

Applying Dependency of Attributes for Business Data Mining in Information Systems: A Theoretical Framework

Rokhmat Rokhmat¹, Yunus Indra Purnama¹, Eka Novita Sari² and Tutut Herawan²

¹*Faculty of Information Technology and Business
Universitas Teknologi Yogyakarta
Kampus UTY, Jalan Lingkar Utara, Yogyakarta, Indonesia*
²*AMCS Research Center
Yogyakarta, Indonesia*

rokhmat_uty@yahoo.com, indra_uty@yahoo.com, eka@amcs.co, tutut@amcs.co

Abstract

This paper presents the applications of dependency of attributes in information systems for data mining from business datasets. Firstly, we present the theoretical framework for data clustering on small business dataset. It is based on a construction of a hierarchical rough set approximation in an information system for data splitting. The hierarchy is defined by the notion of a nested sequence of indiscernibility relations that can be defined from the dependency of attributes. Secondly, an application of such hierarchy for mining maximal association from a business transactional data is presented. It is shown that the dependency provides clear and provable theoretical approach for data clustering and maximal association rules mining.

Keywords: *Rough set theory; Dependency of Attributes; Hierarchical rough set approximation; Data Clustering; Maximal Association Rules Mining.*

1. Introduction

An attribute dependency states that the value of an attribute is uniquely determined by the values of some other attributes. The objective of discovering attribute dependency is to find the relationship among attributes in information systems [1, 2]. Formally, in an information system $S = (U, A, V, f)$, attribute D is called totally depends on attribute C , denoted $C \Rightarrow D$, if each value of D is associated exactly one value of C . Otherwise, D is depends partially on C . One of the methods for discovering attribute dependencies is using rough set theory [3-5]. The discovery of attribute dependencies using rough set theory has been received considerable interest e.g. [6-17]. However, to this date, not many researches on the exploration of attributes dependency for data mining, such as data classification and association rule mining [18].

In this paper, we propose a technique for business data clustering and maximal association rules mining under rough set-based attributes dependency in an information system. It is based on a concept of hierarchical rough set approximation. The Yao's hierarchical rough set approximation induced by a special class of equivalence relations based on (Pawlak) approximation space [19-21] has been extended to hierarchical rough set approximations of an information system, induced by dependency of attributes. Our approach starts with the notion of a nested sequence of indiscernibility relation that can be defined from the dependency of attributes in an information system. Based on a nested sequence of indiscernibility relation, a nested rough set approximation can be constructed. Further on, the notion of a nested rough set approximation is then used for constructing hierarchical rough set approximation in an information system [14]. With this approach, a hierarchical approximation (hierarchical granulation) can be easily constructed. We

describe that the dependency provides clear and provable theoretical approach for data clustering and maximal association rules mining

The rest of this paper is organized as follows. Section 2 describes a fundamental concept of rough set theory. Section 3 describes a construction of a hierarchical rough set approximations using dependency of attributes. Section 4 describes the applications of such hierarchy for data clustering and maximal association rules mining. Finally, the conclusions of our works are described in section 5.

2. Rough Set Theory

Pawlak introduced the rough set theory to deal the problems of vagueness and uncertainty [5]. The rough set theory is not an alternative to classical set theory but it is embedded in it. Fuzzy and rough set theories are not competitive, but complementary to each other [7, 15]. Rough set theory has attracted attention of many researchers and practitioners all over the world [22-25]. Motivation for rough set theory has come from the need to represent a subset of a universe in terms of equivalence classes of a partition of the universe. The original goal of the rough set theory is induction of approximations of concepts. The idea consists of approximation of a subset by a pair of two precise concepts called the lower approximation and upper approximation. Intuitively, the lower approximation of a set consists of all elements that surely belong to the set, whereas the upper approximation of the set constitutes of all elements that possibly belong to the set. The difference between the upper approximation and the lower approximation is a boundary region. The region consists of all elements that cannot be classified uniquely to the set or its complement using available knowledge. Thus any rough set, in contrast to a crisp set, has a non-empty boundary region. In the following subsection, the basic concept of rough set theory in terms of data is presented.

2.1. Information System

Data are often presented as a table, columns of which are labeled by attributes, rows by objects of interest and entries of the table are attribute values. By an information system, 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 , $f : U \times A \rightarrow V$ is a total function such that $f(u, a) \in V_a$, for every $(u, a) \in U \times A$, called information (knowledge) function. An information system is also called a knowledge representation systems or an attribute-valued system. An information system can be intuitively expressed in terms of an information table (see Table 1).

Table 1. An Information System

U/A	a_1	a_2	...	a_k	...	$a_{ A }$
u_1	$f(u_1, a_1)$	$f(u_1, a_2)$...	$f(u_1, a_k)$...	$f(u_1, a_{ A })$
u_2	$f(u_2, a_1)$	$f(u_2, a_2)$...	$f(u_2, a_k)$...	$f(u_2, a_{ A })$
u_3	$f(u_3, a_1)$	$f(u_3, a_2)$...	$f(u_3, a_k)$...	$f(u_3, a_{ A })$
\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots
$u_{ U }$	$f(u_{ U }, a_1)$	$f(u_{ U }, a_2)$...	$f(u_{ U }, a_k)$...	$f(u_{ U }, a_{ A })$

In many applications, there is an outcome of classification that is known. This a posteriori knowledge is expressed by one (or more) distinguished attribute called decision

attribute; the process is known as supervised learning. An information system of this kind is called a decision system. A decision system is an information system of the form $D = (U, A \cup \{d\}, V, f)$, where $d \notin A$ is the decision attribute. The elements of A are called condition attributes.

