Knowledge-Based Systems 24 (2011) 186–195

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A soft set approach for association rules mining

Tutut Herawan [a], Mustafa Mat Deris [b,*]

[a] Faculty of Computer System and Software Engineering, Universiti Malaysia Pahang, Gambang 26300, Pahang, Malaysia
[b] Faculty of Information Technology and Multimedia, Universiti Tun Hussein Onn Malaysia, Parit Raja, Batu Pahat 86400, Johor, Malaysia

ARTICLE INFO

Article history:
Received 7 June 2009
Received in revised form 23 August 2010
Accepted 24 August 2010
Available online 9 September 2010

Keywords:
Association rules mining
Maximal association rules mining
Boolean-valued information systems
Soft set theory
Items co-occurrence

ABSTRACT

In this paper, we present an alternative approach for mining regular association rules and maximal association rules from transactional datasets using soft set theory. This approach is started by a transformation of a transactional dataset into a Boolean-valued information system. Since the "standard" soft set deals with such information system, thus a transactional dataset can be represented as a soft set. Using the concept of parameters co-occurrence in a transaction, we define the notion of regular and maximal association rules between two sets of parameters, also their support, confidence and maximal support, maximal confidences, respectively properly using soft set theory. The results show that the soft regular and soft maximal association rules provide identical rules as compared to the regular and maximal association rules.

1. Introduction

Data mining is a generic term which covers research results, techniques and tools used to extract useful information from large databases. Association rule is one of the most popular data mining techniques and has received considerable attention, particularly since the publication of the AIS and Apriori algorithms [1,2]. They are particularly useful for discovering relationships among data in huge databases and applicable to many different domains including market basket and risk analysis in commercial environments, epidemiology, clinical medicine, fluid dynamics, astrophysics, and crime prevention. The association rules are considered interesting if it satisfies certain constraints, i.e. predefined minimum support (*minsupp*) and minimum confidence (*minconf*) thresholds. Many algorithms of association rules mining have been proposed, including the works of [3–5]. The association rules method was developed particularly for the analysis of transactional databases, whose attributes posses Boolean values and has been shown to be a very efficient data structure for association rules mining. In other words, the occurrence of an item can be viewed as a Boolean variable and its value is "1" if it appears in that particular transaction and "0" otherwise. This conforming that a transactional dataset can be converted into a Boolean-valued information system, $S = (U, A, V_{\{0,1\}}, f)$.

Soft set theory [6], proposed by Molodtsov in 1999, is a new general method for dealing with uncertain data. Soft sets are called (binary, basic, elementary) neighborhood systems. As for standard soft set, it may be redefined as the classification of objects in two distinct classes, thus confirming that soft set can deal with a Boolean-valued information system. Molodtsov [6] pointed out that one of the main advantages of soft set theory is that it is free from the inadequacy of the parameterization tools, unlike in the theories of fuzzy set [7], probability and interval mathematics. In recent years, research on soft set theory has been active, and great progress has been achieved, including the works of the using of fundamental soft set theory, soft set theory in abstract algebra and soft set theory for data analysis, particularly in decision making [8–12]. Since the "standard" soft set (F, E) over the universe U can be represented by a Boolean-valued information system, thus a soft set can be used for representing a transactional dataset. Therefore, one of the applications of soft set theory for data mining is for mining association rules. However, not many researches have been done on this application.

In this work, we propose an alternative approach for regular and maximal association rules mining using soft set theory. We use the Boolean-valued information system that has been shown to be a very efficient data structure for association rules mining. The Boolean-valued information system is a conversion from a transactional dataset. We define the regular support, regular confidence and maximal support and maximal confidence of the rules, respectively, based on the concept of co-occurrences and maximal co-occurrences of parameters in a transactional dataset under soft set theory. There are three main contributions of this work. First, we present that a soft set can be used to represent a transactional data via a Boolean-valued information system. Second, we present the applicability of the soft set theory for mining "regular" association rules and maximal association rules. Third, we show that by using soft set theory, association rules and maximal rules

* Corresponding author. Tel.: +60 7 4538001; fax: +60 7 4532199.
E-mail addresses: tutut@ump.edu.my (T. Herawan), mmustafa@uthm.edu.my (M.M. Deris).

0950-7051/$ - see front matter Crown Copyright © 2010 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.knosys.2010.08.005

discovered are identical to that rules in association rules and maximal association rules approaches.

The rest of this paper is organized as follows. Section 2 describes fundamental concept of "regular" association rules and maximal association rules mining. Section 3 describes the fundamental concept of information system and soft set theory. Section 4 describes a transformation of a transaction table into a soft set via a Boolean-valued information system. Section 5 describes soft set approach for association rules and maximal association rules mining. Section 6 describes the results of the proposed approaches. Finally, the conclusion of this work and future work are described in Section 7.

2. Preliminaries

2.1. Association rules

Let $I = \{i_1, i_2, \ldots, i_{|A|}\}$, for $|A| > 0$ refers to the set of literals called *set of items* and the set $D = \{t_1, t_2, \ldots, t_{|U|}\}$, for $|U| > 0$ refers to the transactional dataset, where each transaction $t \in D$ is a list of distinct items $t = \{i_1, i_2, \ldots, i_{|M|}\}$, $1 \leqslant |M| \leqslant |A|$ and each transaction can be identified by a distinct identifier *TID*. Let, a set $X \subseteq t \subseteq I$ called an *itemset*. An itemset with k-items is called a *k-itemset*. The *support* of an itemset X, denoted sup (X) is defined as a number of transactions contain X. An *association rule* between sets X and Y is an implication of the form $X \Rightarrow Y$, where $X \cap Y = \phi$. The itemsets X and Y are called *antecedent* and *consequent*, respectively. The *support* of an association rule $X \Rightarrow Y$, denoted sup $(X \Rightarrow Y)$, is defined as a number of transactions in D contain $X \cup Y$. The *confidence* of an association rule $X \Rightarrow Y$, denoted cfi $(X \Rightarrow Y)$ is defined as a ratio of the numbers of transactions in D contain $X \cup Y$ to the number of transactions in D contain X. Thus, $\mathrm{cfi}(X \Rightarrow Y) = \frac{\sup(X \Rightarrow Y)}{\sup(X)}$.

A huge number of association rules can be found from a transactional dataset. To find the interesting association rules in a transactional dataset, we must define a specified minimum support (called *minsup*) and specified minimum confidence (called *minconf*). The itemset $Y \subseteq I$ is called *frequent itemset* if sup $(X) \geqslant$ minsup. It is known that a subset of any frequent itemset is a frequent itemset, a superset of any infrequent itemset is not a frequent itemset. Finally, the association rule $X \Rightarrow Y$ holds if conf $(X \Rightarrow Y) \geqslant$ minconf.

