Małgorzata KUTYŁOWSKA[1]

PREDICTION OF WATER CONDUITS FAILURE RATE – COMPARISON OF SUPPORT VECTOR MACHINE AND NEURAL NETWORK

PRZEWIDYWANIE WSKAŹNIKA AWARYJNOŚCI PRZEWODÓW WODOCIĄGOWYCH – WEKTORY NOŚNE ORAZ SIECI NEURONOWE

Abstract: This paper presents the possibility of applying support vector machines (SVMs) and artificial neural networks (ANNs), based on radial basis functions to predict the failure rate of water conduits. The SVM method is an algorithm for carrying out regression and classification, taking into account a nonlinear decision space. This hyperplane divides the whole area in such a way that objects of different affiliation are separated from one another. In the case of ANNs, each of the neurons models a Gaussian response surface. The information from the inputs is transmitted to a basis function and each neuron calculates the Euclidean distance between the input, reference and output vectors. The failure rate of distribution pipes and house connections was predicted on the basis of operational data for the years 2001–2012. In both the methods the independent variables were: the length, diameter and year of construction of the distribution pipes and the house connections. The computations were carried out using the Statistica 12.0 software. The SVM-RBF model for the house connections and the distribution pipes had respectively 14 SVMs (including 7 localized SVMs) and 56 SVMs (including 46 localized SVMs). The ANN-RBF model contained 8 and 27 hidden neurons for respectively the distribution pipes and the house connections.

Keywords: regression methods, pipelines, modelling, radial basis functions

1. Introduction

Water distribution systems are the critical underground infrastructure components because they perform the strategic functions in broadly understood municipal engineering. In the current terrorist threat context [1], the necessity to ensure the proper protection and management of water supply systems is increasingly often highlighted. These are undoubtedly vital issues which together with reliability analyses, water demand analyses and the properly planned modernization of the pipelines and the whole

[1] Faculty of Environmental Engineering, Wroclaw University of Science and Technology, Wybrzeże S. Wyspiańskiego 27, 50–370 Wrocław, Poland, phone: +48 71 320 40 84, email: malgorzata.kutylowska@pwr.edu.pl

water supply infrastructure should be and currently are the subject of numerous studies and projects. Research on the technical condition, failure frequency and operational reliability of water conduits and the associated water losses is highly advanced in Poland and abroad [2–7]. The research findings indicate that such studies need to be continued in order to gain deeper knowledge in this field, especially with regard to mathematical modelling, which owing to the development of computing techniques is constantly improved and uses increasingly more accurate modelling methods [8]. Mathematical prediction and modelling have become very popular in broadly understood environmental engineering [9–13], which is an inducement to apply them to the study of the failure frequency of water distribution systems. So far such regression methods as: support vector machines (SVM), artificial neural networks (ANN) [14], K-nearest neighbours (KNN) [15] and regression and classification trees [16] have been used to determine failure rates.

The main aim of this study was to demonstrate that it is possible to apply a regression method based on support vector machines to predict the failure rate of water supply pipelines and to compare the results obtained in this way with the results of modelling this rate by means of RBF (radial basis functions) artificial neural networks. Besides also some prediction examples using other kernel functions of SVM models (linear, polynomial and sigmoidal) will be presented to achieve wider comparison area. Thanks to such a comparison it will be possible to determine the optimal modelling method for the currently operated water supply networks. A variant of the failure rate prediction method presented here has been successfully applied by Aydogdu and Firat [17], which induced the author to take up this subject as applied to the Polish water distribution network. The results presented here are complementary to the results of modelling the failure rate by SVM, reported in [18] where all the kinds of kernel functions were analyzed and other predictors were used. It is also examined whether the water supply pipeline failure rate prediction methods proposed in this paper can be an alternative to typical mathematical models.

1.1. Support vector machines

The support vector machines (SVM) method is an algorithm for regression and classification with a nonlinear decision space taken into account. This hyperplane divides the whole area in such a way that objects of different affiliation are separated from one another. It is also necessary to keep a maximum margin of error, *ie* the distance from the separating plane. The number of support vector machines determines the complexity of the relations between dependent and independent variables [19]. In the case of a qualitative analysis of such a dependent variable as the failure rate of water conduits, no classification, but regression is performed. Four kinds of SVMs, characterized by four types of kernel functions: linear, polynomial, sigmoidal and radial basis functions (RBF), are distinguished [19]. The notion of kernel functions derives from investigations of linear vector spaces. In the case of the problem considered here, RBF-SVM models, *ie* based on solely radial basis functions, were built and compared with ANN models. Nevertheless predictions results using other kernel functions are also displayed to show wider modelling point of view. But the straight comparison (SVM vs. ANN) could be only performed between models based on the same kernel functions (RBF). In the course of an regression analysis a relation between the dependent variable and the independent variables (predictors) is sought. This relation should possibly most accurately generate a dependent variable value for new cases (testing sample data), *ie* ones which the SVM model has not "seen" before, having been trained on a training sample. The mapping function $\varphi(x)$ is called kernel function which meets Mercer condition and feature map for Mercer kernel is as follows [20]:

