BULETINUL Universităţii Petrol – Gaze din Ploieşti	Vol. LXI No. 2/2009	59 - 66	Seria Matematică - Informatică - Fizică

The Efficiency Level Analysis for the Wastewater Mechanical Treatment Process using Data Mining and Fuzzy Logic

Mădălina Cărbureanu

Petroleum –Gas University of Ploieşti, Informatics Department, Bd. Bucureşti, 39, Ploieşti
e-mail: mcarbureanu @upg-ploiesti.ro

Abstract

In order to improve the wastewater quality in accordance with the current European and Romanian standards, in Ploieşti it will be built a new wastewater treatment plant. The existing wastewater treatment plant has only the mechanical step, which it will be improved and extended with a biological and a sludge treatment step. In the present paper it is determined and analyzed the efficiency level of the existing mechanical step in the process of reducing the biochemical consumption of oxygen and removing the floating materials from the wastewater collected by the plant, using data mining techniques and fuzzy logic.

Key words: *wastewater treatment plant, mechanical step, efficiency, data mining, fuzzy logic*

Introduction

The techniques used in this paper are data mining and fuzzy logic, so we begin with a few words about these.

Data mining is defined as “the process of discovering patterns in data” or “it is about solving problems by analyzing data already presented in databases” [6].

The data mining techniques are divided into two categories: classical data mining techniques (nearest neighbor, clustering and statistics) and next generation data mining techniques (classifying and decision trees, neural artificial networks, rules induction, genetic algorithms and logistic regression). These techniques are presented in [1].

From the techniques mentioned above, we chose to use the decision trees, which accomplish the instances classification by covering the tree from up to bottom. In the making process of a decision tree is often used a special measure, named information gain, necessary to select the testing attribute.

The knowledge’s extraction from a database is used to build decision trees, through data examination and rules induction. In order to build such a structure are used a batch of algorithms, such as: CHAID (Chi-squared Automatic Interaction Detection), CART (Classification and Regression Trees), Quest and C5.0. Many procedures of a decision tree are based on ID3 algorithm, developed by J. Ross Quinlan, at Sydney University in 1979 [3]. The

ID3 algorithm is an inductive artificial intelligence technique which generates classification trees.

For the first part of the application we used the Weka (Waikato Environment Knowledge Analysis) package, described in detail in [1].

Fuzzy logic was defined by Lotfi Zadeh from the Berkeley University, in 1965 [7]. If the classical logic uses two numerical values, 0 (false) and 1(true), fuzzy logic uses logical values in [0,1] interval, where 0 indicates complete falsity and 1 indicates the complete truth. As a consequence, we can define the membership level of an object to a set, where the membership level can take values between 0 and 1. Using fuzzy logic, we can represent imprecise concepts, such as: *small*, *medium*, *big*, etc., concepts named linguistic values [11].

A fuzzy system has a certain number of inputs and outputs, which are crisp values that are real numbers and not fuzzy sets. Through the fuzzification process, the crisp inputs are converted into fuzzy sets and through the defuzzification process the fuzzy conclusions are converted into crisp outputs. Also, through the inference mechanism are obtained fuzzy conclusions using the fuzzy rules [2], [5].

There are two inference methods, respectively: Mamdani and Sugeno. The Sugeno inference method can be applied for systems with just one output, while the Mamdani inference can be applied for systems with many outputs. Because the proposed application has a number of outputs greater than 1, we used the Mandani method.

For the second part of the application, we used the Fuzzy Logic Toolbox implemented in Matlab 7.1.

Further, we will present the current situation of the wastewater treatment plant from Ploieşti.

A wastewater treatment plant removes the pollutants from the urban wastewater's containing a mixture of menage and industrial wastewater. The plant is administrated by the local public administration or by private companies under the local authority's subordination.

The existing wastewater treatment plant is located in the south-east limit of the Ploieşti city, having like emissary the Dâmbu brook. The plant has only a mechanical step on the wastewater treatment line and anaerobe digestion and dewatering on drying platforms, on the sludge treatment path.

The mechanical treatment goal is to protect the wastewater treatment plant equipment's, by removing those materials which can cause blockages or plant equipment's excessive usage. Furthermore, the various objects and big materials removed at the beginning of the wastewater treatment process saves valuable spaces in the plant, with positive effects on the investments and operating costs. The wastewater treatment plant from Ploieşti is composed by the following objects [8]:

- On the wastewater treatment path:
  - Gravity inlet of the brut effluents via two sewers, one which collects the effluents from the East and south-east part of Ploieşti and one which collects the effluents from the west and south-east part of the agglomeration;
  - Transversal storm overflow, which by-passes the pre-treatments for the effluents during the wet weather periods;
  - Vertical bar screen of eight centimetres mesh;
  - Curved bar screen of two centimetres mesh;
  - Horizontal longitudinal grit chamber;
  - Measure channel of Parshall type;
  - Four circular setting tanks;
  - Discharging channel of the settled effluents into Dâmbu.