The association rules are said to be strong if it meets the minimum confidence threshold. However, while association rules provide means to discover many interesting associations, they fail to discover other, no less interesting associations, which also hidden in the data. Maximal association rules introduced by Feldman et al. [13] is a variant of association rules which is designed to handle the above problem. It allows the discovery of associations pertaining to items that most often do not appear alone, but rather together with closely related items, and hence associations relevant only to these items tend to obtain low confidence. These rules are very important in discovering maximal association, particularly from documents text collection. The idea is inspired from the fact that many interesting rules in databases cannot captured by regular rules. Feldman et al. 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 [14]. While association rules 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 [15]. Using only maximal association rules, many interesting regular associations may and will be lost. The initial step to discover maximal rules is a partition on the set of items from a transactional dataset so-called a taxonomy and categorization of items.

2.2. Taxonomy and category

Let $I = \{i_1, i_2, \ldots, i_{|A|}\}$ be a set of items. A *taxonomy* T of I is a partition of I into disjoint subsets, i.e., $T = \{T_1, T_2, \ldots, T_n\}$. Each member of T is called a *category*. For an item i, we denote $T(i)$ the category that contain i. Similarly, if X is an itemset all of which are from a single category, then we denote this category by $T(X)$.

Example 1. There is a dataset consisting of the 10 transactions [13]; 2 articles referring to Countries "Canada, Iran, USA" and refers to Topics "crude, ship"; 1 article referring to "USA" and refers to "earn"' 2 articles referring to "USA" and refers to "jobs, cpi"; 1 article referring to "USA" and refers to "earn, cpi"; 1 article referring to "Canada" and refers to "sugar, tea"; 2 articles referring to "Canada, USA" and refers to "trade, acq" and 1 article referring to "Canada, USA" and refers to "earn". We can present such transactions in the following table.

We can create a taxonomy based on Table 1, which is contains two categories "countries" and "topics", i.e., T = {countries, topics}, where

$$\text{countries} = \{\text{Canada},\ \text{Iran},\ \text{USA}\}$$

and

$$\text{topics} = \{\text{crude},\ \text{ship},\ \text{earn},\ \text{jobs},\ \text{cpi},\ \text{sugar},\ \text{tea},\ \text{trade}\}.$$

2.3. Maximal association rules

To illustrate the notion of maximal association rules, let we consider the idea which are quoted directly from [14]. In maximal association rule $X \overset{\max}{\Rightarrow} Y$, we are interested in capturing the notion that whenever X appears alone then Y also appears, with some confidence. For this, we must fist define the notion of alone. We do so with respect to the categories of T as follows.

For a transaction t, a category T_i and an itemset $X \subseteq T_i$, we say that X is *alone* in t if $t \cap T_i = X$. That is, X is alone in t if X is the largest subset of T_i which is in t. In this case we also say that X is *maximal* in t and that t M-supports X. For a database D, the M-support of X in D, denoted $S_D^{\max}(X)$ is the number of transaction $t \in D$ that M-support X.

A *maximal association rule* or *M*-association rule is a rule of the form $X \overset{\max}{\Rightarrow} Y$, where X and Y subsets distinct categories, $T(X)$ and $T(Y)$, respectively. The M-support of the M-association rule $X \overset{\max}{\Rightarrow} Y$, denoted by $S_D^{\max}(X \overset{\max}{\Rightarrow} Y)$ is defined as

$$S_D^{\max}(X \overset{\max}{\Rightarrow} Y) = |\{t : tM - \text{supports } X \text{ and } t \text{ supports } Y\}|.$$

That is, $S_D^{\max}(X \overset{\max}{\Rightarrow} Y)$ is the number of transactions in D that M-support X and also support Y in the regular sense. The intuitive meaning

Table 1
A data of transactions from [13].

TID	Items
1	Canada, Iran, USA, crude, ship
2	Canada, Iran, USA, crude, ship
3	USA, earn
4	USA, jobs, cpi
5	USA, jobs, cpi
6	USA, earn, cpi
7	Canada, sugar, tea
8	Canada, USA, trade, acq
9	Canada, USA, trade, acq
10	Canada, USA, earn

of the M-association rule $X \overset{\max}{\Rightarrow} Y$ is that whenever a transaction M-supports X, then Y also appears in the transaction, with some probability. However, in measuring this probability we are only interested in those transactions where some element of $T(Y)$ (the category of Y) appears in the transaction. Accordingly, the maximal confidence is defined as follows.

Let $D(X,T(Y))$ be the subset of the database D consisting of all the transactions that M-support X and contain at least one element of $T(Y)$. The confidence of the M-association rule $X \overset{\max}{\Rightarrow} Y$, denoted by $C_D^{\max}\left(X \overset{\max}{\Rightarrow} Y\right)$ is defined as

$$C_D^{\max}(X \overset{\max}{\Rightarrow} Y) = \frac{S_D^{\max}(X \overset{\max}{\Rightarrow} Y)}{|D(X,T(Y))|}.$$

As in regular association rule, to find the interesting association rules in a transactional dataset, we must define a specified minimum M-support (called *min M-sup*) and specified minimum confidence (called *min M-conf*). The maximal itemset $X \subseteq I$ is called *frequent maximal itemset* if $S_D^{\max}(X) \geqslant \min M\sup$. In this case, a subset of any frequent maximal itemset is not necessarily a frequent maximal itemset.

We define the M-factor of the M-association rule $X \overset{\max}{\Rightarrow} Y$ to be the ratio between the M-confidence of the M-association rule $X \overset{\max}{\Rightarrow} Y$ and the confidence of the corresponding regular association $X \Rightarrow Y$. Specifically, let D' be the sub set of transactions that contain at least one item of $T(Y)$. Then, the M-factor of the M-association rule $X \overset{\max}{\Rightarrow} Y$ is defined as

$$M - \text{factor}(X \overset{\max}{\Rightarrow} Y) = \frac{C_D^{\max}(X \overset{\max}{\Rightarrow} Y)}{C_{D'}(X \Rightarrow Y)}.$$

Note that in defining the M-factor, the denominator is the confidence of the regular association rule $X \Rightarrow Y$ with respect to D'. The reason is that since the M-confidence is defined with respect to D', the comparison to regular associations must also be with respect to this set. In general, we seek M-association rules with a higher M-factor, which are the more interesting rules.

Example 2. We consider the transaction data in Table 1. From the notions of support of an itemset, we have the following supported sets (not all).

From Table 1, we let two categories, i.e., T_1 = countries and T_2 = topics, from the notions M-support, we have the following all supported maximal item sets.

From Figs. 1 and 2, the maximal association rules captured with min$M$sup = 2 and min$M$conf = 0.5 are given as follows.

From Fig. 3, if we consider the rules as regular association, then for the first rule we have the support, sup = 2 with cfi = 22% and M-factor 2.27. For the second rule, we have sup = 2 with cfi = 40% and M-factor 1.65. Meanwhile, for the last rule, we have sup = 2 with cfi = 100% and M-factor 1.