The starting point of rough set theory is the indiscernibility relation, which is generated by information about objects of interest. The indiscernibility relation is intended to express the fact that due to the lack of knowledge we are unable to discern some objects employing the available information. Therefore, generally, we are unable to deal with single object. Nevertheless, we have to consider clusters of indiscernible objects. The following definition precisely defines the notion of indiscernibility relation between two objects.

2.2. Indiscernibility Relation

The notion of indiscernibility relation between two objects can be defined precisely.

Definition 2.1. Let $S = (U, A, V, f)$ be an information system and let B be any subset of A . 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. It is well known that, an equivalence relation induces unique partition. The partition of U induced by $IND(B)$ in $S = (U, A, V, f)$ denoted by U / B and the equivalence class in the partition U / B containing $x \in U$, denoted by $[x]_B$.

Given arbitrary subset $X \subseteq U$, in general, X as union of some equivalence classes in U might be not presented. It means that, it may not be possible to describe X precisely in information system. A set X might be characterized by a pair of its approximations, called lower and upper approximations. It is here that the notion of rough set emerges.

2.3. Set Approximations

The indiscernibility relation will be used next to define approximations, basic concepts of rough set theory. The notions of lower and upper approximations of a set can be defined as follows.

Definition 2.2. Let $S = (U, A, V, f)$ be an information system, let B be any subset of A and let X be any subset of U . The B -lower approximation of X , denoted by $\underline{B}(X)$ and B -upper approximations of X , denoted by $\overline{B}(X)$, respectively, are defined by

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

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 $\alpha_B(X) = |\underline{B}(X)| / |\overline{B}(X)|$, where $|X|$ denotes the cardinality of X . For empty set \emptyset , $\alpha_B(\emptyset) = 1$ is defined. Obviously, $0 \leq \alpha_B(X) \leq 1$. If X is a union of some equivalence classes of U , then $\alpha_B(X) = 1$. Thus, the set X is *crisp* (precise) with respect to B . And, if X is not a union of some equivalence

classes of U , then $\alpha_B(X) < 1$. Thus, the set X is *rough* (imprecise) with respect to B [26]. This means that the higher of accuracy of approximation of any subset $X \subseteq U$ is the more precise (the less imprecise) of itself.

Another important issue in database analysis is discovering dependencies between attributes. Intuitively, a set of attributes D depends totally on a set of attributes C , denoted $C \Rightarrow D$, if all values of attributes from D are uniquely determined by values of attributes from C . In other words, D depends totally on C , if there a functional dependency between values of D and C . The formal definition of attributes dependency is given as follows.

Definition 2.3. Let $S = (U, A, V, f)$ be an information system and let D and C be any subsets of A . Attribute D is functionally depends on C , denoted $C \Rightarrow D$, if each value of D is associated exactly one value of C .

2.4. Dependency of Attributes

Since information system is a generalization of a relational database. A generalization concept of dependency of attributes, called a *partial dependency* of attributes is also needed.

Definition 2.4. Let $S = (U, A, V, f)$ be an information system and let D and C be any subsets of A . The dependency attribute D on C in a degree k ($0 \leq k \leq 1$), is denoted by $C \Rightarrow_k D$, where

$$k = \frac{\sum_{X \in U/D} |C(X)|}{|U|}. \quad (1)$$

Obviously, $0 \leq k \leq 1$. If all set X are crisp, then $k = 1$. The expression $\sum_{X \in U/D} |C(X)|$, called a lower approximation of the partition U/D with respect to C , is the set of all elements of U that can be uniquely classified to blocks of the partition U/D , by means of C . 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 .

3. A Construction of Hierarchical Rough Set Approximation in Information Systems using Dependency of Attributes

One such application of rough set theory is in granular computing where the concept of approximation is used to solve some classification problems. A cluster (granule) usually consists of elements that are drawn together by similarity, proximity, or functionality [27, 28]. In granular computing, the granules i.e., classes or clusters of a universe, are used in the processes of problem solving. When a problem involves incomplete, uncertain, or vague information, it may be difficult to differentiate distinct elements and one is forced to consider granules for the purpose of differentiation. The granulated view of the universe is based on a binary relation representing the type of similarities between elements of a universe [21]. Marek and Rasiowa [27] considered gradual approximations of sets based on a descending sequence of equivalence relations. Yao [19, 20] suggested the use of hierarchical granulations for the study of stratified rough set approximations.

To this, Yao [21] proposed hierarchical granulations induced by a special class of equivalence relations.

In this section we present a construction of a hierarchical rough set approximation in an information system using dependency of attributes. To start off, the nested rough set approximation has to be based on the dependency of attributes. Further, the notion of a nested rough set approximation is used for constructing hierarchical rough set approximations.

3.1. Nested Rough Set Approximations

In this sub-section, the construction of nested rough set approximations in an information system using dependency of attributes is presented. We may consider all possible subsets of attributes to obtain different degrees of dependency of attributes.

Definition 3.1. A sequence $\langle s_n \rangle$ is a function which domain is contained in the set of all natural numbers N and range is contained in the set of its terms, $\{s_1, s_2, \dots, s_n\}$.

