SUPPORT VECTOR MACHINES AND PROBABILISTIC NEURAL NETWORKS IN THE ASSESSMENT OF THE RISK OF DAMAGE TO WATER SUPPLY SYSTEMS IN MINING AREAS

Janusz RUSEK *

Department of Engineering Surveying and Civil Engineering
Faculty of Mining Surveying and Environmental Engineering
AGH University of Science and Technology in Krakow
al. Mickiewicza 30, 30-059 Krakow, Poland

Summary

The paper analyzes the usefulness of the two approaches: *SVM* (*Support Vector Machine*) and *PNN* (*Probabilistic Neural Network*) to assess the risk of damage to water supply systems resulting from the impacts of the industrial environment in the *Upper Silesian Coal Basin* (*GZW*). Two classification models in the form of *SVM* and *PNN* networks were created as part of the study, using data on technical and material solutions, identified damage and intensity of mining impacts for the analyzed water supply systems. Selection of the optimal parameters of the structures of both models was carried out using a genetic algorithm.

Keywords: water supply system, mining impacts, probabilistic neural network *PNN*, support vector machines *SVM*

1. INTRODUCTION

Assessment of the risk of damage to water supply systems located in mining areas is a multidimensional issue. This results from the necessity to take into account both mining impacts, as well as geometry, material parameters, technology and technical condition [1, 2]. The risk of damage is defined in the probabilistic notation [3], and therefore it becomes necessary to map this process so as to be able to identify the fragment of the system which is at risk of damage, while defining the probability of such an event. The problem of assessing the risk of damage to the water supply system can therefore be formulated as the classification task.

---

* rusek@agh.edu.pl

The approaches allowing for the classification of a threat, together with determining its probability, include *Probabilistic Neural Networks* [4]) and *Support Vector Machines*, related to the neural networks [5]).

These systems are included in the methods of *Supervised Machine Learning*. Therefore, the basis for creating such models are learning data.

In this paper, the data describing the extent of damage to the water supply systems were collected for the mining area of *Upper Silesian Coal Basin* (235 cases). The database contains information about the technical and material solutions, geometry and the encountered impacts of mining.

2. RESEARCH METHODOLOGY

Artificial neural networks are universal approximators and classifiers of multi-dimensional problems [6]. Due to the fact that these structures comprise numerous independent computational units interacting in parallel with each other, and that each unit (artificial neuron) is able to separate signals in the non-linear form, it is possible to model complex non-linear processes. Additionally, networks with properly selected number of hidden layers can perform mapping over any specified area of the input space [7]. It is also important, that besides the learning, validation and testing data, it is not necessary to predetermine the initial form of the ultimate mapping.

In the case of the classification problem analyzed in this paper, the two approaches, of *Support Vector Machines* (*SVM*) and *Probabilistic Neural Networks* (*PNN*), were selected for the studies. Both methods allow to solve the problem of classification, while making the result more detailed by adding the value of probability. These approaches differ in structure and manner of learning [7]. *SVM* is defined as the so-called robust method, where the main objective is to reach a compromise between the correct classification of the patterns and the level of generalization. In contrast, probabilistic neural networks aim to finally identify the result of the classification of a given pattern in the ranking process, depending on the level of activation of the neurons indicating the specific categories of the output variable [4]. The confrontation of these two approaches may allow to identify the approach which is the most effective in the context of assessing the risk of damage to building structures subjected to mining impacts.

2.1. Theoretical basis for Support Vector Machines

In *SVM* networks, the primary task is to separate patterns between the categories of the dependent variable [6]. With reference to the collection of patterns $\{\mathbf{x}_k, \mathbf{d}_k\}$, k=1…N, where $\mathbf{x}_k$

is the input vector, and the value of $\mathbf{d}_k$ is its assigned category (1 or -1), the objective is to find the optimal decision hyperplane serving as a separator of the pattern data for the assigned categories:

$$y(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \tag{1}$$

However, the application of a linear description for the separating hyperplane is often inefficient and refers to the cases where the data are actually linearly separable.

