

A Decision Support System for the Prediction of Wastewater Pumping Station Failures Based on CBR Continuous Learning Model

Bruno Trstenjak

Department of Computer Engineering,
Polytechnic of Medimurje in Cakovec,
Cakovec, Croatia
btrstenjak@mev.hr

Bruno Palasek

Technical Department,
Medimurje Vode d.o.o.,
Cakovec, Croatia
bruno.palasek@medjimurske-vode.hr

Jurica Trstenjak

Department of Computer Engineering,
Polytechnic of Medimurje in Cakovec,
Cakovec, Croatia
jtrstenjak@mev.hr

Abstract—Nowadays the communities are facing the problem of waste and wastewater. While wastewater systems have become more complex, the need for development of sustainable solution for wastewater management emerged. Therefore, the development of a Decision Support System (DSS) for wastewater disposal management became necessary. This paper presents a new DSS for predicting the failure of wastewater pumping stations, the system architecture and its implementation. The prediction model is based on the Case Based Reasoning (CBR) classification method. The standard CBR classification technique has been upgraded with an algorithm for continuous learning. The paper describes the system structure, its connection to the wastewater system, the internal processes involved in the prediction process and the implemented algorithm for continuous learning. Furthermore, the features used in the prediction are indicated along with the achieved results and the method of results evaluation. The test and obtained results indicate that the proposed DSS is efficient and capable of providing very good results in the prediction process.

Keywords—case based reasoning; continuous learning; decision support system; prediction; wastewater pumping station

I. INTRODUCTION

Nowadays the problem of waste and wastewater is apparent, as it is finding a way of sufficient waste disposal. In order to achieve maximum reliability of the wastewater system, various decision support systems (DSS) have been developed. A DSS aims at supporting the functionality and optimization of the entire system or achieving individually reliability of the elements of the system. Because of the complexity of the wastewater system, it has become increasingly difficult to find a sustainable solution for wastewater management. Authors in [1] presented different approaches and methods used in the process of developing DSSs. They explore a systematic approach of developing DSSs, which includes analysis of the treatment problem(s), knowledge acquisition and representation, and the identification and evaluation of the criteria controlling the selection of optimal treatment systems. Authors in [2] describe the development and operation of a DSS for waste water disposal management. The systems uses concepts such as components, community context, and data

structures. The authors presented two software modules, design generation module and decision aid module, which give different alternatives in system design process. In the development of advanced wastewater DSS, modern IT technology is often used. Authors in [3] presented a new approach in DSS design by using the Machine Learning technique of specifically supervised Decision Tree Algorithm. This developed system provides the best decision for the reuse of wastewater after treatment. This system is deployed on cloud environment, which provides good manageability and fast computational service. Besides the different solutions that are focused on the management of the entire wastewater system, a variety of DSSs has been developed, focused on wastewater system elements. Authors in [4] presented a data-driven model with a new neural network algorithm for scheduling pumps in a wastewater treatment process, an optimization model to optimize the maintenance and operational schedules. Authors in [5] presented a study that aimed to help decision-makers by developing an intuitionistic fuzzy set (IFS) theory based group multi-attribute decision analysis method. Ten criteria in environment, economy, and technology dimensions were employed to achieve sustainability measurement of wastewater treatment processes. The developed multi-criteria sustainability assessment method allows different experts to conduct sustainability measurement. The results reveal that this method can determine a sustainability sequence of different wastewater treatment processes.

This paper presents a DSS for predicting the failure of wastewater pumping stations, the system architecture and its implementation. The primary objective in this research is focused on assessing the correctness of a wastewater pumping station, to predict the correctness of a wastewater pump as its base elements. The system performs an assessment of the probability of malfunction of a wastewater pumping station for the period of the next week. Such information provides timely intervention and failure prevention for a given wastewater system segment. The DSS implementation is applicable to various types of wastewater pumps and allows easy building and expansion of the system. A universal character system, which can be easily adapted to any wastewater system regardless of their architecture, has been developed.

Corresponding author: Bruno Trstenjak

II. RESEARCH METHODOLOGY AND PROPOSED DSS

Nowadays, many machine learning techniques can be used in the prediction process. Perhaps the most widespread of them is the Neural Networks [6]. In the process of forming the architecture of the wastewater pumping decision support system (WPDSS), CBR was selected as the base technique.