Definition 3.2. Let $S = (U, A, V, f)$ be an information system and let A_1, A_2, \dots, A_n be any subset of A . A sequence of indiscernibility relations induced by $A_i, i = 1, 2, \dots, n$, denoted $\langle IND(A_i)_{i=1, \dots, n} \rangle$ is said to be nested if $IND(A_1) \subseteq IND(A_2) \subseteq \dots \subseteq IND(A_n)$.

In this case, we can say that an equivalence relation induced by A_i is coarser than an equivalence relation induced by A_j , where $1 \leq j \leq n$. It is clear that, if $IND(A_1) \subseteq IND(A_2)$, then U / A_1 is finer than U / A_2 .

Definition 3.3. The rough approximations of a subset $X \subseteq U$, denoted $\langle \langle \underline{A}_n(X), \overline{A}_n(X) \rangle \rangle_{i=1, \dots, n}$ is said to be nested if

$$\underline{A}_n(X) \subseteq \dots \subseteq \underline{A}_1(X) \subset X \subset \overline{A}_1(X) \subseteq \dots \subseteq \overline{A}_n(X).$$

Proposition 3.4. Let $S = (U, A, V, f)$ be an information system and let A_1, A_2, \dots, A_n be any subsets of A . If A_i depends on A_{i+1} , for $i = 1, 2, \dots, n-1$, then $\langle \langle \underline{A}_n(X), \overline{A}_n(X) \rangle \rangle_{i=1, \dots, n}$ is a nested rough approximation of a subset $X \subseteq U$.

Proof. Let $A_1, A_2, \dots, A_n \subseteq A$. If A_i depends on A_{i+1} , then $IND(A_{i+1}) \subseteq IND(A_i)$ and every equivalence class induced by $IND(A_{i+1})$ is a union of some equivalence class induced by $IND(A_i)$, it means that for every $x \in U$, $[x]_{A_i} \subseteq [x]_{A_{i+1}}$ and thus U / A_{i+1} is coarser than U / A_i . On the other side, the dependability of A_i on A_{i+1} , for $i = 1, 2, \dots, n-1$, implies $\langle IND(A_i)_{i=1, \dots, n} \rangle$ to be a nested sequence of indiscernibility relations. Since, a nested sequence of indiscernibility relations determines a nested rough approximations of a subset $X \subseteq U$, $\langle \langle \underline{A}_n(X), \overline{A}_n(X) \rangle \rangle_{i=1, \dots, n}$, then the proof is completed.

Proposition 3.5. Let $S = (U, A, V, f)$ be an information system. If $\langle \langle \underline{A}_n(X), \overline{A}_n(X) \rangle \rangle_{i=1, \dots, n}$ is a nested rough approximations of a subset $X \subseteq U$, then $\alpha_{A_{i+1}}(X) \leq \alpha_{A_i}(X)$, for $i = 1, 2, \dots, n-1$.

Proof. It follows from Definition 3.3, we have

$$\underline{A}_n(X) \subseteq \dots \subseteq \underline{A}_1(X) \subset X \subset \overline{A}_1(X) \subseteq \dots \subseteq \overline{A}_n(X).$$

for every $X \subseteq U$. Consequently

$$\frac{|\underline{A}_n(X)|}{|\overline{A}_n(X)|} \leq \frac{|\underline{A}_{n-1}(X)|}{|\overline{A}_{n-1}(X)|} \leq \dots \leq \frac{|\underline{A}_2(X)|}{|\overline{A}_2(X)|} \leq \frac{|\underline{A}_1(X)|}{|\overline{A}_1(X)|}$$

$$\alpha_{A_n}(X) \leq \alpha_{A_{n-1}}(X) \leq \dots \leq \alpha_{A_2}(X) \leq \alpha_{A_1}(X).$$

Example 3.1. We illustrate our approach and compare it with Yao's nested rough set approximation approach. Based on the information system as in Table 2 in [21] Yao's rough set approximations is been presented.

Based on Table 1, Yao considered the sequence of subsets collection of attributes $A = \{A_1, A_2, A_3, A_4\}$, $B = \{A_1, A_2\}$, $C = \{A_1\}$, and $D = \emptyset$. Yao used the notion of a nested sequence of equivalence relation $R_A \subseteq R_B \subseteq R_C \subseteq R_D$. Thus, for each subset of attributes, Yao will then obtain the following partition (granulation structure)

$$U / D = \{a, b, c, d, e, f\}, U / C = \{\{a, b, c, d\}, \{e, f\}\}, U / B = \{\{a\}, \{b, c\}, \{d\}, \{e, f\}\},$$

$$U / A = \{\{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{f\}\}.$$

As for a subset $X(\text{Class} = +) = \{c, d, e, f\}$ of U , the accuracy in each level granulation of the rough set approximation obtained by Yao can tabulated as in Table 3.

Table 2. An Information System

Object	A_1	A_2	A_3	A_4	Class
a	1	1	1	1	-
b	1	2	1	0	-
c	1	2	0	0	+
d	1	3	1	1	+
e	0	1	0	1	+
f	0	1	1	1	+

Table 3. Set Approximations and their Accuracy in Each Level Partition

App of X w.r.t.	Low	Upp	Acc
A	$\{c, d, e, f\}$	$\{c, d, e, f\}$	1.000
B	$\{d, e, f\}$	$\{b, c, d, e, f\}$	0.600
C	$\{e, f\}$	U	0.333
D	ϕ	U	0.000

However, our approach uses the notion of dependency of attributes in an information system. When compared with Yao's approach, we only use the set of attributes A, B and C . From Table 2, we have, $C \Rightarrow_{k=1} B$, $B \Rightarrow_{k=1} A$, and $A \Rightarrow_{k=1} \text{Class}$. The nested rough approximations from those attributes dependencies of a subset $X(\text{Class} = +) = \{c, d, e, f\}$ is given by

$$\langle \underline{A}(X), \bar{A}(X) \rangle, \langle \underline{B}(X), \bar{B}(X) \rangle, \langle \underline{C}(X), \bar{C}(X) \rangle$$

with

$$\underline{C}(X) \subseteq \underline{B}(X) \subseteq \underline{A}(X) \subseteq X \subseteq \bar{A}(X) \subseteq \bar{B}(X) \subseteq \bar{C}(X).$$

Meanwhile, the accuracy in each approximation can be summarized as in Table 4.