$$k(x, y) = \varphi(x)\,\varphi(y) \tag{1}$$

The kernel functions are described by the equations (2–5), respectively for linear, polynomial, sigmoidal and RBF [20]:

$$k(x, y) = (x \cdot y) \tag{2}$$

$$k(x, y) = \big(s(x \cdot y) + \gamma\big)^d \tag{3}$$

$$k(x, y) = \tanh\big(s(x \cdot y) + \gamma\big) \tag{4}$$

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \tag{5}$$

where: γ – a learning rate,
x – independent variable,
y – dependent variable.

The prediction function is calculated from the relation [19]:

$$y(X) = w^T \varphi(x) + b \tag{6}$$

where: b – bias,
w – weight vector,
φ – mapping function.

1.2. RBF artificial neural networks

Since the theory of artificial neural networks is described in detail in the literature on the subject [21], only the most basic information on RBF artificial neural networks is provided here. Unlike the multilayer perceptron, RBF ANNs contain radial neurons (performing a function radially changing around a given centre in the vicinity of which nonzero values are assumed). Each such neuron models a Gaussian response surface. The information from the inputs is transmitted to the radial basis function and each neuron calculates the Euclidean distance between the input, reference and output vectors. It is essential that there are enough radial neurons so that they can accurately correlate the function with the sought solution. The solutions based on RBF artificial neural networks are slow-speed and require a considerable storage area, which is sometimes a serious limitation [19]. RBF ANNs are trained as follows. Training proceeds in two steps [19]: first RBFs are arranged using the input signals and then weights between the RBFs and the output neurons are determined. Consequently, no iterative process is required, which is evidence of the absence of typical training epochs. The characteristic feature of RBF artificial neural networks is that the RBFs are determined on the basis of the input vector and after the weights are added up the result is fed to the output. The location and width of the basis functions and the weights linking them with the output signals are of major importance in RBF artificial neural networks [19].

2. Experimental methodology

The failure rate (λ, fail./(km · a)) of the distribution pipes and the house connections in a selected Polish town was predicted using the RBF-SVM method and the RBF-ANN method. The two approaches were based on radial basis functions. Moreover some prediction results using other kernel functions (linear, polynomial and sigmoidal) are displayed to make a comparison more accurate. Operational data for the years 2001–2012 obtained from the town's water company were used for modelling. In the case of SVM modelling, the whole data set was randomly divided into two equal (50%) subsets. The training sample and the testing sample had 147 data each for the house connections and respectively 124 and 125 data for the distribution pipes. First a model was built using the training data and then it was tested on an "unseen" sample. In the case of ANNs, the procedural algorithm was slightly different due to the peculiarities of this kind of modelling. The artificial neural network learning process consisted of several stages: a training stage (50% of the data), followed by a testing stage (25% of the data) and finally, the validation (25% of the data) of the created models. In the considered case, the whole data set (294 data for the house connections and 249 data for the distribution pipes) was used to train the ANN. The division into a training sample, a testing sample and a validating sample was random. Using the ANN algorithm one can also make prognosis based on unseen data. Such a prognosis was made using a separately created set of operational data. Models were built for separately the distribution pipes and the house connections. The computations were carried out using the Statistica 12.0 software.

Since the SVM method is a kind of nonparametric regression, the correlations between the dependent variables (the predicted value) and the independent variable need not to be known. *V*-fold cross validation was used to find the optimal model parameters. In this type of cross validation, data are divided into V randomly selected disjoint parts. Using the *V*-1 parts of data as training examples the dependent variable is predicted and the prediction error is calculated on the basis the residual sum of squares. The procedure is executed for all the *V* data segments. Then a model quality measure is determined on the basis of the averaged errors of the particular cycles. The optimal model parameters are selected during a quality analysis. The parameters determined in the course of the *V*-fold cross validation are: gamma, capacity, epsilon and the number of support vector machines (including localized vectors) [19]. Tenfold ($V = 10$) cross validation was used in the considered problem, whereby it was possible to select proper values for such parameters (learning constants) as capacity (C) and epsilon (ε), since they are not *a priori* known. In the case of artificial neural networks, model parameters (*eg* the number of hidden neurons and the type of activation functions) are determined during ANN training using a proper training algorithm. Between ten and twenty ANN models, for which the number of hidden neurons ranged from 1 to 30, were tested. The model characterized by the smallest mean-square error and the best fit between the real data and the predicted ones was selected. The results presented later in this paper are for this selected optimal ANN model.