A. Case Based Reasoning (CBR)

The hybrid prediction model is based on the CBR classifier. The principle of CBR is based on solving new problems by observing their similarity with previously solved problems. The CBR method uses a problem-solving approach analogous to the way of problem solving by humans when they recall from past experiences. Each CBR system contains an embedded library of the cases that were resolved in the past. Each case represents a description of the problem with its associated solution. CBR method with a built-in function of similarities tries to find the most similar case from the library. The retrieved cases from the library are used to suggest a solution. If the proposed solution is not satisfactory, the method tries to revise the selected cases and find a new solution. The method adds a new revised case to the case library and thereby expands the knowledge base. The whole execution cycle of the algorithm can be divided into four main steps: Retrieve, Reuse, Revise, and Retain [7]. CBR performs measurements of similarity on local and global level. Local similarity refers to the measurement of similarity between pairs of features. Global similarity refers to comparison of similarity between all the features that make up the object. Measuring similarity can be shown by [8]:

$$\text{Similarity}(T, S) = \frac{\sum_{i=1}^n f(T_j, S_i) \times w_i}{\sum_{i=1}^n w_i} \quad (1)$$

where T indicates the target case, S is a source case, n is the number of features in each case, i the individual feature from 1 to n , f is the similarity function for features i in cases T and S and w presents the importance weighting of feature i .

B. Continuous Learning

Continuous learning is focused on constantly improving the prediction model and achieving higher prediction accuracy [9, 10]. After the prediction of the next week is conducted, the results for the end of the working week of the wastewater system are kept. WPDSS starts the process of comparing the recent results with the prediction results. The WPDSS calculates the overall prediction accuracy and checks if it has fallen below the set threshold. If it is below the set threshold, the process of improving the CBR model starts automatically. In the process of improvement, the implemented algorithm calculates the corrective factor [11]. The corrective factor defines the degree of correction for the weight features values, as shown in (1). This weight change itself can be constant or it can decay as the learning proceeds. The corrective factor can be calculated, as shown in (2)-(3) [12]:

$$\text{Addition, increase: } W_i(t+1) = W_i(t) + \Delta i \frac{F_c}{K_c} \quad (2)$$

$$\text{Addition, decrease: } W_i(t+1) = W_i(t) - \Delta i \frac{F_c}{K_c} \quad (3)$$

K_c and F_c indicate the number of times that a case has been correctly and incorrectly retrieved. The ratio F_c/K_c reduces the influence of the weight update as the number of successful retrievals increases. The value Δi determines the initial weight change. The process of upgrading the classification model can be also started after each weekly period, regardless of the achieved prediction accuracy. This approach achieves fine-tuning of the system to any changes.

C. System Architecture

To achieve the set goals, a web based DSS has been developed. The architecture of the system enables the execution of parallel processes and connection to the existing wastewater system. Implementation and testing of the DSS was carried out on the wastewater system of the Medimurje County, Croatia. Testing was performed on a sample of over forty wastewater pumping stations. The existing wastewater pumping system uses several types of pumps made by two manufacturers: WILO (EMU FA10.94E - 294, T20-4/27H,) and Grundfos (SEV.80.80.110.A.2.51D, SEG.40.09.2.50B, SEG.40.31.2.50B). The diversity of wastewater pumps is the result of their long-term development, due to the system expansion and various reconstructions of wastewater pumping stations. Figure 1 shows the architecture of the new WPDSS and its basic components.

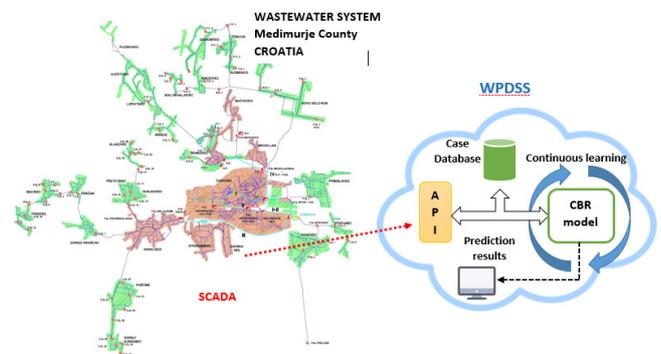


Fig. 1. The WPDSS architecture

The system is linked to the wastewater system and its SCADA (Supervisory Control and Data Acquisition) management system. SCADA is an industrial computer-based control system employed to gather and analyse real-time data to keep track, monitor, and control industrial equipment in different types of industries [3]. The data collected by SCADA are daily sent to the WPDSS through a specially defined procedure. The data sent by SCADA refer to the measured values during the operation of each wastewater pump. SCADA and WPDSS are connected using an application programming interface (API). At the end of the week the WPDSS generates weekly and two-week data cases from these data. The formed data cases are used in the prediction process for the next period. The central component in the WPDSS is the CBR model. The model is based on CBR machine learning technique [13]. For

