DATA MINING PROCESS FOR UNBIASED PERSONNEL ASSESSMENT IN POWER ENGINEERING COMPANY

Taras Zavaliy[1], Iouri Nikolski[2]

***Abstract**: The use of data mining algorithms in human resources management is somewhat underestimated. We show that simple analytical tools and techniques can improve HR management process in large corporations. The psychological testing results, which we analyzed, were used for assessing staff in the power engineering company. The data set contained 188 records and 38 attributes such as employee's job title, job place, age, experience, reaction speed, memory capacity, concentration ability, skills, motivations etc. These empirical data were complemented with evaluations from administration. The main goal of our experiments was to build and compare several rule-based classification models using the Rosetta software system. These models were built for two different decision attributes. The model induction process involved model approximation in order to deal with low quality of the data. This process revealed important factors of personnel efficiency. Comparing different models using ROC curves, AUC values and success rates we found an optimal solution and showed that biased information in the data set is affecting regularities in data and should not be used for decision-making.*

INTRODUCTION

The real-life problem considered in this report lies in HR management field. The data set which we analyzed contained results of psychological testing conducted at the power engineering company in order to improve safety and quality of work. The empirical data collected by psychologists were complemented with feedback information from personnel management. The problem of objectivity and quality of personnel monitoring aroused. The decision-making process in this case was also of interest. We used rough sets for data mining and induced several classification models. By comparing them with each other we showed that biased information in the data set is affecting regularities in data and should not be used for decision-making.

Methods of data mining are commonly used for finding regularities in data, building prediction models, solving classification or clusterization tasks. The use and evolution of these methods is an active research issue (Witten, Frank, & Hall, 2011), (Mitra, Pal, & Mitra, 2002). They have been used for dealing with real-life situations, processing large volumes of noisy, imprecise data. Soft computing methods, i. e. fuzzy sets, rough sets, genetic algorithms, implement and exploit notions of imprecision, noise, uncertainty, and approximate reasoning to achieve acceptable solutions of the modeling task. Thus, the rough set approach for data mining is based on the rough set theory (Pawlak, 1982), (Komorowski, Polkowski, & Skowron, 1998) that naturally deals with noise and redundancy in data by reducing the amount of features needed for successful model construction.

[1] Taras Zavaliy, Phd student, Information Systems and Networks Department, Lviv Polytechnic National University, S. Bandery Str., 12, Lviv, 79013, Ukraine, taras.zavaliy@gmail.com
[2] Iouri Nikolski, Prof, Information Systems and Networks Department, Lviv Polytechnic National University, S. Bandery Str., 12, Lviv, 79013, Ukraine, y_nikol@yahoo.com

DATA MINING PROCESS

The core of our data mining process is the rough set approach. A basic principle of the rough set-based learning is to discover redundancy and dependencies between the features of a problem represented in the form of a *decision table* − data table with predefined decision attribute(s). With the rough set method we extract the minimal subset of attributes called *reduct* or *minimal hitting set*. Attributes from the reduct preserve all dependencies in the decision table and do not affect the original discernibility of examples in the table. One of the methods for constructing reduct is Boolean reasoning (Ohrn, 2000), the algorithms, which implements Boolean reasoning, is Johnson algorithm (Ohrn, ROSETTA Technical Reference Manual, 2001). It searches for the minimal hitting set and allows approximate solutions being build. To control approximation scale we used the *hitting fraction* (*HF*) parameter. It denotes a fraction of subsets s_i of attributes, which should be used for extracting the reduct. Adjusting hitting fraction is a mechanism for compensating low quality of data and improving performance of the model as we showed in our experiments (Zavaliy, Nikolski, & Pasichnyk, 2007).

Our classification model consists of a set of rules *RUL* extracted from data and a set *P* of parameters assigned to them. We also adjust parameter *HF* for building approximate models. In the case of binary classification task a threshold τ for classifying example to a particular class is used. Thus, classification model is defined as follows

$$M = \langle RUL, P, HF, \tau \rangle. \qquad (1)$$

Several numerical parameters can be associated with a decision rule $\alpha \to \beta$ (IF α THEN β) (Ohrn, 2000). These parameters are *Support*(α→β), *Accuracy*(α→β) and *Coverage*(α→β). They not only describe the quality of the rules, but may impact the classification outcome directly. In the two-class classification process one of the class, say X_1, casts some normalized number of votes. If this number is greater than the predefined threshold τ, the example gets classified to X_1, otherwise it is marked as belonging to X_0. After applying induced model to the voting classification of k test examples we obtain a *success rate* (*SR*) of correct predictions. The overall 7-step process was implemented in scripting language of the Rosetta toolkit.