To increase the separation chances for the patterns **x**, Cover's theorem is used (e.g. [7]). It involves the projection of the original patterns **x** into the so-called feature space of higher dimension.

Assuming that the projection of the original patterns **x** into the feature space is performed using a certain transformation $\varphi(\mathbf{x})$, a separation hyperplane is obtained in the following form:

$$y(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b = 0 \tag{2}$$

where: $\mathbf{x} \in R^n$ – is the vector of the input data in the n-dimensional space,

$\varphi(\cdot): R^n \rightarrow R^{n_h}$ – is a certain transformation converting raw input data into the so-called *feature space,*

$\mathbf{w}^T$ – is the vector of the weights.

The mapping $\varphi(\cdot): R^n \rightarrow R^{n_h}$ is given in an implicit way, and results from the application of a specific type of the kernel function (e.g. [5]).

The final description of the *SVM* classifier, according to (e.g. [8]) can be written as:

$$y(\mathbf{x}) = \sum_{k=1}^{N_{SV}} \alpha_k d_k K(\mathbf{x}, \mathbf{x}_k) + b \tag{3}$$

The factor $K(\mathbf{x}_k,\mathbf{x}_j)$ appearing in the equation (3) is the kernel of the system, which is predetermined explicitly and it is the result of assembling the implicit functions $\varphi(\cdot)$ [7]:

$$K(\mathbf{x}_k, \mathbf{x}_j) = \varphi(\mathbf{x}_k)\varphi(\mathbf{x}_j) \tag{4}$$

The form of the kernel is selected arbitrarily from all the functions which meet the assumptions of Mercer's Theorem [9].

The study used the classification method *C-SVC* (*C-Support Vector Classifier*), implemented in the *LIBSVM* package [9].

The main problem associated with the construction of the *SVM* classifier is determining the optimum values of the parameters *C* and γ. The parameter *C* is a regularization constant present in the formulation of the so-called *loss function* which determines the learning process (e.g. [8]). On the other hand, the parameter γ determines the width of the adopted kernel functions (5) (e.g. [5]).

$$K(\mathbf{x}_k, \mathbf{x}_j) = \exp\left(-\frac{(\mathbf{x}_k - \mathbf{x}_j)^2}{\gamma^2}\right) \overset{\sigma=\frac{1}{\gamma^2}}{=} \exp\left(-\sigma(\mathbf{x}_k - \mathbf{x}_j)^2\right) \tag{5}$$

In the study, basing on the method proposed by [9], the selection of the parameters C and γ was carried out, using the method of gradientlessness optimization based on the genetic algorithm [10, 11].

2.2. Theoretical basis of Probabilistic Neural Networks

Probabilistic Neural Networks can be used both for the classification and regression problems. Unlike other artificial neural networks (e.g. *MLP - Multilayer Perceptron* or *RBF - Radial Basis Neural Network*), the advantage of *PNN* is the possibility to interpret its structure as the estimated probability density distribution for the dependent variable. Also, the process of building such networks is different and it does not require learning, which is present in most of the other feedforward networks. This results from the lack of weights of the synaptic connections, which are present in the structures of other neural networks. The role of *weighing* in the projection of a given input vector **X** during the network simulation is performed by the Gaussian kernel functions, arranged in the training patterns [11].

Probabilistic Neural Network consists of four computational layers: input layer, pattern layer (component kernel functions estimating the probability density for each category), summation layer and output layer [4].

With a determined network, it is possible to simulate the system response to a given multi-dimensional input vector $\mathbf{X} = (x_1, ..., x_n)^T \in R^n$.

Then, in the pattern layer, for a given input vector **X,** the values of each kernel function $F_{k,i}(\mathbf{X})$ are determined. In this way, the activation value of the individual kernel functions is obtained. Computational units in the pattern layer are divided into k = 1..K groups, corresponding to the specific categories at the network output.