The notions of an information system, a Boolean-valued information system, and a soft set are presented in the following Section. The definition and an example of a soft set are quoted

$$\begin{aligned}
&\sup\{\text{Canada}\}=6,\ \sup\{\text{USA}\}=9,\ \sup\{\text{Iran}\}=2,\\
&\sup\{\text{Canada, USA}\}=5,\ \sup\{\text{Canada, Iran}\}=2,\\
&\sup\{\text{Iran, USA}\}=2,\ \sup\{\text{Canada, Iran, USA}\}=2\\
&\sup\{\text{crude}\}=2,\ \sup\{\text{ship}\}=2,\ \sup\{\text{earn}\}=3,\\
&\sup\{\text{jobs}\}=2,\ \sup\{\text{cpi}\}=3,\ \sup\{\text{sugar}\}=1,\\
&\sup\{\text{tea}\}=1,\ \sup\{\text{trade}\}=2,\ \sup\{\text{acq}\}=2,\\
&\sup\{\text{crude, ship}\}=2,\ \sup\{\text{jobs, cpi}\}=2,\\
&\sup\{\text{earn, cpi}\}=1,\ \sup\{\text{sugar, tea}\}=1,\ \sup\{\text{trade, acq}\}=2
\end{aligned}$$

Fig. 1. The supported itemsets.

$$\begin{aligned}
&M\sup\{\text{Canada}\}=1,\ M\sup\{\text{USA}\}=4,\\
&M\sup\{\text{Canada, USA}\}=3,\ M\sup\{\text{Canada, Iran, USA}\}=2,\\
&M\sup\{\text{crude, ship}\}=2,\ M\sup\{\text{earn}\}=2,\ M\sup\{\text{jobs, cpi}\}=2,\\
&M\sup\{\text{earn, cpi}\}=1,\ M\{\text{sugar, tea}\}=1,\ M\{\text{trade, acq}\}=2
\end{aligned}$$

Fig. 2. The supported maximal itemsets.

$$\begin{aligned}
&\{\text{USA}\}\Rightarrow\{\text{jobs, cpi}\} \text{ with } S_D^{\max}=2 \text{ and } C_D^{\max}=50\%\\
&\{\text{Canada, USA}\}\Rightarrow\{\text{acq, trade}\} \text{ with } S_D^{\max}=2 \text{ and } C_D^{\max}=66\%\\
&\{\text{Canada, Iran, USA}\}\Rightarrow\{\text{crude, ship}\} \text{ with } S_D^{\max}=2 \text{ and } C_D^{\max}=100\%
\end{aligned}$$

Fig. 3. The maximal rules.

directly from [6,16]. The relation of a soft set with a Boolean-valued information system also described in this section.

3. Soft set theory

Throughout this section U refers to an initial universe, E is a set of parameters, $P(U)$ is the power set of U.

3.1. Information system

The syntax of an information system is very similar to relations in relational database. Entities in relational databases are also represented by tuples of attribute values. An *information system* as is a 4-tuple (quadruple) $S = (U,A,V,f)$, where $U = \{u_1,u_2,\ldots,u_{|U|}\}$ is a non-empty finite set of objects, $A = \{a_1,a_2,\ldots,a_{|A|}\}$ 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 an information 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 and can be intuitively expressed in terms of an information table (see Table 2).

In the above table, note that each maps $t = f(u,a) : U \times A \rightarrow V$ is a tuple

$$t_i = (f(u_i,a_1), f(u_i,a_2), f(u_i,a_3), \ldots, f(u_i,a_{|A|})),$$
where $i = 1,2,3,\ldots,|U|$.

The tuple t is not necessarily associated with entity uniquely. In an information table, two distinct entities could have the same tuple representation (duplicated/redundant tuple), which is *not permissible* in relational databases. Thus, the concept of information systems is a generalization of the concept of relational databases. In an information system $S = (U,A,V,f)$, if $V_a = \{0,1\}$, for every $a \in A$, then S is called a *Boolean-valued information system*.

3.2. Soft set theory

Definition 1. A pair (F,E) is called a soft set over U, where F is a mapping given by

$$F : E \rightarrow P(U).$$

Table 2
An information system.

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

In other words, a soft set over U is a parameterized family of subsets of the universe U. For $e \in E$, $F(e)$ may be considered as the set of e-elements of the soft set (F,E) or as the set of e-approximate elements of the soft set. Clearly, a soft set is not a (crisp) set. To illustrate this idea, let we consider the following example.

Example 3. Let we consider a soft set (F,E) which describes the "attractiveness of houses" that Mr. X is considering to purchase. Suppose that there are six houses in the universe U under consideration,

$$U = \{h_1, h_2, h_3, h_4, h_5, h_6\}$$

and

$$E = \{e_1, e_2, e_3, e_4, e_5\}$$

is a set of decision parameters, where e_1 stands for the parameters "expensive", e_2 stands for the parameters "beautiful", e_3 stands for the parameters "wooden", e_4 stands for the parameters "cheap", e_5 stands for the parameters "in the green surrounding".

Consider the mapping

$$F : E \to P(U),$$

given by "houses $(\cdot)$", where $(\cdot)$ is to be filled in by one of parameters $e \in E$.

Suppose that

$$F(e_1) = \{h_2, h_4\}, \quad F(e_2) = \{h_1, h_3\}, \quad F(e_3) = \{h_3, h_4, h_5\},$$
$$F(e_4) = \{h_1, h_3, h_5\}, \quad F(e_5) = \{h_1\}.$$

Therefore, $F(e_1)$ means "houses (expensive)", whose functional value is the set $\{h_2,h_4\}$. Thus, we can view the soft set (F,E) as a collection of approximations as below

$$(F,E) = \left\{ \begin{array}{l} \text{expensive houses} = \{h_2, h_4\}, \\ \text{beautiful houses} = \{h_1, h_3\}, \\ \text{wooden houses} = \{h_3, h_4, h_5\}, \\ \text{cheap houses} = \{h_1, h_3, h_5\}, \\ \text{in the green surrounding houses} = \{h_1\} \end{array} \right\}$$

Each approximation has two parts, a predicate p and an approximate value set v. For example, for the approximation "expensive houses = $\{h_2,h_4\}$", we have the predicate name of expensive houses and the approximate value set or value set if $\{h_2,h_4\}$.

Thus, a soft set (F,E) can be viewed as a collection of approximations below:

$$(F,E) = \{p_1 = v_1, p_2 = v_2, p_3 = v_3, \ldots, p_n = v_n\}.$$

In recent years, research on soft set theory has been active, and great progress has been achieved. The soft set is a mapping from parameter to the crisp subset of universe. From such case, we may see the structure of a soft set can classify the objects into two classes (yes/1 or no/0). Thus we can make a one-to-one correspondence between a Boolean-valued information system and a soft set, as stated in Proposition 1 (see Table 3).

Table 3
Tabular representation of a soft set in the above example.