the purpose of prediction, the standard CBR classification technique has been upgraded with the algorithm of continuous learning. The constant learning algorithm is responsible for the automatic improvement of the system and its CBR prediction model. The WPDSS system uses its case database for its work. After the prediction has been made, the WPDSS enables presentation and analysis. The prediction result informs the user about the likelihood of malfunction of some of the wastewater pumping stations in the period of the next week. At the end of every week, based on the actual state of the wastewater pumping station, the system performs a prediction accuracy analysis. If the system indicates lower prediction accuracy, the system starts redefining certain parameters of the CBR model. All these changes contribute to a better functioning of the prediction model.

III. OVERVIEW OF THE WORK PRINCIPLE

The work of the entire WPDSS system can be represented by a series of internal processes. Figure 2 shows the most

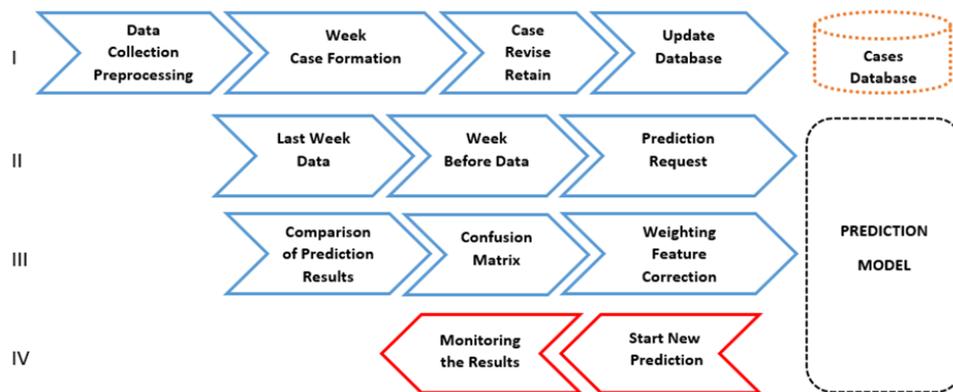


Fig. 2. WPDSS internal processes

The second group (II) represents the processes for preparing data, and forming the cases that will be used in prediction. The WPDSS collects data during the working week and at the end of the week it forms a weekly case. The new prediction process includes the features of the last two weeks of the wastewater stations' works. The internal process merges the data from the last two weeks and sends them to the prediction model. The determined prediction result refers to the next week period.

The third group (III) of internal processes are responsible for continuous learning. One of its novelties is the modification of CBR as a technique for continuous learning. The work philosophy of the modified CBR technique is very similar to the back propagation Neural Network. The network allows corrections of the neurons' weight based on prediction errors. A similar, but not identical, philosophy has been used in this new approach to improve the prediction accuracy of the CBR model. Based on the comparison of the prediction results and the real states of the wastewater stations, the system measures the prediction accuracy for the current week. If the prediction accuracy is lower than the allowable level, the system triggers the internal correction process for new weight values used in (1). After the weight values are corrected, a new CBR

prediction starts. The correction of weight values will not start if the system achieves high prediction accuracy. At the beginning of the system operation, the weights of the features are set to initial neutral value.

The last group (IV) of internal processes consists of the processes responsible for presenting the prediction results. At the end of the week, a new prediction process starts automatically. As the system is implemented in cloud environment via a communication protocol, the result of prediction for each wastewater station is sent to the SCADA system or a remote point. Internal processes allow monitoring of the system performance, analysis of previous prediction results or the data of a new prediction.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Dataset Description

The implementation and testing of the WPDSS was carried out in cooperation with the water system in the Medjmurje County, Croatia. For the purposes of testing the WPDSS, data from 40 wastewater pumping stations were used. The selected wastewater pumping stations are geographically distributed throughout the region. Input prediction data were used in case

format. Each case format consisted of three components: static, dynamic, and environment. The static data refer to the characteristics of the wastewater pumps set by the manufacturer. Dynamic data refer to the measurements of the wastewater pump operation obtained from the SCADA. To all these features, the system adds the environment features. These

features describe the work environment of the wastewater pumping station. The flexibility of the WPDSS allows easily changing the number of the features which are involved in the prediction process. Table I shows the features used in the prediction process.