The individual kernel functions are Gaussian curves written in the following form:

$$F_{k,i}(\mathbf{X}) = \frac{1}{(2\pi\sigma)^{n/2}} \exp\left(-\frac{\|\mathbf{X} - \mathbf{X}_{k,i}\|^2}{\sigma^2}\right) \tag{6}$$

where: σ – width (fuzzy parameter) of the kernel function,

$\mathbf{X}_{k,i} \in R^n$ – pattern in the input space constituting the center of the kernel function $F_{k,i}$

In the summation layer, for each specified sub-group of pattern neurons, representing K of different categories within a given category ($k = 1 \ldots K$), summation of the neuron activation values is performed. The summation process can be written as:

$$G_k(\mathbf{X}) = \sum_{i=1}^{M_k} w_{ki} F_{k,i}(\mathbf{X}) \quad k \in \{1, \ldots, K\} \tag{7}$$

where:
M_k – number of neurons from the layer of the pattern assigned to the recognition of the category k

$w_{k,i}$ – non-negative weighing factors, meeting the assumption: $\sum_{i=1}^{M_k} w_{k,i} = 1$.

The final result of the classification of the pattern **X** presented at the *PNN* input is obtained based on the comparison of the determined G_k values in the summation layer and the selection of that k category, for which G_k had the greatest value [4, 11]:

$$C(\mathbf{X}) = \arg \max_{1 \le k \le K} (G_k) \tag{8}$$

3. STUDY RESULTS

The study was based on the database regarding damage to the water supply system located in the Upper Silesian Coal Basin. The total number of all the analyzed cases was 235. The set of input data comprised: the variable describing the material of the pipeline (*M*), the variable describing the resistance category of the analyzed section of the network (*KO*), the variable describing the current technical condition (*ST*) and the index of continuous surface deformation ε_{max} - maximum surface deformation at a given section of the network.

The data set was divided into the training set (165 patterns) and the test set (70 patterns). With such a division, two classification models were built, which described the risk of damage (*PNN* model and *SVM* model).

3.1. The SVM model

The basis for creating the classification model using the *SVM* approach was the *LIBSVM* library dedicated to the *Matlab* environment [9]. The *C-SVC* (*C-Support Vector Classifier*) method was used, available as part of the package, and radial kernel functions were adopted.

To solve the problem, it was required to determine the arbitrarily chosen constants *C* (regularization parameter) and γ (kernel width). In order to determine the optimal set of these parameters, the method of gradientlessness optimization was applied, using the genetic algorithm [11, 12]. Table 1 summarizes the results of testing the conformity of model classification for the training set and the test set. Table 2 summarizes the basic characteristics of the created model: the number of support vectors and the parameters determined through optimization.

Tab. 1. Conformity of classification for the SVM model

The percentage of correctly classified cases for the training set (165 patterns)	The percentage of correctly classified cases for the test set (70 patterns)
97.58 %	91.43 %

Tab. 2. Characteristics of the created model

The number of support vectors N_{SV}	Parameter C	Parameter γ
28	29.64	1.82

The created *SVM* model is characterized by a high degree of pattern recognition, both from the training and test sets - Table 1. In addition, a clear reduction in the number of support vectors representing the core of the model structure (reduction of the original set of 165 learning patterns to 28 - Table 2) justifies the good generalizing properties (generalizing the acquired knowledge to the cases of the test set, which are not used in the learning process).

Table 3 demonstrates the results of classification in the form of the confusion matrix. On the diagonal of the matrix there is the sum of the cases correctly classified within a given category (I – undamaged, II - damaged). The extreme fields contain the cases classified incorrectly (category I instead of II, and vice versa). The analysis of the confusion matrix allows for the inference of model predispositions to falsify the results for the benefit of specific categories of the output variable (extreme fields).

As it appears from the values demonstrated in Table 3, the created model in its indications does not favor any of the categories of the output variable.