In both the methods the independent variables were: length (L_r, L_p), diameter (D_r, D_p) the year of construction (Y_r, Y_p) of the distribution pipes and the house connections. The same independent variables had been used by Aydogdu and Firat [17] to model the failure rate of water conduits by means of a combination of SVM and fuzzy logic. It should be noted that the average length of the pipelines was used in the calculations. This approach to failure rate determination (using the average length) is suggested in the literature on the subject [2]. The real (experimental) failure rate was calculated from the well-known relation [2, 3]:

$$\lambda = \frac{N(t)}{L \cdot \Delta t} \tag{7}$$

where: $N(t)$ – the number of failures of linear objects in time interval Δt, units;
L – the average length of pipelines in time Δt, km;
Δt – the observation time, year.

3. Results and discussion

The parameters of the built SVM-RBF and ANN-RBF models for the different types of water conduits are presented in Table 1. The validation error was one of the considerations for selecting an SVM model most accurately predicting the failure rate. The validation error for the distribution pipes and the house connections amounted to respectively 0.08 and 0.11. Nevertheless, the failure rate prediction on the basis of the testing sample was not satisfactory from the predicted/real data fit point of view. Moreover, the number of SVMs for the distribution pipes was high and as much as 82% of them were localized SVMs, *ie* with weights equal to ± the capacity value (Table 1), indicating a more complicated model structure. In the case of any kind of modelling, one should answer the question whether the aim is to obtain a perfect data fit at any cost, *ie* at the expense of model architecture complication, or rather to reveal the correlations between the dependent and independent variables. The latter approach enables one to find out whether the independent variables (predictors in the case of the SVM method) show significant correlations with the dependent variable, even if the estimate carries a predefined admissible error.

Table 1

Parameters of SVM-RBF and ANN-RBF models

Type of conduit/parameter	Distribution pipes	House connections
SVM model		
Gamma	0.333	0.333
Capacity (*C*)	3	1
Epsilon (ε)	0.2	0.5
Number of support vectors (localized)	56 (46)	14 (7)
Cross-validation error	0.081	0.110
ANN model		
Number of hidden neurons	8	27
Activation functions: hidden/output layer	Gaussian/linear	Gaussian/linear
Training algorithm	RBFT	RBFT
Correlation coefficient (learning/prognosis step)	0.956/0.859	0.997/0.897
Determination coefficient (learning/prognosis step)	0.914/0.737	0.994/0.805

In the case of the ANN models, Pearson's correlation coefficient (R), a determination coefficient (R^2) and a relative mean-square prediction error (amounting to about 20% for the distribution pipes and the house connections) would be compared. The error is rather high in comparison with the results of failure rate modelling by, *eg* artificial neural networks based on the multilayer perceptron [22]. But in the cited paper there was an additional input signal – the pipeline material, which (besides the perceptron structure) could have contributed to the closer convergence between the real and predicted data. Also in [23], where hourly water demand histograms were predicted, RBF ANNs were found to be less useful than the multilayer perceptron. Despite the fact that there were three times more hidden neurons in the house connections model than in the distribution pipes model (Table 1), the prediction results are worse and characterized by larger discrepancies between the real and predicted data (Figs 1 and 2). Because of the nature of RBF ANNs, the activation functions and the training method were pre-imposed, which also can have a bearing on modelling quality in comparison with, *eg* artificial neural networks using the multilayer perceptron, where it is possible to use

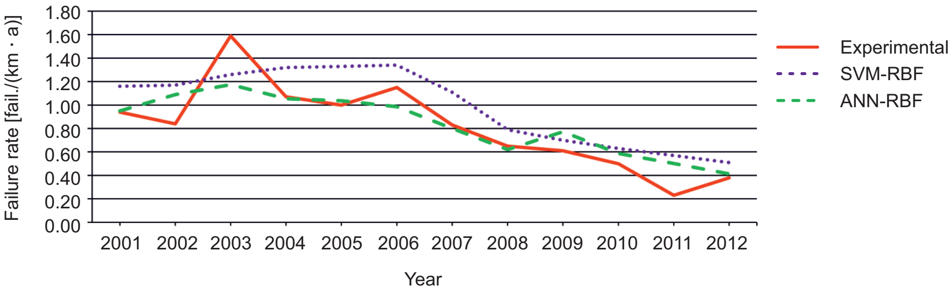

Fig. 1. Prediction results of house connections' failure rate – testing-SVM/prognosis-ANN

