Feature Selection and Rule Extraction Based on Variable Precision Rough Set

I A Kesuma¹, M Zarlis¹ and E B Nababan”

¹Department of Information Technology, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan 20155, Indonesia

E-mail: ernabrn@usu.ac.id

Abstract. Main factors in feature selection is dimensionality in data and decision making on partial information handling as precision classification errors. This study will demonstrate the feature selection using the rough set method, especially in handling missing information along with an ambiguous decision system with a precision variable-based approach to finding significant features and extracting the required rules. The experiment was carried out with 12 Dermatology datasets from the UCI Machine Learning Repository consisting of 11 conditional attributes and a decision variable. The experimental results show the results of the dominant conditional feature selection of 45% of the overall existing conditional features, along with more concise rules based on selected features.

1. Introduction

Various methods of data mining have been developed to deal with more dynamic data processing problems such as the decision tree development concept, neural networks, association rules, clustering methods, and so on. However, some method approaches do not have to be used in the data processing process which tends to be qualitative because of limitations in data handling, in addition to the type of population approach that may require some assumptions in statistics [1]. The rough set method is one that looks promising in handling qualitative information data and provides an approach based on individual object models [2]. But in its development, the rough set has a deficiency when applied to data extraction and problem in classification analysis. This is because the information available can sometimes only be fulfilled for partial classification.

2. Rough Set

Rough set was first introduced by Pawlak in 1982 as a mathematical method of artificial intelligent application in dealing with problems of or incompleteness of information [3]. The purpose of the rough set method is to get a short rule estimate from an information table [4]. Rough set method has the benefits or advantages in data analysis information as follows:

• Finding patterns of irregularity in information.
• Can determine a minimum reduction or reduction of data.
• Improvement of information data sensitivity patterns.
• Establish a pattern of rules for decision making.
• Able to process computing in parallel.
• Easy to understand computational methods.
2.1. Information and Decision System

In data mining term, an information system is known as dataset. The dataset consists of lines that represent data, and columns that represent attributes or variables of data. Information System (IS) is denoted as follows.

\[ IS = (U, A) \]  

where \( U = \{ e_1, e_2, e_3, \ldots, e_m \} \) are object of elements and \( A = \{ a_1, a_2, \ldots, a_n \} \) are conditional attributes.

Decision system is an information system with additional attributes named with decision attributes, in data mining known as targets that represent the results of known classifications. The decision system is the function described as below.

\[ D = (U, A \cup d) \text{ where } d \notin A \]  

An example of the information system and decision system is shown in Table 1.

**Table 1. Example of Information System and Decision System**

| object | a   | b   | c     | decision |
|--------|-----|-----|-------|----------|
| P1     | No  | No  | Normal| No       |
| P2     | No  | No  | High  | No       |
| P3     | No  | No  | Very High| Yes    |
| P4     | No  | Yes | High  | Yes      |
| P5     | No  | Yes | Very High| Yes    |
| P6     | Yes | Yes | High  | Yes      |
| P7     | Yes | Yes | Very High| Yes    |
| P8     | No  | No  | High  | No       |
| P9     | Yes | No  | Very High| Yes    |
| P10    | Yes | No  | High  | No       |

In the table, the data set in the object column is called the universal set of \( U \). The \( a, b, \) and \( c \) columns are a set of condition attributes of \( A \). Element \( A \) is called conditional attributes. While values from decision column is called as decision attributes, it usually not restricted and the value is in binary (e.g. true or false).

2.2. Indiscernibility Relation

The rough set method uses the indiscernibility concept in the attribute selection process. Classified as Indiscernibility Relation, if the condition attribute of an object has the same value on the condition attribute of another object. For example in table 1, object P1, P2, P3, P4, P5, and P8 have the value 'no' in the conditional \( a \) attribute.

2.3. Equivalence Class

Table 1 shows that the attribute value of both conditions and decisions in P2 has an attribute value equal to the P8 value, so that both can be collected as Equivalence Class objects. While the value of other object attributes, looks different from each other, so it cannot be collected in one Equivalence Class.
2.4. Discernibility Modulo D
Discernibility Matrix Modulo D is a table that results in the comparison of attribute values between objects by observing the condition attribute value and the decision attribute value. Will produce a value if the attributes that are compared are different, and vice versa if the attributes that are compared are the same, then it does not produce a value.

