ $k=Prod(z)$	

4 RESULTS

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

Results in Table 2. are notes error computation formulas such as RMSE, NRMSE and R 2.

In Fig. 2, predicted results show high accuracy because the predicted data are fitted to actual ones and according to

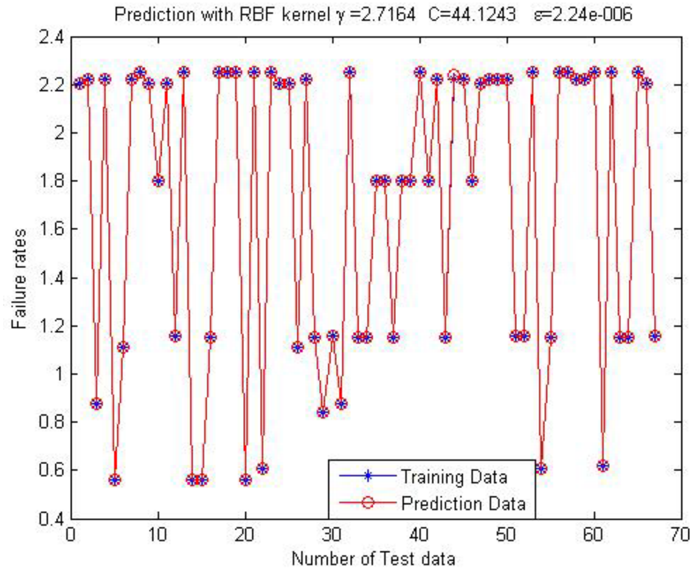

Fig. 2. SVR-PSO result with RBF kernel function

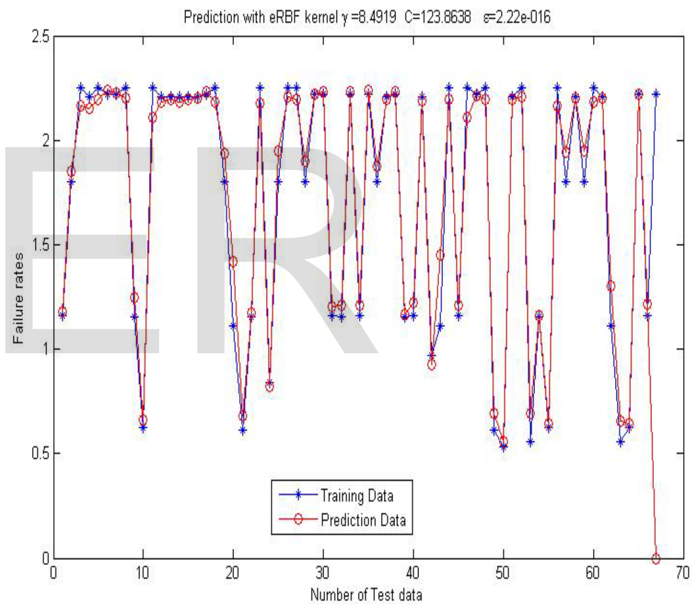

Fig. 3. SVR-PSO result with eRBF kernel function

Table 2
SVR-PSO algorithms with related kernel function

Kernel Type	Time(s)	R^2	RMSE	NRMSE
RBF	5193.7	0.999984220828	0.0024826349012	0.006973621957
eRBF	4116.5	0.988466254208	0.0759893943896	0.201073852214
Polynomial	6071.8	0.861547297833	0.3166679380075	0.862418408819
Spline	3987.9	0.686025106862	0.3672989623209	0.972129041987

5 Conclusion

In this paper, a hybrid SVR-PSO 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 attempted 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 Particle Swarm Optimization 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

[1] A. Dehghan, K. J. McManus, and E. Gad, "Statistical analysis of structural failures of water pipes," Proceedings of the Inst of Civil Eng - Water Management, vol. 161, pp. 207 –214, 2008.

[2] O. Giustolisi, "Using a multi-objective genetic algorithm for SVM construction," Journal of Hydroinf, vol. 8, pp. 125-139, 2006.

[3] C. F. Juang, "A Hybrid of Genetic Algorithm and Particle Swarm Optimization for Recurrent Network Design," IEEE TRANSACTIONS ON SYS, vol. 34, pp. 997-1006, 2004.

[4] J. Kennedy, R. Eberhart, "Particle Swarm Optimization," IEEE Int Conf on Neural Networks, pp. 1942–1948, 1995.

[5] Z. Qi, Y. Tian, Y Shi," Robust twin support vector machine for pattern classification," Pattern Recognition, vol. 46, pp. 305-316, 2013.

[6] A. Smola, B. Scholkopf, "A tutorial on Support Vector Regression," J Statistics and Computing, vol. 14 , pp.199 -222 , 1998 .

[7] J. Soltani, M. Pour Tabari, "determination of effective parameters in pipe Failure rate in water distribution system using the combination of Artificial Neural Networks and Genetic Algorithm", Journal of Water & Wastewater, vol. 83, pp. 2-18, 2012,(In Persian).

[8] M. Tabesh, J. Soltani, R. Farmani, D.A. Savic, "Assessing Pipe failure Rate and Mechanical Reliability of water Distribution Networks Using Data Driven Modeling," Journal of Hydroinf, vol. 11, pp. 1-17, 2009.

[9] R. Johnson,"Rainfall-Runoff Modelling Using Artificial Neural Networks,", J Hydrologic Eng, vol. 4, pp. 232-239, 1999.

[10] V. N. Vapnik, "Principles of risk minimization for learning theory,". Advances in Neural Information Processing Sys, vol. 4, pp. 831-838 1992.

[11] V. N. Vapnik, O. Chapelle, "Bounds on error expectation for support vector machines," Neural Computation, vol. 12, pp. 2013– 2036,2000.

[12] V. N. Vapnik,"The Nature of Statistical Learning Theory," Springer Verlag", USA ,2010.