MIDDLE POMERANIAN SCIENTIFIC SOCIETY OF THE ENVIRONMENT PROTECTION
ŚRODKOWO-POMORSKIE TOWARZYSTWO NAUKOWE OCHRONY ŚRODOWISKA

Annual Set The Environment Protection
Rocznik Ochrona Środowiska

Volume/Tom 20. Year/Rok 2018 ISSN 1506-218X 528-537

Assessing the Diameters of Water Pipes Using the k-Nearest Neighbours Method in the Calculations of Water Distribution Systems

Jacek Dawidowicz[*], Wojciech Kruszyński[*], Dariusz Andraka[*], Andrzej Czapczuk[**]
[*]Białystok University of Technology
[**]F.B.I. TASBUD S.A., Science and Research Division

1. Introduction

Water distribution systems provide water in cities and in rural areas. The basic element through which water reaches the consumer are the water pipes, hence their correct design is extremely important. The choice of pipe diameter requires hydraulic calculations. Computer programmes may choose diameters, but usually, it is the designer's task. The aim of the work is to assess the correctness of the selection of water-pipe diameters assuming that the flows on the calculated sections have been determined correctly. This paper proposes a classifier, based on the k-Nearest Neighbours method, which, on the basis of a reliable flow, will assess the appropriateness of the diameter chosen. For this paper, calculations with a steady-state model, were carried out during the maximum water consumption times of the aforementioned consumers. Flows for these conditions are the basis for gauging the dimensions of the diameters of the water supply pipes (Knapik & Bajer 2010, Lansey & Mays 2000).

2. An overview of the application of artificial intelligence methods in the design of water distribution systems

The design and calculation of technical objects requires knowledge-based activities that are difficult to describe, using the classical computational algorithm. Artificial intelligence methods could become a great support in this type of task. Artificial neural networks, expert systems and the *k*-Nearest Neighbours method, or the Support Vector Machine are worthy of mention here. Computer programmes, used in calculations, may support the designer in his/her task, to some extent, through appropriate suggestions and prompts. In this article, the possibility of using the *k*- Nearest Neighbours method, to assess the diameters of water pipes, is analysed. The above method was used to assess pressure losses during the hydraulic calculations of water distribution systems (Biedugnis & Czapczuk 2018). In this article, (Oliveira & Boccelli 2017) the *k*- Nearest Neighbours method was used to forecast short-term, water demand periods. The *k*-NN approach is a pattern-recognition algorithm, where the predicted values are directly determined by previous observations which were most similar.

A review of artificial intelligence methods, in the design and operation of water supply systems, is provided in the article (Czapczuk et al. 2015). In the calculation of hydraulic water distribution systems, artificial neural networks are usually used. Suggestions for using them in the calculations themselves (Czapczuk et al. 2017, Brkić & Ćojbašić 2016, Besarati et al. 2015, Bubtiena et al. 2011), in taring simulation models (Meirelles et al.2017), as well as to evaluate the results of the hydraulic calculations obtained, (Dawidowicz 2017) are to be found. A method for estimating the diameter of water pipes using artificial neural networks of the multilayer perceptron type is presented in the paper (Dawidowicz 2018). One can also cite examples of expert systems used in the design stage (Dawidowicz 2012) as well as the actual operation (Changa et al. 2011). The application of RBF neural networks for the assessment of the water flow rate in the pipework is presented in the paper (Czapczuk & Dawidowicz 2018).

3. Introduction to the k-Nearest Neighbours Method

The k-Nearest Neighbours algorithm (*k*-NN) is used to predict the value of a random variable or classification (Triantaphyllou & Felici 2006). The method belongs to the group of lazy algorithms, that is, those algorithms which do not form the internal representation of the training data but look for a solution, only when a test pattern appears. In this work, the *k*-NN method was used to classify the diameter of water pipes. The task is to classify a new object, based on information as to which object or objects the new object is adjacent to. An important parameter in the method, is the value of *k*, that is, the number of teaching examples in the immediate environment. In the case of classification, the method takes into account, the values of most of the examples in the neighbourhood, that is, it sets the value of a new example by voting.

Choosing a neighbourhood, that is, the value of the *k* parameter, is essential. With a low *k* value, there will be a large variation in classification. Higher *k* values allow areas to be divided up, smoothly and noise to be removed; however, they can also lead to errors in the classification of rarer patterns. The problem then arises of generalising them, correctly. On the basis of the analyses, a *k* value must be chosen to minimise the likelihood of mis-classification. The *k*-Nearest Neighbour method proposes an optimal *k* value, based on the Cross-Check method.