2.5. Reduct and Decision Rule
In determining the rules of the decision, there are attributes that affect the results of the classification, this is called the reduct attribute. While attributes that do not affect the results of classification are called non-reduct attributes, and these attributes may be reduced or ignored. In the process of selecting a minimum attribute from a set of condition attributes, prime implicant boolean functions are used.

2.6. Variable Precision Rough Set
Variable Precision Rough Set (VPRS) model was introduced by Ziarko as a continuation of the classic rough set model, which was proposed to analyze more functional data patterns. Ziarko introduced a partial classification as a \( \beta \) (beta) precision parameter. The value of \( \beta \) is defined as a misclassification and ranges in the value of \( 0 \leq \beta < 0.5 \) [5]. The VPRS procedure consists of 4 steps as follows [6]:
- Selection of beta (\( \beta \)) precision parameter values
- Determine the \( \beta \)-reduct set
- Delete redundant objects
- Rule extraction.

The VPRS model is a development model for the rough set classic method with the additional assumption that the relative degree of error is in the classification. The relative degree of misclassification is denoted as follows.

\[
c(C,Y) = \begin{cases} 
1 - \frac{|C \setminus Y|}{|C|}; & |C| > 0 \\
0; & |C| = 0
\end{cases}
\] (3)

Member of the conditional attribute set \( C \) is part of the member of the decision set \( Y \). A comparison of the number of members is called exactly the same if the relative degree of error is equal to 0. Whereas the value of the relative error value is greater than 0, then it is an \( \beta \)-reduct candidate. Determining the value of \( \beta \) aims to choose the condition of the relative degree of minimum error between the set of candidate \( \beta \)-reducts that have been produced. The value of \( \beta \)-reduct is determined by the following equation.

\[
\beta(C,Y) = \max (m_1, m_2)
\] (4)

Where

\[
m_1 = 1 - \min (c(C,Y) > 0.5)
\] (5)

And

\[
m_2 = \max (c(C,Y) < 0.5)
\] (6)

3. Methodology
Figure 1 describe the methodology of research. It shows that there are two main processes, the first is the pre-processing data stage and the second is the application phase of the Variable Precision Rough Set method. The pre-processing stage is needed to filter and find value data that are incomplete and inconsistent. The application section of the VPRS method follow the basic rules of the classical rough set method witch adapted into precision-based methods according to the steps described previously.
4. Result and Discussion

4.1. Dataset Pre-processing
The dataset used in this study is the Dermatology dataset from the UCI Machine Learning Repository. The Dermatology dataset consists of 12 attributes, 11 are conditional symptom attributes and 1 attribute of the target decision that is the diagnosis [7]. Dataset has been modified in such a way as to use Variable Precision Rough Set method in dealing with data loss in an information system. Table 2 shows that there are still some invaluable attribute values (null values). This can be referred to as incomplete information value. In this study, the value of incomplete information is symbolized by the character "*".

| Object | a | b | c | d | e | f | g | h | i | j | k | Decision |
|--------|---|---|---|---|---|---|---|---|---|---|---|----------|
| X1     | 2 | 1 | 0 | 2 | 2 | 0 | 0 | * | 0 | 0 | 0 | 4        |
| X2     | 2 | 3 | 3 | 0 | 0 | 0 | 0 | * | 2 | 2 | 0 | 1        |
| X3     | 2 | 2 | 0 | 3 | 0 | 0 | 0 | * | 1 | 0 | 0 | 2        |
| X4     | 2 | 1 | 0 | 2 | 2 | 0 | 0 | * | 0 | 0 | 0 | 4        |
| X5     | 2 | 2 | 0 | 3 | 0 | 0 | 0 | 0 | 1 | * | * | 3        |
| X6     | 2 | 1 | 2 | 1 | 3 | 0 | 3 | * | 0 | 0 | 3        |
| X7     | 2 | 3 | 2 | 0 | 0 | 0 | 0 | * | 2 | 2 | 0 | 2        |
| X8     | 2 | 2 | * | 3 | 0 | 0 | 0 | 0 | 1 | * | * | 1        |
| X9     | 2 | 3 | 2 | 0 | 0 | 0 | 0 | * | 2 | 2 | 0 | 1        |
| X10    | 2 | 3 | 2 | * | 0 | 0 | 0 | 0 | 2 | 2 | 0 | 1        |