Table 4. Set Approximations and their Accuracy in each Level Partition using Attributes Dependency

App of X w.r.t.	Low	Upp	Acc
A	$\{c, d, e, f\}$	$\{c, d, e, f\}$	1.000
B	$\{d, e, f\}$	$\{b, c, d, e, f\}$	0.600
C	$\{e, f\}$	U	0.333

From Table 4, with lower degree of dependency, coarser partition will be obtained, at the same time obtaining lesser accuracy in the rough approximation. Thus, with this approach, from Table 4, we can easily define the notions of a nested rough set approximation in an information system and at the same time obtain the identical accuracy of approximation as in Table 3.

3.2. Hierarchical Rough Set Approximations

In this sub-section, a construction of hierarchical rough set approximations using dependency of attributes is presented. It is based on nested rough set approximations. A hierarchy may be viewed as a successive top-down decomposition of a universe U . Alternatively, a hierarchy may also be viewed as a successive bottom-up combination of smaller clusters to form larger clusters [21]. Follows Definition 3.1, a hierarchical rough set approximation is constructed in Figure 1.

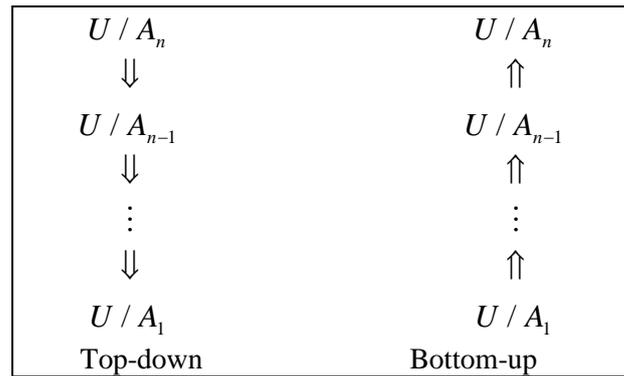


Figure 1. A Hierarchical Rough Set Approximation

The highest level of a hierarchical rough set approximation is the partition induced by $IND(A_n)$ and the lowest of the hierarchy is the partition induced by $IND(A_1)$. Consequently, for every $x \in U$, we have the nested clusters, in this case equivalence classes, contain x , i.e. $[x]_{A_1} \subseteq [x]_{A_2} \subseteq \dots \subseteq [x]_{A_n}$.

Let $C = \{X \subseteq U : X \text{ is a class in each partition } U / A_i\}$, thus every class induced by $IND(A_{i+1})$ is a union of some class induced by $IND(A_i)$.

4. Applications in Business Data Mining

We explain the applications of the proposed hierarchical rough set approximation concept for data clustering and maximal association rules mining through business datasets, respectively derived from [29] and a dataset from (<http://www.research.att.com/lewis/reuters21578.html>).

4.1. Data Clustering

In the clustering process, there are no predefined classes and no examples that would show what kind of desirable relations should be valid among the data that is why it is perceived as an unsupervised process. According to the method adopted to define clusters, one of the clustering algorithms is hierarchical clustering. It proceeds successively by either merging smaller clusters into larger ones, or by splitting larger clusters [30]. From a hierarchical rough set approximation as described in Figure 1, we show how attributes dependency can be used to classify objects from a dataset.

From Table 5, there are ten objects $U = \{1,2,3,4,5,6,7,8,9,10\}$ with six categorical-valued attributes $A = \{a_1, a_2, a_3, a_4, a_5, a_6\}$. We let three subsets of A i.e. $B = \{a_2, a_3, a_6\}$, $C = \{a_2, a_3\}$ and $D = \{a_2\}$. To this, we have the following partitions, induced by their indiscernibility relations, respectively:

$$U / B = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{7\}, \{8\}, \{9\}, \{10\}\},$$

$$U / C = \{\{1\}, \{2\}, \{3,5,8\}, \{4\}, \{6,9\}, \{7\}, \{10\}\},$$

And

$$U / D = \{\{1,4\}, \{2\}, \{3,5,7,8\}, \{6,9,10\}\}.$$

Table 5. An Information System from [29]

U/E	a_1	a_2	a_3	a_4	a_5	a_6
1	Big	Blue	Hard	Indefinite	Plastic	Negative
2	Medium	Red	Moderate	Smooth	Wood	Neutral
3	Small	Yellow	Soft	Fuzzy	Plush	Positive
4	Medium	Blue	Moderate	Fuzzy	Plastic	Negative
5	Small	Yellow	Soft	Indefinite	Plastic	Neutral
6	Big	Green	Hard	Smooth	Wood	Positive
7	Small	Yellow	Hard	Indefinite	Metal	Positive
8	Small	Yellow	Soft	Indefinite	Plastic	Positive
9	Big	Green	Hard	Smooth	Wood	Neutral
10	Medium	Green	Moderate	Smooth	Plastic	Neutral