TABLE I. LIST OF FEATURES USED IN THE PREDICTION PROCESS

No	Static features	Dynamic features	Environment features
1	Manufacturer	Weekly average current (A)	Location of pumping station
2	Year of manufacture	Maximum daily current (A)	Installation depth (m)
3	Rated current (A)	Weekly average flow (l/s)	Number of households
4	Rated current at 3/4 load	Maximum daily flow (l/s)	Altitude of the pumping station
5	Rated current at 1/2 load	Pump alarm was on	Weather conditions
6	Max flow (l/s)	Weekly average number of starts	
7	Rated power (kW)	Maximum daily starts	
8	Max starts per hour (No)	Maximum daily starts per hour	
9	Maximum particle size (mm)	Weekly average working hours	
10	Rated speed (rpm)	Maximum daily working hours	
11	Maximum installation depth (m)	Maximum water-in-oil (%)	
12	Type of impeller	Service period (%)	
13	Maximum operating pressure (bar)		
14	Water-in-oil sensor (Y/N)		
15	Temp. sensor (Y/N)		

B. Evaluation of Prediction Accuracy

For several months before starting the weekly prediction, the WPDSS was only collecting information from the wastewater pumping stations. In this way, the system formed the initial case dataset for the prediction process. When the WPDSS launches a new prediction process, the prediction results indicate at which pumping station certain operation problems can be expected. At the beginning of the system formation, all features are assigned equal weights. During continuous learning, these values are changing. A feature assigned with a higher weight value has a more significant influence on the functioning of the pumping stations. During the test period, features that do not have any or very little influence on pumping station operation will be detected. These features will be excluded from the case data model or will be replaced with new features. In the created application, the user is presented with the prediction results in a table form as shown in Table II. The Table contains the following information: the names of the wastewater pumping stations, their pump labels, and the failure risk level (low, moderate, high) for each pumping station from the list. This information indicates the probability of failure for the next period. The number of faults detected at the wastewater station in the previous period. Finally, the Table shows the recommendations to perform urgent services at any wastewater station for the next week. During the development and testing period, information about recommended activities was intensively used. The values of recommended activities were used as guidelines for the prevention of failure. On the wastewater pumping stations which WPDSS labeled as “mandatory pump service”, weekly surveillance was carried out although according to maintenance schedule this was not planned. This approach resulted in a reduction of the total number of station failures in the wastewater system up to 40%.

TABLE II. WATER STATIONS LIST AND THEIR FAILURE RISK LEVEL

Station name/ type of pump	Failure risk level	Failures/ year	Recommended activity
P1_Mackovec EMU FA10.94E -	LOW RISK	0	Maintenance by schedule
P1_PP2Mihaljevec SEV.80.80.110.A.	MODERATE RISK	3	Regular station service
P1_PP4Mihaljevec SEG.40.12.2.50B	LOW RISK	1	Maintenance by schedule
P1_CS1Lopatinec SEG.40.31.2.50B	HIGH RISK	2	Mandatory pump service
...

In order to obtain complete details about the characteristics of the CBR model, additional analysis of model performance using the Receiver Operating Characteristic (ROC) curve was performed [14]. Figure 3 shows the ROC curves of measurements of classification accuracy over a few datasets.