- On the sludge treatment path:
  - Sludge pumping station;
  - Sludge drying beds;
  - Storage of dewatered sludge site;
  - One bio-gas reservoir.

At the mechanical step level is measured the efficiency in removing the biochemical consumption of oxygen (CBO5) and total floating materials (MTS) pollutants from wastewater. The biochemical consumption of oxygen represents the oxygen quantity used by micro-organisms for the oxidative degradation of the organically substances contained, at standard temperature and time (five days). The floating materials are insoluble substances in the wastewater, which can be removed through filtration, whizzing or sedimentation. The removing of MTS and partially CBO5 is achieved in mechanical step, through the vertical and curved bar screens and grit chamber. The total removal of CBO5 it will be achieved in the biological step thought organically substances decomposition with the help of aerobe bacteria.

The current emissary is Dâmbu brook with a length about 45 kilometers, from which 8.3 kilometers are on the Ploieşti territory and with the discharge in Teleajen river.

For improving the wastewater quality, in order to answer the qualitative conditions for evacuating in emissary, it must be taking into consideration the NTPA 001/2002, regulation which establish the charging limits with pollutants of wastewater at the evacuation in the natural receivers [9]. Taking into account the deterioration of the existing mechanical step of the plant and the necessity of changing the emissary, the new plant must have an improved mechanical step, added with a biological and a sludge treatment step.

In the biological step is achieved the total removal of CBO5, the removal of biological nitrogen, phosphorus and chemical consumption of oxygen (CCO-Cr). The biological treatment is a modification of the sludge actuation process which ensures not just the removal of CBO5 and MTS, but also an advanced system of nutrients retaining for effective reducing of nitrogen and phosphorus. This process combines areas in which are developing biological micro-organisms which remove the nutrients from the wastewater. This biological step will have the following components [8]:

- Two distribution chambers for a equal wastewater flows distribution;
- The sludge tank for the decomposing of CBO5;
- Secondary clarifying tank for sedimentation;
- Technological water pumping station for ensuring the water necessary for plant working;
- Wastewater pumping plant which discharges the treated wastewater in Teleajen river.

As a result of the wastewater treatment process is obtained biological sludge which is treated for chemical content neutralization and for reducing the volume necessary for storing. The biological sludge extracted from biological step has increased water content, which is reduced thought successive steps of dewatering and thickness.

For monitoring and optimization, the new plant will use the HydroDat system [8]. This is an operational management system with hardware and software components, used for operating parameters and damages registering, archiving, visualization and evaluation, in accordance with legal stipulations.

The proposed application has two parts:

- The first one consists in applying data mining techniques (decision trees) and data mining algorithms (ID3) to obtain the necessary decision rules for the fuzzy inference system ;
- The second one consists in building the fuzzy inference system architecture (FIS) and the analysis of the efficiency level for the wastewater mechanical treatment process using fuzzy logic.

The Developed Application

Because the existing wastewater treatment plant from Ploieşti possesses only the mechanical step, we decided to develop an application which goal is to determine the efficiency level of this step. To do that, we chose to use Weka and Matlab 7.1. software. The physical indicators measured in the existing wastewater mechanical treatment step are CBO5 (biochemical consumption of oxygen) and MTS (floating materials). The values for these two indicators are presented in table 1 and are given by the wastewater treatment plant technical documentation, representing values sampled at the plant input and output [8], [10].

Table 1. The indicators values

Indicator	Sampling point	2009						
		February	March	April	May	June	July	August
CBO5(mg/l)	Input	88.3	91.47	91.38	89.23	90.72	90.79	92.85
	Output	42.51	43.31	38.85	37.89	38.33	37.74	39.78
MTS(mg/l)	Input	229.46	239.58	241.96	235.79	237.1	236.85	236.92
	Output	143.12	143.98	131.58	119.84	119.93	131.26	131.55

The efficiency of the plant mechanical treatment step represents the removing level of the wastewater charge with CBO5 and MTS. The condition for defining the mechanical step efficiency is given by the pollutants removing level, which must be for CBO5>20% and for MTS>50%. For determine the efficiency level of the wastewater mechanical treatment step from Ploieşti, we used the values presented in table 2, given by technical documentation [8].