The degree dependencies of those set of attributes are

$$B \Rightarrow_{k_1} C, \text{ where } k_1 = \frac{\sum_{X \in U / C} |\underline{B}(X)|}{|U|} = \frac{|\{1,2,3,4,5,6,7,8,9,10\}|}{|\{1,2,3,4,5,6,7,8,9,10\}|} = 1$$

and

$$C \Rightarrow_{k_1} D, \text{ where } k_1 = \frac{\sum_{X \in U / D} |\underline{C}(X)|}{|U|} = \frac{|\{1,2,3,4,5,6,7,8,9,10\}|}{|\{1,2,3,4,5,6,7,8,9,10\}|} = 1.$$

Thus, a nested rough set approximation of a set $X \subseteq U$ can be obtained as follows

$$\langle \underline{B}(X), \bar{B}(X) \rangle, \langle \underline{C}(X), \bar{C}(X) \rangle, \langle \underline{D}(X), \bar{D}(X) \rangle$$

with

$$\underline{D}(X) \subseteq \underline{C}(X) \subseteq \underline{B}(X) \subseteq X \subseteq \bar{B}(X) \subseteq \bar{C}(X) \subseteq \bar{D}(X).$$

The hierarchical of rough set approximations is obtained as follows.

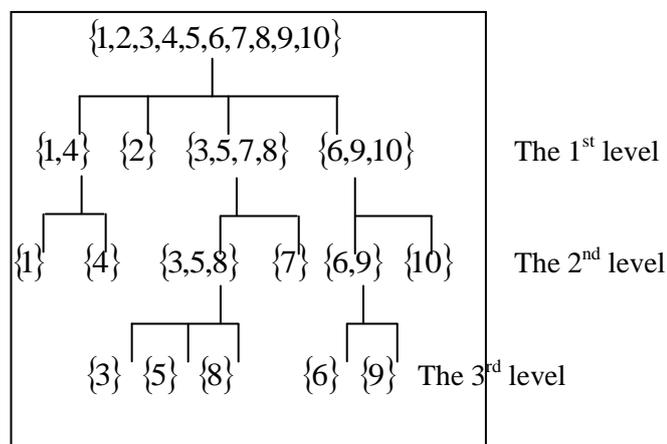


Figure 2. A Hierarchy of the Rough Set Approximations

From Figure 2, the set of classes is given by

$$C = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{7\}, \{8\}, \{9\}, \{10\}, \{1,4\}, \{6,9\}, \{3,5,8\}, \{6,9,10\}, \{3,5,7,8\}, U\}$$

Hence, we can classify the objects based on all possible partitions as follows.

$$\begin{aligned} \wp_1 &= \{\{1,2,3,4,5,6,7,8,9,10\}\} \\ \wp_2 &= \{\{1,4\}, \{2\}, \{3,5,7,8\}, \{6,9,10\}\} \\ \wp_3 &= \{\{1,4\}, \{2\}, \{3,5,7,8\}, \{6,9\}, \{10\}\} \\ \wp_4 &= \{\{1,4\}, \{2\}, \{3,5,7,8\}, \{6\}, \{9\}, \{10\}\} \\ \wp_5 &= \{\{1,4\}, \{2\}, \{3,5,8\}, \{6,9,10\}, \{7\}\} \\ \wp_6 &= \{\{1,4\}, \{2\}, \{3,5,8\}, \{6,9\}, \{7\}, \{10\}\} \\ \wp_7 &= \{\{1,4\}, \{2\}, \{3,5,8\}, \{6\}, \{7\}, \{9\}, \{10\}\} \\ \wp_8 &= \{\{1,4\}, \{2\}, \{3\}, \{5\}, \{6,9,10\}, \{7\}, \{8\}\} \\ \wp_9 &= \{\{1,4\}, \{2\}, \{3\}, \{5\}, \{6,9\}, \{7\}, \{8\}, \{10\}\} \\ \wp_{10} &= \{\{1,4\}, \{2\}, \{3\}, \{5\}, \{6\}, \{7\}, \{8\}, \{9\}, \{10\}\} \\ \wp_{11} &= \{\{1\}, \{2\}, \{3,5,7,8\}, \{4\}, \{6,9,10\}\} \\ \wp_{12} &= \{\{1\}, \{2\}, \{3,5,7,8\}, \{4\}, \{6,9\}, \{10\}\} \\ \wp_{13} &= \{\{1\}, \{2\}, \{3,5,7,8\}, \{4\}, \{6\}, \{9\}, \{10\}\} \\ \wp_{14} &= \{\{1\}, \{2\}, \{3,5,8\}, \{4\}, \{6,9,10\}, \{7\}\} \\ \wp_{15} &= \{\{1\}, \{2\}, \{3,5,8\}, \{4\}, \{6,9\}, \{7\}, \{10\}\} \\ \wp_{16} &= \{\{1\}, \{2\}, \{3,5,8\}, \{4\}, \{6\}, \{7\}, \{9\}, \{10\}\} \\ \wp_{17} &= \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6,9,10\}, \{7\}, \{8\}\} \\ \wp_{18} &= \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6,9\}, \{7\}, \{8\}, \{10\}\} \\ \wp_{19} &= \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{7\}, \{8\}, \{9\}, \{10\}\} \end{aligned}$$

Figure 3. The Possible Partitions (Clusters)

From Figure 3, we can classify the objects based on a pre-defined number of classes. This is subjective and is pre-decided based either on user requirement or domain knowledge.