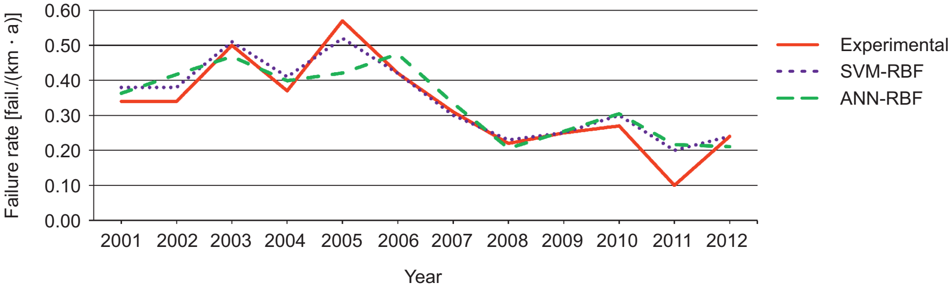

Fig. 2. Prediction results of distribution pipes' failure rate – testing-SVM/prognosis-ANN

several different functions, such as the sigmoidal function, the exponential function and so on [22].

The failure rate prediction results for the learning sample are presented in Table 2 while the ones for the testing sample (the SVM model) and the prognosis stage (the ANN model) are shown in Figs 1 and 2.

Table 2

Results of failure rate prediction – learning step

Year	House connections			Distribution pipes		
	Experimental	ANN-RBF	SVM-RBF	Experimental	ANN-RBF	SVM-RBF
2001	0.94	0.95	1.17	0.34	0.36	0.38
2002	0.84	0.84	1.16	0.34	0.36	0.38
2003	1.59	1.58	1.26	0.50	0.47	0.48
2004	1.07	1.07	1.32	0.37	0.39	0.41
2005	1.00	1.00	1.32	0.57	0.48	0.52
2006	1.15	1.15	1.33	0.42	0.42	0.42
2007	0.83	0.80	1.12	0.31	0.30	0.33
2008	0.65	0.63	0.79	0.22	0.21	0.22
2009	0.61	0.62	0.70	0.25	0.25	0.24
2010	0.50	0.50	0.63	0.27	0.26	0.24
2011	0.23	0.33	0.57	0.10	0.21	0.20
2012	0.38	0.38	0.51	0.24	0.25	0.24

It is necessary to draw this distinction between the testing sample for the SVM model and the prognosis stage for the ANN model since the testing sample data and the prognosis stage data were unknown to the given model. Only in this way, by analyzing the results obtained for the data set unseen by the model, one can assess the quality of the model and its applicability to dependent variable (failure rate) prediction. Moreover, in the Table 3 and 4 the results prediction in learning step and main parameters of the models using other kernel functions (linear, polynomial and sigmoidal) are displayed, respectively. The main aim of this work was to compare the modelling results of ANN and SVM models based on radial basis function. The information shown in the Table 3 and 4 should be rather treated as supplement. The analysis of the Table 3 shows that for distribution pipes SVM model based on sigmoidal function has the best convergence with experimental data in learning step (R^2 = 0.85). Slightly different situation is observed concerning house connections. For these conduits sigmoidal and polynomial functions were responsible for the best agreement (learning step) between predicted and experimental values of failure rate, R^2 equalled to 0.87 and 0.84, respectively. The whole comparison of prediction results using all kernel functions was described in [18] for another water distribution system. In this work only basic information concerning other kernel functions are stated.

Table 3

Results of failure rate prediction (other kernel functions) – learning step

Year	House connections			Distribution pipes		
	SVM-linear	SVM-polynomial	SVM-sigmoidal	SVM-linear	SVM-polynomial	SVM-sigmoidal
2001	1.04	1.04	0.95	0.43	0.43	0.39
2002	1.04	1.04	0.84	0.43	0.43	0.40
2003	1.18	1.21	1.20	0.47	0.47	0.51
2004	1.22	1.07	1.10	0.45	0.43	0.39
2005	1.24	1.06	0.93	0.49	0.48	0.53
2006	1.25	1.14	1.22	0.45	0.43	0.40
2007	0.99	1.03	0.85	0.33	0.40	0.37
2008	0.66	0.78	0.65	0.22	0.27	0.24
2009	0.58	0.63	0.52	0.21	0.26	0.22
2010	0.50	0.52	0.50	0.20	0.27	0.12
2011	0.43	0.33	0.50	0.15	0.22	0.04
2012	0.38	0.37	0.51	0.17	0.20	0.23

An analysis of Table 2 clearly shows that there is better agreement between the ANN-RBF model training results and the real failure rate λ values than in the case of the SVM-RBF model used for the house connections. For the distribution pipes the differences in failure rate predictions between the two modelling methods are not so significant and one can say that the two methods are equally effective, as indicated by the fact that coefficients R = 0.96 and R^2 = 0.92 are identical for both methods. Considerable errors (over 100%) occur in the estimates of the failure rate for the distribution mains only in 2011, which is undoubtedly due to the fact that this rate is very much different from the rates for the other analyzed years. A similar error occurs in the predictions of the failure rate for the house connections in 2011. Therefore the question arises: what should be done in the case of divergent data? Should they be included in order not to disrupt the continuity of the analysis of operational data, as it was done in this paper, or rather completely rejected? However, one should take into account the kind of analyzed problem. The modelling of the technical condition and failure rate of a water distribution network should not omit or exclude some years from the analysis simply because of divergent data. The information about the failure rate level in a given year is based on pipeline failures which really occurred. Therefore if some years were neglected in the analysis, this would result in an incomplete picture of the reality. Whereas models based on all the available data make it possible to obtain a complete picture, albeit not always a very accurate one.

