

Data mining for industrial facilities

V.L. Kolesnikov¹, P.P. Urbanovich^{1,2}, A.I. Brakovich¹

1- Belarusian State Technological University, Minsk, Belarus,
e-mail: brakovich@yandex.ru; 2 - Catholic University of Lublin, Poland

The modern technology of data mining is based on the templates (patterns) that reflect the multi-dimensional fragments of relationships in the data. The source of the information in this paper is an industrial enterprise Unix, which produces sanitary types of paper from recycled paper. The problem is that the quality of raw materials is extremely unstable over time. Random changes in these external factors lead to the need for rapid intuitive correction of technological parameters. From the many data mining techniques we have chosen an information-theoretic algorithm of classification trees constructing with unlimited number of descendants of a node [1,2]. The rapid process of studying, rules generation of hard to formalize areas, rules extraction in a natural language, clear classification model are the advantages of this algorithm. The most important production problem is to receive sanitary types of paper with maximum absorption [1]. Therefore, in this article authors have described the data mining of this particular target attribute. The construction of the trees on the process data was carried out by using an *SQLServer.DMClientXLAddIn* and tabs "Mining" of *Microsoft Excel*.

In construction of a decision tree the method of cut off branches was used in which the cleavage were only those nodes which have higher target attribute values. Similarly, the trees were constructed and produced technological modes to ensure the best value of all the target attributes. It's required to estimate the values of control actions, which provide high quality products not for each indicator separately but all together.

To select the optimal technological mode for all target attributes fuzzy system with fuzzification and defuzzification in the input and the output is used. As fuzzification system is composed of fuzzy rules setting up the membership function. Attention is drawn to the fact that maximization requires only one attribute (wicking), the other should be minimized. For defuzzification multiplicative convolution of individual criteria is used. It provides a scalar criterion in the form of the multiplication (1) [3]:

$$D_i = \left[\prod_{j=1}^m (d_{ij})^{\mu_j} \right]^{1 / \sum_{j=1}^m \mu_j}$$

where i - number of rows in table of observations; j - partial criterion; m - number of individual criteria; w_j - statistical weight of j criterion. After sets of fuzzification and defuzzification optimal values of process conditions were identified. They were the best for requirements of five criteria for three month observation period.

Thus, a method of decision making under uncertainty conditions for the optimum process for producing sanitary paper by data mining in supervision was developed.

References

- [1] Kolesnikov V.L.. System analysis of production systems in the printing industry, BSTU, 2011.
- [2] Brakovich A., Urbanovich P. Mathematical model of control points placing for monitoring polluting substances emissions in the atmosphere on the basis of the NP-full coloring graph task decision // *Przegląd Elektrotechniczny*, № 7, 2010, p. 87-89.
- [3] Brakovich A., Kolesnikov V., Urbanovich P. Mathematical foundations of composite desirability function for evaluation of the product quality in the relationship with anthropogenic impacts on the environment // *Informatyka, automatyka, pomiary w gospodarce i ochronie srodowiska*, №4a, 2012, p. 36-39.