A ROUGH SETS BASED EVALUATION MODEL FOR BOT PROJECTS BIDDERS

Yonghe Yu^{1,2}, Miyuan Shan¹

1.School of Management, Hunan University, Changsa 410082, China

2.Department of Civil Engineering, Guangxi University of Technology, Liuzhou 545006, China

Abstract: Bidding has been used as an optimal means in selecting investor of build, operate and transfer (BOT) projects in recently. An accurate evaluation on bidders is the key point of a successful bidding. However, evaluation on bidders is a tough and comprehensive work and the result of the evaluation is always affected by evaluators' understanding on standards set for evaluation, evaluators' expertise, experience and preference. In this paper, a hybrid and rough sets based method is presented to facilitate the discretization for continuous attributes and feature selection in evaluation. After a quick review on basic notions of rough sets, general procedures of the method are illustrated by an example, which shows the method's validity.

Key words: Evaluation Model, BOT Projects, Rough Sets

1. INTRODUCTION

The build, operate and transfer (BOT) concession model has been seen as a major trend during the recent two decades in the privatization of public sector infrastructure projects. This is particularly so in the rapidly developing regions. Though the recent financial and currency turmoil in Asia have caused a slowdown in their economies, most promoters and lenders of BOT projects are still confident of the role of the private sector in funding infrastructure projects. This is because the BOT mechanism offers host governments an opportunity to accelerate infrastructure development without incurring large public expenditure and borrowing. At the same time it also creates new business openings for the private sector, especially the

construction industry, to participate more positively in the world economic development and contribute to the quality of life of the community. However, the road to winning a BOT concession is by no means easy. Pre-award negotiation is often protracted and arduous. Post-award implementation is usually subjected to unforeseen uncertainties. In the recent financial crisis, a few mega-sized infrastructure projects, such as the Bakun Dam in Malaysia and power plants in Indonesia, were canceled or suspended indefinitely due to funding difficulties [1].

The host governments must take a long view in such contract negotiation, because of the long duration and high capital costs of infrastructure projects and changing priority of the bidders. Upon successful construction of the projects, the actual concession period can last for 10-50 years depending on the type of project. During these operational phases, the maintenance of throughput and tariff rates in line with original forecast remains an area of great uncertainty and challenge to the promoters. BOT projects are a high-risk and high return business. A substantial risk premium, significantly exceeding the normal minimum attractive rate of return, is usually expected by both host governments and bidders.

Evaluation of tender proposals for privatized infrastructure projects that are let under the build, operate and transfer (BOT) concept involves decision making by different government officials from different departments and by their consultants. Choosing the best proposal would depend on the three elements: the quality of the definition of specific criteria, the quality of evaluation of the available alternatives, and the quality of the understanding of what these alternatives can produce. The purpose of the evaluation and selection process is to identify what needs to be done, develop the specific criteria for its accomplishment, and identify the advantages [2].

Many factors should be considered when making decisions on selecting the final contractor. During the decision-making, many factors concerning bidders cannot be measured by tenderer directly and some of them are incomplete and irrelevant, which makes it difficult for tenderer to judge. Even though, an evaluation indicator can be set up but the attributes values are difficult to deal with. Some values are continuous, which makes it more difficult to input the model to calculate and judge.

Our research focuses on a hybrid rough sets based evaluation model to help the hosts governments to evaluate the bidders in a comprehensive and high efficiency way. This paper consists of five sections. Section 2 proposes the evaluation model including the motivation, relevant theories and general

algorithms and procedures of the model. An instance is presented to illustrate how to use the model in details in Section 3. Conclusions and discussion for further research hints are included in the last section.

2. PROPOSED METHODOLOGIES

2.1 Motivations

Due to the difficulties encountered in calculating the evaluation value for bidder of BOT projects, our proposed model cuts to the points from the perspective of two basic data mining issues: discretization and feature selection.

Defining classification rules is an important procedure in data mining. Discretization and feature selection are effective techniques in dealing with continuous attributes for rule generating. Many classification rule generation systems can only deal with discrete data, such as AQ, ID3. But continuous attributes seem to be common when mining data. Efforts have been made on discretization methods by using some novel techniques such as fuzzy theory, entropy, genetic algorithm and so on.