Tab. 3. Confusion matrix for the SVM classifier - the training and the test sets

Confusion matrix – SVM (C-SVC)									
Output class	I	98 59.4 %	3 1.8 %	→∑ 97.0 % →∑ 3.0%	*Output class*	I	34 48.6 %	1 1.4 %	→∑ 97.1 % →∑ 2.9 %
	II	1 0,6 %	63 38,2 %	→∑ 98.4 % →∑ 1.6 %		II	5 7.1 %	30 42.9 %	→∑ 85.7 % →∑ 14.3 %
		↓∑ 99.0 %	↓∑ 95.5 %	→↓∑ 97.6 %			↓∑ 87.2 %	↓∑ 96.8 %	→↓∑ 91.4 %
		↓∑ 1.0%	↓∑ 4.5 %	→↓∑ 2.4 %			↓∑ 12.8 %	↓∑ 3.2 %	→↓∑ 8.6 %
		I	II	Training set 165			I	II	Test set 70
		Target class					*Target class*		

Based on the presented results, it can be concluded that the percentage level of the conformity for the classification of both the training set and the test set is acceptably high, which justifies the good properties of fitting to the learning data as well as of generalization of the knowledge to the field of new patterns. Additionally, in the case of incorrect classification, the model does not exhibit excessive inertia directed to the individual categories.

3.2. The PNN model

Similarly, as for the *SVM* network (*C-SVC*), also for the *PNN* network a classification model for the input variables: *M*, *KO*, *ST* and ε_{max} was created.

The process of building such a network takes into account an arbitrary selection of the parameter defining the width of the Gaussian kernel function σ. Therefore, as in the case of *SVM* network learning, the optimal selection of the parameter σ was performed, using the genetic algorithm [11, 12]. Table 4 contains the summary of the main characteristics of the created *PNN* model, and Table 5 - the confusion matrix.

Tab. 4. Conformity of classification for the PNN model

The width of the kernel function γ	The percentage level of conformity	
	Training set	Test set
0,20	97.57 %	90.00 %

Table 4 demonstrates that the *PNN* model has a high degree of conformity in relation to the learning data (97.57% of correctly classified patterns). A slightly lower level of conformity was obtained for the test set. This illustrates the good fitting quality of the model, as well as the lack of overfitting [7]).

Tab. 5. Confusion matrix for the PNN classifier – the training and the test sets

Confusion matrix – PNN									
Output class	I	98 59.4 %	3 1.8 %	→∑ 97.0 % →∑ 3.0%	*Output class*	I	34 48.6 %	1 1.4 %	→∑ 97.1 % →∑ 2.9 %
	II	1 0.6 %	63 38.2 %	→∑ 98.4 % →∑ 1.6 %		II	6 8.6 %	29 41.4 %	→∑ 82.9 % →∑ 17.1 %
		↓∑ 99.0 % ↓∑ 1.0%	↓∑ 95.5 % ↓∑ 4.5 %	→↓∑ 97.6 % →↓∑ 2.4 %			↓∑ 85.0 % ↓∑ 15.0 %	↓∑ 96.7 % ↓∑ 3.3 %	→↓∑ 90.0 % →↓∑ 10 %
		I	II	Training set 165			I	II	Test set 70
		Target class					*Target class*		

The analysis of the distribution of the model indications, presented in the form of the confusion matrix (Table 5) allows to conclude that the model does not favor significantly any category of the output variable.

3.3. The risk of damage to the water supply system in probabilistic notation

In the course of the study, simulation of the created classification models was carried out in order to complete the classification results by the value of probability of the occurrence of the obtained indications.

In the case of the *SVM* model, the procedure of determining the probability of obtaining a given classification result is a two-stage process. The first stage includes the result of the *SVM* model classification. The second stage involves fitting of the sigmoid function, reflecting the posterior probability, to the classification results generated by the *SVM* model created in the first stage [13, 14].

On the other hand, determination of the probability for the classification results obtained by the *PNN* approach involves averaging the values of the Gaussian activated kernels present in the penultimate layer of the network [4, 11]. Therefore, the classification of a given case and assigning probability to it, is a one-stage process.

Figure 1 illustrates in the graphical way the results of the simulation how *SVM* models work (*C-SVC*) and *PNN* on the set of data consisting of the training set and the test set. The results have been presented with respect to the ε_{max} variable, describing the impact of the continuous surface deformation. On the graphs, categorized classification results (damaged and undamaged) have been marked, as well as the corresponding probability values.