Table 4

Parameters of SVM models based on other kernel functions

Type of conduit/parameter	Distribution pipes	House connections
SVM-linear		
Gamma	—	—
Capacity (C)	10	2
Epsilon (ε)	0.4	0.3
Number of support vectors (localized)	36 (27)	50 (46)
Cross-validation error	0.094	0.112
SVM-polynomial		
Gamma	0.333	0.333
Capacity (C)	10	5
Epsilon (ε)	0.4	0.3
Number of support vectors (localized)	40 (33)	52 (48)
Cross-validation error	0.124	0.136
SVM-sigmoidal		
Gamma	0.333	0.333
Capacity (C)	10	10
Epsilon (ε)	0.1	0.1
Number of support vectors (localized)	116 (110)	123 (119)
Cross-validation error	0.093	0.119

The analysis of the Table 4 shows that SVM models based on other kernel functions are more complicated due to *eg* higher capacity value, higher value of parameter ε (linear and polynomial functions) and more support vectors (also localized). The aim of the modelling is not only to get convergent prediction results, but also to achieve relatively simple model structure and its parameters. The choosing of the optimal model (concerning SVM modelling) should be based not only on agreement between experimental and predicted values, but also on analyzing the cross-validation error and the model architecture described by *eg* number of support vectors (also localized) and capacity. The enough huge data base is also the problem during the modelling. In some cases the number of available operating parameters or number of registered cases (received from water utilities) is too low to build the model responsible for prediction of failure rate with satisfactory convergence with real data. The models proposed in this paper consist of basic operating data. In the future it seems to be reasonable to widen the vector of independent variables to create more general models.

Similar correlations (as the ones described above concerning RBF models) between real and predicted data were obtained at the testing stage (the SVM model) and the prognosis stage (the ANN model), as shown in Figs 1 and 2. In the case of the house connections, more convergent results are generated by the ANN model, but for some years (*eg* 2003, 2006 and 2009) the differences are much larger than the ones observed at the learning stage. Despite many divergences, the trend in the variation of the predicted values is similar to the trend in the variation of experimental values. In the years 2006–2008 a similar pattern is observed for the SVM model, but most of the λ values are much higher than the real ones.

The prediction of the failure rate of the distribution pipes (Fig. 2) by the SVM method and the ANN method was characterized by acceptable agreement between the predicted and experimental results in both cases. The Pearson's correlation coefficient for the SVM model and the ANN model amounted to respectively 0.96 and 0.86, indicating that the SVM method is slightly better for predicting the failure rate of distribution pipes than the ANN method. The opposite is true for house connections. It should also be noted that the results of predicting the failure rate of the distribution pipes and the house connections (Table 2, Figs 1 and 2) by means of the SVM-RBF model are very similar for both the learning sample and the testing sample. Whereas the results of learning and prognosis by the ANN-RBF method show larger discrepancies for both types of pipelines. Even though at the learning stage the agreement between the real and predicted results is satisfactory, the prediction process (using new data) carries a larger (but still acceptable from the engineering point of view) error. This is evidence of greater effectiveness of RBF-based training by means of SVMs than ANNs. However, at the current stage of the research it cannot be explicitly indicated which of the methods is better and should be widely adopted in the modelling of the failure rate of water distribution networks. Further research in this area is needed, also on operational data from other water companies, permitting more in-depth analyses and broader generalizations.

The prediction results on testing sample concerning house connections and distribution pipes using other kernel functions by means of SVM model are displayed in the Figs 3 and 4. Similarly as in learning step, polynomial and sigmoidal functions were

