Computer Assisted Mechanics and Engineering Sciences, 14: 767–776, 2007.

An optimization of heuristic model of water supply network

Ryszard Wyczółkowski, Bogdan Wysogląd
Silesian University of Technology, Department of Fundamentals of Machine Design
ul. Konarskiego 18A, 44-100 Gliwice, Poland

(Received in the final form September 18, 2007)

In the paper an intelligent monitoring system of local water supply system is described. The main task of this system concerns water leakages detecting. For inputs, this system uses information from few pressure or flow sensors, mounted on the pipeline network, the output is a piece of information about leakage detection and localization.

A heuristic model of water supply network makes the main part of intelligent diagnostic system. The model was built with the use of artificial neural networks. This paper presents the structure and optimization of a heuristic model. The authors took advantage of methods of artificial intelligence and methods known from model-based process diagnostics to increase the accuracy with which system detects of water leakages.

Keywords: water supply systems, diagnostics, genetics algorithm, artificial neural network

1. Introduction

Water supply systems are one of the most essential parts of the urban and rural technical infrastructure. It is necessary for them to be reliable, especially because of counteraction of water loss. Finding leaks is one of the typical problems connected with water pipelines maintenance. This task is not simple enough, because leaking water quite often can run deep into ground and therefore pipe failure does not show up on the ground surface. Bearing this in mind one can expect that a diagnostic system, supporting leakage finding would be very useful, especially on an industrial area with coal mining, where leakages are often encountered. Additionally, traditional methods of leakage finding, based on leakage noise detecting and analysing, are less efficient with, nowadays very popular, plastic pipes, which are poor sound conductors.

In fact, mathematical dependencies between flow and pressure loss in a pipe are known, so it is possible to use it for leakage detection. However theoretically, if we knew water consumption in all the points of network, it would be possible to calculate pressure and flow in any required point of the network. Comparison of calculated and measured parameters could allow finding leakages and other causes of water loss.

Moreover, to establish such a kind of a monitoring system it is necessary to measure "on line" all legal water consumption. Although it is possible to decrease number of inputs to the water supply system (and then decrease number of measurement points needed), it must result in much worse accuracy of the monitoring system. That is why this idea of monitoring system, based on certain model, is not quite good enough for practical implementation.

2. Problem description

2.1. Concept of monitoring system

To avoid necessity of using measuring system which is big, complex and spread out at significant area of the country, with big number of on-line measuring points, the concept of diagnostic system, which uses approximate approach for modeling the pipeline network and recognizes a leak of water was suggested [12]. The idea of this system is based on methods known from model-based process diagnostics where a model of the object being monitored is used for fault detection [6]. Taking measuring flow and/or pressure in chosen points on pipeline networks, the diagnostic system based on appropriate trained artificial neural network (ANN), would suggest if any leakage exists, and where it is possibly located. The structure of the developed diagnostic system is depicted in Fig. 1.

