A Framework on Rough Set-Based Partitioning Attribute Selection

Tutut Herawan and Mustafa Mat Deris

Faculty of Information Technology and Multimedia,
Universiti Tun Hussein Onn Malaysia,
Parit Raja, 86400, Batu Pahat, Johor, Malaysia
tututherawan@yahoo.com, mmustafa@uthm.edu.my

Abstract. In this paper, we focus our discussion on the rough set-based partitioning attribute selection. Firstly, we point out that the statement of MMR technique is an extension of Mazlack's technique is unreasonable. We prove that the mean roughness of MMR technique is only the opposite of that Mazlack's TR technique. Secondly, we observe that the suggestion of MMR to achieve lower computational complexity using the roughness measurement based on relationship between an attribute $a_i \in A$ and the set defined as $A-\{a_i\}$ instead of calculating the maximum with respect to all $\{a_j\}$ where $a_i \neq a_j$, $1 \le i, j \le |A|$ only can be applied to a special type of information system and we illustrate this with an example. Finally, we propose an alternative technique for selecting partitioning attribute using rough set theory based on dependency of attributes in an information system. We show that the proposed technique is a generalization and has lower computational complexity than that of TR and MMR.

Keywords: Rough set theory; Dependency of attributes.

1 Introduction

Recently, there has been works in the area of applying rough set theory in the process of selecting partitioning attribute. Mazlack *et al.* [1] proposes two techniques to select partitioning attribute: i.e., Bi-Clustering (BC) technique based on bi-valued attributes and Total Roughness (TR) technique. Mazlack *et al.* suggested that BC technique will be attempted first in order to achieve low dissonance inside the partition. With this technique, there are three different approaches of selecting partitioning attribute, i.e. arbitrary, imbalanced, and balanced. For balanced or unbalanced partitioning approaches, it is likely that two kinds of problems may occur. First, it may have several candidates of bi-partitioning attributes. Then, a decision has to be made to which one should be chosen as the partitioning attribute. Second, no two-valued attribute can be found to form balance partitioning. At this point, partitioning on multiple valued attributes will be considered. Therefore, for selecting partitioning attribute for data set with multiple-valued attributes, Mazlack *et al.* proposed a technique using the average

D.-S. Huang et al. (Eds.): ICIC 2009, LNAI 5755, pp. 91–100, 2009.

of the accuracy of approximation (accuracy of roughness) in the rough set theory [2-4] called Total Roughness (TR). In other words, it is based on the average of mean roughness of an attribute with respect to the set of all other attributes in an information system, where the higher the total roughness is, the higher the accuracy of selecting partitioning attribute. In [5], Parmar *et al.* proposes a new technique called Min-Min Roughness (MMR) for categorical data clustering. In this technique, bi-valued and multi-valued attributes are equally treated and it's accuracy of approximation is measured using the well-known Marczeweski-Steinhaus metric [6-8] applied to the lower and upper approximations of a subset of the universe in an information system. However, the mean roughness MMR is only the opposite of that TR. Consequently, they produce the same accuracy and complexity for selecting partitioning attribute. In addition, with MMR technique, the complexity is however still an issue due to all attributes are considered to obtain the partitioning attribute. Further, we observe that the suggestion of MMR to achieve lower computational complexity using the roughness measurement based on relationship between an attribute $a_i \in A$ and the set defined as $A - \{a_i\}$ instead of calculating the maximum with respect to all $\{a_j\}$ where $a_i \neq a_j$, $1 \leq i, j \leq |A|$ only can be applied to a special type of information system and we illustrate this with an example. Therefore, there is a need for a technique in selecting partitioning attribute to improve the TR and MMR. One way to select partitioning attribute is to discover the relationship among attributes based on their dependencies. In this paper, a technique called Maximum Dependency of Attributes (MDA) is proposed. It is based on the dependency of attributes using rough set theory in an information system. A test case is considered to evaluate and compare the performance of MDA with TR and MMR techniques. We show that the proposed technique provides better performance with that of TR and MMR techniques.

