"HENRI COANDA"
AIR FORCE ACADEMY
ROMANIA

"GENERAL M.R. STEFANIK"
ARMED FORCES ACADEMY
SLOVAK REPUBLIC

INTERNATIONAL CONFERENCE of SCIENTIFIC PAPER
AFASES 2012
Brasov, 24-26 May 2012

# THE MONITORING OF WASTEWATER pH USING A DATA MINING TECHNIQUE

**Mădălina CĂRBUREANU ***

*Mechanical and Electrical Engineering, Petroleum-Gas University, Ploiesti, Romania

***Abstract:*** *The monitoring of wastewater pH value at the output of a plant neutralization process it is a very important task because it directly influence the quality of the plant effluent and therefore the quality of the emissary. A strong basic (alkaline) or acid pH determines the alteration of the emissary characteristics and also of the streams wherewith it is in contact. At plant neutralization process output the pH must have a neutral composition this way being respected the technical normative in domain. In the present paper, it is used a data mining technique, namely classification (C5.0 algorithm implemented in See5 system) for monitoring the pH value (acid or basic) at the neutralization process output. For implementing the results obtained through this technique, it is also developed a system for pH monitoring using CBuilder programming environment.*

***Keywords:*** *data mining, classification, rules, pH neutralization*

## 1. INTRODUCTION

The pH neutralization process from a wastewater treatment plant it is a very complex one and more it is a non-linear process that takes place in the plant chemical step. The pH at the process output can be: basic (alkaline), acid or neutral. At the process output the pH must be a neutral one to follow the pH admissible limits imposed through technical normative in domain, such as NTPA-001/2002 [4].

In literature, there are presented many applications of data mining techniques in wastewater treatment domain. Some examples in this sense are:

1. TELEMAC (Telemonitoring and Advanced Telecontrol of Yield Wastewater Treatment Plants) [1];
2. GESCONDA - an intelligent data analysis system for knowledge discovery and management in environmental databases, [2], etc.

The structure of the paper it is organized as follows:

1. The classification implementation where through the application of C5.0 algorithm are obtained the decision rules necessary to develop the proposed system;

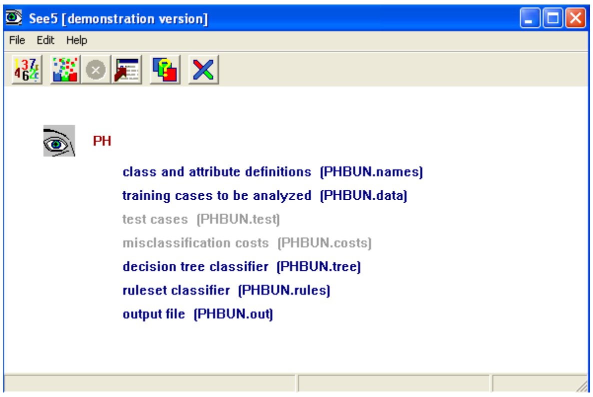
2. The proposed monitoring system development where it is presented the monitoring system developed using the classification results;

3. Conclusions where are emphasized the most important aspects of classification and the utility of the developed system.

## 2. THE CLASSIFICATION IMPLEMENTATION

C5.0 algorithm it is an improved version of ID3 and C4.5 data mining algorithms, introduced by Quinlan to solve the classification problems and to outrun all the limitation of the previous algorithms [3]. The classifiers obtained through the usage of this algorithm are under decision tree form or under rule set form. In this case we choose to use the decision tree form.

In the current paper we used the Windows version of C5.0 data mining tool, namely See5 demonstration version as we can observe in figure 1.



class and attribute definitions [PHBUN.names]
training cases to be analyzed [PHBUN.data]
test cases [PHBUN.test]
misclassification costs [PHBUN.costs]
decision tree classifier [PHBUN.tree]
ruleset classifier [PHBUN.rules]
output file [PHBUN.out]

Fig. 1. See5 interface

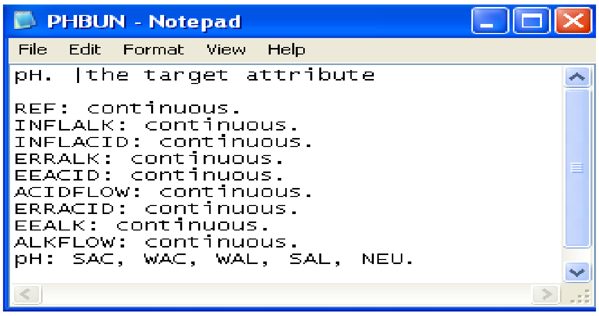
As we can observe in figure 1, See5 uses certain type of files such as [5]:

1. .names – file in which are described the application attributes;
2. .data – file that contains the cases used for building the classifier;
3. .tree – file that contains the classifier under decision tree form;
4. .rules – file that contains the classifier under decision rules form;
5. .out – the report obtained through the classifier generation.