U	e_1	e_2	e_3	e_4	e_5
h_1	0	1	0	1	1
h_2	1	0	0	0	0
h_3	0	1	1	1	0
h_4	1	0	1	0	0
h_5	0	0	1	1	0
h_6	0	0	0	0	0

Proposition 1. *If (F,E) is a soft set over the universe U, then (F,E) is a Boolean-valued information system $S = (U,A,V_{\{0,1\}},f)$.*

Proof. Let (F,E) be a soft set over the universe U, we define a mapping

$$F = \{f_1, f_2, \ldots, f_n\},$$

where

$$f_1 : U \to V_1 \quad \text{and } f_1(x) = \begin{cases} 1, & x \in F(e_1), \\ 0, & x \notin F(e_1), \end{cases}$$
$$f_2 : U \to V_2 \quad \text{and } f_2(x) = \begin{cases} 1, & x \in F(e_2), \\ 0, & x \notin F(e_2), \end{cases}$$
$$\vdots$$
$$f_n : U \to V_n \quad \text{and } f_n(x) = \begin{cases} 1, & x \in F(e_n), \\ 0, & x \notin F(e_n). \end{cases}$$

Thus, if $A = E$, $V = \bigcup_{e_i \in A} V_{e_i}$, where $V_{e_i} = \{0,1\}$, then a soft set (F,E) can be considered as a Boolean-valued information system $S = (U,A,V_{\{0,1\}},f)$. □

Obviously, for the reverse process, an information system of Boolean-value can be represented as a soft set.

4. Transformation of a transactional dataset into a soft set

In this section we discuss a transformation of a transaction table into a soft set via a Boolean-valued information system.

4.1. Transforming a transactional dataset into a soft set

Let $I = \{i_1,i_2,\ldots,i_{|A|}\}$ be a set of items and $D = \{t_1,t_2,\ldots,t_{|U|}\}$ be a transactional dataset. For a Boolean-valued information systems, $S = (U,A,V_{\{0,1\}},f)$, we have the following transformation.

For every $a \in A$ and $u \in U$, we define the map (bitmap) f: $U \times A \to \{0,1\}$ such that $f(u,a) = 1$ if a appears in t, otherwise $f(u,a) = 0$. Then, we have a Boolean-valued information system as a quadruple $S = (U,A,V_{\{0,1\}},f)$. The information system $S = (U,A,V_{\{0,1\}},f)$ is referred to as a transformation of a transactional dataset into a Boolean-valued information system.

A Boolean-valued information system $S = (U,A,V_{\{0,1\}},f)$, is said to be *deterministic* if for every $a \in A$ and for every $u \in U$, $|f(u,a)| = 1$. The data structures of a transactional dataset and a Boolean-valued information system give the same attribute-value of objects. For the soft set theory, it is known that a Boolean-valued information system is an efficient data structure for storing information. Therefore the transactional dataset schema has to be transformed into a Boolean-valued information system (bitmap table). The first column (U) holds the primary key attribute *TID*, each of the remaining columns contain a Boolean-valued attribute i.e. an item from the transactional dataset. This structure is efficient, because the column-address of a Boolean-valued attribute doesn't change. Therefore it is possible to check very fast each tuple, by direct access for the appropriate column, whether the searched item is or is not available in the corresponding tuple of the original dataset. If it appear, then there is a '1', otherwise a '0'.

From Proposition 1, since every Boolean-valued information system can be represented as a soft set, and from Fig. 4, we have a transformation of a transactional dataset into a soft set as follows.

5. Soft set theory for association rules mining

In this section we present the applicability of soft set theory for association rules and maximal association rules mining. The

$$\left.\begin{array}{l} i_1 \rightarrow a_1 \\ i_2 \rightarrow a_2 \\ \vdots \;\; \vdots \;\; \vdots \\ i_{|A|} \rightarrow a_{|A|} \end{array}\right. \Leftrightarrow I = \{i_1, i_2, \cdots, i_{|A|}\} \rightarrow A = \{a_1, a_2, \cdots, a_{|A|}\}$$

and

$$\left.\begin{array}{l} t_1 \rightarrow u_1 \\ t_2 \rightarrow u_2 \\ \vdots \;\; \vdots \;\; \vdots \\ t_{|U|} \rightarrow u_{|U|} \end{array}\right. \Leftrightarrow D = \{t_1, t_2, \cdots, t_{|U|}\} \rightarrow U = \{u_1, u_2, \cdots, u_{|U|}\}$$

Fig. 4. The transformation of a transactional dataset into a Boolean-valued information system.

pre-requisite of using soft set approach for maximal association rules mining is the transactional dataset need to be transformed into a soft set, where each item is regarded as a parameter (attribute). In the proposed approach, we use the notion of co-occurrence of parameters for association rules mining as used in [17].

Throughout this section, and from Fig. 5, the pair (F,E) refers to the soft set over the universe U representing a Boolean-valued information system $S = (U, A, V_{\{0,1\}}, f)$ from the transactional dataset, $D = \{t_1, t_2, \ldots, t_{|U|}\}$.

Definition 2. Let (F,E) be a soft set over the universe U and $u \in U$. An items co-occurrence set in a transaction u can be defined as

$$\mathrm{Coo}(u) = \{e \in E : f(u,e) = 1\}.$$

Obviously, $\mathrm{Coo}(u) = \{e \in E : F(e) = 1\}$.

Example 3. The soft set representing Table 1 is given below.

From Fig. 6, we have the following co-occurrence of each item.

5.1. Soft association rules mining

Definition 4. Let (F,E) be a soft set over the universe U and $X \subseteq E$. A set of attributes X is said to be supported by a transaction $u \in U$ if $X \subseteq \mathrm{Coo}(u)$.

Definition 5. Let (F,E) be a soft set over the universe U and $X \subseteq E$. The support of a set of parameters X, denoted by sup (X) is defined by the number of transactions U supporting X, i.e.

$$I = \{i_1, i_2, \cdots, i_{|A|}\} \Rightarrow A = \{a_1, a_2, \cdots, a_{|A|}\} \Rightarrow E = \{e_1, e_2, \cdots, e_{|E|}\},$$
$$D = \{t_1, t_2, \cdots, t_{|U|}\} \Rightarrow U = \{u_1, u_2, \cdots, u_{|U|}\}$$

and

$$f_i : U \rightarrow V_i \text{ and } f_i(x) = \begin{cases} 1, & x \in F(e_i) \\ 0, & x \notin F(e_i) \end{cases}$$

Thus,

$$D = \{t_1, t_2, \cdots, t_{|U|}\} \Rightarrow (F, E)$$

Fig. 5. The transformation of a transactional dataset into a soft set.

$$(F,E) = \left\{\begin{array}{l} \text{Canada} = \{1,2,7,8,9,10\}, \text{Iran} = \{1,2\}, \text{USA} = \{1,2,3,4,5,6,8,9,10\}, \\ \text{crude} = \{1,2\}, \text{ship} = \{1,2\}, \text{earn} = \{3,10\}, \text{jobs} = \{4\}, \text{cpi} = \{3,10\}, \\ \text{tea} = \{7\}, \text{sugar} = \{7\}, \text{trade} = \{8,9\}, \text{acq} = \{8,9\} \end{array}\right\}$$

Fig. 6. The soft set representing Table 1.