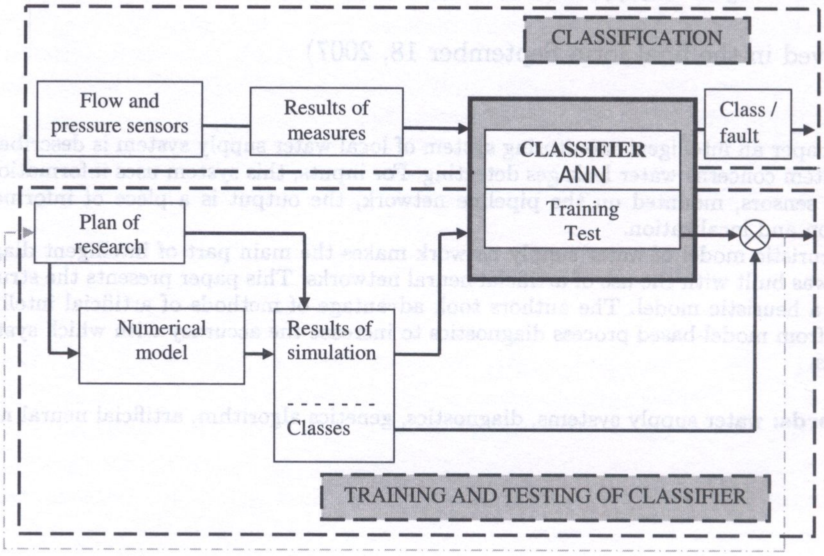

Fig. 1. The structure of the developed diagnostic system

For practical verification of the suggested method, in collaboration with the local water supply system holder, a prototype diagnostic system is being developed. The considered system will work within one district of town. The monitored network has about 25 km of pipelines with different diameters and supplying about one thousand of water consumers.

2.2. Numeric model of water supply network

Since the artificial neural network was planned to be used as a model of the real network, the problem of preparing training data should have been solved. For it is difficult and inconvenient to simulate leakage and collect data from real object, the numerical model of the network was built. In order to build this model of the EPANET simulation environment [4] was used. Running this model it was possible to calculate flows and pressure in all the points of the network with a leakage located in any point of it or without leakage at all.

To describe temporary water consumption for every water consumer, an accountant data was used. For each user an average consumption was calculated. To describe consumption changes during all day, a daily time pattern consumption, described in [7] was established. To simulate random changes of water consumption, this consumption calculated for each time and each user was randomly changed within the range ±20%. In this way, simulations for all days was differ from each other.

3. OPTIMIZATION OF LOCATION OF THE SENSORS

3.1. The proposed method of optimization

The first problem which should be solved to apply this conception in practice is to find the best localization of sensor which should be installed on water pipeline. In the first stage of the diagnostic system creation, because of economical and technical reasons, the owner of the supply system decided to limit the number of used flow sensor to six. The number of pressure sensors, which are cheaper and much easier to mount, was not limited.

The task that must be solved is an optimal choice of a location of 6 sensors in the water supply network. The values of pressures and flows measured by these sensors make up input vector of the neural networks, which recognize a leak of water and its localization.

To find the best location of sensors an optimization with genetic algorithm was used. The algorithm of the solution to the problem is shown below.

1. Choose a few dozen of measuring points.
2. Randomly choose a few dozen of points in which water leaks will be simulated. The ANN classifier will recognize these points.
3. Generation of data (training and test) with the use of numerical pipeline network model.
4. Represent the problem variable domain as a chromosome of a fixed length. Define a fitness function. Randomly generate an initial population of chromosomes.
5. Calculate the fitness of each chromosome. As the measure of chromosome's fitness the efficiency measure of the ANN classifier was used (correct classification ratio [2]).
6. Typical operation for GA (selection, crossover, mutation) to create a new population.
7. Go to point 5, and repeat the process until the termination criterion is satisfied.
8. Verification of usefulness of the ANN classifier chosen in point 7 with the use of a new set of data.

3.2. Generation of data

Because of this huge number of possibilities, permitted sensor localizations were reduced to main junction only — where the main pipes crossed. In all main junctions one potential pressure sensor and flow sensors for each connected pipes were localized. It provided 16 pressure and 45 flow sensors.

As a potential leakage points a 33 location were chosen randomly. The points were spread on all network. The localization of measuring points and simulated points of water leaks is shown in Fig. 2.

Calculations referred to all potential sensor pressures and flows for each leakage. Simulations were repeated for every hour for 30 days.

In this way 24480 sets of training data, for different leakages and hours of the day were collected. Each set contains data from all chose sensor localizations (16 values of pressure and 45 values of flow), and information about where leakage was (or no leakage information). This data was used for training and testing different ANNs with any subsets of initial set of sensors as an input and leakage localization as output of classifiers.