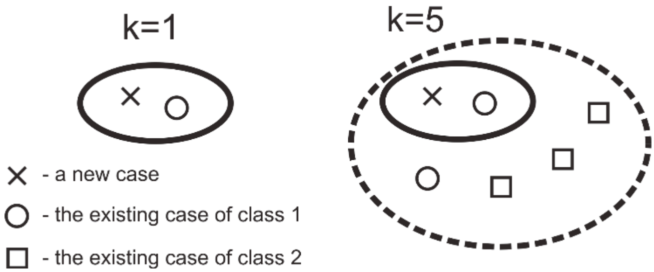

Fig. 1. Principle of the *k*-NN method for different neighbourhood values
Rys. 1. Zasada działania metody k-NN dla różnych wartości sąsiedztwa

If we consider $k = N$, where N is the number of all the elements of the set of teaching patterns, then the result of the classification will always be determined by the class most-represented in this set of teaching examples.

The Euclidean measure is usually used to evaluate the distance between the points describing the teaching examples:

$$D(x,p) == \sqrt{(x-p)^2} \tag{1}$$

where:
x – the new case to be classified,
p – one of the teaching cases.

4. Training examples for the assessment of the diameters of water lines by the k-Nearest Neighbours method

To use the *k*-Nearest Neighbours method, a set of teaching patterns is required, containing input data which define the object and are stored in vector X, as well as target values *d*. This corresponds to the teaching strategy under supervision.

A set of teaching examples was obtained on the basis of the hydraulic calculations of examples of water distribution systems. Diameters for the pipelines were chosen for conditions of maximum water consumption Q_{hmax}. Hydraulic calculations were made, taking into account the internal diameters of the pipelines. Each calculation option was checked and corrected where calculation irregularities appeared. Due to the large amount of data, a procedure was developed, in order to convert the results of the calculations of individual segments to an appropriate format and then to save them as a set of teaching examples.

11961 training examples were obtained containing the input variable in the form of a nominal flow, through water supply line $Q_{m,}$ corresponding to the output variable DN, that is, the nominal diameter.
Nominal diameters were adopted as follows:

- DN90, DN110, DN160, DN225 for PE100 polyethylene pipes of the SDR17 series (EN 12201-2:2011),
- DN250, DN300, DN350, DN400, DN450, DN500 for ductile iron pipes (EN 545: 2010).

The amalgamated teaching examples have been divided into two subsets: teaching (75%) and testing (25%). The test group in the teaching process is not used at all; rather, it is intended for the independent determination of the accuracy of the *k*-NN method.

5. Development of the k-Nearest Neighbours model to assess the diameter of water pipes

On the basis of the set of teaching examples, a model was constructed and the diameters of the water pipes were classified using the k-Nearest Neighbours method, using various neighbourhood values.

The calculations for the neighbourhood in the scope of k = 1-10 were made in the work. Tables 1-3 summarise the results of the classification of the pipe diameters for the neighbourhood, where $k = 1$, $k = 5$ and $k = 10$. The Euclidean measure was used to calculate the distance between teaching examples; the best results are obtained in the vicinity of $k = 5$.

Detailed classification results are included in the Confusion Matrix. This is a square matrix, in which information, as to which classes the individual examples actually belong, are given in rows, as well as information as to which classes they were classified into, by the classifier, are given in columns. The diagonal contains examples that were correctly categorised, while those located beyond the diagonal were classified incorrectly. At the same time, examples lying beyond the diagonal, point to those classes which were incorrectly classified.

The matrix of classification errors, established by the *k*-Nearest Neighbours method, for neighbourhood $k = 5$, is shown in Table 4. Analysis of the results indicates the high quality of the classifier, which can select or assess the diameter of the pipeline, on the basis of the flow.

Table 1. Summary of the results for the classification of $k = 1$, pipe diameters
Tabela 1. Podsumowanie wyników klasyfikacji średnic rurociągów dla k = 1

	In total	Accurate	Incorrect	Relevant (%)	Invalid (%)
DN110	651	632	19	97.1	2.9
DN160	456	442	14	96.9	3.1
DN225	383	378	5	98.7	1.3
DN250	365	354	11	97.0	3.0
DN300	304	296	8	97.4	2.6
DN350	326	310	16	95.1	4.9
DN400	204	203	1	99.5	0.5
DN450	200	200	0	100.0	0.0
DN500	101	101	0	100.0	0.0

Table 2. Summary of the results for the classification of k = 5, pipe diameters
Tabela 2. Podsumowanie wyników klasyfikacji średnic rurociągów dla k = 5

	In total	Accurate	Incorrect	Relevant (%)	Invalid (%)
DN110	652	635	17	97.4	2.6
DN160	457	445	12	97.4	2.6
DN225	380	377	3	99.2	0.8
DN250	370	360	10	97.3	2.7
DN300	296	295	1	99.7	0.3
DN350	323	309	14	95.7	4.3
DN400	211	207	4	98.1	1.9
DN450	200	200	0	100.0	0.0
DN500	101	101	0	100.0	0.0

