EFFICIENCY ASSESSMENT OF THE RULE AND MODEL BASES IN DECISION SUPPORT SYSTEMS

B.A. Zhelezko, O.A. Siniavskaya
Belorussian State Economic University
Minsk, BELARUS
e-mail: boriszh@yandex.ru

Abstract

In the article new indices for efficiency assessment of the intelligence rule and model bases in decision support systems (DSS) are suggested. These indices are rule base certainty and coverage, rating class efficiency, rating efficiency. Formulae for these coefficients calculation are deducted, received with use of rough sets theory and DSS efficiency assessment experience. Recommendations on their practical using are given.

The problem of information systems efficiency estimation is very important and difficult problem because it is difficult to allocate what part of the received effect is resulted due to this system, without other factors accounting. Decision support systems are the special type of information systems, intended for the decision maker assistance in their professional work by means of data, knowledge and model bases using for the decisions substantiation. The purpose of the given work is DSS efficiency estimation methodology development with rough sets theory using which will allow DSS model and rules bases application formalizing.

Rough sets theory is a new mathematical approach to intelligent analysis and data mining. The theory was suggested by polish scientist Zdzislaw Pawlak in 1982 [1]. Rough sets have many successful applications in practice, for example, in medicine, pharmaceutics, banking and finance, market analysis, environment control, etc [2, 3].

One of the rough set theory possibilities is decision rules efficiency assessment. In terms of this theory *decision rules* are the expressions in the "if…then…" form: $\Phi \rightarrow \Psi$. Antecedent is denoted as Φ, consequent as Ψ, antecedent and consequent sets as *For(C)* and *For(D)*correspondingly, where "*For*" is a *formulae* [1, 4].

For decision rules evaluation and interpretation and conclusions obtaining on their base a number of following quantity indicators are used [4].

Support of a rule is a maximum number of objects satisfying both antecedent and consequent:

$$supp_S(\Phi, \Psi) = card(\|\Phi \wedge \Psi\|_S). \tag{1}$$

Certainty factor is a frequency of occurrence of the objects having property Ψ in set of the objects having property Φ:

$$cer_S(\Phi, \Psi) = \frac{card(\|\Phi \wedge \Psi\|_S)}{card(\|\Phi\|_S)}. \tag{2}$$

This coefficient is widely used also in data mining and known in data mining terms as *confidence coefficient.*

Coverage factor is a frequency of occurrence of the objects having property Φ in set of the objects having property Ψ:

$$cov_S(\Phi, \Psi) = \frac{card(\|\Phi \wedge \Psi\|_S)}{card(\|\Psi\|_S)}. \tag{3}$$

Strength of a rule is the ratio of the *support of a rule* to a number of objects in the decision table:

$$\sigma_S(\Phi, \Psi) = \frac{supp_S(\Phi, \Psi)}{card(U)}. \tag{4}$$

The offered approach of decision rules efficiency estimation has been developed and generalized with reference to DSS. For an estimation of DSS rules and knowledge bases following indices are offered.

Rule base certainty is the relation of quantity of situations in which rules from rule base ($N_{\Phi\Psi}$) were carried out to total quantity of situations when rule antecedents were carried out (N_{Φ}), including those situations when consequents were not carried out:

$$CRT = \frac{N_{\Phi\Psi}}{N_{\Phi}}. \tag{5}$$

Rule base coverage is the relation of quantity of situations in which rules from rule base ($N_{\Phi\Psi}$) were carried out to total quantity of situations when rule consequents was carried out (N_{Ψ}), including those situations when antecedents was not carried out:

$$CVR = \frac{N_{\Phi\Psi}}{N_{\Psi}}. \tag{6}$$

Offered indices may be used for the whole rule base efficiency assessment ad also for the separate rules efficiency assessment. These are the conditional probabilities which express how exactly DSS knowledge base describes the problem area. Rule base certainty shows the probability of efficient decision concluding by means of DSS. Rule base coverage shows the probability of decision cause deduction from the DSS rule base.

For the DSS which allow rating following efficiency indices are offered.

Rating class efficiency is the relation of quantity of objects which were classified by DSS as belonging to i-th rating class ($NRDSS_i$) to quantity of objects which belonging to this rating class in real situation (NR_i):

$$ERclass_i = \frac{NRDSS_i}{NR_i}. \tag{7}$$

Index 7 is calculating during DSS testing process with using of the retrospective data about problem area and by means of received results comparison with the real situation. For example, if DSS allow receiving the rating of the enterprises by the bankruptcy risk factor, then for the i-th class which include the enterprises with low bankruptcy risk, $ERclass_i$ index is calculated as the relation of the number of enterprises classified by DSS as successful (without bankruptcy possibility), to the number of enterprises which were not bankrupts in real situation.

Given index is the estimation of the rating model correctness and it may be a base for the definition of necessity of the model correction or its part modification. If it is impossible to correct the model, this index allows calculating the probability of the object relation to certain rating class.

Rating efficiency is the relation of quantity of objects which were correctly classified by DSS ($NDSS_{rating}$) to total quantity of objects in the rating (N_{rating}):

$$ER = \frac{NDSS_{rating}}{N_{rating}}. \tag{8}$$

Index 8 characterized the model base efficiency and index 7 characterized the separate logical models efficiency.

Thus, in the article the methodology of DSS efficiency estimation [5] is developed on the base of rough sets theory methods. Four indices were offered due to assessment of the intellectual part of the DSS, which give not only DSS efficiency characteristic, but also indicate situations when model and rules bases would be modified for the decision deduction process improving.

References

[1] Pawlak, Z. Rough sets and intelligent data analysis / Z. Pawlak// Information Sciences. – 2002. – #47. – P. 1-12.

[2] Beynon M.J., Peel M.J. Variable precision rough sets theory and data discretization: an application to corporate failure prediction // Omega. – 2001. – #29. – P. 561-576.

[3] Greco S., Matarazzo B., Slowinski R. Rough sets theory for multicriteria decision analysis // European Journal of Operational Research. – 2001. – #129. – P. 1-47.

[4] Zhelezko B.A., Siniavskaya O.A. Rough sets theory applications in decision-making problems // Człowiek i jego decyzje / red. K.A. Kłosiński, A. Biela. – Wydawnictwo KUL, Lublin, 2009. – Vol.1. – P. 61-70.

[5] Zhelezko B.A., Siniavskaya O.A., Ahrameiko A.A. Estimation of quality of objects with complex structure under uncertainty conditions // Computer Data Analysis and Modeling: Robustness and Computer Intensive methods: Proc. of the Seventh International Conference (September 6-10, 2004, Minsk). Vol. 2. – Minsk: Academy of Administration at the President of The Republic of Belarus, 2004. – P. 67-70.