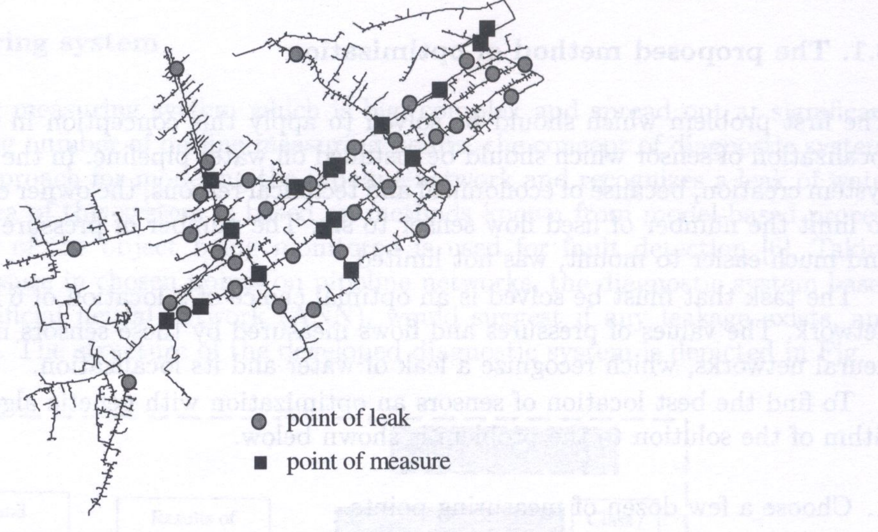

Fig. 2. Localization of measuring points and simulated points of water leaks

3.3. Verification of usefulness of data

In the first step, the usefulness of data to build ANN classifier was checked. To check data usefulness, three multi classes classifiers were built. The input vectors were different for each classifiers and contained:

$m = 61$ values of pressures and flows of water,

$m = 45$ values of flows of water,

$m = 16$ values of pressures of water.

To build classifier a two-layer network, with tan-sigmoid transfer function in the hidden layer and a log-sigmoid transfer function in the output layer was applied. It makes a useful structure for classification problems [11]. During the research a *Neural Network Toolbox* of MatLab was used.

The network had $n = 34$ output neurons because there are 33 points of leak. The 34-th output represents state of the water network without leak. On the basis of information available in books [9], concerning the recommended number of neurons, one used $k = 3m$ neurons in the hidden layer. Before training the preprocessing routine procedure was used to scale the inputs so that they are in the range $[-1, 1]$.

The data were to divide into training and test subsets. One-fourth of the data for the test set, and three-fourths for the training set. The backpropagation training algorithms was used for training. The training was stopped after 300 iterations.

The result of training refers to the neural networks, which recognized leaks. In the Table 1 the results of research of three classifiers are depicted. The best results one could obtained in case of classifier which had input vector composed only from values of water flows in pipes.

Table 1. The results of verification of usefulness of data

Input vector include:	Number of inputs	Number of outputs	Correct classification ratio
pressures and flows	61	34	87.8 %
pressures	16	34	67.4 %
flows	45	34	92.8 %

From the economical point of view the usage of only pressure sensors would be the best solution, because it is easy and cheaper to install a pressure sensor in a pipe than a flow sensor. But during the preliminary research it was observed that the changes of pressures in measuring points of network both with leakage and without leakage were very small and because of measurement inaccuracy, these small pressures changes can be practically insignificant.

The number of pressure sensors, which are cheaper and much easier to mount, were not limited, but as the first examination showed, they are not useful enough for leakage localization (in the considered network). Basing on the results of research obtained in Section 3 a decision to take only water flows during further research was taken.

4. Genetic algorithm

4.1. The structure of neural classifier

In order to find the best localization for flow sensors, with the use of genetic algorithms, a two-layer network with tan-sigmoid transfer function in the hidden layer and a log-sigmoid transfer function in the output layer were applied. The network had an input vector containing 6 values of flows, 34 output neurons and 18 neurons in the hidden layer.

Practical application of genetic algorithm requires the fitness calculating for each testing chromosome. It was expected that more than 1000 values of correct classification ratios of the ANN classifiers should be calculated. In this case appropriate training algorithm of ANN has a significant influence on efficiency of the research.