For the results of the *SVM* model classification (*C-SVC*), the obtained probability level oscillated at about 0.9. The probability values for the results of the *PNN* model classification are characterized by a greater scatter (0.55 to 1.0). Given that the probabilities determined by the *PNN* method are “encoded” in the model structure, comparing with the two-stage procedure for determining the probabilities by the *SVM* method (*C-SVC*), it can be concluded that the *PNN* approach produces more reliable results.

4. SUMMARY AND CONCLUSIONS

The studies presented in this work aimed to demonstrate the possibilities of using Support Vector Machines and Probabilistic Neural Networks to assess the risk of damage to public utilities subjected to mining impacts.

Analysis of the obtained results leads to the following conclusions:

- both *SVM* structures of and *Probabilistic Neural Networks* can be used to assess the risk of damage to public utilities subjected to mining impacts, as evidenced by high percentage levels of conformity obtained in the course of learning and testing the created models, both methods, besides the implementation of the standard classification task, offer the possibility to define the obtained results more precisely by adding the probability values, which can be interpreted as a measure of the risk of damage to the water supply system,

- the advantage of the *PNN* approach is the possibility to determine probabilities for the classification results directly from the analysis of its structure. Unlike the *SVM* method, for which the procedure for determining the probabilities for the classification results is carried out as an additional stage after the creation of the model.

The obtained results of the analysis prove that there is the possibility of using the methods of *Support Vector Machines* and *Probabilistic Neural Networks* as tools in assessing the risk of damage do objects subjected to environmental impacts.

