Framework for finding maximal association rules in mobile web service environment using soft set

Krishna Kumar Mohbey*
Department of Computer Science,
Central University of Rajasthan,
Ajmer, 305817, India
Email: kmohbey@gmail.com
*Corresponding author

G.S. Thakur
Department of Mathematics and Computer Applications,
National Institute of Technology Bhopal,
Bhopal 462051, India
Email: ghanshyamthakur@gmail.com

Abstract: Electronic commerce is very popular nowadays. It is a fast and convenient way to transfer information and communicate with people. E-commerce uses various web services to perform a specific task. When a particular user accessed web services, they are sequentially stored into a database that is called web service sequences. Association rules are used to correlate different web services for knowledge prediction. In this paper, we design a framework for generating maximal association rules of accessed web service sequences using soft set. Soft set uses binary values for their standard representation. This framework converts web service sequences into Boolean-valued information system using the concept of coexistence attributes in a sequence. We define the concept of maximal association rules between attribute sets. Here, maximal support and confidence are also defined using soft set. Experimental results show that the proposed soft-set-based framework provides identical rules when compared with other maximal association rules and rough-set-based rules.

Keywords: web services sequence; maximal association rule; soft set; coexist services; Boolean value system.


Biographical notes: Krishna Kumar Mohbey is an Assistant Professor of Computer Science at Central University of Rajasthan, India. He received his Bachelor’s degree in Computer Application from MCRPV Bhopal (2006), Master’s in Computer Application from Rajiv Gandhi Technological University Bhopal (2009) and PhD from Department of Mathematics and Computer Applications from National Institute of Technology Bhopal, India (2015). His areas of interest are data mining, mobile web services, big data analysis and user behaviour analysis.
1 Introduction

Behaviour analysis, mathematical modelling and computing of problems with uncertainty are one of the newest areas in interdisciplinary research involving computational cleverness and decision-making. It is seen that uncertainty arising from various domains has very different nature and cannot be captured within a single framework. In addition to behaviour analysis, probability theory and statistics, we currently have some advanced soft computing methods such as fuzzy sets (Zadeh, 1965), rough sets (Pawlak, 2002) and soft sets (Molodtsov, 1999) for dealing with different kinds of uncertainty. For handling uncertainty, Molodtsov (1999) initiated the concept of soft sets via a set-valued mapping. A distinguishing feature of soft sets, which is different from probability theory, fuzzy sets, rough set and interval mathematics, is that precise quantity such as probability and membership grade is not essential. Fuzzy set theory is based on the method of gradualness and describes fuzzy concepts using membership functions. It provides an effective way for modelling vagueness and ambiguity in human reasoning and intelligent decision-making process. The invisible relation generated from the collected data is the mathematical basis of rough sets, which treat uncertainty using the method of granulation expressed by rough lower and upper approximations (Pawlak, 2002). Molodtsov’s soft sets (Molodtsov, 1999) provide us a new way of coping with uncertainty from the viewpoint of parameterisation. There are lots of researches going on soft set. It includes fundamental soft set theory, abstract algebra using soft set, data analysis using the soft set and decision-making through soft set (Maji et al., 2002; Chen et al., 2005; Roy and Maji, 2007; Kong et al., 2008; Zou and Xiao, 2008). Soft sets have potential applications in various fields such as the smoothness of functions, game theory, operations research, Riemann integration, probability theory and e-commerce (Molodtsov, 1999; Feng et al., 2014; Caron, 2015). Since then, research on soft sets has been very active and received much attention from researchers worldwide. Nowadays, e-commerce is playing the vital role in our daily life. It is composed of different web services for different purposes. These services are the part of e-commerce, which are used to transmit information from one place to another in short time, accessing useful information from the internet and responsible for online transactions. Web services are lightweight applications or programs, which can be accessed using the internet (Mohanty et al., 2010; Sang et al., 2012). These web services are mostly accessed by Smartphone or laptop devices. The services that are accessed by mobile devices are called mobile services. A particular user may access a series of web services at different time at different locations or a single location. These web services are useful for different perspectives. By these services, mobile became very useful, easier and faster (Mohbey and Thakur, 2015). To extract the
interesting pattern of services, data mining techniques played an important role. Some mostly accessed mobile services are Facebook, WhatsApp, Viber, Line, Gmail, Truecaller, online Dictionary, online Ticket Booking, Online Banking, etc. There are various companies that provide specific web services to the user (Erkoc and Romo-Fragoso, 2015). By the use of these web services, user behaviour can be predicted with various kinds of constraints and utilities (Papadopoulos et al., 2013). Utility mining has emerged as one of the most significant research issues in data mining field. In utility-based pattern mining, each service has external utility or importance such as profit or maximum uses, and internal importance, which indicate non-binary values of services in mobile e-commerce environment such as number of times particular service is accessed (Papadopoulos et al., 2013). In e-commerce environment, user behaviour prediction has arisen as an emerging issue in data mining. Over the past years, some studies have employed various behaviour pattern analyses, such as management and development of websites, planning of e-commerce environment and cross-marketing in business environments (Mohbey and Thakur, 2015). Web services accessing sequence are important for predicting behaviour as well as providing advertisements to users at different locations. When a particular user accessed services, accessed sequence is generated (Mohbey and Thakur, 2013). Figure 1 shows a simple scenario of web services sequence. SID1 shows that when a user visits office, restaurant, park and home, he/she uses video, message, WhatsApp and chat in a sequence. Some web service accessed sequences are stored in Table 1. These generated sequences are used in association mining between various mobile services. One can predict a variety of knowledge from these associations of mobile web services. Data mining is also an emerging field, which covers research techniques, tools and results to extract useful information from large sequences of data (Arduin et al., 2013; Kim, 2009). Association rule is one of the most popular techniques in data mining, which is used to find the correlation between different parameters.