Generally, most of the discretization methods fall into two categories: local and global[3]. Local methods discretize only one attribute at a time, while the global ones discretize all attributes simultaneously. In general local discretization is relatively easy to implement and has been adopted in many applications. However, it is considered as a sub-optimal approach because of the lost of valuable information of other attributes[4, 5]. On the other hand, global discretization can usually get better results than the local one for taking interact of all attributes into account. But the costly computation limits its widely application in practice.

Another issue related to data mining is concerned with feature selection. Generally, a lot of features (attributes, variables) about the target object are stored and kept for later analysis. However, these features gathered to identify objects are often not all useful or equally informative: some of them may be redundant or irrelevant for the identification. Many candidate features are included since the relevant features are unknown. Unfortunately, the presence of more information in the form of additional features does not always result in a better revelation of the hidden natural patterns in the data. On the contrary, the presence of irrelevant features can mask the underlying

natural patterns in the data, and the operators will have to look through all the data to ascertain the quality of objects during data mining. This kind of job is normally tedious, and will remarkably reduce the quality of data mining. Under such a situation, a feature selection technique can help discover those important features that best describe the target object and thus can simplify the mining procedure. Feature selection addresses this concern by the selection of a subset of features that is relevant. The importance of feature selection in a broader sense is not only to simplify the procedure of data mining, but also to reduce the search space and save future computation time and costs. Hence, to some extent, feature selection can be viewed as a pre-pruning process to choose minimally sized features that are necessary and sufficient to describe target concepts [6]. In recent decades, many feature selection methods have been developed (overviews can be found in Refs.[7,8]). These methods have different strengths and drawbacks, based on the specific feature selection criterion.

2.2 Rough sets

The rough sets philosophy is founded on the assumption that with every object of the universe of discourse we associate some information (data, knowledge). Rough set theory overlaps with many other theories, especially with fuzzy set theory, evidence theory and Boolean reasoning methods, discriminant analysis, which was first proposed by Poland scholar Z. Pawlak in 1982 and is regarded as a new soft computing method to deal with imprecise, vagueness and uncertainty data. Due to its advantage, which include the elimination of the need for additional information about data and the ability to extract rules directly from data itself, this theory has been used in more and more domains in less than two decades, such as medical diagnosis, stock market forecast, pattern recognition, machine learning, decision making and etc [9, 10].

A decision table is the basic kind of knowledge in Rough set theory. Let $S=(U, A \cup \{d\})$ be a decision table with n objects, where U is a non-empty, finite set called the universe and A is a non-empty, finite set of condition attributes, i.e. $a: U \rightarrow v_a$ for $a \in A$, where v_a is called the value of a , and $d \notin A$ is a distinguished attributes called decision. And the other concepts of rough set theory such as equivalence class, indiscernibility relation, approximation of sets, can be found in Refs [9,10,11].

2.3 Discretization

General discretization could be described as follows:

Let $[a, b]$ be the value domain of an continuous attributes A . Discretization of A is a process of searching for an optimal (with respect to a given evaluation measure) cut points set $\{c_1, c_2, \dots, c_m\}$, where $a < c_1 < c_2 < \dots < c_m < b$, which divides $[a, b]$ into m disjoin intervals $[a, c_1]$, $[c_1, c_2]$, ..., $[c_m, b]$. Then the continuous values of A are transformed into m different discrete values $v_1, v_2 \dots v_m$.

$$DV(obj) = \begin{cases} v_1 & \text{if } a \leq V(obj) \leq c_1 \\ v_i & \text{if } c_{i-1} < V(obj) \leq c_i \\ v_m & \text{if } c_{m-1} < V(obj) \leq c_m \end{cases}$$

$i=2,3 \dots m-1$. $V(obj)$ is the original continuous value of an object on A and $DV(obj)$ is the discretized value of A .

The Rough Sets Theory cannot be used to deal with the continuous attributes, which is the disadvantage of this method[12]. The crucial problem now is how to discretize them therefore. Usually, the value of attribute changes monotonically with deterioration of the state. Boundary values to divide its domain into intervals corresponding to different state are selected by experience. However, when experience to discretize the domain is not present, other discretization methods have to be used. There are many discretization methods, such as equal-width-intervals, equal-frequency-intervals and Minimal Entropy Method. In this paper, MD-discretization method is used to discretize the attributes . The algorithm is presented below.