4.2. Tolerance Class and Decision Class
The initial stage before determining the precision parameter is grouping object data that has the same value in the conditional attribute. Based on table 2, equivalence class is obtained in table 3 as follows.

Table 3. Equivalence class

| Equivalence Class | Object Class | Total Object |
|-------------------|--------------|--------------|
| E1                | X1, X4       | 2            |
| E2                | X2, X9, X10  | 3            |
| E3                | X3, X8, X5   | 3            |
| E4                | X2           | 1            |
| E5                | X6           | 1            |
While the decision class is obtained by grouping attribute values of the same decision. Table 4 is a list of decision classes.

| Notation | Decision Class Value | Object Class | Total Object |
|----------|----------------------|--------------|--------------|
| D₁       | 1                    | X₂, X₈, X₉, X₁₀ | 4            |
| D₂       | 2                    | X₃, X₇        | 2            |
| D₃       | 3                    | X₅, X₆        | 2            |
| D₄       | 4                    | X₁, X₄        | 2            |

4.3. Result

Based on the comparison of the degree of relation between equivalence classes and decision class, value of $\beta$ can be calculated with the equation (3), so that the comparison of the degree of each attribute is like table 5.

| Equivalent Class | Decision Class |
|------------------|----------------|
|                  | D₁  | D₂  | D₃  | D₄  |
| E₁               | 1   | 1   | 1   | 0   |
| E₂               | 1/3 ($\beta$) | 2/3 | 1   | 1   |
| E₃               | 1   | 2/3 | 2/3 | 1   |
| E₄               | 0   | 1   | 1   | 1   |
| E₅               | 1   | 1   | 0   | 1   |

This is consistent with the definition of misclassification which ranges from $0 \leq \beta < 0.5$ [5]. This value refers to the degree value of the equivalence class relation that has the same decision class attribute but a different conditional attribute, namely E₂ where there is a set of objects consisting of X₇, X₉ and X₁₀ (see table 6).

In addition, Discernibility Modulo D is needed to obtain a minimum subset of attributes as a whole, which is by comparing the contents of an object's attributes with each other object. Where the results of the comparison of the attribute values are then converted into the Conjunctive Normal Form (CNF) with simplification operations based on the member set which refers to the resulting $\beta$ value. The results of the simplification are as follows.

$$R_{\text{E}\beta} = x₇ \cap x₉ \cap x₁₀$$

$$= (b \cap c \cap h) \cup (c \cap d) \cup (c \cap e \cap h) \cup (c \cap h \cap i) \cup (c \cap h \cap j)$$

It appears that the set of attributes consisting of {b, c, h}, {c, d}, {c, e, h}, {c, h, i} and {c, h, j}. From these results we can find a subset of the minimum attributes with a minimal and non-recurring frequency of reduct occurrence consisting of {b}, {d}, {e}, {i}, and {j}. Table 6 shows of information result after reduct attribute respectively.

| Object | b | d | e | i | j | Decision |
|--------|---|---|---|---|---|----------|
| X₁     | 1 | 2 | 2 | 0 | 0 | 4        |
| X₂     | 3 | 0 | 0 | 2 | 2 | 1        |
| X₃     | 2 | 3 | 0 | 1 | 0 | 2        |
| X₄     | 1 | 2 | 2 | 0 | 0 | 4        |
| X₅     | 2 | 3 | 0 | 1 | * | 3        |
| X₆     | 1 | 3 | 1 | * | 0 | 3        |
| X₇     | 3 | 0 | 0 | 2 | 2 | 2        |
Two aspects that need to be reviewed from Table 6, the first aspect is the redundancy of objects and the contradictions of missing values (*) in the condition attribute. When viewed from the pattern of redundancy attribute values and decision values produced, it can be said that X_4 object with X_8 is a redundant object. Then one object from the comparison can be removed. Likewise with X_2 objects with X_5 or X_7.