For See5 system we build two files, such as: pH.data and pH.names. The pH.data supplies information on the training cases from which See5 will extract patterns. It contains a number of twenty four entries for a number of parameters such as:

1. REF – the pH reference value imposed through legislation;
2. INFLALK – the value of the alkaline pH in the plant influent;
3. INFLACID – the value of the acid pH in the plant influent;
4. ERRALK – the error value defined as the difference between pH reference and the alkaline pH value in the plant influent;
5. EEACID – the opening degree for the acid ($H_2SO_4$) pump (%);
6. ACIDFLOW – the acid flow necessary to bring a basic pH to a neutral one;
7. ERRACID – the error value defined as the difference between pH reference and the acid pH value in the plant influent;
8. EEALK – the opening degree for the alkaline/basic (NaOH) pump (%);
9. ALKFLOW – the alkaline flow necessary to bring an acid pH to a neutral one;
10. pH– the pH value (strong acid, weak acid, strong alkaline, weak alkaline, neutral ) at the process output.

The pH.names file describes the application attributes and classes as we can observe in figure 2.

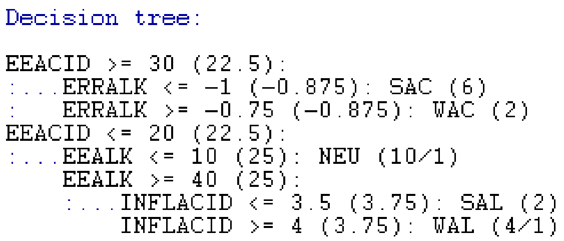


```
pH.   |the target attribute

REF: continuous.
INFLALK: continuous.
INFLACID: continuous.
ERRALK: continuous.
EEACID: continuous.
ACIDFLOW: continuous.
ERRACID: continuous.
EEALK: continuous.
ALKFLOW: continuous.
pH: SAC, WAC, WAL, SAL, NEU.
```

Fig. 2. pH.names structure

As we have mentioned, the classifier it is generated under a decision tree or under a rule set form. We generated the classifier under decision tree form, as we can observe in figure 3.



```
Decision tree:

EEACID >= 30 (22.5):
:...ERRALK <= -1 (-0.875): SAC (6)
:   ERRALK >= -0.75 (-0.875): WAC (2)
EEACID <= 20 (22.5):
:...EEALK <= 10 (25): NEU (10/1)
    EEALK >= 40 (25):
    :...INFLACID <= 3.5 (3.75): SAL (2)
        INFLACID >= 4 (3.75): WAL (4/1)
```

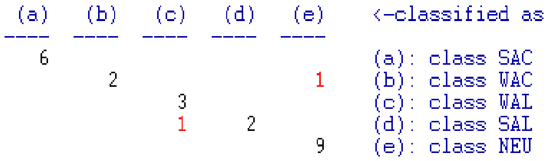
Fig. 2. Decision tree

In table 1 are presented the statistical results obtained through the generation of the decision tree.

Table 1. Statistical results

| Classifier form | Decision tree |
|---|---|
| Error | 8.3% |
| Misclassified instances | 2 |
| Correct classified instances | 22 |
| Rules number | 5 |

In figure 3 it is presented the confusion matrix where it is emphasized the performance on the training cases that shows the kinds of error made.



```
 (a)   (b)   (c)   (d)   (e)    <-classified as
 ----  ----  ----  ----  ----
    6                           (a): class SAC
          2                 1   (b): class WAC
                3               (c): class WAL
                1     2         (d): class SAL
                            9   (e): class NEU
```

Fig. 3. Confusion matrix

In figure 3, the decision tree misclassifies:
1. one of the weak acid (WAC) cases as neutral (NEU);
2. one of the strong alkaline (SAL) cases as weak alkaline (WAL).

It is useful to know also each attribute contribution to the classifier (attribute usage), as we can observe in table 2.

Table 2. Attribute usage

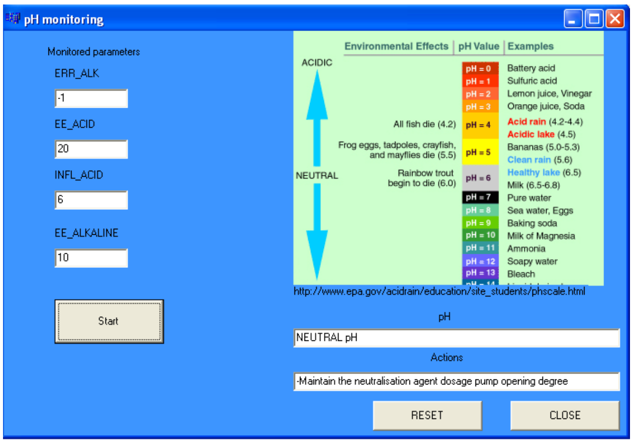
| Attribute | Attribute usage (%) |
|---|---|
| EEACID | 100% |
| EEALK | 71% |
| INFLACID | 38% |
| ERRALK | 33% |

From table 2, we observe that the most important contribution to classification is that of EEACID attribute (100%).