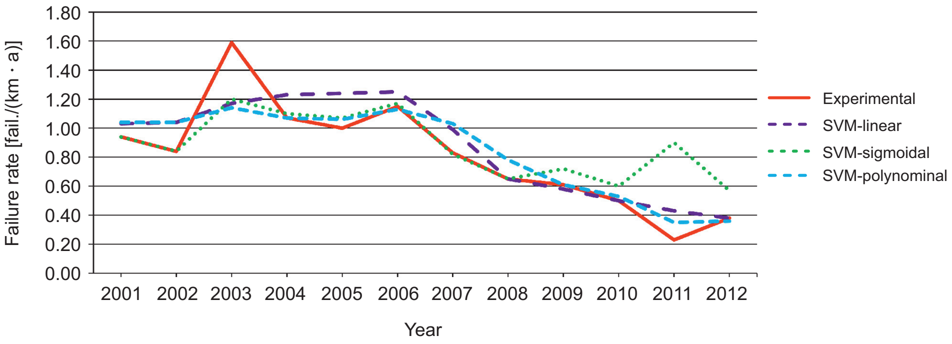

Fig. 3. Prediction results of house connections' failure rate – testing

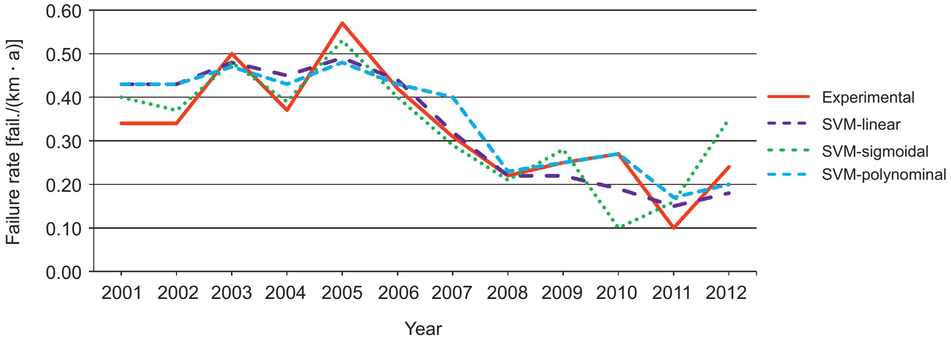

Fig. 4. Prediction results of distribution pipes' failure rate – testing

responsible for the optimal convergence between predicted and experimental values of indicator λ. It is necessary to remember that modelling only tries to imitate the reality. It is obvious that prediction results could not be exactly the same as experimental data. The problem is what error level is permitted and assumed at the very beginning of the modelling. Operating data have sometimes a lot of mistakes. If it is possible one should cooperate with exploiters to explain inaccuracy and to complete lack of parameters.

Such basic independent variables as: pipeline length, diameter and construction year were used to model the failure rate of the water supply pipelines. The weights of the connections between neurons were determined for the selected ANN model. The ANN-RBF model, describing the failure rate of the house connections, was characterized by the weakest connection between hidden neuron no. 19 and the output signal. The weight of this connection amounted to –0.09. The highest connection weight value (1.00) was observed between the input neuron (diameter) and hidden neuron no. 8. This supports the thesis, advanced in numerous publications [2, 3, 24, 25], that the failure rate of a pipeline is correlated with its diameter. The ANN-RBF model for the distribution pipes had a connection between hidden neuron no. 2 and the output signal (λ), with a weight of –4.13. This means that the situation was similar as in the case of the house connections. The strongest connection was characterized by a weight of 0.90 and occurred between the input signal (length) and hidden neuron no. 3. This means that the total length of a pipeline of this type for a given year plays a significant role in the modelling of the failure rate of distribution pipes. In the case of the SVM model, weights which should be considered jointly with another model parameter, *ie* capacity, were determined for the particular SVMs. If the weight of a given SVM is close to the value of ±C, this means that this SVM lies relatively near the hyperplane. If the weight is equal to ±C, the given SVM lies directly on the hyperplane boundary and it is then referred to as a localized SVM. The SVM-RBF model for the house connections had capacity C = 1 and the weights of its 14 SVMs were in a range of –1.00–1.00. There were 7 localized SVMs. The other SVMs had weights ranging from –0.86 to 0.76. The SVM-RBF model for the distribution mains had capacity C = 3 and the weights of 56 SVMs were in a range of –3.00–3.00. There were 46 localized SVMs. The weights of the other SVMs ranged from –2.94 to 2.76.

4. Conclusions

This paper has presented the results of the prediction of the failure rate of the distribution pipes and house connections in one of the Polish towns, based on operational data for the years 2001–2012, by means of support vector machines and artificial neural networks. The subject seems to be of importance for the correct and quick estimation of the reliability level. The created SVM-RBF and ANN-RBF models can be useful in cases when it is necessary to determine the failure rate in order to take a quick decision concerning the planned repairs of conduits. The obtained prediction results indicate that both the methods can be used to estimate the failure rate of municipal systems. But, similarly as in the case of other methods, a relatively large database must be available in order to identify the relevant correlations (training/ learning) and then to test the model. One should bear in mind that each modelling carries a prediction error. When selecting an optimal model one should not only consider the achievement of the best possible convergence, but also assess the effect of an inaccurate estimate. The consequences of a failure of, *eg* distribution pipes are incomparably more massive and severe than any damage to house connections.