A continuous attribute $a \in A$ defines a sequence $v_1^a < v_2^a \dots < v_{n_a}^a$, where

$$\{v_1^a, v_2^a, v_{n_a}^a\} = \{a(x) : x \in U\} \quad \text{and } n_a \leq n.$$

Then the set of all possible cuts on a is denoted by

$$C_a = \{(a, \frac{v_1^a + v_2^a}{2}), (a, \frac{v_1^a + v_2^a}{2}), \dots, (a, \frac{v_1^a + v_2^a}{2})\}$$

The set of possible cuts on all attributes is denoted by

$$C_A = \bigcup_{a \in A} C_a$$

Algorithm 1 MD-discretization

Input: the consistent decision table $S = (U, A \cup \{d\})$, $A = \{a_1, a_2 \dots, a_m\}$

Output: the semi-minimal set of cuts D consistent with S

Method:

$D = \Phi$; CA = initial set of cuts on S ;

$L = \{(x, y) \in U^2 : d(x) \neq d(y)\}$;

While ($L \neq \Phi$) **do**

Begin

Choose the cut $C_{\max} \in C_A$ that discerns the largest number of pairs of objects in L (i.e. the cut corresponding to the column with the largest number of "1")

Insert $C_{\max}$ into D and remove it from C_A .

Remove all pairs of objects from L discerned by $C_{\max}$.

End

2.4 Feature selection

In this section, the computation of feature selection in decision table is discussed. The resulting selection is a minimal set of attributes that enable to represent the whole set of attributes in a decision table to be considered when generation the rules from tables. The problem of computing all features from a decision system has been proved to be NP-Hard. Several new methods have been proposed such as MIBARK algorithm, feature selection-based algorithm and etc to improve the efficiency of basic feature identification. In this paper, a more effective algorithm is used.

Definition1 Let $\Psi = \{Y_1, Y_2, \dots, Y_k\}$ be a partition of U , $P \subseteq A$.

$$\gamma_P(\Psi) = \frac{\sum_{i=1}^k \text{card}(\underline{PY}_i)}{\text{card}(U)}$$

The coefficient is called the quality of approximation of partition Ψ by set of attributes P , or in short, quality of sorting. It expresses the ratio of all P -correctly sorted objects to all objects in the system.

Definition2 Let $R \subseteq A$, $\gamma_R(\Psi) = \gamma_A(\Psi)$ if not exist another set $R' \subset R$, where $\gamma_{R'}(\Psi) = \gamma_R(\Psi)$, then R is a reduction of A . The core of A denoted by $\text{CORE}(A)$, is the set attributes, which appears in all reduction of A .

Algorithm 2 Feature-Selection

Input: the consistent decision table $S = (U, A \cup \{d\})$, $A = \{a_1, a_2, \dots, a_m\}$, $\Psi = \{Y_1, Y_2, \dots, Y_k\}$ is a partition $U/\{d\}$

Output: a set of reduction P covering U

Method:

Step1: For $i=1$ to m do

Begin

Compute $\gamma_{a_i}(\Psi)$, $Q = \text{CORE}(A)$, $P = Q$, $T = A \setminus P$.

If $\gamma_R(\Psi) = \gamma_A(\Psi)$ **then** P is a reduction P covering U , **return end.**

End

Step2: Compute $\gamma_{a_g}(\psi) = \max(\gamma_{a_i}(\psi) : a_i \in T)$,

Let $P = P \cup \{a_g\}$, $T = T \setminus \{a_g\}$, go to step3.

Step3: If $\gamma_R(\psi) = \gamma_A(\psi)$ then go to step4 else $\gamma_R(\psi) < \gamma_A(\psi)$, go to step2 end.

Step4: If $P = CORE(P)$ then P is a reduction P covering U, return else go to step5 end.

Step5: For any condition attribute $a_j \in P \setminus CORE(P)$, $P = P \setminus \{a_j\}$, go to step4.

2.5 Inconsistency checking

To assure the effectiveness of discretization and feature selection, inconsistency checking is a necessary procedure, which was used in many literatures. The followings are the general procedures of inconsistency checking and its relationship with decision tables under the context of rough sets.