The resilient backpropagation (Rprop) training algorithm is one of the fastest algorithms on pattern recognition problems [10]. Its performance also degrades as the error goal is reduced. The memory requirements for this algorithm are relatively small in comparison to the other algorithms.

4.2. Representation of the problem domain as a chromosome

At the beginning we represented the problem domain as a chromosome which consists 34 genes represented by 0 or 1. Each gene represents one point of the model of the water supply network in which pressure and flow of water can be measured. If a gene number i is equal 1 that means that value of flow measured in the point number i is an input to the ANN. As mentioned above, it was assumed that the number of measuring points is equal 6. It causes each chromosome contains 6 genes represented by 1 and 28 genes represented by 0.

At the beginning population of $N = 50$ chromosomes was chosen. As a fitness function, the efficiency measure of the ANN classifier (correct classification ratio) [2] was used. During this research the roulette wheel selection [8], one of the most commonly used chromosome selection techniques, was applied.

4.3. Crossover and mutation operators

Because each chromosome ought to contain exactly 6 genes represented by 1, typical crossover and mutation operators were modified for using in this described problem.

At first, the crossover operator randomly chooses 3 crossover points where two parent chromosomes 'break', and then exchange the chromosome parts. As a result, two new offspring's are created. If number of genes represented by 1 is different then 6, the chromosome is deleted and crossover operation is repeated.

The mutation operator changes the value of a randomly selected gene in a chromosome. After first operation chromosome contains 5 or 7 genes represented by 1.

In the next step, if chromosome contains 5 genes represented by 1, the mutation operator randomly selects gene in a chromosome which value is 0 and changes its value for 1. And accordingly, if a chromosome contains 7 genes represented by 1, the mutation operator randomly selects gene in a chromosome which value is 1 and changes its value for 0.

4.4. Results

During the research crossover with the probability p_c equal to 0.7 and mutation with the probability p_m equal to 0.01 was used. The values chosen for p_c and p_m are fairly typical in GAs [3]. A common practice in GAs is to specify the number of generation. It is assumed that the desired number of generations is 50 while the research.

The most serious problem in the use of GAs is connected with the quality of the results, in particular whether or not an optimal solution is reached. It depends on the number of chromosomes in population and the value of p_m . Figure 3 shows plots of the best and average values of the fitness function across 50 generations (performance graphs). The best and average curves represented here are typical for GAs [7]. The curves rise rapidly at the beginning of the run, and when the population converges on the nearly optimal solution, flatten at the end.

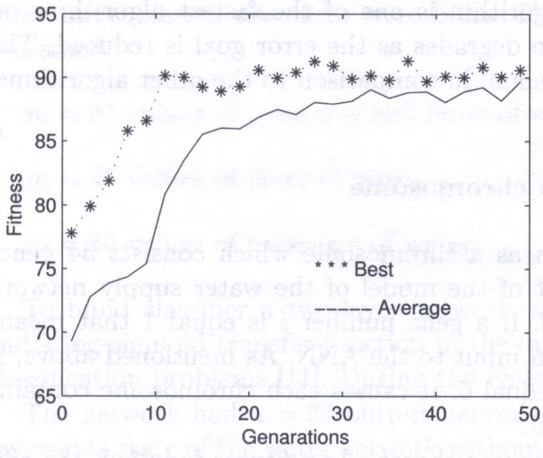

Fig. 3. Performance graphs created in a population of 50 chromosomes, mutation rate $p_m = 0.01$