Table 3. Summary of the results for the classification of k = 10, pipe diameters
Tabela 3. Podsumowanie wyników klasyfikacji średnic rurociągów dla k = 10

	In total	Accurate	Incorrect	Relevant (%)	Invalid (%)
DN110	651	632	19	97.1	2.9
DN160	461	443	18	96.1	3.9
DN225	381	375	6	98.4	1.6
DN250	364	357	7	98.1	1.9
DN300	300	297	3	99.0	1.0
DN350	315	306	9	97.1	2.9
DN400	217	212	5	97.7	2.3
DN450	201	200	1	99.5	0.5
DN500	100	100	0	100.0	0.0

Table 4. Matrix of errors in the test subset for the k = 5 neighbourhood
Tabela 4. Macierz pomyłek dla sąsiedztwa k = 5 dla podzbioru testowego

	DN 110	DN 160	DN 225	DN 250	DN 300	DN 350	DN 400	DN 450	DN 500
DN110	635	5	0	0	0	0	0	0	0
DN160	17	445	0	0	0	0	0	0	0
DN225	0	7	377	1	0	0	0	0	0
DN250	0	0	3	360	0	0	0	0	0
DN300	0	0	0	9	295	3	0	0	0
DN350	0	0	0	0	1	309	4	0	0
DN400	0	0	0	0	0	11	207	0	0
DN450	0	0	0	0	0	0	0	200	0
DN500	0	0	0	0	0	0	0	0	101

6. Conclusions

The above method for classifying the diameter of water pipes was developed for the purposes of calculating hydraulic water distribution systems. Based on the data from hydraulic calculations, the *k*-NN algorithm can evaluate the diameter of the ducts on individually calculated sections of the water supply network and can then propose an appropriate value or accept the existing value. The *k*-NN method obtained, shows a high accuracy index in the classification of the diameters of the pipes in the $k = 5$ neighbourhood. It should be remembered, however, that there may be cases of incorrect classification, hence the solution should be treated as an additional tool in the selection or assessment of the diameters adopted. The final decision on the selection of the diameter, is down to the person performing the calculations.

The method is also quite simple from the point of view of theoretical assumptions; this is undoubtedly one of the advantages of this solution. The method for classifying water pipe diameters may only be used, however, for the automatic selection of pipe diameters, prior to undertaking the appropriate hydraulic calculations or evaluating the results obtained, because the target levels of pressure losses in the pipelines have not been achieved. Undoubtedly, this kind of module would be an interesting addition to any computational programme.

The analyzes were carried out as part of the work No S/WBiIŚ/02/2014 and financed from the funds of the Ministry of Science and Higher Education

References

Besarati, S. M., Myers, P. D., Covey, D. C., & Jamali, A. (2015). Modeling friction factor in pipeline flow using a GMDH-type neural network. *Cogent Engineering, 2(1)*, 1-14.

Biedugnis, S., & Czapczuk, A. (2018). The application of the 'K-nearest neighbour' method to evaluate pressure loss in water supply lines. *Technical Transactions, 115(1)*, 141-149, doi: 10.4467/2353737XCT.18.011.7962.

Brkić, D., & Ćojbašić, Ž. (2016). Intelligent flow friction estimation. *Computational intelligence and neuroscience 2016*, 1-10.

Bubtiena, A.M., Elshafie, A.H., & Jafaar, O. (2011). Application of artificial neural networks in modeling water networks. *Proceedings IEEE 7th International Colloquium on Signal Processing and Its Applications,* Penang, Malaysia, 50-57, doi: 10.1109/CSPA.2.011.5759841.

Changa, Ni-Bin, Pongsanonea, N.P., & Ernestb, A. (2011). Comparisons between a rule-based expert system and optimization models for sensor deployment in a small drinking water network. *Expert Systems with Applications, 38(8)*, 10685-10695.

Czapczuk, A., & Dawidowicz, J. (2018). The Application of RBF Neural Networks for the Assessment of the Water Flow Rate in the Pipework. *Proceedings of the 2nd International Conference on Artificial Intelligence: Technologies and Applications (ICAITA 2018), Advances in Intelligent Systems Research*, *146*, Atlantis Press, 46-49, doi:10.2991/icaita-18.2018.12.

Czapczuk, A., Dawidowicz, J., & Piekarski, J. (2015). Metody sztucznej inteligencji w projektowaniu i eksploatacji systemów zaopatrzenia w wodę. *Rocznik Ochrona Środowiska, 17(2)*, 1527-1544 (in Polish).