Figure 1  Example of web service accessing sequence (see online version for colours)
To find frequent patterns from transaction databases, various studies have been proposed. Apriori is the pioneer for mining frequent itemsets from transaction databases (Agrawal et al., 1993; Mittal et al., 2015). After the publication of AIS and Apriori algorithm (Agrawal et al., 1993; Agrawal and Srikant, 1994), association rules received considerable attention. They are generally used for generating a relationship between large sequences of data. Association rules are applicable in different domains, including market basket analysis, risk analysis, fluid dynamics, crime preventions, stock market analysis, etc. A particular association rule is considered as interesting if it satisfies certain constraints. These constraints are predefined minimum support and predefined minimum confidence thresholds. Various association rule mining algorithms have been proposed (Herawan and Deris, 2011; Deris et al., 2006; Lim and Lee, 2010; Chen and Weng, 2009). The regular association rules are based on the concept of frequent sets of attributes, which occurs in many sequences (Agrawal et al., 1993; Amir et al., 2005; Deris et al., 2006; Lim and Lee, 2010; Herawan et al., 2010), whereas maximal association rules are based on frequent maximal sets of attribute, which appear maximally in many sequences. Like the regular association rules, maximal association rules are of the form \( X \Rightarrow Y \) with parameter set \( X \) and \( Y \). However, while a regular association rule \( X \Rightarrow Y \), if \( X \) then \( Y \) with some confidence. It means that if someone buys \( X \) he/she also buys \( Y \) with some confidence. A maximal association rule \( X \Rightarrow Y \) says that if \( X \) maximally, then \( Y \) maximally with some confidence. Here, maximally indicates to the maximum value among all the other transactions. Similarly, in web service environment, those services that are more accessed together are said to be maximally associated web services. The association rule method was developed for the analysis of transactional databases, whose parameters have Boolean values. The values of a parameter are generally stored in the Boolean data structure. In other words, if an item is available to transaction, its value is 1 in Boolean variable otherwise 0. The same Boolean value representation can also be applied to sequences of mobile services. If a web service is available in a sequence, its Boolean variable value will be 1, otherwise 0. Soft set is called (binary, basic, elementary) neighbourhood system. As for standard soft set,
it may be redefined as the classification of objects in two distinct classes, thus it confirms that soft set can deal with Boolean values. Standard soft set, i.e., \((F, E)\) over the universe \(U\), can be represented in Boolean value, thus a soft set can be used for representing web services sequence. Therefore, soft sets are useful in finding association rules between different web services. However, not many researchers have done work on web service association mining using soft set (Herawan and Deris, 2011).

In this paper, we proposed a framework for finding maximal association rules using soft sets. We also used Boolean values to represent web services sequence because it is an efficient data structure for finding association rules. In our proposed work, we have found Boolean values from accessing web service sequences, firstly, and then applied maximal association mining. We have defined regular support, confidence and maximal support and confidence, respectively, on association rules with soft set and coexistence between web service parameters. In this paper firstly, we determine that soft set can be used to represent a web services sequence via a Boolean value. Second, we used a soft set for generating maximal association in web services sequences. Third, it is identified that soft-set-based maximal association rules are identical to simple association rules. The proposed soft-set-based framework is better than other approaches because it uses parameterised concept, while other approaches like fuzzy set needed membership function and rough set uses granulation method.

The rest of the paper is organised as follows. Section 2 describes preliminaries, i.e., fundamental concepts of association, maximal association rules and taxonomy with categories. Section 3 describes the concept of a Boolean value system, soft set concept and representation of Boolean values. Section 4 describes the conversion of web services sequence into a soft set using Boolean values. Section 5 describes the association rule and maximal association rule generation using soft sets. Section 6 describes the experimental analysis and the results of the proposed approach. Finally, the conclusions of this proposed work are discussed in Section 7.