OBJECTIVES

The main goal of experiments was to analyze and optimize the process of personnel assessment. We had to build classification models from data about employees in order to compare assessments made by psychologists and managers. The model induction process involved model approximation in order to deal with low quality and redundancy of the data. Approximate models were induced with hitting fraction set to be less than 1, and precise model had hitting fraction set to 1. Comparing different models using *ROC* curves, *AUC* and success rates revealed the optimal solution and led to some assumptions about quality of the data. Feature reduction, which is in the core of rough sets, revealed most relevant attributes. These attributes should be used to objectively assess employee professional aptitude.

EXPERIMENTS

The decision table with psychological testing results contained 188 examples and 38 attributes. Conditional attributes contained symbolic information such as employee's job title, job place, age, experience, as well as specific testing results – reaction speed, memory capacity, concentration ability, skills, motivations etc.

Two decision attributes, "aptitude" and "reliability", were defined by psychologists and managers correspondingly. The table had no missing or undefined values. In the course of experiments six models were induced for two decision attributes using three different approximation levels. Training and test data sets contained 94 different examples each.

Table 1. Parameters of constructed models.

Model	HF	Reduct length	Rules count	*SR*	*AUC*
Model #1	1.0	4	60	0.6	0.68
Model #2	0.99	3	32	0.78	-
Model #3	0.96	2	13	0.82	0.81
Model #4	1.0	5	86	0.05	-
Model #5	0.98	3	45	0.35	-
Model #6	0.91	2	16	0.5	0.54

Models #1-3 were built for the decision attribute "aptitude", and models #4-6 were built to predict "reliability" attribute. The results of models evaluation are summarized in Table 1. Sample rules with *coverage*(α)>10% are listed below.

1. age_group(1) $\wedge$ integrated_score(good) $\rightarrow$ aptitude(good);
2. age_group(2) $\wedge$ integrated_score(good) $\wedge$ ambitions(average) $\rightarrow$ aptitude(good);
3. job_place(PES) $\wedge$ attention(average) $\rightarrow$ reliability(good).

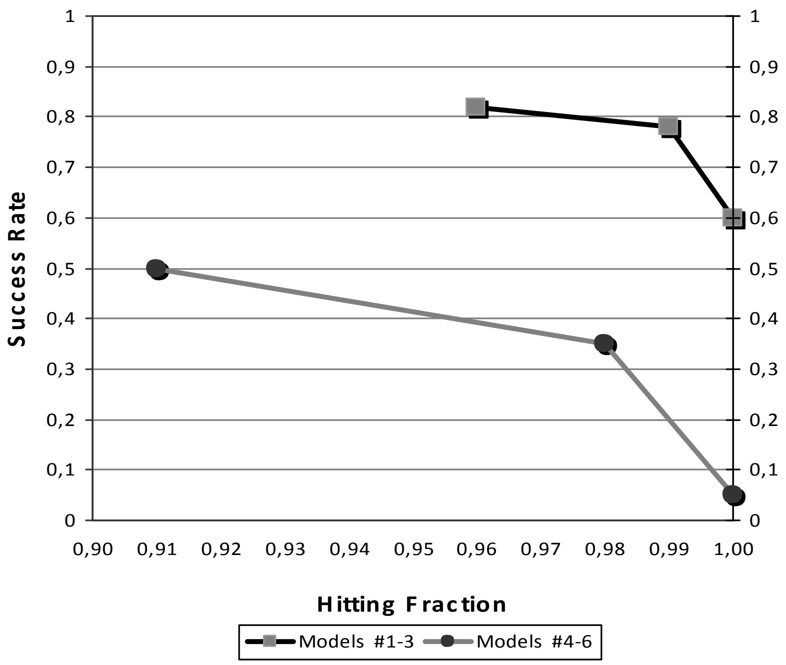

Fig. 1. Comparison of classification success rates for two different decision attributes.

Figure 1 represents the performance of six classifiers in classifying 94 test examples. It is clear that approximate models with *HF*<1 show better results, but models #4-6 show very poor performance in predicting "reliability" attribute. This may be caused by the fact that "reliability" attribute is defined using subjective judgments and is not in correlation with the rest of the data.