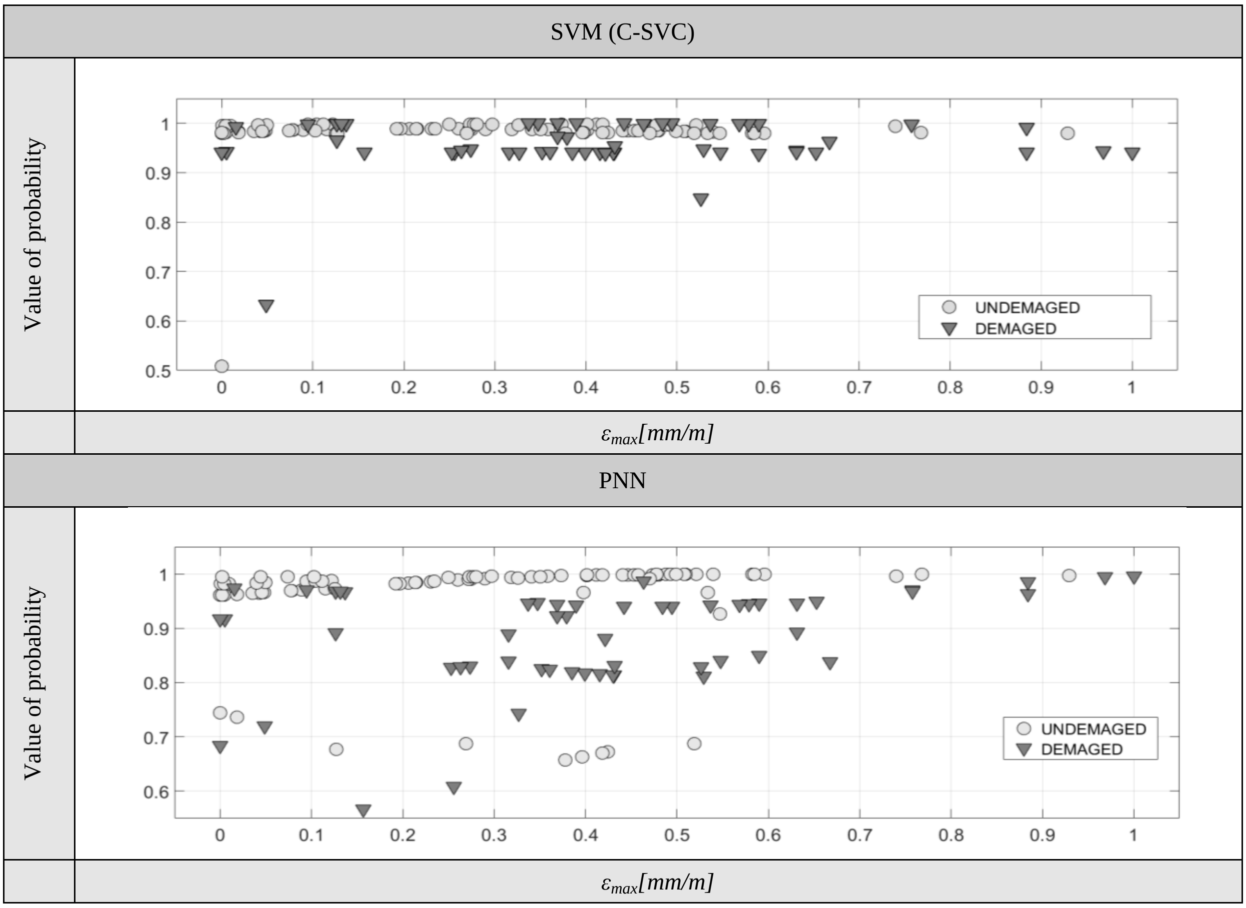

Fig. 1. Simulation results of the created *SVM* and *PNN* models

-

The article was prepared as part of the AGH statutory research No. 11.11.150.005.

References

1. FIREK K. Proposal for classification of prefabricated panel building damage intensity rate in mining areas. Archives of Mining Sciences. Publisher Institute of Rock Mechanics PAN, Vol. 54, Iss. 3, Kraków., **2009**.

2. KWIETNIEWSKI J., RAK J.R. Niezawodność infrastruktury wodociągowej i kanalizacyjnej w Polsce (Reliability water and sewerage infrastructure in Poland). Civil and water engineering committee PAN. Warsaw, **2010**.
3. SUNDARARAJAN, C.R. Probabilistic Structural Mechanics Handbook: Theory and Industrial Applications. Springer US, Texas, **1995**.
4. SPECHT D. F. Probabilistic neural networks. Neural Networks 3: 109–118, **1998**.
5. BURGES C.J.C. A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery, 2, 121–167 (1998). Kluwer Academic Publishers, Boston, **1998.**
6. OSOWSKI S. Neural networks for information processing. Publishing house of Warsaw Technical University, Warsaw, **2013**.
7. ŁĘSKI J. Neuro-fuzzy systems. WNT, Warsaw, **2008**.
8. SCHÖLKOPF B., SMOLA A.J. Learning with Kernels. MIT Press. Cambridge, Massachusetts, **2002**.
9. CHANG C-C., LIN C-J. LIBSVM: a Library for Support Vector Machine. Software available at: http://www.csie.nyu.edu.tw/~cjlin/libsvm. **2008**.
10. AFFENZELLER M., WAGNER S., WINKLER S., BEHAM A. Genetic Algorithms and Genetic Programming: Modern Concepts and Practical Applications. Chapman and Hall/CRC. 2009.
11. RUTKOWSKI L. Methods and techniques of artificial intelligence. Wydawnictwo Naukowe PWN, Warsaw, **2012**.
12. RUSEK J. Computational intelligence methods concerning in the problem of modelling technical wear of buildings in mining areas. Geomatics and Environmental Engineering, Quarterly, AGH University of Science and Technology, Vol. 6, No. 3, Cracow, **2012**.
13. FRANC V., ZIEN A., SCHÖLKOPF B. Support Vector Machines as Probabilistic Models In Proceedings of the 28th International Conference on Machine Learning, pages: 665-672, International Machine Learning Society, Madison, WI, USA, ICML, **2011**.
14. PLATT J. C. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In A. Smola et al. (ed.), Advances in Large Margin Classifiers. MIT Press, Cambridge, **2000**.