The rules of the decision tree presented in figure 2, written under *if-then* form are:
1. IF EEACID>=30 AND ERRALK<=-1 THEN pH=SAC;
2. IF EEACID>=30 AND ERRALK>=-0.75 THEN pH=WAC;
3. IF EEACID<=20 AND EEALK<=10 THEN pH=NEU;
4. IF EEACID<=20 AND EEALK>=40 and INFLACID<=3.5 THEN pH=SAL;
5. IF EEACID<=20 AND EEALK>=40 AND INFACID>=4 THEN pH=WAL.

## 3. THE PROPOSED MONITORING SYSTEM

Using the rules obtained through classification, we developed using C++Builder 6 a system for monitoring the pH of wastewater at the output of a plant neutralization process. The proposed system interface it is presented in figure 4.



pH monitoring

Monitored parameters

ERR_ALK: -1

EE_ACID: 20

INFL_ACID: 6

EE_ALKALINE: 10

pH: NEUTRAL pH

Actions: -Maintain the neutralisation agent dosage pump opening degree

Start
Environmental Effects
pH Value
Examples
ACIDIC
NEUTRAL
pH = 0 Battery acid
pH = 1 Sulfuric acid
pH = 2 Lemon juice, Vinegar
pH = 3 Orange juice, Soda
All fish die (4.2) pH = 4 Acid rain (4.2-4.4) Acidic lake (4.5)
Frog eggs, tadpoles, crayfish, and mayflies die (5.5) pH = 5 Bananas (5.0-5.3) Clean rain (5.6)
Rainbow trout begin to die (6.0) pH = 6 Healthy lake (6.5) Milk (6.5-6.8)
pH = 7 Pure water
pH = 8 Sea water, Eggs
pH = 9 Baking soda
pH = 10 Milk of Magnesia
pH = 11 Ammonia
pH = 12 Soapy water
pH = 13 Bleach
http://www.epa.gov/acidrain/education/site_students/phscale.html
RESET
CLOSE

Fig. 4. The system interface

The pH character (neutral, weak acid, strong acid, weak alkaline or strong alkaline) at a plant neutralization process is influenced by a number of parameters such as:
1. the composition of the plant influent (acid or basic (alkaline));
2. the error defined as the difference between the reference value of pH (pH=7, neutral pH) and the influent composition (acid or basic) at the plant input;
3. the opening degree of agent neutralization dosage pumps (pump for acid or basic

neutralization agent dosage) that depends on the error value, etc.

The proposed pH monitoring system offers to the user the following facilities:

1. for the monitored parameters, the proposed system provide to the user (that can be , for instance, a plant operator), information regarding the value of the wastewater pH at the process output (if the value of pH is in the admissible values established through technical normative in domain - NTPA 001/2002 );
2. the system also provides a set of actions that can be taken depending on the pH value at the process output, such as:
   a. if the value of pH at the neutralization process output in neutral then the suggested action is for instance to maintain the functioning parameters (the opening degree) for the acid or basic neutralization agent dosage pumps;
   b. if the value of pH at the process output is acid then the suggested action is to increase the opening degree of pump for dosing the basic/alkaline solution;
   c. if the value of pH at the process output is basic then the suggested action is to increase the opening degree for acid dosage pump.

The utility of such type of system developed using a data mining technique (such as classification) consist in the fact that it can helps the plant operators in monitoring the values of pH at the plant output and it also suggest the immediate solution at the problems that can occur.

## 4. CONCLUSIONS & ACKNOWLEDGMENT

The pH neutralization process it is of present interest due to its complexity and to the various factors that influence it. By constructing classifiers (under decision tree or rule sets form) the number of monitored parameters it is reduced using the attribute usage criterion, fact that facilitates the achieving of an improved and more focused monitoring.

The development of a system for pH monitoring can supply to plant operator valuable information regarding the functioning parameters of the plant pH neutralization process fact that offers the possibility to take immediate measures to prevent dangerous situations, such as: the pH of the plant effluent is acid/strong acid or basic/strong basic with dangerous effects on the environment.

The pH neutralization process it will be discussed in a future paper, where it will be developed a fuzzy controller for an automatic system dedicated to pH control.

## REFERENCES

1. Dixon, M., Gallop, J.R., Lambert, S.C., Healy, J.V., Experience with data mining for the anaerobic wastewater treatment process, *Environmental Modeling & Software*, Vol. 22, pp. 315-322 (2007).
2. Gibert, K., Sanchez-Marre, M., Rodriques-Roda, I., GESCONDA: An intelligent data analysis system for knowledge discovery and management in environmental databases, *Environmental Modeling & Software*, Vol. 21, pp. 115-120, (2006).
3. Khosrow-Pour, M., *Emerging Trend and Challenges in Information Technology Management*, Idea Group Inc, Hershey , USA, Available : http://www.dl-provider .com/download-k: Download %20 Emerging%20Trends%20and%20Challenges%20in%20Information%20Technology %20Management%20%20.html?aff.id=1266&aff.subid=1 (February , 2012).
4. NTPA -001/2002, Available: http: // www .aqua-biotec.ro/NTPA001.pdf (March, 2012).
5. RULEQUEST RESEARCH, See5: An Informal Tutorial, Available: http:// www. rulequest.com/see5-win.html (February, 2012).