Table 2. The efficiency values

Indicator	Sampling point	2009						
		February	March	April	May	June	July	August
CBO5	Efficiency	51.86%	52.65%	57.48%	57.54%	56.84%	58.43%	56.40%
MTS	Efficiency	37.63%	39.90%	45.62%	49.18%	49.42%	44.58%	44.47%

For developing the proposed application, from Matlab we used Fuzzy Logic Toolbox necessary to build the fuzzy inference system. For this system we need, first of all, some decision rules, which are obtained using an algorithm from the decision trees category, implemented in Weka, named ID3[4]. The ID3 algorithm divides the training set into a number of different sets. It searches the most relevant attribute and uses like method for restrict the searching area, a measure named entropy. It is a greedy algorithm which develops a decision tree from up to bottom, at each nod selecting that attribute which best classifies the local training examples. The best attribute is that with maximum information gain. For achieving the classification and for a better data manipulation, the ID3 algorithm requires that for each variable to be established three intervals, presented in table 3. To do that, we used the average of the values from tables 1 and 2. For instance, the average for CBO5I is 90.68 mg/l, and the values which are greater than this value, belongs to *big* interval. For *medium* and *small* intervals, we chose a value of 87 mg/l, representing the smallest value for CBO5I, registered in 2009, according with the technical documentation [8].

Table 3. The indicators intervals

Indicator	Small	Medium	Big
CBO5I	<87	[87;90.68]	>90.68
CBO5E	<38	[38;39.77]	>39.77
EFFCBO5	<20(noneff)	[20;55.89](mediumeff)	>55.89(bigeff)
MTSI	<228	[228;236.8]	>236.8
MTSE	<120	[120;131.6]	>131.6
EFFMTS	<35(noneff)	[35;44.40](mediumeff)	>44.40(bigeff)

Using the training data from table 1 and 2 and the intervals established in table 3, we obtained a database for CBO5, implemented in Weka, which is presented in figure 1.

```
@relation mechanic.symbolic
@attribute CBOI {small,medium,big}
@attribute CBOE {small,medium,big}
@attribute efficiency {noneff,mediumeff,bigeff}
@data
medium,big,mediumeff
big,big,mediumeff
big,medium,bigeff
medium,small,bigeff
medium,medium,noneff
medium,big,noneff
big,big,noneff
big,medium,mediumeff
big,small,bigeff
```

Fig. 1. CBO5.arff database

As we can observe in figure 1, we have three attributes: *CBOI* represents the biochemical consumption of oxygen value at the mechanical step entrance, *CBOE* is the biochemical consumption of oxygen value at the mechanical step emergence and *efficiency* representing the removing level of CBO5 from the wastewater in mechanical step. .In a similar way was built the database for MTS.

It must be made the following observation: it was necessary to built two distinct databases for the two indicators (CBO5 and MTS) because they are separately measured at the entrance and emergence of the mechanical step, being established different limits in the measurement of the mechanical step efficiency in removing CBO5 and MTS from the wastewater (MTS>50% and CBO5>20%).

Applying the ID3 algorithm, we obtained the following decision tree for CBO5. Similarly, was obtained the MTS decision tree.

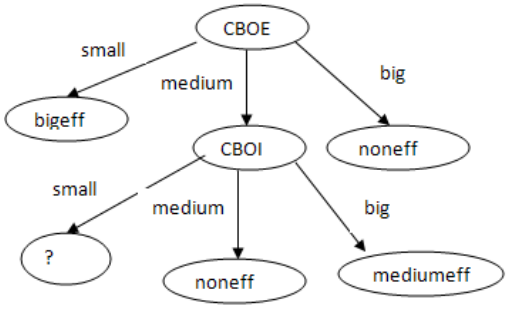

Fig. 2. CBO5 decision tree

Interpreting the two decision trees obtained using ID3, we have a batch of decision rules, presented in figure 3.

1. If (CBO5E is small) then (EFFCBO5 is big) (1)
2. If (CBO5E is medium) and (CBO5I is medium) then (EFFCBO5 is non) (1)
3. If (CBO5E is medium) and (CBO5I is big) then (EFFCBO5 is medium) (1)
4. If (CBO5E is big) then (EFFCBO5 is non) (1)
5. If (MTSE is small) then (EFFMTS is big) (1)
6. If (MTSE is medium) and (MTSI is medium) then (EFFMTS is non) (1)
7. If (MTSE is medium) and (MTSI is big) then (EFFMTS is medium) (1)
8. If (MTSE is big) then (EFFMTS is non) (1)

Fig. 3. Decision rules

Having at disposal the rules presented in figure 3, we developed a system with fuzzy logic to determine the mechanical step efficiency level for the existing wastewater treatment plant. The structure of the fuzzy inference system (FIS), with the afferent inputs (*CBO5E*, *CBO5I*, *MTSE* and *MTSI*) and outputs (*EFFCBO5* and *EFFMTS*), is presented in figure 4.