4.2. Maximal Association Rules Mining

Association rules mining has become one of the hot and critical disciplines in knowledge discovery [31]. It aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction databases or other data repositories. The association rules are said to be strong if it meets the minimum confidence threshold. Concisely, support measures the number of records that contain the association rules, while confidence measures the percentage of times that the association rules are found. However, while association rules provide means to discover many interesting associations, they fail to discover other, no less interesting associations, which are also hidden in the data (Amir, *et al*, 2005). Maximal association rules mining proposed by Feldman *et al*. [32] is a variant of association rules which is designed to handle the above problem. The idea is inspired from the fact that many interesting rules in databases cannot be captured by regular rules.

To illustrate the notion of maximal association rules, let us consider the idea which are quoted directly from [26].

Let $I = \{i_1, i_2, i_3, \dots, i_n\}$ be a set of items and $D = \{t_1, t_2, \dots, t_m\}$ is transaction database over I . A taxonomy T of I is a partition of I into disjoint sets $T = \{T_1, T_2, \dots, T_k\}$. Further, each elements of T is called category. Based on Table 5, we can define a taxonomy T as $T = \{\text{Countries}, \text{Topics}\}$, where $\text{Countries} = \{\text{USA}, \text{Canada}, \text{France}\}$ and $\text{Topics} = \{\text{Corn}, \text{Fish}\}$. According to [26], for a transaction t and a category T_i , an itemset $X \subseteq T_i$ is said to be *maximal* in t if $t \cap T_i = X$. Thus, X is maximal in t if X is the largest subset of T_i which is in t . A *maximal association rule* is a rule of the form $X \Rightarrow^{\max} Y$, where X and Y are maximal subsets in distinct categories, $T(X)$ and $T(Y)$, respectively. The *support* of the rule $X \Rightarrow^{\max} Y$, denoted by $\text{MSupp}\left(X \Rightarrow^{\max} Y\right)$ is defined as

$$\text{MSupp}\left(X \Rightarrow^{\max} Y\right) = |\{t : t \text{ maximal supports } X \cup Y\}|.$$

The *confidence* of the rule $X \Rightarrow^{\max} Y$, denoted by $C_D^{\max}\left(X \Rightarrow^{\max} Y\right)$ is defined as

$$\text{MConf}\left(X \Rightarrow^{\max} Y\right) = \frac{\text{MSupp}\left(X \Rightarrow^{\max} Y\right)}{|X|}.$$

Feldman *et al.* [32] noted that maximal association rules are not designed to replace regular association rules, but rather to complement them. Every maximal association rule is also regular association, with perhaps different support and confidence [26]. While association rules [31] are based on the notion of frequent itemsets which appears in many records, maximal association rules are based on frequent maximal itemsets which appears maximally in many records[33].

Maximal association rules have been studied by many authors. Bi *et al.* [34] and Guan *et al.*

[33] proposed the same technique for discovering maximal association rules using rough set theory. The pre-requisite of their approach is based on a transformation of a transactional data set into a categorized data set. To this end, they need techniques to partition a given set of attributes to a set of concepts against a taxonomy. The technique starts with partitioning on each attribute; and then mutually computing lower approximations to find supported sets of attributes. For computing the support of the maximal rules, they used the notion of set approximation.

Let $S = (U, A, V, f)$ be an information system. Suppose that $A_i, A_j \in A$ be two distinct attributes, refer to *categories*. The partition of U generated by A_i and A_j are given as $U / A_i = \{X_1, \dots, X_n\}$ and $U / A_j = \{Y_1, \dots, Y_n\}$, respectively.

As defined in [33, 34], the supported sets from the attribute A_i with respect to Y_1, Y_2, \dots, Y_n are obtained by computing the lower approximation as

$$\text{supp}(Y_i) = \bigcup_{X_i \in U/a_i} X_i : X_i \subseteq Y_i, 1 \leq i \leq n.$$

And, the supported sets from the attribute A_j with respect to X_1, X_2, \dots, X_n are obtained by computing the lower approximation as

$$\text{supp}(X_i) = \bigcup_{Y_i \in U/a_j} Y_i : Y_i \subseteq X_i, 1 \leq i \leq n.$$

However, to compare confidence values obtained using both methods of the maximal association rule method and their method, they used the same measuring method with the maximal association method as proposed in [26], rather than the conventional formula of the rough set method. It is here, we pointed out that the proposed approach is not solely uses the rough set theory.

In the last part of this work, we show how attributes dependency in an information system can be used clearly to capture the maximal rules. We will further elaborate our proposed hierarchical rough set approximation as described in Figure 1 for discovering maximal association in document collection and compare it with the rough-set based approach of [33, 34]. The data is presented in a Boolean-valued information system derived from the widely used Reuters-21578 (<http://www.research.att.com/lewis/reuters21578.html>). It is a labeled document collection, i.e. a benchmark for text categorization, as follows: Assume that there are 10 articles regarding corn which relate to the USA and Canada and 20 other articles concerning fish and the countries USA, Canada and France.

**Table 6. A Data Set from
 (<http://www.research.att.com/lewis/reuters21578.html>)**

<i>U/E</i>	USA	Canada	France	Corn	Fish
u_1	1	1	0	1	0
...
u_{10}	1	1	0	1	0
u_{11}	1	1	1	0	1
...
u_{30}	1	1	1	0	1

From Table 6, the maximal itemsets are captured as follow.