The rest of this paper is organized as follows. Section 2 describes rough set theory. Section 3 describes the analysis and comparison of TR and MMR techniques. Section 4 describes the Maximum Dependency of Attributes (MDA) technique. A comparison test of MDA with BC, TR and MMR techniques is described in section 5. Finally, the conclusion of this work is described in section 6.

2 Rough Set Theory

An information system as in [4] is a 4-tuple (quadruple) $S = (U, A, V, f)$, where U is a non-empty finite set of objects, A is a non-empty finite set of attributes, $V = \bigcup_{a \in A} V_a$, V_a is the domain (value set) of attribute a, $f : U \times A \to V$ is a total function such that $f(u,a) \in V_a$, for every $(u,a) \in U \times A$, called information (knowledge) function.

Definition 1. *Two elements $x, y \in U$ are said to be B-indiscernible (indiscernible by the set of attribute $B \subseteq A$ in S) if and only if $f(x,a) = f(y,a)$, for every $a \in B$.*

Obviously, every subset of A induces unique indiscernibility relation. Notice that, an indiscernibility relation induced by the set of attribute B, denoted by $IND(B)$, is an equivalence relation. The partition of U induced by $IND(B)$ is denoted by U/B and the equivalence class in the partition U/B containing $x \in U$, is denoted by $[x]_B$. The notions of lower and upper approximations of a set are defined as follows.

Definition 2. (See [4].) *The B-lower approximation of X, denoted by* $\underline{B}(X)$ *and B-upper approximations, denoted by* $\overline{B}(X)$ *of X, respectively, are defined by*

$$\underline{B}(X)=\{x \in U \mid [x]_B \subseteq X\} \text{ and } \overline{B}(X)=\{x \in U \mid [x]_B \cap X \neq \phi\}. \quad (1)$$

The accuracy of approximation (accuracy of roughness) of any subset $X \subseteq U$ with respect to $B \subseteq A$, denoted $\alpha_B(X)$ is measured by

$$\alpha_B(X)=\left|\underline{B}(X)\right| / \left|\overline{B}(X)\right|, \quad (2)$$

where $|X|$ denotes the cardinality of X. The higher of accuracy of approximation of any subset $X \subseteq U$ is the more precise (the less imprecise) of itself. The accuracy of roughness in equation (2) can also be interpreted using the well-known Marczeweski-Steinhaus (MZ) metric [6-8]. By applying the Marczeweski-Steinhaus metric to the lower and upper approximations of a subset $X \subseteq U$ in information system S, we have

$$D\left(\underline{R}(X), \overline{R}(X)\right)=1-\frac{\left|\underline{R}(X) \cap \overline{R}(X)\right|}{\left|\underline{R}(X) \cup \overline{R}(X)\right|}=1-\frac{\left|\underline{R}(X)\right|}{\left|\overline{R}(X)\right|}=1-\alpha_R(X). \quad (3)$$

Definition 3. *Let* $S=(U, A, V, f)$ *be an information system and let D and C be any disjoint subsets of A. Dependency attribute D on C in a degree k* $(0 \leq k \leq 1)$, *is denoted by* $C \Rightarrow_k D$. *The degree k is defined by*

$$k=\frac{\sum_{X \in U/D}\left|\underline{C}(X)\right|}{|U|}. \quad (4)$$

Attribute D is said to be fully depends (in a degree of k) on C if $k=1$. *Otherwise, D is partially depends on C.*

Thus, D fully (partially) depends on C, if all (some) elements of the universe U can be uniquely classified to equivalence classes of the partition U/D, employing C. Based on Definition 3, we can select the partitioning attributes based on the maximum value of k.

3 The Analysis and comparison of TR and MMR Techniques

In this section, we analyze and compare the Total Roughness (TR) and Min-Min Roughness (MMR) techniques.