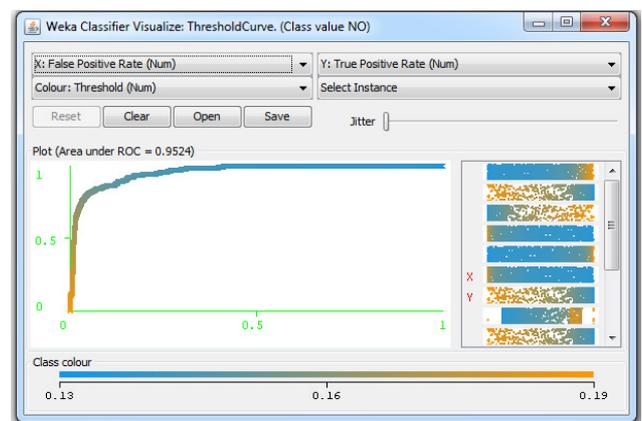


Fig. 3. ROC curve of WPDSS performance

The shapes of the ROC curves indicate a very good performance of the CBR model with prediction accuracy between 85% and 97%. The analysis of the structure of the recorded data in the case database indicated that the system is working with unbalanced data. The reason for this is the high quality of wastewater stations and their regular maintenance. Certain studies have shown that unbalanced data may have a detrimental effect on the prediction accuracy, which has not been manifested in this research. Parallel to the evaluation of the prediction model, we have conducted an evaluation of the complete system. The communication channel between SCADA and WPDSS was tested and the reliability of various scripts for daily data transmission to the web platform was checked.

V. CONCLUSION

This paper introduced a new approach for the prediction of correct functioning of wastewater pumping stations. A new wastewater pumping decision support system (WPDSS) was presented along with its architecture, its internal processes, and the way of performing the prediction process. The WPDSS works in web environment and uses a different machine learning method for the prediction process. The development and implementation of the WPDSS system were carried out in cooperation with the wastewater system in the Medimurje County, Croatia. The preliminary evaluation results indicate that the architecture of the system is well-conceived. Additional research will surely provide new guidance in developing and improving the overall system performance.

REFERENCES

- [1] M. A. Hamouda, W. B. Anderson, P. M. Huck, "Decision support systems in water and wastewater treatment process selection and design: a review", *Water Science & Technology*, Vol. 60, No. 7, pp. 1757-1770, 2009
- [2] A. S. Patil, N. J. Kulkarni, "Decision support system for waste water management: a review", *International Journal of Innovative Research in Advanced Engineering*, Vol. 1, No. 3, pp. 24-29, 2014
- [3] A. K. Bhavsar, J. S. Shah, "Cloud based decision support system for waste-water management using supervised decision tree algorithm", *International Journal of Computer Sciences and Engineering*, Vol. 5, No. 10, pp. 367-372, 2017
- [4] Z. Zhang, X. He, A. Kusiak, "Data-driven minimization of pump operating and maintenance cost", *Engineering Applications of Artificial Intelligence*, Vol. 40, pp. 37-46, 2015
- [5] J. Z. Ren, H. W. Liang, "Multi-criteria group decision-making based sustainability measurement of wastewater treatment processes", *Environmental Impact Assessment Review*, Vol. 65, pp. 91-99, 2017
- [6] W. X. Hu, "The application of artificial neural network in wastewater treatment", *IEEE 3rd International Conference on Communication Software and Networks*, Xi'an, China, May 27-29, 2011
- [7] Y. Guo, J. Hu, Y. Peng, "Research on CBR system based on data mining", *Applied Soft Computing*, Vol. 11, No. 8, pp. 5006-5014, 2011
- [8] M. M. Richter, R. Weber, "Case-Based Reasoning: A Textbook", in: *Basic CBR Elements*, Springer Science & Business Media, 2013, pp. 17-34
- [9] H. Huang, H. Qin, Z. Hao, A. Lim, "Example-based learning particle swarm optimization for continuous optimization", *Information Sciences*, Vol. 182, No. 1, pp. 125-138, 2012
- [10] F. P. A. Lima, M. L. M. Lopes, A. D. P. Lotufo, C. R. Minussi, "An artificial immune system with continuous-learning for voltage disturbance diagnosis in electrical distribution systems", *Expert Systems with Applications*, Vol. 56, pp. 131-142, 2016
- [11] E. F. Vazquez, "Updating weighting matrices by Cross-Entropy", *Investigaciones Regionales*, Vol. 21, pp. 53-69, 2011
- [12] A. Bonzano, P. Cunningham, B. Smyth, "Using introspective learning to improve retrieval in CBR: A case study in air traffic control", in: *Case-Based Reasoning Research and Development*, Springer, 2005
- [13] C. K. Riesbeck, R. C. Schank, *Inside Case-Based Reasoning*, Psychology Press, 2013
- [14] S. Bernard, C. Chatelain, S. Adama, R. Sabourin, "The multiclass ROC front method for cost-sensitive classification", *Pattern Recognition*, Vol. 52, pp. 46-60, 2015