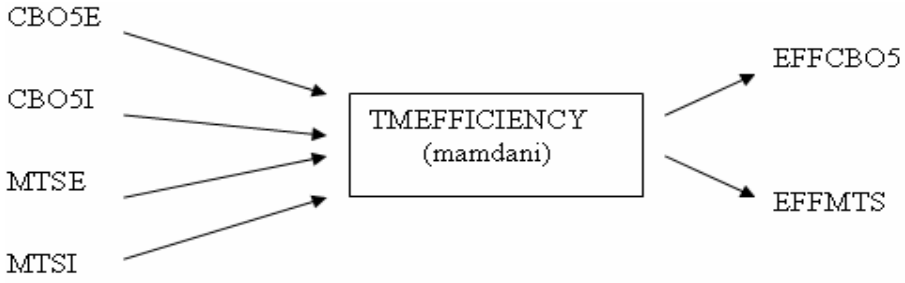

Fig. 4. FIS Architecture

The Rule Viewer presented in figure 5, displays a roadmap of the whole fuzzy inference process. The first four columns of plots show the if-part of each rule and the last two show the then-part of each rule. The last two plots in the last columns of plots represent the aggregate weighted decision for the given inference system. The decision depends on the input values for the system. The defuzzified output is displayed as a bold vertical line on these plots.

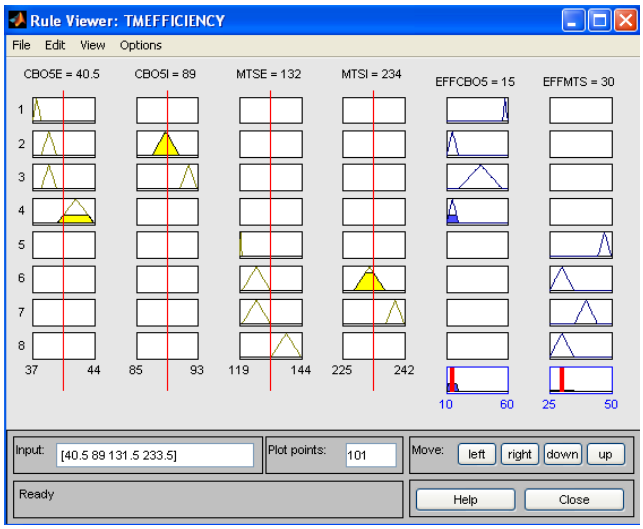

Fig. 5. The application Rule Viewer

We want to determine the mechanical step efficiency level in the mechanical treatment process of CBO5 and MTS. In tables 4 and 5 are presented some of the values obtained for *EFFCBO5* and *EFFMTS*.

Table 4. The efficiency level for CBO5 treatment

CBO5E	CBO5I	EFFCBO5
37.6(small)	-	57.9(bigeff)
39.10(medium)	91.4(big)	37.9(mediumeff)
38.70(medium)	89.20(medium)	15(noneff)
39.70(medium)	90.6(medium)	15(noneff)
38(small)	-	57.80(bigeff)

Table 5. The efficiency level for MTS treatment

MTSE	MTSI	EFFMTS
120(small))	-	47.2(bigeff)
125(medium)	231(medium)	30(noneff)
130(medium)	239(big)	39.7(mediumeff)
124(medium)	234(medium)	30(noneff)
131(medium)	240(big)	39.70(mediumeff)

The Control surface from figure 6 represents, in a compact way, all the information in the process, and shows the range of possible defuzzified values for CBO5 at the plant output.