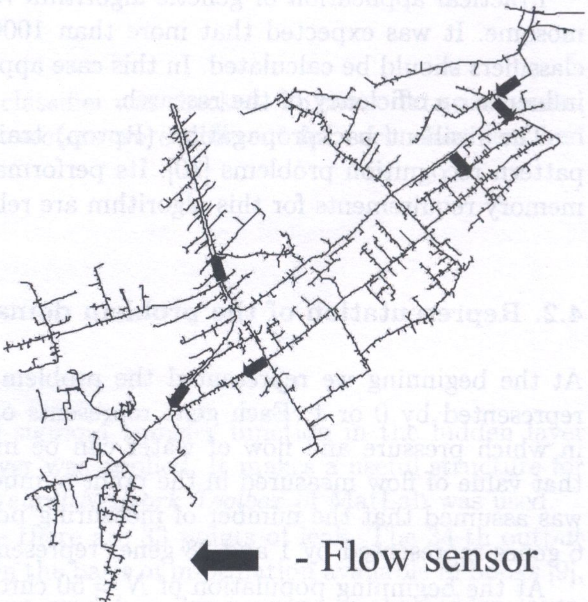

Fig. 4. The chosen flow sensor location

In the Table 2 the obtained results of research are presented. The best result of fitness function (a correct classification ratio) achieved in case of classifier which had input vector composed from 6 values of water flows in pipes was equal to 90.8%. In comparison with the result obtained in the case of classifier which had input vector composed from 45 values (correct classification ratio was equal to 92.8%) the final result of research one can consider as satisfactory.

Table 2. The best result of fitness function

	Value	Generation number
Max. fitness	90.8 %	37
Max. average fitness	89.4 %	50

The genetic algorithm chose the best subset of six sensors. The chosen location of six flow sensors was shown on Fig. 4.

5. Optimization of approximate model of water supply network

As it was described above, the ANN as a model of water supply network was established. The state of the network (no leakage, leakage in first point of network, leakage in second point of network, etc) was taken as an output of this ANN.

The described classifier was prepared to distinguish only 33 points of leakage (and leakages located near them). It was necessary to modify ANN classifier (model of water supply network) to recognize leakages that are located in any points of network. Increasing number of simulated points of leakage during ANN training seems to be the simplest solution to do it. But in this way a number of neurons in ANN will be significantly risen.

The problem was to decide how the potential leakage location should be pointed. It was decided to divide the network into some separated areas (zones) and point only some area when leakage is located.

Two different divisions of pipeline network ware done with use different criteria of division. In the first division the network was partitioned into 39 zones depicted in Fig. 5. Main junctions of pipeline network were central points of each zones.

The achieved results show that in the most cases ANN pointed proper zones. But many leakages, located in absolutely different places were not distinguished. Figure 6 depicts histogram of obtain errors. The numbers on X-axis means: 0 – proper zone was pointed, 1 – nearest zone to the proper zone was pointed, 2 – next to nearest zone was pointed, etc.

During the second division the network was partitioned into 22 zones depicted in Fig. 7. Main junctions of pipeline network ware a borders points of zones (neighbors areas were separated by nodes of water supply system).

Obtained results of research, values of a correct classification ratio for each zone, were presented in Fig. 8. Leakages located in most zones were pointed good enough, but in the case of few zones the results appeared poor.

Fig. 5. The first division of pipeline network

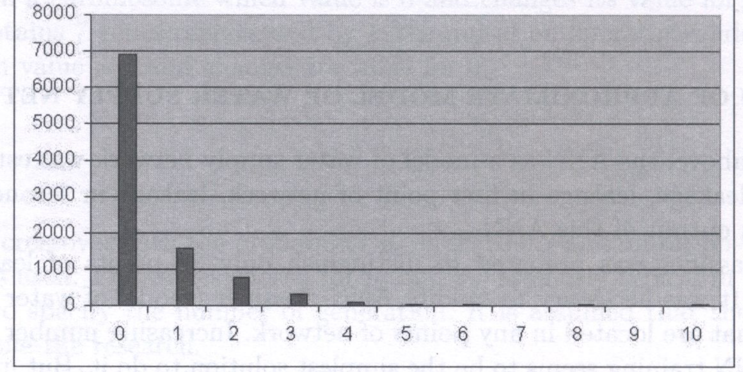

Fig. 6. Histogram of errors (10.000 trials)