The results of prediction concerning other kernel functions are also satisfactory. The SVM models for the distribution pipes using linear, polynomial and sigmoidal kernel functions were characterized by the weights (without localized vectors) in the range –8.14 to 6.63, –9.70 to 9.04 and –9.25 to 9.95, respectively. In relation to house connections the weights without localized vectors) varied in the range –1.86 to –0.23, –3.26 to 2.69 and –9.89 to 3.09, respectively.

The optimal SVM-RBF model had gamma coefficient amounted to 0.33 for both the distribution pipes and the house connections while capacity C and the number of SVMs were respectively 3 and 4 times greater in the case of the model describing the failure rate of the distribution pipes. The error of the *V*-fold cross validation amounted to 0.081 and 0.110 for the model describing the failure rate of respectively the distribution pipes and the house connections. The input signals and the predictors, respectively, in both the ANN method and the SVM method were the length, diameter and year of laying the water conduit in the ground. The optimal ANN-RBF model contained 8 and 27 hidden neurons for respectively distribution pipes and house connections. The correlation and determination coefficients are slightly higher at the stage of learning the artificial neural network than during prognosis. The obtained modelling results can be in still better agreement with the real data. This would increase the quality of the decisions concerning, *eg* the planning of replacements and renewals of pipeline sections. The accuracy of the modelling can be increased by adding more predicators of, *eg* the conduit material, the number of failures or the pressure prevailing in the pipeline, to the vector of independent variables. However, it is not always possible to obtain such data due to the fact that water companies do not record all their operational information. It should be noted that the situation is improving by the year and the operators increasingly often use the GIS database, which facilitates data acquisition and analysis. Therefore, further studies aimed at determining such independent variables which will be proper parameters for the correct prediction of the failure rate by SVMs and ANNs are needed.

Acknowledgment

The work was realized within the allocation No. B50519 awarded for Faculty of Environmental Engineering Wroclaw University of Science and Technology by Ministry of Science and Higher Education. The grant was allocated for scientific researches of young scientists in years 2015–2016.

References

[1] Pietrucha-Urbanik K. Glob Network Environ Sci Technol J. 2014;16(5);893-900. http://journal.gnest.org/sites/default/files/Submissions/gnest_01414/gnest_01414_published.pdf

[2] Hotloś H. Ilościowa ocean wpływu wybranych czynników na parametry i koszty eksploatacyjne sieci wodociągowych [Quantitative assessment of the effect of some factors on the parameters and operating costs of water-pipe networks]. Wrocław: Wrocław University of Technology Publishing House; 2007.

[3] Kwietniewski M, Rak J. Niezawodność infrastruktury wodociągowej i kanalizacyjnej w Polsce [Reliability of water supply and wastewater disposal infrastructure in Poland]. Warszawa: Monographs of the Civil Engineering Committee at the Polish Academy of Sciences, Studies in Engineering No. 67; 2010.

[4] Iwanek M, Kowalski D, Kwietniewski M. Badania modelowe wypływu wody z podziemnego rurociągu podczas awarii [Model studies of a water outflow from an underground pipeline upon its failure]. Ochr Środ. 2015;37(4):13-17. http://www.os.not.pl/docs/czasopismo/2015/4-2015/Iwanek_4-2015.pdf

[5] Zimoch I, Szymik-Gralewska Zastosowanie zintegrowanej metody analizy niezawodnościowo-ekonomicznej w zarządzaniu przewymiarowaną infrastrukturą wodociągową [Application of integrated reliability-economic analysis in management of oversized water supply infrastructure]. Ochr Środ. 2015;37(4):25-30. http://www.os.not.pl/docs/czasopismo/2015/4-2015/Zimoch_4-2015.pdf

[6] Tscheikner-Gratl F, Sitzenfrei R, Rauch W, Kleidorfer M. Struct Infrastruct Eng. 2016;12(3):366-380. DOI: 10.1080/15732479.2015.1017730.

[7] Piratla KR, Yerri SR, Yazdekhasti S, Cho J, Koo D, Matthews JC. Proc Eng. 2015;118:727-734. DOI: 10.1016/j.proeng.2015.08.507

[8] Scheidegger A, Leitao JP, Scholten L. Water Res. 2015;83:237-247. http://dx.doi.org/10.1016/j.watres.2015.06.027.

[9] Cieżak W, Cieżak J. Environ Prot Eng. 2015;41(2):179-186. DOI: 10.5277/epe150215.

[10] Kaźmierczak B, Wdowikowski M. Periodica Polytechnica Civ Eng. 2016;60(2):3-5-312. DOI 10.3311/PPci.8341.

[11] Tchórzewska-Cieślak B. Environ Prot Eng. 2011;37(3):111-118. http://epe.pwr.wroc.pl/2011/3_2011/12tchorzewska.pdf

[12] Zimoch I, Łobos E. Application of the Theil statistics to the calibration of a dynamic water supply model. Environ Prot Eng. 2010; 36(4):105-116. http://epe.pwr.wroc.pl/2010/zimoch_4-2010.pdf

[13] Kolasa-Więcek A. Ecol Chem Eng S. 2013;20(2):419-428. DOI: 10.2478/eces-2013-0030.