Let a decision table $S = (U, A \cup \{d\})$, $U = \{U_1, U_2, \dots, U_m\}$ is the universe which consists of a finite set of objects, $A = \{A_1, A_2, \dots, A_n\}$ is the set of conditional attributes and d is the decision attribute (this is also called the class of objects). $A_i(U_j)$ denotes the value of object U_j on attributes A_i and $d(U_j)$ is the class of U_j .

An object $U_i \in U$ is called inconsistent in S if an object $U_j \in U$ can be found which satisfied the following conditions:

$A_k(U_i) = A_k(U_j)$ and $d(U_i) \neq d(U_j)$, $i, j \in \{1, 2, \dots, m\}$, $k = 1, 2, \dots, n$.

We define $Iobj(S)$ as the set of all inconsistent objects in S , and then inconsistency of S can be defined as

$$INC(S) = \frac{|Iobj(S)|}{|U|}$$

A given δ is the criteria to judge the inconsistency of discretization and feature selection.

3. AN INSTANCE

3.1 Inconsistency checking

To accelerate the economic development of Guangxi, the local government invested more than 4 billion on constructing highway to connect the Guangxi with Guangdong province, Hunan province and Yunnan province. In order to improve the investment efficiency and guarantee the fairness of competition, bidding is employed as a common method to select the contractors.

3.2 Attributes of Evaluation Model

Attribute 1: project cost (measured in million yuan) $U(1)$ — includes the cost of labors and facilities;

Attribute 2: time limit for project (measured in days) $U(2)$ — indicates how long the project will take;

Attribute 3: scheme of technology $U(3)$ — refers to methodology and experiment routines used in construction;

Attribute 4: sets of rules for quality control $U(4)$ — to judge whether it is convincing to guarantee the project's quality;

Attribute 5: time needed for preparing the materials and staffs (measured in days) $U(5)$;

Attribute 6: proportion of senior engineers $U(6)$ — indicates the proportion of senior engineers among all the professionals involved;

Attribute 7: capacity to prepay by cash $U(7)$ — to judge the financial situation of the bidders;

Attribute 8: history record and experiences $U(8)$ — to be judged by how many similar contracts the bidders have taken;

Attribute 9: expertise of management and professionals $U(9)$ — to judge the qualification of the bidders' human resources.

3.3 Calculus

In order to test our method, we performed an instance. Suppose $S=(U, A \cup \{d\})$ (as shown in table 1) is a decision table for evaluation, where $A=\{a_1, a_2, \dots, a_4\}$ and $U=\{u_1, u_2, \dots, u_9\}$. Attributes a_2 and a_4 are continuous attributes, and a_1 and a_3 are discrete attributes.

Table 1. The decision table

U	a1	a2	a3	a4	d
U1	3	1.1	3	2.5	1
U2	1	1.1	2	3.5	0
U3	2	1.8	2	1.5	0
U4	2	1.5	1	1.5	1
U5	3	1.5	1	2.5	0
U6	2	1.6	2	3.5	1
U7	3	1.6	1	3.5	1
U8	3	1.3	2	0.5	1
U9	1	1.3	1	0.5	0

The set of all cuts consists of four cuts on the attribute a_2 and three cuts on the attribute a_4 .

$$C_A = \{(a_2, 1.2), (a_2, 1.4), (a_2, 1.55), (a_2, 1.7)\} \cup \{(a_4, 1), (a_4, 2), (a_4, 3)\}$$

Table 2 The new decision table S_1

L	f(a2,1.2)	f(a2,1.4)	f(a2,1.55)	f(a2,1.7)	f(a4,1)	f(a4,2)	f(a4,3)
(u1,u2)	0	0	0	0	0	0	1
(u1,u3)	1	1	1	1	0	1	0
(u1,u5)	1	1	0	0	0	0	0
(u1,u9)	1	0	0	0	1	1	0
(u2,u4)	1	1	0	0	0	1	1
(u2,u6)	1	1	1	0	0	0	0
(u2,u7)	1	1	1	0	0	0	0
(u2,u8)	1	0	0	0	1	1	1
(u3,u4)	0	0	1	1	0	0	0
(u3,u6)	0	0	0	1	0	1	1
(u3,u7)	0	0	0	1	0	1	1
(u3,u8)	0	1	1	1	1	0	0
(u4,u5)	0	0	0	0	0	1	0
(u4,u9)	0	1	0	0	1	0	0
(u5,u6)	0	0	1	0	0	0	1
(u5,u7)	0	0	1	0	0	0	1
(u5,u8)	0	1	0	0	1	1	0
(u6,u9)	0	1	1	0	1	1	1
(u7,u9)	0	1	1	0	1	1	1
(u8,u9)	0	0	0	0	0	0	0