$\{t_1, t_2, \dots, t_{10}\} \cap \text{Countries} = \{\text{USA, Canada}\}$
$\{t_{11}, t_{12}, \dots, t_{30}\} \cap \text{Countries} = \{\text{USA, Canada, France}\}$
$\{t_1, t_2, \dots, t_{10}\} \cap \text{Topics} = \{\text{Corn}\}$
$\{t_{11}, t_{12}, \dots, t_{30}\} \cap \text{Topics} = \{\text{Fish}\}$

Figure 4. The Maximal Itemsets

Meanwhile, the supports of maximal itemsets are as follow.

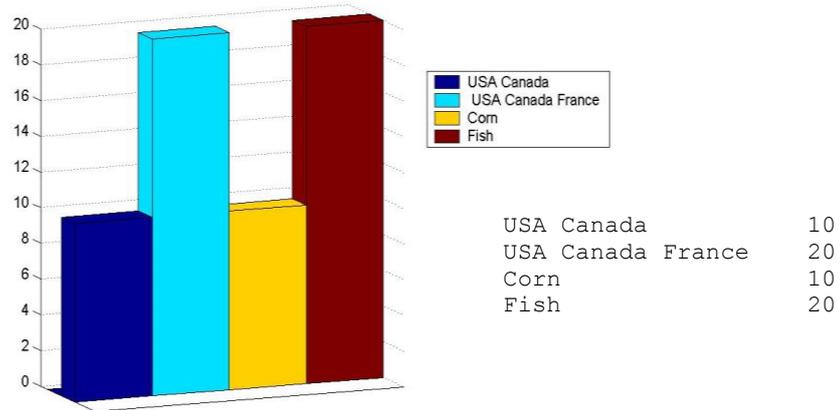


Figure 5. The Supports of Maximal Itemsets

From Figures 4 and 5, we have the partitions of category Countries as

$$U / \text{Countries} = \{\{1,2, \dots, 10\}, \{11,12, \dots, 30\}\},$$

and the partitions of category Topics as

$$U / \{\text{Topics}\} = \{\{1,2, \dots, 10\}, \{11,12, \dots, 30\}\}.$$

The hierarchical of rough set approximations are obtained as follows.

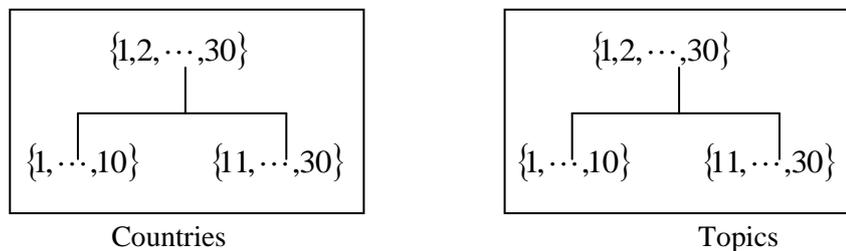


Figure 6. A Hierarchy of the Rough Set Approximations

From Figure 6, we get the maximal association between two categories as given in Figure 7.

$$\begin{aligned} & \{USA, Canada\} \xRightarrow{\max} \{Corn\}, \\ & \{Corn\} \xRightarrow{\max} \{USA, Canada\} \\ & \{USA, Canada, France\} \xRightarrow{\max} \{Fish\}, \\ & \{Fish\} \xRightarrow{\max} \{USA, Canada, France\}. \end{aligned}$$

Figure 7. The Maximal Rules

We notice from Figure 7 that, the maximal rules captured are equivalent with that of [32-34]. The maximal supports and confidences of the rules are 100% equal with their totally dependency degree.

5. Conclusion

In this paper the notion of dependency of attributes in information systems has been used. We have shown that it can be employed to define a nested sequence of indiscernibility relations. Subsequently, the notion of a nested sequence of indiscernibility relations can be used to define nested rough set approximation in information systems. Further, we have also shown that the notion of a nested rough set approximation can be used for constructing hierarchical rough set approximations. For the applications, firstly we have presented how hierarchical rough set approximations can be applied for business data clustering through a business information system. Finally, we have presented an application of such hierarchy for capturing maximal association rules in business documents collection. It is shown that our approach clearly and properly capture the maximal rules.

Acknowledgement

This work is supported by Universitas Teknologi Yogyakarta Research Grant Ref number O7/UTY-R /SK/O/X/2013.