$$\begin{array}{l}
\mathrm{Coo}(u_1) = \{\text{Canada, Iran, USA, crude, ship}\} \\
\mathrm{Coo}(u_2) = \{\text{Canada, Iran, USA, crude, ship}\} \\
\mathrm{Coo}(u_3) = \{\text{USA, earn}\} \\
\mathrm{Coo}(u_4) = \{\text{USA, job, cpi}\} \\
\mathrm{Coo}(u_5) = \{\text{USA, job, cpi}\} \\
\mathrm{Coo}(u_6) = \{\text{USA, earn, cpi}\} \\
\mathrm{Coo}(u_7) = \{\text{Canada, tea, sugar}\} \\
\mathrm{Coo}(u_8) = \{\text{Canada, USA, trade, acq}\} \\
\mathrm{Coo}(u_9) = \{\text{Canada, USA, trade, acq}\} \\
\mathrm{Coo}(u_{10}) = \{\text{Canada, USA, earn}\}
\end{array}$$

Fig. 7. The co-occurrence of items in a transaction.

$$\begin{array}{l}
\sup\{\text{Canada}\} = |\{u_1, u_2, u_7, u_8, u_9, u_{10}\}| = 6, \\
\sup\{\text{USA}\} = |\{u_1, u_2, u_3, u_4, u_5, u_6, u_8, u_9, u_{10}\}| = 9, \\
\sup\{\text{Iran}\} = |\{u_1, u_2\}| = 2, \\
\sup\{\text{Canada, USA}\} = |\{u_1, u_2, u_8, u_9, u_{10}\}| = 5 \\
\sup\{\text{Canada, Iran}\} = |\{u_1, u_2\}| = 2, \\
\sup\{\text{Iran, USA}\} = |\{u_1, u_2\}| = 2, \\
\sup\{\text{Canada, Iran, USA}\} = |\{u_1, u_2\}| = 2, \\
\sup\{\text{crude}\} = |\{u_1, u_2\}| = 2, \\
\sup\{\text{ship}\} = |\{u_1, u_2\}| = 2, \\
\sup\{\text{earn}\} = |\{u_3, u_6, u_{10}\}| = 3, \\
\sup\{\text{jobs}\} = |\{u_4, u_5\}| = 2, \\
\sup\{\text{cpi}\} = |\{u_4, u_5, u_6\}| = 3, \\
\sup\{\text{sugar}\} = |\{u_7\}| = 1, \ \sup\{\text{tea}\} = |\{u_7\}| = 1, \\
\sup\{\text{trade}\} = |\{u_8, u_9\}| = 2, \ \sup\{\text{acq}\} = |\{u_8, u_9\}| = 2, \\
\sup\{\text{crude, ship}\} = |\{u_1, u_2\}| = 2, \\
\sup\{\text{jobs, cpi}\} = |\{u_4, u_5\}| = 2, \\
\sup\{\text{earn, cpi}\} = |\{u_6\}| = 1, \ \sup\{\text{sugar, tea}\} = |\{u_7\}| = 1, \\
\sup\{\text{trade, acq}\} = |\{u_8, u_9\}| = 2
\end{array}$$

Fig. 8. The supported sets.

$$\sup(X) = |\{u : X \subseteq \mathrm{Coo}(u)\}|,$$

where $|X|$ is the cardinality of X.

Example 6. From Definitions 4 and 5, the supported set obtained from Fig. 7 can be shown as in Fig. 8.

Definition 7. Let (F,E) be a soft set over the universe U and X, $Y \subseteq E$, where $X \cap Y = \phi$. An association rule between X and Y is an implication of the form $X \Rightarrow Y$. The itemsets X and Y are called antecedent and consequent, respectively.

Definition 8. Let (F,E) be a soft set over the universe U and X, $Y \subseteq E$, where $X \cap Y = \phi$. The support of a association rule $X \Rightarrow Y$, denoted by sup $(X \Rightarrow Y)$ is defined by

$$\sup(X \Rightarrow Y) = \sup(X \cup Y) = |\{u : X \cup Y \subseteq \mathrm{Coo}(u)\}|$$

Definition 9. Let (F,E) be a soft set over the universe U and X, $Y \subseteq E$, where $X \cap Y = \phi$. The confidence of a association rule $X \Rightarrow Y$, denoted respectively by conf $(X \Rightarrow Y)$ and conf $(X \Rightarrow Y)$ is defined by

Pseudo-code: Soft association rules

```
function Soft association rule(A)
[m n]=size(A); value=[];comb=1;
while combination~=n;
    c=combinations(1:n,comb);[a b]=size(c);
    for l=1:a;
        combination=A(:,v=c(l,:));
        if length(v)==1
            attribute=find(combination==length(v))';
        else
        attribute =find(sum(combination')'==length(v))';
        end
        if length(attribute)>=%;
        check=v;
[m2 n2]=size(A);combination2=1;
while combination2~=n2;
    c2= combinations(1:n2,combination2);[a2 b2]=size(c2);
    for l=1:a2;
        combination2=A(:,v2=c2(l,:));
        if length(v2)==1
            attribute2=find(combination2==length(v2))';
        else
        attribute2=find(sum(combination2')'==length(v2))';
        end
        if length(attribute2)>=%;
       ck=ismember(check,v2);
       if sum(ck)==0
           v22=A(:,[check  v2]);
           combinationv22=sum(v22')' ;
           attribute=find(combinationv22==length([check v2])');
           if length(atr)>=%
               sup1=length(attribute);
               sup=length(attribute);
               conf=sup./sup1;
           end
       end
        end
    end;
    combination2=combination2+1;
end
       end
    end;
    combination=combination+1;
end
```

Fig. 9. The pseudo-code for soft association rules mining.

$$\mathrm{conf}(X \Rightarrow Y) = \frac{\sup(X \cup Y)}{\sup(X)} = \frac{|\{u : X \cup Y \subseteq \mathrm{Coo}(u)\}|}{|\{u : X \subseteq \mathrm{Coo}(u)\}|}.$$

The pseudo-code for mining association rules using soft set theory is given in Fig. 9.

5.2. Taxonomy and categorization using soft set theory

In this sub-section, we present the taxonomy and categorization using soft set theory as follows. (See Figs. 10 and 11)

$M\ \sup\{\text{Canada}\} = |\{u_7\}| = 1,$
$M\ \sup\{\text{USA}\} = |\{u_3, u_4, u_5, u_6\}| = 4,$
$M\ \sup\{\text{Canada, USA}\} = |\{u_8, u_9, u_{10}\}| = 3,$
$M\ \sup\{\text{Canada, Iran, USA}\} = |\{u_1, u_2\}| = 2,$
$M\ \sup\{\text{crude, ship}\} = |\{u_1, u_2\}| = 2,\ M\ \sup\{\text{earn}\} = |\{u_3, u_{10}\}| = 2,$
$M\ \sup\{\text{jobs, cpi}\} = |\{u_4, u_5\}| = 2,\ M\ \sup\{\text{earn, cpi}\} = |\{u_6\}| = 1,$
$M\{\text{sugar, tea}\} = |\{u_7\}| = 1,\ M\{\text{trade, acq}\} = |\{u_8, u_9\}| = 2$

Fig. 10. The supported maximal sets.