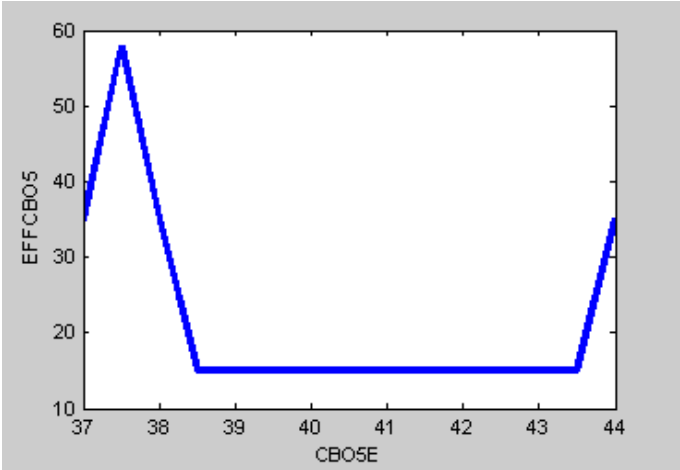

Fig. 6. The Control surface

The graph presented in figure 6 highlights the relation between biochemical consumption of oxygen emergence values and the efficiency level of the mechanical step in CBO5 treatment. For instance, if the CBO5E value is 37.6 mg/l, that is small, then EFFCBO5 is 57.9%, that means a big efficiency in the CBO5 mechanical treatment or when CBO5E is 39.77 mg/l, that is big then EFFCBO5 is 15% that means non-efficiency.

Conclusions

Using data mining techniques and fuzzy logic, we developed an application for determining the efficiency level of the wastewater mechanical treatment step based on input and output values of the main indicators analyzed on this step. Fuzzy logic has applications in wastewater treatment plant, due to the fact that supplies results which can be easily interpreted and is a coherent instrument for imprecise information processing. Furthermore, using fuzzy logic the complex process from the wastewater plant is optimally translated into a system which allows a high level of process transparency.

Taking into consideration that the condition for defining the mechanical step is given by the pollutants removing efficiency level, which must be for CBO5>20% and for MTS>50% and analyzing the results presented in tables 4 and 5, we can observe that the maximum efficiency level for CBO5 is 57.90% and for MTS is 47.20%.

We can conclude that in order to increase the efficiency level in removing MTS, the mechanical step must be modernized with a higher technology. Also, for answering to the Romanian and European specific legislation, which establish a minimum level of 75% in the CBO5 treatment level efficiency, it must be build the biological step for the present wastewater treatment plant.

At the moment, these two objectives are the main priorities for the local public administration, pursuing the decreasing of the pollution level of the new emissary Teleajen river.

References

1. Cărbureanu, M. - *The Human Development Index Prediction using Data Mining Techniques, MSe thesis,* Petroleum-*Gas University of Ploieşti, 2008*
2. Passino, K.M., Yourkovich, S., - *Fuzzy Control*, Addison-Wesley Longman, Inc., California, 1998
3. Quinlan, J.R. - *Machine Learning, Induction of decision trees,* volume.1, Springer, Netherlands, 1986
4. Schrodt, P. A., - *Predicting Interstate Conflict Outcomes Using a Bootstrapped ID3 Algorithm,* Political Analysis, Oxford Journals, 1990
5. Văduva, I., Grigore, A. - *Introducere în modelarea fuzzy*, Editura Universității din Bucureşti, 2004
6. Witten, I.H., Frank, E., - *Data Mining. Practical Machine Learning Tools and Techniques,* Elsevier Inc.,San Francisco, 2005
7. Zadeh, L.A., Jacprzyk, L., - *Fuzzy Logic for the management of uncertainty,* John Wiley &Sons , New York, 1992
8. *** - *The technical document for the objective - Wastewater collecting and treatment system modernization from Ploiesti city, RASP Ploieşti*, 2009
9. *** - *The NTPA-001/2002 normative for establishing the loading limits with pollutants for industrial and city wastewater at the evacuation into the natural receivers*, published in the Romania Official, part I, no. 187, March 20, 2002
10. *** - *National Institute of Statistics*, http://www.insse.ro, accessed November 10, 2009
11. *** - *Fuzzy Logic with correlation,* http://thor.info.uaic.ro/~alaiba/club-ai/index.php?title=Logica_fuzzy_cu_corela%C5%A3ie, accessed November 12, 2009

Analiza nivelului de eficienţă a procesului de tratare mecanică a apei uzate folosind data mining şi logica fuzzy

Rezumat

Pentru îmbunătăţirea calităţii apei uzate în concordanţă cu standardele naţionale şi europene în vigoare, în Ploieşti va fi construită o nouă staţie de epurare a apelor uzate. Staţia existentă dispune numai de o treaptă mecanică, ce va fi modernizată şi extinsă prin construcţia treptei biologice şi a treptei de tratare a nămolului. În cadrul acestui articol este determinat şi analizat nivelul de eficienţă a treptei mecanice existente în procesul de reducere a consumului biochimic de oxigen şi de îndepărtarea materiilor în suspensie din apa uzată colectată de staţie, utilizând tehnici de data mining şi logica fuzzy.