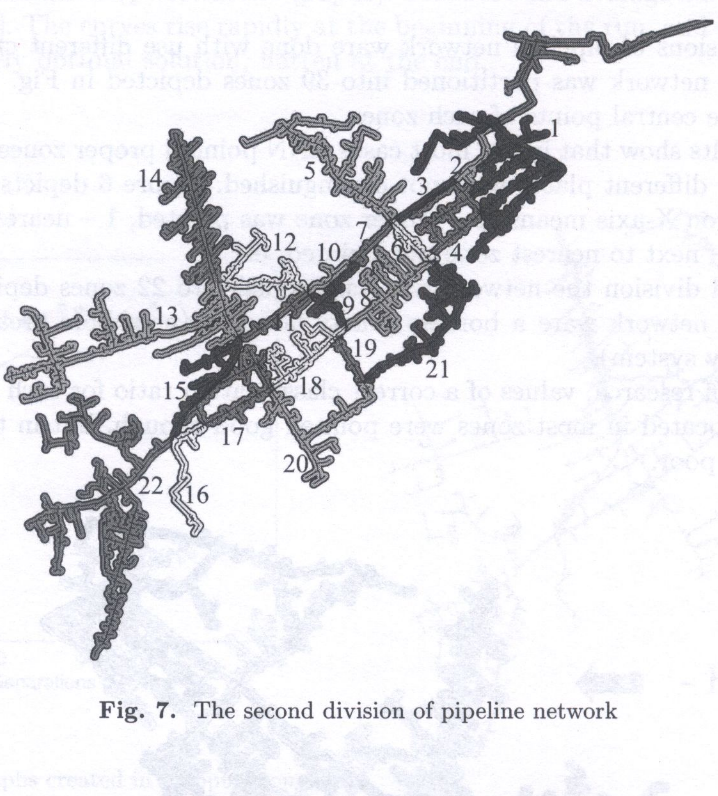

Fig. 7. The second division of pipeline network

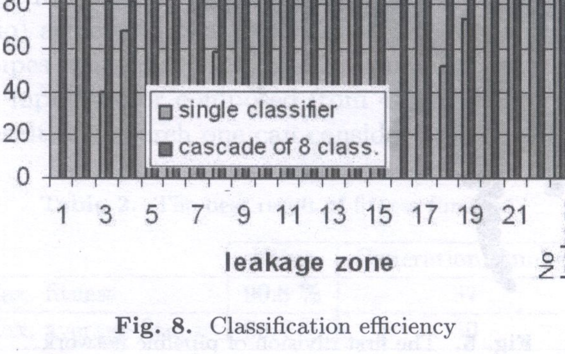

Fig. 8. Classification efficiency

Obtained results shown, that leakages located in most zones were pointed good enough, but in the case of few zones the results appeared poor.

To improve this situation instead one ANN (multi classes classifier) a cascade of ANNs was applied. For seven zones which were not "recognized" good enough, a separate ANN (binary classifiers, which recognizing leakage in distinguish zone) was prepared and trained. When main classifier was not able to recognize the state of the network satisfactory, binary classifiers completed diagnosis. The comparison of results obtained for one ANN and the cascade of ANNs was shown in Fig. 8.

In Fig. 9 histogram of classification errors for ANNs cascade was presented — it is essential, that system points proper area of leakage localizations or the nearest area at least.