The second aspect review is the contradiction in the value of incomplete data. Object X_9 with X_{10} can be said to be partially redundant. This is because the value of the attribute d is X_{10}. Likewise with X_3 and X_8, even though the decision value produced is different, if viewed from the condition attribute value pattern can be said partially still has the potential to produce the same pattern of rules because the value in the attribute column j. This becomes contradictory where two rules are the same but can produce different decisions. Therefore, it is necessary to do a comparison process with the Discernibility Modulo D method to produce a consistent pattern of conditions. The following table is the information system after information reduct.

Table 7. Information system with information reduct

| Object | b | d | e | i | j | Decision |
|--------|---|---|---|---|---|----------|
| X_1   | * | 2 | 2 | 0 | * | 4        |
| X_2   | 3 | 0 | * | 2 | 2 | 1        |
| X_3   | 2 | * | 0 | 1 | 0 | 2        |
| X_4   | * | 2 | 2 | 0 | * | 4        |
| X_5   | * | * | * | * | * | 3        |
| X_6   | 1 | 3 | 1 | * | * | 3        |
| X_7   | * | 0 | * | * | * | 2        |
| X_8   | * | * | * | * | * | 1        |
| X_9   | 3 | 0 | * | 2 | 2 | 1        |
| X_{10} | * | * | * | * | * | 1        |

In the table, the value of "*" is not used to establishment of rules. There are three objects that have a value of "*" in all condition attributes: X_5, X_8, and X_{10}, they are can be eliminated so that the pattern of rules can be extract as in table 8.

Table 8. Rule extraction

| No | Rules |
|----|-------|
| 1  | IF d=2 AND e=2 AND i=0 THEN Decision=4 |
| 2  | IF b=3 AND d=0 AND i=2 AND j=2 THEN Decision=1 |
| 3  | IF (b=2 OR j=0) AND (e=0 OR j=0) AND (i=1 OR j=0) THEN Decision=2 |
| 4  | IF d=2 AND e=2 AND i=0 THEN Decision=4 |
| 5  | IF (b=1 OR d=3) AND e=1 THEN Decision=3 |
| 6  | IF d=0 THEN Decision=2 |
| 7  | b=3 AND d=0 AND i=2 AND j=2 THEN Decision=1 |
5. Conclusion and Future Work

The conclusion of the Variable Precision Rough Set method application is that the dominant conditional feature selection result is 45% of all existing conditional features. While conditional attributes that do not include reduct are reduced by 55%. It has indicated that the Discernibility Modulo D concept is used as in the rough set classic method, but it refers to the value of the $\beta$ precision variable based on the comparison of relation degrees as a form of variable method by developing precision relation in Variable Precision Rough Set method.

The suggestions that can be proposed from this research are:

- This research can still be developed with the application of the Variable Precision Rough Set application with reference to the specific $\beta$ precision range threshold value.
- This research still opens the opportunity to analyze misclassification variables in the Variable Precision Rough Set method.

References

[1] Fan Y N, Tseng T L, Chern C C and Huang C C 2009 Rule Induction Based on an Incremental Rough Set Expert Systems With Applications 36 Int. J. Elsevier pp. 11439-11450.

[2] Kusiak A 2001 A Feature Transformation Methods in Data Mining IEEE Transaction on Electronics Packaging Manufacturing vol 24(3) pp. 214-221.

[3] Pawlak Z 1982 Rough Set Int. J. of Comp. and Information Science 11, pp. 341-351.

[4] Pawlak Z 1996 Rough Set and Data Analysis Proceedings of The 1996 Asian, 1996, pp. 1-6. References.

[5] Ziarko W 1993 Variable Precision Rough Set Model J. of Comp. System and Science Vol. 46 No. 1 pp. 39-59.

[6] Su C T and Hsu J H 2006 Precision Parameter in the Variable Precision Rough Sets Model: An Application Omega Vol. 34 No. 2 pp. 149–157.