3.1 The TR Technique

The definition of information system is based on the notion of information system as stated in section 2. From the definition, suppose that attribute $a_i \in A$ has *k*-different values, say β_k, $k = 1,2,\cdots,n$. Let $X(a_i = \beta_k)$, $k = 1,2,\cdots,n$ be a subset of the objects having *k*-different values of attribute a_i. The roughness of TR technique of the set $X(a_i = \beta_k)$, $k = 1,2,\cdots,n$, with respect to a_j, where $i \neq j$, denoted by $\mathrm{R}_{a_j}(X|a_i = \beta_k)$, is defined by

$$\mathrm{R}_{a_j}(X|a_i = \beta_k) = \frac{\left|\underline{X_{a_j}}(a_i = \beta_k)\right|}{\left|\overline{X_{a_j}}(a_i = \beta_k)\right|} \quad k = 1,2,\cdots,n \tag{5}$$

From TR technique, the mean roughness of attribute $a_i \in A$ with respect to attribute $a_j \in A$, where $i \neq j$, denoted $\mathrm{Rough}_{a_j}(a_i)$, is evaluated as follows

$$\mathrm{Rough}_{a_j}(a_i) = \frac{\sum_{k=1}^{|V(a_i)|} \mathrm{R}_{a_j}(X|a_i = \beta_k)}{|V(a_i)|}, \tag{6}$$

where $V(a_i)$ is the set of values of attribute $a_i \in A$.

The total roughness of attribute $a_i \in A$ with respect to attribute $a_j \in A$, where $i \neq j$, denoted $\mathrm{TR}(a_i)$, is obtained by the following formula

$$\mathrm{TR}(a_i) = \frac{\sum_{j=1}^{|A|} \mathrm{Rough}_{a_j}(a_i)}{|A| - 1}. \tag{7}$$

As stated in Mazlack *et al.* [1], the highest value of TR, the best selection of partitioning attribute.

3.2 The MMR Technique

The definition of information system is based on the notion of information system as stated in section 2. From the definition, suppose that attribute $a_i \in A$ has *k*-different values, say β_k, $k = 1,2,\cdots,n$. Let $X(a_i = \beta_k)$, $k = 1,2,\cdots,n$ be a subset of the objects having *k*-different values of attribute a_i. The roughness of MMR technique of the set $X(a_i = \beta_k)$, $k = 1,2,\cdots,n$, with respect to a_j, where $i \neq j$, denoted by $\mathrm{R}_{a_j}(X|a_i = \beta_k)$, is defined by

$$\mathrm{MMR}_{a_j}(X|a_i = \beta_k) = 1 - \frac{\left|\underline{X_{a_j}}(a_i = \beta_k)\right|}{\left|\overline{X_{a_j}}(a_i = \beta_k)\right|}, \; k = 1,2,\cdots,n. \tag{8}$$

It is clear that MMR technique uses MZ metric to measure the roughness of the set $X(a_i = \beta_k)$, $k = 1,2,\cdots,n$, with respect to a_j, where $i \neq j$.

The mean roughness of MMR technique is defined by

$$\text{MMRough}_{a_j}(a_i) = \frac{\sum_{k=1}^{|V(a_i)|} \text{MMR}_{a_j}(X|a_i = \beta_k)}{|V(a_i)|}. \tag{9}$$

According to Parmar *et al.* [5], the least mean roughness, the best selection of partitioning attribute.

Proposition 4. *The value of roughness of MMR technique is the opposite of that TR technique.*

Proof. Since MMR technique uses MZ metric to measure the roughness of the set $X(a_i = \beta_k)$, $k = 1,2,\cdots,n$, with respect to a_j, where $i \neq j$, i.e.,

$$\text{MMR}_{a_j}(X|a_i = \beta_k) = 1 - \frac{\left|\underline{X_{a_j}(a_i = \beta_k)}\right|}{\left|\overline{X_{a_j}(a_i = \beta_k)}\right|},$$

then from (5) and (8), we have

$$\text{MMR}_{a_j}(X|a_i = \beta_k) = 1 - \text{R}_{a_j}(X|a_i = \beta_k). \tag{10}$$