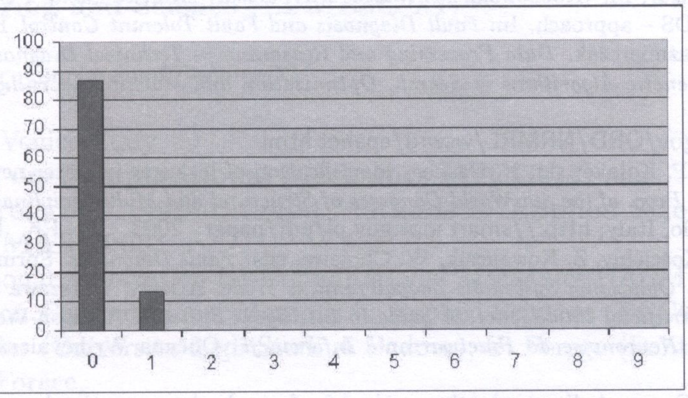

Fig. 9. Histogram of errors for ANNs cascade

6. Conclusions

Because the water network model used to collecting data was not fully calibrated, presented results should be considered as a preliminary research. At the begin, there were not any sensors installed on the network, and it was necessary to say where they should be located. So in first stage, the model was build, and location of sensors was selected as it was described. The measurements from each installed sensor was compared with simulation results. During paper was being prepared, three sensor was installed, and they measured flows corresponded to simulated ones. In the nearest future the process of flow sensor installation will be finished, so it will be possible to finish calibration and carry out first practical tests of the monitoring system.

The simulation results show the uselessness of the water pressure measurements to the water leaks detection. It may be caused by proportion between flow value and pipelines diameters. In this specific network, water flows are rather small and causes only insignificant pressure changes. In other networks, especially with long transit pipelines, this inference may not be true.

In comparison with other methods [1, 5], proposed one has many advantages. The method pointing where the sensors should be located. To build classifier, there is no need to linearize dependence between flow and hydraulics resistance. Using ANN as a classifier, we optimized model globally, not as a sum of local models. There is no need to simplification physical model of water network to decrease number of pipes and junctions, before classifier's building.

7. Further system development

In the next step, after collecting data from installed sensors, it will be possible to improve the system to detect less intensive leakages. The idea is to predict the future flow measured by the given sensor with few hours horizon time. Recurrent errors of prediction can help to determine an additional flow, connected only with potential leakage, and not connected with water consumption.

ACKNOWLEDGMENTS

The research presented in the paper has been partially supported by EFS under Sectoral Operational Programme — Improvement of the Competitiveness of Enterprises, years 2004–2006 under grant "System monitorowania i zarządzania eksploatacją sieci wodociągów i kanalizacji miasta Rybnik" and the Ministry of Education and Scientific Research/Information under grants No. 4 T07B 018 27.

REFERENCES

[1] A. Borowa, K. Mazur, M. Grochowski, M.A. Brdyś, K. Jezior. MultiRegional PCA for leakage detection and localization I DWDS - approach. In: *Fault Diagnosis and Fault Tolerant Control.* EXIT, Warsaw, 2007.

[2] W. Cholewa, J. Kaźmierczak. *Data Processing and Reasoning in Technical Diagnostics.* WNT, Warsaw, 1995.

[3] D.E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison-Wesley, Reading, 1989.

[4] http://www.epa.gov/ORD/NRMRL/wswrd/epanet.html

[5] J. Holnicki-Szulc, P. Kołakowski, N. Nasher. Identification of leakages in water networks — Virtual distortion method approach. *Proc. of the 5th World Congress of Structural and Multidisciplinary Optimization,* May 19–23, 2003, Lido di Jesolo, Italy; http://smart.ippt.gov.pl/pdf/paper_2003_JH_PK_NN.pdf.

[6] J. Korbicz, J.M. Kościelny, Z. Kowalczuk, W. Cholewa, eds. *Fault Diagnosis.* Springer, 2004.

[7] W. Mielcarzewicz. *Obliczanie Systemów Zaopatrzenia w Wodę.* Arkady, Warszawa 2000.

[8] M. Negnevitsky. *Artificial Intelligence. A guide to Intelligent Systems.* Addison-Wesley, 2004.

[9] S. Osowski. *Sieci Neuronowe do Przetwarzania Informacji.* Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa, 2000.

[10] M. Riedmiller, H. Braun. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. *Proceedings of the IEEE International Conference on Neural Networks,* 1993.

[11] L. Rutkowski. *Metody i Techniki Sztucznej Inteligencji.* PWN, Warszawa, 2005.

[12] R. Wyczółkowski, W. Moczulski. Concept of intelligent monitoring of local water supply system. *Materials of AI-METH 2005 - Artificial Intelligence Methods,* November 16–18, 2005, Gliwice, Poland.