$\{\text{USA}\} \Rightarrow \{\text{jobs, cpi}\}$ with $M\ \sup = 2$ and $M\mathrm{conf} = 50\%$
$\{\text{Canada, USA}\} \Rightarrow \{\text{acq, trade}\}$ with $M\ \sup = 2$ and $M\mathrm{conf} = 66\%$
$\{\text{Canada, Iran, USA}\} \Rightarrow \{\text{crude, ship}\}$ with $M\ \sup = 2$ and $M\mathrm{conf} = 100\%$

Fig. 11. The maximal association rules.

Let (F,E) be a soft set over the universe U. A *taxonomy* T of E is a partition of E into disjoint subsets, i.e., $T = \{E_1, E_2, \ldots, E_n\}$. Each member of T is called a *category*. For an item i, we denote $T(i)$ the category that contain i. Similarly, if X is an itemset all of which are from a single category, then we denote this category by $T(X)$.

5.3. Soft maximal association rules

In Section 5.1, we have defined the notion of an association rule between two sets of parameters, its support and confidence using soft set theory based on the co-occurrence of parameters/items concept in a transaction. In this sub-section, we will show how soft set theory can be properly used for mining maximal association rules. We will show that by using soft set theory, rules discovered are identical and faster to maximal association rules in [13,14] and rough set-based maximal association rule in [15,17,18].

```
Pseudo-code: Soft maximal association rules
function Soft maximal association rules
support(A,category,tol,Mconfid)
n=1; attribute1=1;AD=A;R=0; AC=AD;ca=clock;
while n~=length(category)+1;
    attribute2=attribute1+category(n)-1;n2=category(n);
    while n2~=0;
        c=combinations(akat:bkat,n2);[a b]=size(c);
    for l=1:a;
        v=c(l,:);combination=A(:,v);
        if length(v)==1
            attribute=find(combination==length(v))' ;
        else
        komb2=sum(combination')';attribute=find(komb2==length(v))';
        end
        if length(attribute)>=tol;check=v;attribute;A([attribute],[check])=0;
 AB=AD;ns=1; attribute1=1;
while ns~=length(category)+1;
    attribute2= attribute1+category(ns)-1;ns2=category(ns);
    while ns2~=0;
        cs=combinations(attribute1:attribute2,ns2);[as bs]=size(cs);
    for l=1:as;
        vs=cs(l,:); combinations=AB(:,vs);
        if length(vs)==1
            attributes=find(combinations==length(vs))' ;
        else
 combination2s=sum(combinations')';attributes=find(combination2s==length(vs))';
        end
        if length(attributes)>=tol;
       AB([attributes],[vs])=0; ck=ismember(check,vs);
       if sum(check)==0
       v22=AC(:,[check  vs]);combinationv22=sum(v22')';length([check  vs]);
           attribute=find(combinationv22==length([check  vs])');
           if length(attribute)>=tol
               sup1=length(attribute); sup=length(atr);conf=sup./sup1;
               AC([attribute],[cek vs])=0;
              if conf>=Mconfid
               R=R+1;from=check,to=vs,conf
               disp('=========================');
             end
           end
       end
          end
      end
      ns2=ns2-1;
    end,
    attribute1=attribute2+1; ns=ns+1;
end
    end
end
      n2=n2-1;
    end,
    attribute1=attribute2+1; n=n+1;cb=clock;
end
time=etime(cb,ca);
disp('Response time');disp(time);
```

Fig. 12. The pseudo-code for soft maximal association rules mining.

Definition 10. Let (F,E) be a soft set over the universe U and $X \subseteq E_i$. A set of attributes X is said to be maximal supported by a transaction u if

$X = \text{Coo}(u) \cap E_i$.

Definition 11. Let (F,E) be a soft set over the universe U and $X \subseteq E_i$. The maximal support of a set of parameters X, denoted by sup (X) is defined by the number of transactions U maximal supporting X, i.e.

$$M\sup(X) = |\{u : X = \text{Coo}(u) \cap E_i\}|,$$

where $|X|$ is the cardinality of X.

Obviously, $M\sup(X) = |\{e : e \in X \wedge F(e) = 1\}|$.

Table 4
A data of transactions from [19].

U/A	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

$$(F,E)=\begin{Bmatrix}\text{USA}=\{1,\cdots,30\},\text{Canada}=\{1,\cdots,30\},\text{France}=\{11,\cdots,30\},\\ \text{corn}=\{1,\cdots,10\},\text{fish}=\{11,\cdots,30\}\end{Bmatrix}$$

Fig. 13. The soft set representing Table 4.

$$\text{Coo}(u_1)=\cdots=\text{Coo}(u_{10})=\{\text{USA, Canada, corn}\}$$
$$\text{Coo}(u_{11})=\cdots=\text{Coo}(u_{30})=\{\text{USA, Canada, France, fish}\}$$

Fig. 14. The co-occurrence of items in a transaction from Table 4.

```
USACanada           10
USACanadaFrance     20
Corn                10
Fish                20
```

Fig. 15. The maximal supported sets from Table 4.

Example 12. Let T_1 = countries and T_2 = topics, from the notions *M*-support, we have the following supported maximal sets.

Definition 13. Let (F,E) be a soft set over the universe U and two maximal itemsets $X, Y \subseteq E_i$, where $X \cap Y = \phi$. A maximal association rule between X and Y is an implication of the form $X \overset{\max}{\Rightarrow} Y$. The itemsets X and Y are called maximal antecedent and maximal consequent, respectively.

Definition 14. Let (F,E) be a soft set over the universe U and two maximal itemsets $X, Y \subseteq E_i$, where $X \cap Y = \phi$. The maximal support of a maximal association rule $X \Rightarrow Y$, denoted by $M\sup(X \overset{\max}{\Rightarrow} Y)$ is defined by

$$M\sup(X \overset{\max}{\Rightarrow} Y) = M\sup(X \cup Y) = |\{u : X \cup Y = \text{Coo}(u) \cap E_i\}|$$

Definition 15. Let (F,E) be a soft set over the universe U and two maximal itemsets $X, Y \subseteq E_i$, where $X \cap Y = \phi$. The confidence of a maximal association rule $X \overset{\max}{\Rightarrow} Y$, denoted respectively by $M\text{conf}(X \overset{\max}{\Rightarrow} Y)$ and conf $(X \Rightarrow Y)$ is defined by

$$M\text{conf}(X \overset{\max}{\Rightarrow} Y) = \frac{M\sup(X \cup Y)}{M\sup(X)} = \frac{|\{u : X \cup Y = \text{Coo}(u) \cap E_i\}|}{|\{u : X = \text{Coo}(u) \cap E_i\}|}.$$

Example 16. The maximal association rules captured with min-*M*sup = 2 and min*M*conf = 0.5 are given as follows.

The pseudo-code for mining maximal association rules using soft set theory is given in Fig. 12.