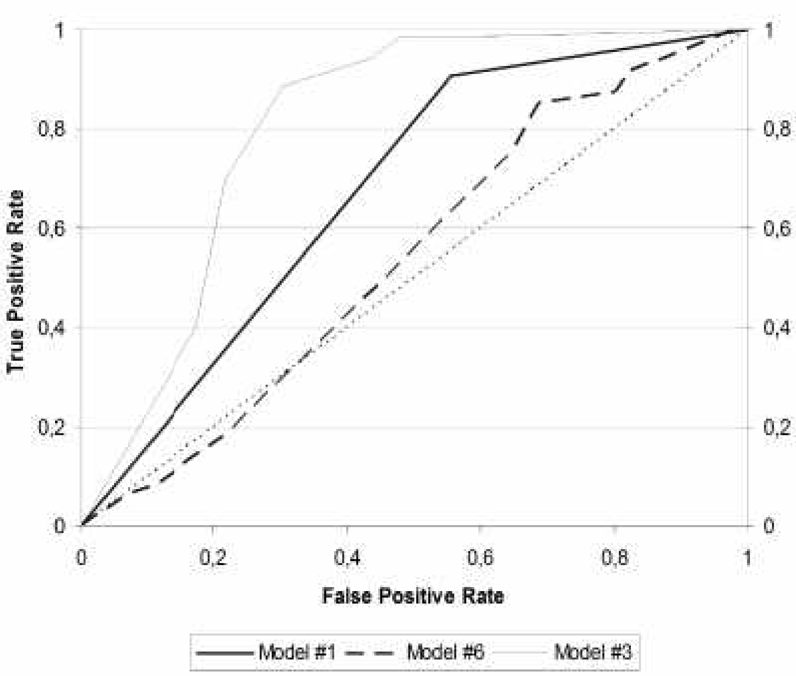

Fig. 2 ROC curves representing the ability of constructed models to recognize class X_1=*good*.

Figure 2 shows ROC curves plotted for three models – precise model #1, approximate model #3 and approximate model #6. It is obvious that approximate model #3 is better in recognizing examples from the class X_1=*good*. The curve for approximate model #3 dominates the rival curve for model #1. Comparing the *AUC* values also shows superiority of approximate model − 0.81 against 0.68 (see Table 2). At the same time, model #6 built to predict "reliability" attribute shows very low performance (*AUC*=0.54) in recognizing examples from the X_1=*good* class. This too can be explained by the fact that "reliability" and "successfulness" attributes in the data set were defined by the personnel management, not by psychologists. Therefore, these attributes are disconnected from the rest of the data and are not involved in meaningful regularities.

We also revealed that only 6-10 conditional attributes out of 19-36 are present in desisioon rules and influence classification outcome. After approximation the amount of conditional attributes in rules was reduced to only two, improving model performance. These results were interesting to psychologists and managers, allowing them to modify their personnel assessment process.

CONCLUSIONS

The data used for rule extraction experiments were not of the highest quality and 94 training examples, obviously, were not sufficient for high-quality model induction. Yet it was possible to achieve 82% success rate for one of the constructed models. All of the classification models were compared using ROC curves, *AUC* value and success rate.

The outcome of the research shows that applying the rough set approach in human resource management may lead to reasonable results. Redundancy of attributes in data table or insufficient amount of training examples can be compensated by approximate solutions. Rough set based technique allowed us to significantly reduce feature set and extract more general rules from data. These rules formed classification models which showed good performance classifying test examples. Information about importance of features will be used for decision-making about future psychological testing and personnel monitoring in the company.

LITERATURE

1. Komorowski, J., Polkowski, L., & Skowron, A. (1998). Rough Sets: A Turtorial. y A. Skowron, & S. K. Pal, *Rough Fuzzy Hybridization: A New Trend in Decision-Making* (cc. P. 3-98). Springer-Verlag.
2. Mitra, S., Pal, S. K., & Mitra, P. (2002). Data mining in soft computing framework: a survey. *IEEE Transactions on Neural Networks, Vol. 13.*, 3-14.
3. Ohrn, A. (2000). *Discernibility and Rough Sets in Medicine: Tools and Applications, PhD thesis.* Norwegian Univercity of Science and Technology, Dep. of Computer and Information Science.
4. Ohrn, A. (2001). *ROSETTA Technical Reference Manual.* Отримано з www.lcb.uu.se: http://www.lcb.uu.se/tools/rosetta/materials/manual.pdf
5. Pawlak, Z. (1982). Rough Sets. *International Journal of Computer and Information Sciences, vol. 11/5*, 341-356.
6. Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: practical machine learning tools and techniques, 3-d ed.* Morgan Kaufmann Publishers.
7. Zavaliy, T. I., Nikolski, I. V., & Pasichnyk, V. V. (2007). Data reduction in analysis of psychological testing results (in Ukraininan). *Vidbir i obrobka informaciji, №27 (103)*, 93-98.