The new decision table S1 consists of 7 attributes and 20 objects(table 2). The MD-discretization algorithm choose the cut (a2,1.4) first because it discerns 10 pairs of objects, then it chooses the cuts (a2, 1.2) , (a2, 1.55), (a4, 2) and (a4, 3). The result of discretization is the set of cuts $S=\{(a2, 1.2), (a2, 1.4), (a2, 1.55), (a4, 2), (a4, 3)\}$ using algorithm1.

$$\Psi = \{Y_1, Y_2\} = \{\{u_1, u_4, u_6, u_7, u_8\}, \{u_2, u_3, u_5, u_7\}\}$$

$$A=\{A1, A2, A3, A4\}=\{\{a1\}\{a2\}\{a3\}\{a4\}\}$$

$$A1Y1 = \Phi, A2Y1 = \{2, 9\}$$

$$\gamma_{a1}(\psi) = \frac{\sum_{i=1}^2 \text{card}(\underline{A_1}Y_i)}{\text{card}(U)} = 2/9$$

Using the same method to compute $\gamma_{a2}(\psi)$, $\gamma_{a3}(\psi)$ and $\gamma_{a4}(\psi)$. The result of feature reduction is the set of $P=\{a1, a3, a4,\}$ using the algorithm2, and the fitting rate of condition attribute is 20%. The result shows that condition attribute a2 is redundant or irrelevant for the evaluation system.

4. CONCLUSION

We have presented a hybrid and rough set based discretion and feature selection method with an inconsistency checking for evaluation. By employing rough set, the efficiency of discretion and feature selection is improved and more valuable information in raw data are reserved. An instance is raised to illustrate the method's validity, which implies further application of this method.

ACKNOWLEDGEMENTS

This research is supported by the National Nature Science Foundation of China(No. 70372039) and Guangxi Nature Science Foundation (No. 0481005).

REFERENCES

- [1] Neal J. Financing of BOT projects. M.Sc. lectures, Nanyang Technological University, Singapore, 1993.
- [2] Robert L.K Tiong, Jahjdul Alum. Evaluation of proposals for BOT projects. Int. J. project management, 1997,15(2): 67-72.
- [3] R Susmaga, Analyzing. Discretizations of continuous attributes given a monotonic discrimination function. Intelligent Data Analysis, 1997,1: 157-179.

- [4] S.D. Bay. Multivariate discretization of continuous variables for set mining. Proceedings of the sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Boston, MA, August), 2000: 315-319.
- [5] S.J. Hong. Use of contextual information for feature ranking and discretization. *IEEE Transactions on knowledge and data engineering*, 1997,9: 718-730.
- [6] Lian Yin Zhai et al, Feature extraction using rough set theory and genetic algorithms — an application for the simplification of product quality evaluation, *Computers & Industrial Engineering*, 2002: 43.
- [7] B. Walczak, D.L. Massart, Chapter calibration in wavelet domain, in: B. Walczak (Ed.), *Wavelets in Chemistry*, Elsevier, Amsterdam, 2000: 323–349.
- [8] R. Kohavi, G. John. Wrappers for feature subset selection. *Artificial Intelligence*. 1997: 273–324.
- [9] Z. Pawlak, *Rough Sets, Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishing, 1991.
- [10] R. Slowinski, *Intelligent Decision Support, Handbook of Applications and Advances of Rough Sets Theory*, Kluwer Academic Publishing, 1992.
- [11] S. Tsumoto, H. Tanaka, Automated discovery of medical expert system rules from clinical databases based on Rough Sets, in: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining KDD'96* , AAAI Press, Menlo Park, 1996.
- [12] Yao, Y. Y., Wong, S. K. M., & Lin, T. Y, A review of rough sets models. In T. Y. Lin, & N. Cercone (Eds.), *Rough sets and data mining—Analysis for imprecise data*, Boston, MA: Kluwer, 1997.
- [13] Ye dongyi, An improvement to Jelonek's attributes reduction algorithm, *Acta electronica sinica*, 2000,28(12).