6. Experimental results

In this section, we compare the proposed soft maximal association rules mining with the algorithms of [13–15,17,18]. The proposed approach is elaborated through two datasets derived from [19] and [3]. The algorithm of the proposed approach is implemented in MATLAB version 7.6.0.324 (R2008a). All of the algorithms are executed sequentially on a processor Intel Core 2 Duo CPUs. The total main memory is 1 gigabyte and the operating system is Windows XP Professional SP3.

Table 5
The mapped air pollution data.

Data	Items
$CO_2 \geqslant 0.02$	*a*
$O_3 \leqslant 0.007$	*b*
$PM_{10} \geqslant 80$	*c*
$SO_2 \geqslant 0.04$	*d*
$NO_2 \geqslant 0.03$	*e*

Table 6
The executed transactions.

TID	Item	TID	Item
1	*a, e*	16	*a, c, d, e*
2	*a*	17	*a, c, d, e*
3	*a, d*	18	*a, b, c, d, e*
4	*a*	19	*a, c, d, e*
5	*a, c, d, e*	20	*a, c, d, e*
6	*a, b, c, d, e*	21	*a, c, d, e*
7	*a, b, c, d, e*	22	*a, c, d, e*
8	*a, c, e*	23	*a, e*
9	*a, d*	24	*a*
10	*a, d, e*	25	*a, d, e*
11	*a, d, e*	26	*d*
12	*a, d*	27	*a, d, e*
13	*c, d*	28	*a, c, d, e*
14	*a, c, d, e*	29	*a, c, d, e*
15	*a, c, e*	30	*a, d*

Table 7
The time response improvement of maximal association rules and rough set-based maximal association rules approaches by soft maximal association rules approach.

	Response time improvement
A	87%
B	50%

6.1. A Dataset derived from the widely used Reuters-21578 [19]

A labeled document collection, i.e. a benchmark for text categorization, as follows. Assume that there are 10 articles regarding product corn which relate to the countries USA and Canada and 20 other articles concerning product fish and the countries USA, Canada and France. The information system is given in the following table.

The soft set representing Table 4 is given below:

From Fig. 13, we have the following co-occurrence.

Based on Fig. 14, we can make taxonomy as follows T = {countries, products}, where countries = {USA, Canada, France} and topics = {corn, fish}. The maximal supported sets are (see Fig. 15)

```
Antecedent        Consequent        Msup   Mconf

USACanada         Corn              10     100%
USACanadaFrance   Fish              20     100%
Corn              USACanada         10     100%
Fish              USACanadaFrance   20     100%
```

Fig. 16. The maximal association rules.

$$(F,E)=\left\{\begin{array}{l} \mathrm{a}=\{u_1,u_2,u_3,u_4,u_5,u_6,u_7,u_8,u_9,u_{10},u_{11},u_{12},u_{14},u_{15},u_{16},u_{17},\\ u_{18},u_{19},u_{20},u_{21},u_{22},u_{23},u_{24},u_{25},u_{27},u_{28},u_{29},u_{30}\},\mathrm{b}=\{u_6,u_7,u_{18}\},\\ \mathrm{c}=\{u_5,u_6,u_7,u_8,u_{13},u_{14},u_{15},u_{16},u_{17},u_{18},u_{19},u_{20},u_{21},u_{22},u_{28},u_{29}\},\\ \mathrm{d}=\{u_3,u_5,u_6,u_7,u_9,u_{10},u_{11},u_{12},u_{13},u_{14},u_{16},u_{17},u_{18},u_{19},u_{20},u_{21},u_{22}\},\\ \mathrm{e}=\{u_1,u_5,u_6,u_7,u_8,u_{10},u_{11},u_{14},u_{15},u_{16},u_{17},u_{18},u_{19},u_{20},u_{21},u_{22},u_{23},u_{25},u_{27},u_{28},u_{29}\} \end{array}\right\}$$

Fig. 17. The soft set representing the transaction data in Table 6.

$$\begin{array}{l} \mathrm{Coo}(u_1)=\{a,e\},\ \mathrm{Coo}(u_2)=\{a\},\ \mathrm{Coo}(u_3)=\{a,d\},\ \mathrm{Coo}(u_4)=\{a\},\\ \mathrm{Coo}(u_5)=\{a,c,d,e\},\ \mathrm{Coo}(u_6)=\{a,b,c,d,e\},\ \mathrm{Coo}(u_7)=\{a,b,c,d,e\},\\ \mathrm{Coo}(u_8)=\{a,c,e\},\ \mathrm{Coo}(u_9)=\{a,d,e\},\ \mathrm{Coo}(u_{10})=\{a,d,e\},\\ \mathrm{Coo}(u_{11})=\{a,d,e\},\ \mathrm{Coo}(u_{12})=\{a,d\},\ \mathrm{Coo}(u_{13})=\{d\},\\ \mathrm{Coo}(u_{14})=\{a,c,d,e\},\ \mathrm{Coo}(u_{15})=\{a,c,e\},\ \mathrm{Coo}(u_{16})=\{a,c,d,e\},\\ \mathrm{Coo}(u_{17})=\{a,c,d,e\},\ \mathrm{Coo}(u_{18})=\{a,b,c,d,e\},\ \mathrm{Coo}(u_{19})=\{a,c,d,e\},\\ \mathrm{Coo}(u_{20})=\{a,c,d,e\},\ \mathrm{Coo}(u_{21})=\{a,c,d,e\},\ \mathrm{Coo}(u_{22})=\{a,c,d,e\},\\ \mathrm{Coo}(u_{24})=\{a\},\ \mathrm{Coo}(u_{25})=\{a,d,e\},\ \mathrm{Coo}(u_{26})=\{d\},\ \mathrm{Coo}(u_{27})=\{a,d,e\},\\ \mathrm{Coo}(u_{28})=\{a,c,d,e\},\ \mathrm{Coo}(u_{29})=\{a,c,d,e\},\ \mathrm{Coo}(u_{30})=\{a,d\} \end{array}$$

Fig. 18. The co-occurrence of items in a transaction from Table 6.

a	2
b	3
ad	2
ae	2
dc	1
ade	4
acde	13

Fig. 19. The maximal supported sets from Table 6.

Antecedent	Consequent	Msup	Mconf
b	acde	3	100%

Fig. 20. Maximal association rules obtained.

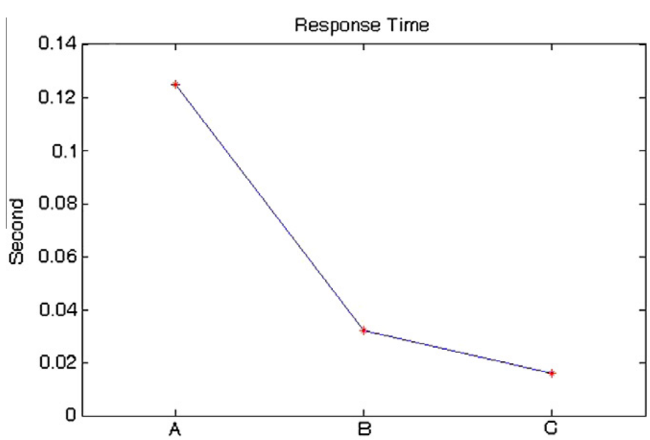

Fig. 21. The comparison of executing time.