[14] Nishiyama M, Filion Y. Can J Civ Eng. 2014;41(10):918-923. DOI: dx.doi.org/10.1139/cjce-2014-0114.

[15] Kutyłowska M. Prediction of failure rate of water pipes using K-nearest neighbours method. Proceedings of the IWA 8th Eastern European Young Water Professionals Conference Gdańsk, Poland, 12-14 May, 2016, 93-94. http://iwa-ywp.eu/wp-content/uploads/2016/06/Book_of_abstracts.pdf

[16] Bevilacqua M, Braglia M, Montanari R. Reliability Eng Syst Saf. 2003; 79(1):59-67. http://www.sciencedirect.com/science/article/pii/S0951832002001801

[17] Aydogdu M, Firat M. Wat Resour Manage. 2015;29:1575-1590. DOI: 10.1007/s11269-014-0895-5.

[18] Kutyłowska M, Orłowska-Szostak M. Przewidywanie wskaźnika awaryjności przewodów wodociągowych za pomocą metody wektorów nośnych [Forecasting failure rate of water pipes using support vector machines]. In: Kuś K, Piechurski F, editors. Nowe technologie w sieciach i instalacjach wodociągowych i kanalizacyjnych. Gliwice: 2016.

[19] Statistica 12.0., Electronic Manual.

[20] Guo YM, Wang XT, Liu C, Zheng YF, Cai XB. Maint and Reliability. 2014;16(1):85-91. http://www.ein.org.pl/pl-2014-01-14.

[21] Suzuki K. Artificial neural networks. Architectures and applications. Chicago: InTech; 2013.

[22] Kutyłowska M. Eng Fail Anal. 2015;47:41-48. http://dx.doi.org/10.1016/j.engfailanal.2014.10.007.

[23] Siwoń Z, Cieżak W, Cieżak J. Modele neuronowe szeregów czasowych godzinowego poboru wody w osiedlach mieszkaniowych [Neural network models of hourly water demand time series in housing areas]. Ochr Środ. 2011;33(2):23-26.
http://www.os.not.pl/docs/czasopismo/2011/2-2011/Siwon_2-2011.pdf

[24] Pelletier G, Mailhot A, Villeneuves JP. J Wat Resour Plann and Manage. 2003;129(2):115-123. DOI: 10.1061/(ASCE)0733-9496(2003)129:2(115).

[25] Tabesh M, Soltani J, Farmani R, Savic D. J Hydroinformatics. 2009;11(1):1-17. DOI: 10.2166/hydro.2009.008.

PRZEWIDYWANIE WSKAŹNIKA AWARYJNOŚCI PRZEWODÓW WODOCIĄGOWYCH – WEKTORY NOŚNE ORAZ SIECI NEURONOWE

Wydział Inżynierii Środowiska
Politechnika Wrocławska, Wrocław

Abstrakt: W pracy przedstawiono możliwość zastosowania metody wektorów nośnych (SVM) oraz sztucznych sieci neuronowych (SSN) opartych na radialnych funkcjach bazowych do przewidywania wskaźnika intensywności uszkodzeń przewodów wodociągowych. Metoda wektorów nośnych jest algorytmem, za pomocą którego dokonuje się regresji i klasyfikacji z uwzględnieniem nieliniowej przestrzeni decyzyjnej. Ta hiperpłaszczyzna dzieli cały obszar w taki sposób, że obiekty o różnej przynależności są od siebie oddzielone. Natomiast w przypadku sieci neuronowych każdy z neuronów modeluje tzw. gaussowską powierzchnię odpowiedzi. Informacje z wejść przekazywane są funkcji bazowej, a każdy neuron oblicza odległość euklidesową między wektorami wejściowymi, wzorcowymi i wyjściowymi. Przewidywanie wskaźnika intensywności uszkodzeń przewodów rozdzielczych i przyłączy wykonano na podstawie danych eksploatacyjnych z lat 2001–2012. W przypadku obydwu metod zmiennymi niezależnymi były: długość, średnica oraz rok budowy przewodów rozdzielczych i przyłączy. Obliczenia przeprowadzono w programie Statistica 12.0. Model SVM dla przyłączy i przewodów rozdzielczych posiadał odpowiednio 14 wektorów nośnych, w tym 7 związanych oraz 56 w tym 46 związanych. Model SSN zawierał 8 i 27 neuronów ukrytych odpowiednio w odniesieniu do przewodów rozdzielczych i przyłączy.

Słowa kluczowe: metody regresyjne, rurociągi, modelowanie, radialne funkcje bazowe