References

- [1] T. Herawan and M. M. Deris, "Rough Set Theory for Topological Spaces in Information Systems", in *Modelling & Simulation, AMS, Third Asia International Conference*, (2009).
- [2] T. Herawan and M. M. Deris, "A Construction of Nested Rough Set Approximations in Information Systems Using Dependency of Attributes", *Power Control and Optimization: Proceedings of the Second Global Conference on Power Control and Optimization*, (2009).
- [3] Z. Pawlak, "Rough sets," *International Journal of Computer & Information Sciences*, vol. 11, (1982), pp. 341-356.
- [4] Z. Pawlak, *Rough sets: Theoretical aspects of reasoning about data* vol. 9: Springer Science & Business Media, (1991).
- [5] Z. Pawlak and A. Skowron, "Rudiments of rough sets," *Information sciences*, vol. 177, pp. 3-27, (2007).
- [6] U. o. R. D. o. C. Science and W. Ziarko, "The discovery, analysis and representation of data dependencies in databases", Regina: Department of Computer Science, University of Regina, (1991).
- [7] W. Ziarko and N. Shan, "Discovering attribute relationships, dependencies and rules by using rough sets", *46th Hawaii International Conference on System Sciences*, (1995).
- [8] W. Ziarko, "Dependencies in structures of decision tables", *Rough Sets and Intelligent Systems Paradigms*, ed: Springer, (2007), pp. 113-121.
- [9] I. Düntsch and G. Gediga, "Statistical evaluation of rough set dependency analysis", *International Journal of Human-Computer Studies*, vol. 46, (1997), pp. 589-604.
- [10] I. Düntsch and G. Gediga, "Algebraic aspects of attribute dependencies in information systems", *Fundamenta Informaticae*, vol. 29, (1997), pp. 119-133.
- [11] T. Herawan and M. M. Deris, "A framework on rough set-based partitioning attribute selection", *Emerging Intelligent Computing Technology and Applications. With Aspects of Artificial Intelligence*, ed: Springer, (2009), pp. 91-100.
- [12] T. Herawan and M. M. Deris, "Rough set theory for selecting clustering attribute", *Power Control and Optimization: Proceedings of the Second Global Conference on Power Control and Optimization*, (2009).
- [13] T. Herawan, M. M. Deris and J. H. Abawajy, "A rough set approach for selecting clustering attribute", *Knowledge-Based Systems*, vol. 23, (2010), pp. 220-231.
- [14] T. Herawan, I. T. R. Yanto and M. M. Deris, "A construction of hierarchical rough set approximations in information systems using dependency of attributes", *Advances in Intelligent Information and Database Systems*, ed: Springer, (2010), pp. 3-15.
- [15] T. Herawan, "Rough set approach for categorical data clustering", *Universiti Tun Hussein Onn Malaysia*, (2010).

- [16] N. Senan, R. Ibrahim, N. M. Nawi, I. T. R. Yanto and T. Herawan, "Rough Set Theory for Feature Ranking of Traditional Malay Musical Instruments Sounds Dataset", *Software Engineering and Computer Systems*, ed: Springer, (2011), pp. 516-529.
- [17] N. Senan, R. Ibrahim, N. M. Nawi, I. T. R. Yanto and T. Herawan, "Rough set approach for attributes selection of traditional Malay musical instruments sounds classification", *Ubiquitous Computing and Multimedia Applications*, ed: Springer, (2011), pp. 509-525.
- [18] M. Magnani, "Technical report on rough set theory for knowledge discovery in data bases", University of Bologna, (2003).
- [19] Y. Yao, "Rough sets, neighborhood systems, and granular computing", *Can an Conf. on Electr Comput. Eng.*, (1999).
- [20] Y. Yao, "Stratified rough sets and granular computing", *Proceedings of the 18th International Conference of the North American Fuzzy Information Processing Society*, (1999).
- [21] Y. Yao, "Information granulation and rough set approximation", *International Journal of Intelligent Systems*, vol. 16, (2001), pp. 87-104..
- [22] I. Yanto, T. Herawan and M. Deris, "RoCeT: Rough set approach for clustering web transactions", *International Journal of Biomedical and Human Sciences*, vol. 16, (2010), pp. 135-145.
- [23] I. T. R. Yanto, T. Herawan and M. M. Deris, "A Framework of Rough Clustering for Web Transactions", *Advances in Intelligent Information and Database Systems*, ed: Springer, (2010), pp. 265-277.
- [24] I. T. R. Yanto, T. Herawan and M. M. Deris, "Data clustering using variable precision rough set", *Intelligent Data Analysis*, vol. 15, (2011), pp. 465-482.
- [25] I. T. R. Yanto, P. Vitasari, T. Herawan and M. M. Deris, "Applying variable precision rough set model for clustering student suffering study's anxiety", *Expert Systems with Applications*, vol. 39, (2012), pp. 452-459.
- [26] A. Amir, Y. Aumann, R. Feldman and M. Fresko, "Maximal association rules: A tool for mining associations in text", *Journal of Intelligent Information Systems*, vol. 25, (2005), pp. 333-345.
- [27] W. Marek and H. Rasiowa, "Gradual approximating sets by means of equivalence relations", *Bulletin of Polish Academy of Sciences, Mathematics*, vol. 35, (1987), pp. 233-238.
- [28] Y. Yao, "Granular computing using neighborhood systems", *Advances in Soft Computing*, ed: Springer, (1999), pp. 539-553.
- [29] D. Parmar, T. Wu and J. Blackhurst, "MMR: an algorithm for clustering categorical data using rough set theory", *Data & Knowledge Engineering*, vol. 63, (2007), pp. 879-893.
- [30] A. K. Jain, M. N. Murty and P. J. Flynn, "Data clustering: a review", *ACM computing surveys (CSUR)*, vol. 31, (1999), pp. 264-323.
- [31] R. Agrawal, T. Imieliński and A. Swami, "Mining association rules between sets of items in large databases", *ACM SIGMOD Record*, (1993), pp. 207-216.
- [32] R. Feldman, Y. Aumann, A. Amir, A. Zilberstein and W. Kloesgen, "Maximal Association Rules: A New Tool for Mining for Keyword Co-Occurrences in Document Collections", *KDD*, (1997), pp. 167-170.
- [33] J. Guan, D. A. Bell and D. Liu, "The rough set approach to association rule mining", *Data Mining, ICDM, Third IEEE International Conference*, (2003).
- [34] Y. Bi, T. Anderson and S. McClean, "A rough set model with ontologies for discovering maximal association rules in document collections", *Knowledge-Based Systems*, vol. 16, (2003), pp. 243-251.