The maximal rules captured as in Fig. 16 are equivalent with that in [13,15,17,18]. Since in this dataset is easy to capture the maximal rules, we do not a comparison in term of executing time. The executing time of our proposed approach through this dataset is 0.014 s.

6.2. The air pollution dataset from [3]

We will further explain an example of mining association rules using soft set theory from a transactional dataset. It based on the observation of the air pollution data taken in Kuala Lumpur on July 2002 as presented and used in [3]. The association rules of the presented results are based on a set of air pollution data items, i.e. {CO_2, O_3, PM_{10}, SO_2, NO_2}. The value of each item is with the unit of part per million (ppm) except PM_{10} is with the unit of micrograms (μgm). The data were taken for every one-hour every day. The actual data is presented as the average amount of each data item per day. For brevity, each data item is mapped to parameters *a*, *b*, *c*, *d*, and *e* respectively, as shown in Table 5. From Table 5, there are five different parameters (items), i.e. $\{a,b,c,d,e\}$ and 30 transactions. Each transaction is defined as a set of data items corresponds to such numbers. The executed transactions are described in Table 6.

From Table 6, we let the set of parameters as the set of items, $E=\{a,b,c,d,e\}$. Thus, the soft set representing Table 6 is: (see Table 7).

From the soft set in Fig. 17, we have the following co-occurrence sets in each transaction (see Fig. 18).

Based on Tables 5 and 6, we can make taxonomy as follows:

$$T=\{\text{dangerous condition},\ \text{good condition}\},$$

where

$$\text{dangerous condition}=\{a,c,d,e\}\quad\text{and}\quad\text{good condition}=\{b\}.$$

The supported maximal itemsets are given as follows:

For capturing interesting maximal rules in the air pollution dataset, we set the minimum *M*-support and minimum *M*-confidence as min*M* sup = 2 and min*M*conf = 50%, respectively. And the rule discovered is given in Fig. 19.

The comparison of response time of traditional [13,14] (A), rough set-based [15,17,18] (B) and soft set-based (C) maximal association rules approaches is given in the following Figure.

The improvement of time response in capturing maximal rules of soft maximal association rules approach for maximal association rules approaches and rough set-based maximal association rules approaches are given in the following table (see Figs. 20 and 21).

7. Conclusion and future works

In this paper, we have presented an alternative approach for "regular" association rules and maximal association rules mining from a transactional dataset using soft set theory. This approach is started by a transformation of a transactional dataset into a soft set. Using the concept of parameters co-occurrence in a transaction, we have defined the notion of support and confidence of "regular" association rules using soft set theory. Meanwhile, for soft maximal association rule, we have defined the notion of its *M*-support and *M*-confidence under soft set theory. We have elaborated the proposed approach through a benchmark dataset for text categorization and a dataset of air pollution in Kuala Lumpur on July 2002. We have shown that the maximal rules discovered using soft set theory under given min*M*sup and min*M*cof thresholds are identical to the maximal association rules and rough set-based maximal association rules mining approaches. However, the proposed approach achieves faster time to capture the rules. With this approach, we believe that some applications using soft set theory for mining various levels of association rules and decision support systems will be easier. For the future research, we investigated the applicability of the rough-set based dependency of attributes [20] and the concept of approximate reducts [21] in an information system for maximal association rules mining.

References

[1] R. Agrawal, T. Imielinski, A. Swami, Mining association rules between sets of items in large databases, in: Proceedings of the ACM SIGMOD International Conference on the Management of Data, 1993, pp. 207–216.
[2] R. Agrawal, R. Srikant, Fast algorithms for mining association rules, in: Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), 1994, pp. 487–499.
[3] M. Mat Deris, N.F. Nabila, D.J. Evans, M.Y. Saman, A. Mamat, Association rules on significant rare data using second support, International Journal of Computer Mathematics 83 (1) (2006) 69–80.
[4] A.H.L. Lim, C.S. Lee, Processing online analytics with classification and association rule mining, Knowledge-Based Systems 23 (3) (2010) 248–255.
[5] Y.L. Chen, C.H. Weng, Mining fuzzy association rules from questionnaire data, Knowledge-Based Systems 22 (1) (2009) 46–56.
[6] D. Molodtsov, Soft set theory-first results, Computers and Mathematics with Applications 37 (1999) 19–31.
[7] L.A. Zadeh, Fuzzy set, Information and Control 8 (1965) 338–353.
[8] P.K. Maji, A.R. Roy, R. Biswas, An application of soft sets in a decision making problem, Computers and Mathematics with Applications 44 (2002) 1077–1083.
[9] D. Chen, E.C.C. Tsang, D.S. Yeung, X. Wang, The parameterization reduction of soft sets and its applications, Computers and Mathematics with Applications 49 (2005) 757–763.
[10] A.R. Roy, P.K. Maji, A fuzzy soft set theoretic approach to decision making problems, Journal of Computational and Applied Mathematics 203 (2007) 412–418.
[11] Y. Zou, Z. Xiao, Data analysis approaches of soft sets under incomplete information, Knowledge Based Systems 21 (2008) 941–945.
[12] Z. Kong, L. Gao, L. Wang, S. Li, The normal parameter reduction of soft sets and its algorithm, Computers and Mathematics with Applications 56 (2008) 3029–3037.
[13] R. Feldman, Y. Aumann, A. Amir, A. Zilberstein, W. Klosgen, Maximal association rules: a new tool for mining for keywords cooccurrences in document collections, in: The Proceedings of the KDD 1997, 1997, pp. 167–170.
[14] A. Amir, Y. Aumann, R. Feldman, M. Fresco, Maximal association rules: a tool for mining associations in text, Journal of Intelligent Information Systems 25 (3) (2005) 333–345.
[15] J.W. Guan, D.A. Bell, D.Y. Liu, The rough set approach to association rule mining, in: The Proceedings of the Third IEEE International Conference on Data Mining (ICDM'03), 2003, pp. 529–532.
[16] P.K. Maji, R. Biswas, A.R. Roy, Soft set theory, Computers and Mathematics with Applications 45 (2003) 555–562.
[17] Y. Bi, T. Anderson, S. McClean, A rough set model with ontologies for discovering maximal association rules in document collections, Knowledge-Based Systems 16 (2003) 243–251.
[18] J.W. Guan, D.A. Bell, D.Y. Liu, Mining Association Rules with Rough Sets, Studies in Computational Intelligence, Springer Heidelberg, 2005. pp. 163–184.
[19] Reuters-21578, <http://www.research.att.com/lewis/reuters21578.html>, retrieved on July 12th, 2010.
[20] T. Herawan, M. Mat Deris, J.H. Abawajy, A rough set approach for selecting clustering attribute, Knowledge-Based Systems 23 (3) (2010) 220–231.
[21] Y. Qian, J. Liang, D. Li, F. Wang, N. Ma, Approximation reduction in inconsistent incomplete decision tables, Knowledge-Based Systems 23 (5) (2010) 427–433.