Thus, the value of mean roughness of MMR technique (9) is also the opposite of that TR technique (6), i.e.,

$$\begin{aligned}
\text{MMRough}_{a_j}(a_i) &= \frac{\sum_{k=1}^{|V(a_i)|} \text{MMR}_{a_j}(X|a_i = \beta_k)}{|V(a_i)|} \\
&= \frac{\sum_{k=1}^{|V(a_i)|} \left(1 - \text{R}_{a_j}(X|a_i = \beta_k)\right)}{|V(a_i)|} \\
&= \frac{\sum_{k=1}^{|V(a_i)|} 1 - \sum_{k=1}^{|V(a_i)|} \text{R}_{a_j}(X|a_i = \beta_k)}{|V(a_i)|} \\
&= \frac{|V(a_i)|}{|V(a_i)|} - \frac{\sum_{k=1}^{|V(a_i)|} \text{R}_{a_j}(X|a_i = \beta_k)}{|V(a_i)|} \\
&= 1 - \text{Rough}_{a_j}(a_i), \text{ for } i \neq j.
\end{aligned} \tag{11}$$

The MMR technique is based on the minimum value of mean roughness in (11), without calculating total roughness (7). □

This analysis and comparison seems to show that TR and MMR are given the same result. On the other hand, with MMR technique, to achieve lower computational complexity in selecting partitioning attribute in an information system $S = (U, A, V, f)$, it is suggested to measure the roughness based on relationship between an attribute $a_i \in A$ and the set defined as $A - \{a_i\}$ instead of calculating the maximum with respect to all $\{a_j\}$ where $a_i \neq a_j$. We observe this technique only can be applied in a very special data set. To illustrate this problem, we consider to the following example.

Example 5. We consider to the data set in illustrative example of Table 2 in [5]. The calculation of MMR is based on formulas in equations (8) and (9). According to Parmar *et al.* attribute a_1 is chosen as the clustering (partitioning) attribute. However, if we consider to measure the roughness of attribute $a_i \in A$ with respect to the set of attributes $A - \{a_i\}$, then we get the value of MMR as in Table 1.

Table 1. The modified MMR of all attributes in data set from [5]

Attribute w.r.t.	Mean Roughness	MMR
a_1	Rough $A - \{a_1\}$ 0	0
a_2	Rough $A - \{a_2\}$ 0	0
a_3	Rough $A - \{a_3\}$ 0	0
a_4	Rough $A - \{a_4\}$ 0	0
a_5	Rough $A - \{a_5\}$ 0	0
a_6	Rough $A - \{a_6\}$ 0	0

Based on Table 1, we cannot select a partitioning attribute. On the other hand, the suggested technique would lead a problem, i.e., after calculation of mean roughness of attribute $a_i \in A$ with respect to the set of attributes $A - \{a_i\}$, the value of MMR usually cannot preserve the original decision. Thus, this modified technique is not relevant to all type of data set.

In section 4, we introduce the Maximum Dependency of Attributes (MDA) technique to deal with the problem of selecting partitioning attribute.

4 Maximum Dependency of Attributes (MDA) Technique

The MDA technique for selecting partitioning attribute is based on the maximum degree of dependency of attributes. The justification that the higher of the degree of dependency of attributes implies the more accurate for selecting partitioning attribute is stated in the Proposition 6.

Proposition 6. *Let* $S=(U,A,V,f)$ *be an information system and let D and C be any subsets of A. If D depends totally on C, then*

$$\alpha_D(X)\le\alpha_C(X),\, \textit{for every } X\subseteq U.$$

Proof. Let D and C be any subsets of A in information system $S=(U,A,V,f)$. From the hypothesis, we have $IND(C)\subseteq IND(D)$. Furthermore, the partitioning U/C is finer than that U/D, thus, it is clear that any equivalence class induced by $IND(D)$ is a union of some equivalence class induced by $IND(C)$. Therefore, for every $x\in X\subseteq U$, we have $[x]_C\subseteq[x]_D$. And hence, for every $X\subseteq U$, we have

$$\underline{D}(X)\subseteq\underline{C}(X)\subset X\subset\overline{C}(X)\subseteq\overline{D}(X).$$

Consequently

$$\alpha_D(X)=\frac{|\underline{D}(X)|}{|\overline{D}(X)|}\le\frac{|\underline{C}(X)|}{|\overline{C}(X)|}=\alpha_C(X).$$ □