Czapczuk, A., Dawidowicz, J., & Piekarski, J. (2017). Application of Multilayer Perceptron for the Calculation of Pressure Losses in Water Supply Lines, *Rocznik Ochrona Środowiska, 19*, 200-210.

Dawidowicz, J. (2012). System ekspertowy do oceny układu systemu dystrybucji wody sporządzony za pomocą wnioskowania indukcyjnego. *Rocznik Ochrona Środowiska*, *14*, 650-659 (in Polish).

Dawidowicz, J. (2017). Evaluation of a pressure head and pressure zones in water distribution systems by artificial neural networks. *Neural Computing & Application*, doi:10.1007/s00521-017-2844-8.

Dawidowicz, J. (2018). A Method for Estimating the Diameter of Water Pipes Using Artificial Neural Networks of the Multilayer Perceptron Type. *Proceedings of the 2nd International Conference on Artificial Intelligence: Technologies and Applications (ICAITA 2018), Book Series: Advances in Intelligent Systems Research, 146*, Atlantis Press, 50-53, doi:10.2991/icaita-18.2018.13.

Knapik, K., & Bajer, J. (2010). *Wodociągi*. Podręcznik dla studentów wyższych szkół technicznych. Wydawnictwo Politechniki Krakowskiej, Kraków (in Polish).

Lansey, K., & Mays, L.W. (2000). Hydraulics of water distribution systems. In Mays, L.W. (Ed.) *Water distribution systems handbook,* McGraw-Hill, New York.

Meirelles, G., Manzi, D., Brentan, B., Goulart, T., & Luvizotto, E. (2017). Calibration Model for Water Distribution Network Using Pressures Estimated by Artificial Neural Networks. *Water Resources Management, 31(13)*, 4339-4351. doi:10.1007/s11269-017-1750-2

Oliveira, P.J.A., & Boccelli, D.L. (2017). k-Nearest Neighbor for Short Term Water Demand Forecasting. *Proceedings World Environmental and Water Resources Congress 2017: Hydraulics and Waterways and Water Distribution Systems Analysis*, 501-510.

Triantaphyllou, E., & Felici, G. (Eds.). (2006). *Data mining and knowledge discovery approaches based on rule induction techniques, 6*, Springer Science & Business Media.

Metoda oceny średnic przewodów wodociągowych za pomocą metody k-Najbliższych Sąsiadów w obliczeniach systemów dystrybucji wody

Streszczenie

Systemy dystrybucji wody dostarczają wodę w miastach i na terenach wiejskich. Podstawowym elementem, przez który woda dociera do odbiorców są przewody wodociągowe, stąd niezwykle istotne jest ich poprawne zaprojektowanie. Dobór średnic rurociągów wymaga przeprowadzenia obliczeń hydraulicznych. Programy komputerowe mogą automatycznie dobierać średnice, ale najczęściej zadanie to należy do projektanta. Obecnie opracowuje się metody, które wspomagałyby projektantów w realizacji powyższych zadań. W niniejszej pracy zaproponowano klasyfikator oparty na metodzie k-Najbliższych Sąsiadów (k-NN), który na podstawie przepływu miarodajnego Q_m będzie oceniał poprawność dobranej średnicy. W tym celu sporządzono 11961 przykładów uczących zawierających zmienną wejściową w postaci przepływu miarodajnego Q_m oraz odpowiadającą mu zmienną wyjściową zdefiniowaną jako średnica nominalna DN. Na podstawie zestawu przykładów uczących skonstruowano klasyfikator za pomocą metody k-Najbliższych Sąsiadów, stosując różne wartości sąsiedztwa. Uzyskana metoda k-NN pokazuje wskaźnik wysokiej dokładności w klasyfikacji średnic rur dla wartości sąsiedztwa k = 5.

Abstract

Water distribution systems provide water in cities and in rural areas. The basic element through which water reaches the consumer are the water pipes, hence their correct design is extremely important. The choice of pipe diameter requires hydraulic calculations. Computer programmes may choose diameters, but usually, it is the designer's task. This paper proposes a classifier, based on the k-Nearest Neighbours method, which, on the basis of a reliable flow, will assess the appropriateness of the diameter chosen. In the work 11961 training examples were obtained containing the input variable in the form of a nominal flow, through water supply line Q_m, corresponding to the output variable DN. On the basis of the set of training examples, a model was constructed and the diameters of the water pipes were classified using the k-Nearest Neighbours method, using various neighbourhood values. The *k*-NN method obtained, shows a high accuracy index in the classification of the diameters of the pipes in the k = 5 neighbourhood.

Słowa kluczowe:
system dystrybucji wody, obliczenia hydrauliczne, metoda k-Najbliższych Sąsiadów, średnice rurociągów

Keywords:
water distribution system, hydraulic calculations, k-Nearest Neighbours Method, diameters of water pipes