Proposition 7. *MDA is a generalization of TR and MMR.*

Proof. In an information system, $S=(U,A,V,f)$, an attribute $a_i\in A$ may have several different values, say β_k, $k=1,2,\cdots,n$. Thus from equations (2) and (5), we can generalize the roughness of the sets as follows

$$R_{a_j}(X)=\frac{|(Xa_j(a_i=\beta_1))|}{U}+\frac{|(Xa_j(a_i=\beta_2))|}{U}+\cdots+\frac{|(Xa_j(a_i=\beta_n))|}{U}$$
$$=\frac{\sum_{U/a_i}|X(a_i=\beta_k)|}{U},\quad k=1,2,\cdots,n. \tag{12}$$

The formula in equation (12) is a degree of dependency of attribute a_j on attribute a_i, where $i\neq j$. Thus, MDA is a generalization of TR and MMR. □

5 A Comparison Test

To this framework, a small-sized data set from [9] is considered to compare and evaluate the accuracy and the complexity of MDA with TR, MMR and in addition with BC techniques. To measure the accuracy of selecting partitioning attribute, we use the formula of mean roughness in equation (6) to represent all techniques. The higher the mean roughness is the higher the accuracy of the selecting partitioning attribute.

In Table 2, there are five categorical attributes: Magazine Promotion (MP), Watch Promotion (WP), Life Insurance Promotion (LIP), Credit Card Insurance (CCI) and Sex (S) and ten objects are considered. Notice that with the BC technique, the attribute with the least distinct balanced-value will be selected as a partitioning attribute without consideration of the maximum value of total roughness of each attributes. Thus, for BC technique, attribute LIP will be chosen as a partitioning attribute. Meanwhile, the TR, MMR and MDA are based on calculations using formula (7), (9) and (4), respectively. The results of the calculation of TR, MMR and MDA of all attributes in Table 2 are summarized in Tables 3, 4 and 5, respectively.

Table 2. The Credit Card Promotion dataset from [9]

#	MP	WP	LIP	CCI	S
1	yes	no	no	no	male
2	yes	yes	yes	no	female
3	no	no	no	no	male
4	yes	yes	yes	yes	male
5	yes	no	yes	no	female
6	no	no	no	no	female
7	yes	no	yes	yes	male
8	no	yes	no	no	male
9	yes	no	no	no	male
10	yes	yes	yes	no	female

Table 3. The value of TR of all attributes

Attribute w.r.t.	TR mean roughness				TR
MP	WP	LIP	CCI	S	0.0875
	0	0.25	0.1	0	
WP	MP	LIP	CCI	S	0
	0	0	0	0	
LIP	MP	WP	CCI	S	0.0625
	0.15	0	0.1	0	
CCI	MP	WP	LIP	S	0.1500
	0.15	0	0.25	0.2	
S	MP	WP	LIP	CCI	0.02500
	0	0	0	0.1	

Table 4. The value of MMR of all attributes

Attribute w.r.t.	MMR mean roughness				MMR
MP	WP	LIP	CCI	S	0.75
	1	0.75	0.9	1	0.9
WP	MP	LIP	CCI	S	1
	1	1	1	1	
LIP	MP	WP	CCI	S	0.85
	0.85	1	0.9	1	
CCI	MP	WP	LIP	S	0.75
	0.85	1	0.75	0.8	0.8
S	MP	WP	LIP	CCI	0.9
	1	1	1	0.9	

As shown in Table 3, the value of TR of LIP, i.e. 0.0625 is lower than that TR of MP, i.e. 0.0875 and CCI, i.e. 0.15. Thus, the decision of BC technique to select LIP as a partitioning attribute is not the best. Based on Table 3, the TR technique selects CCI as the partitioning attribute. From Table 4, MMR selects attribute CCI as partitioning attribute. This is due to the fact that the second mean roughness of attribute CCI, i.e. 0.8 is less than that of the attribute IT, i.e. 0.9. Meanwhile, from Table 5, MDA selects attribute CCI is selected as the partitioning attribute.

Table 5. The degree of dependency of all attributes

Attribute depends on	Degree of dependency				MDA
MP	WP	LIP	CCI	S	0.5
	0	0.5	0.2	0	0.2
WP	MP	LIP	CCI	S	0
	0	0	0	0	
LIP	MP	WP	CCI	S	0.3
	0.3	0	0.2	0	
CCI	MP	WP	LIP	S	0.5
	0.3	0	0.5	0.4	0.4
S	MP	WP	LIP	CCI	0.2
	0	0	0	0.2	

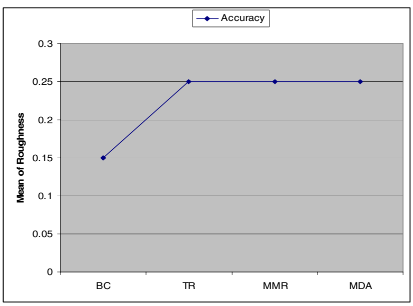

Fig. 1. The accuracy of BC, TR, MMR and MDA techniques

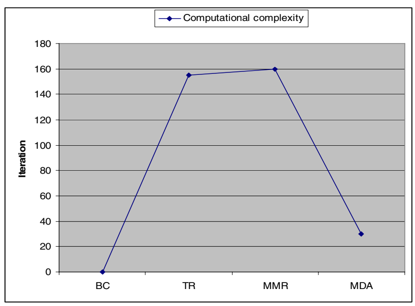

Fig. 2. The computational complexity of BC, TR, MMR and MDA techniques

Based on formula in equation (6), the accuracy of TR, MMR and MDA in selecting partitioning attribute is the same, i.e. 0.25. But, the accuracy of BC technique is 0.15. Additionally, since BC technique only select the partitioning attribute based on least distinct balanced-value of attribute, then BC need no computation. However, for the other techniques, MDA achieve lower computational complexity than that TR and MMR techniques.

6 Conclusion

In this paper, we have pointed out some unreasonable statements of MMR technique. We have proven that the mean roughness of MMR technique is only the opposite of that Mazlack's TR technique and shown that the idea of MMR to achieve lower computational complexity is not error free by an example. In order to solve these problems, MDA (Maximum of Dependency Attributes), an alternative rough set-based technique for selecting partitioning attribute by taking into account the dependency of attributes in an information system is proposed. We have proven that MDA technique is a generalization of TR and MMR techniques which achieve lower computational complexity and has higher accuracy than that BC technique. With this approach, we believe that some applications using dependency attributes in the theory of rough set in information system through this view will be applicable. For the future activities, we use the proposed technique for categorical data clustering and decision support systems in complex domain through larger data set likes the benchmark datasets from some standard UCI database.

Acknowledgement

This work was supported by the FRGS under the Grant No. Vote 0402, Ministry of Higher Education, Malaysia.

References

1. Mazlack, L.J., He, A., Zhu, Y., Coppock, S.: A rough set approach in choosing partitioning attributes. In: Proceedings of the ISCA 13th, International Conference, CAINE-2000, pp. 1–6 (2000)
2. Pawlak, Z.: Rough sets. International Journal of Computer and Information Science 11, 341–356 (1982)
3. Pawlak, Z.: Rough sets: A theoretical aspect of reasoning about data. Kluwer Academic Publisher, Dordrecht (1991)
4. Pawlak, Z., Skowron, A.: Rudiments of rough sets. Information Sciences. An International Journal 177(1), 3–27 (2007)
5. Parmar, D., Wu, T., Blackhurst, J.: An algorithm for clustering categorical data using rough set theory. Data and Knowledge Engineering 63, 879–893 (2007)
6. Yao, Y.Y.: Two views of the theory of rough sets in finite universes. Approximate Reasoning, An International Journal 15(4), 191–317 (1996)
7. Yao, Y.Y.: Constructive and algebraic methods of the theory of rough sets. Information Science, An International Journal 109(1-4), 21–47 (1998)
8. Yao, Y.Y.: Information granulation and rough set approximation. International Journal of Intelligent Systems 16(1), 87–104 (2001)
9. Roiger, R.J., Geatz, M.W.: Data Mining: A Tutorial-Based Primer. Addison Wesley, Reading (2003)