2 Preliminaries

2.1 Association rules

Let \(W_s = \{w_{s_1}, w_{s_2}, w_{s_3}, \ldots, w_{s_A}\}\) for \(|A| > 0\) refers to the set of web services and set \(W = \{s_1, s_2, s_3, \ldots, s_U\}\) for \(|U| > 0\) refers to the web services sequences set, where each sequence \(s \in W\) is the list of distinct web services \(s = \{w_{s_1}, w_{s_2}, w_{s_3}, \ldots, w_{s_M}\}\), \(1 \leq |M| \leq |A|\) and each sequence can be identified by a distinct identifier SID. Let a set \(X \subseteq s \subseteq W_s\) called a web service set. A web service set with \(k\)-web service is called a \(k\)-web service set. The support of a web service set \(X\), denoted by support \((X)\), is defined as a number of sequences that contain \(X\). An association rule between \(X\) and \(Y\) set is an implication of \(X \Rightarrow Y\), where \(X \cap Y = \emptyset\). The web service set \(X\) and \(Y\) are called antecedent and consequent, respectively. The support \((X \Rightarrow Y)\) is defined as a number of sequences \(W\) that contain \(X \cup Y\). The confidence \(X \Rightarrow Y\) is the ratio of the number of sequences in \(W\) that contain \(X \cup Y\) to the number of sequences in \(W\) that contain \(X\). Minimum support and minimum confidence are used to find interesting association rules.
in web services sequence databases. Any web services set \( Y \subseteq Ws \) is called frequent web services set if \( \text{support}(X) \geq \text{minimum_support} \) and the association rule \( X \Rightarrow Y \) holds if confidence \( (X) \geq \text{minimum_confidence} \). If \( \text{minimum_support} \) and \( \text{minimum_confidence} \) are not applied on the transactional dataset, a huge number of association rules can be generated. It is also known that a subset of any frequent itemset is a frequent itemset and a superset of any infrequent itemset is not a frequent itemset.

2.2 Maximal association rules

Maximal association rules are the variety of association rules, which is introduced by Feldman et al. (1997) for discovering fewer interesting associations. If any rule did not follow the minimum confidence constraint, it is not strong. 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. This kind of association can be found in the maximal association rule. Maximal association rules discovered together with closely related web services, which are having low confidence. The idea is inspired by the fact that many interesting rules in a sequence database cannot be discovered by regular association rules. Feldman et al. noted that maximal association rule is not a replacement for regular association rules, but rather to complement them. Every maximal association rule is also a regular association rule with different support and confidence (Amir et al., 2005; Zadeh, 1965). While association rules are based on notion of frequent itemsets that appear in many records, maximal association rules are based on frequent maximal itemsets that appear maximum in various records. To discover maximal association rules, initial set of web services is partially from a sequence database; it is called categorisation and taxonomy of web services. A maximal association rule is a rule of the form \( X \Rightarrow Y \), where \( X \) and \( Y \) subset distinct category \( T_a(X) \) and \( T_a(Y) \), respectively. The \( \text{max_support} \) of the maximal association rule \( X \Rightarrow Y \), denoted by \( \text{sup}_{\text{max}} \), is defined as

\[
\text{sup}_{\text{max}} \ (X \Rightarrow Y) = |\{s : s \text{ support } X \text{ and } s \text{ support } Y\}|
\]  

(1)

\( \text{sup}_{\text{max}} \) is the number of sequences in \( W \) that \( \text{max_support} X \) and also support \( Y \) in the regular sense. The maximal association rule \( X \Rightarrow Y \) says that when a web service sequence \( \text{max_support} X \), then \( Y \) also present in sequence with some probability. The confidence of the maximal association rule, denoted by \( \text{conf}_{\text{max}} \), is defined by

\[
\text{conf}_{\text{max}} \ (X \Rightarrow Y) = \frac{\sup_{\text{max}} \ (X \Rightarrow Y)}{w(\text{sup}(X), \text{sup}(Y))}
\]

(2)

To find frequent maximal web service set, minimum \( \text{max_confidence} \) must be predefined and it follows \( \text{sup}_{\text{max}} \geq \min \text{max_support} \). We also defined \( \text{max_factor} \) of maximal association rule \( X \Rightarrow Y \) is the ratio between the maximal confidence of the maximal association rule \( X \Rightarrow Y \) and the confidence of the corresponding regular association \( X \Rightarrow Y \).
max_factor_{\text{max}}(X \Rightarrow Y) = \frac{\text{conf}_{\text{max}}(X \Rightarrow Y)}{|\text{conf}_{\text{max}}(X \Rightarrow Y)|}, \quad (3)

where $W'$ is a subset of sequences that contain at least one service of $T_d(Y)$.

### 2.3 Category and taxonomy

Let $W = \{w_1, w_2, w_3, \ldots, w_n\}$ be a set of web services. A taxonomy $T_d$ of $W$ is a partition of $W$ into disjoint subsets, i.e., $T_d = \{T_1, T_2, T_3, \ldots, T_n\}$. Each member of the $T_d$ is called a category. For a web service $w$, we denote $T_d(w)$ the category that contains $w$. Similarly, if $X$ is a web service set and all of them are from a single category, then it is denoted by $T_d(X)$ (Herawan and Deris, 2011).

**Example 1:** There is a web service accessed sequence database, consisting of 10 sequences. It is shown in Table 1. On the basis of the table, we can create the following taxonomy, which contains two categories ‘Web services’ and ‘Actions’, i.e., $T_d = \{\text{Web services}, \text{Actions}\}$, where Web services $= \{\text{Facebook, WhatsApp, Gmail}\}$ and Actions $= \{\text{chat, msg, mail, video, photo, news, tag, like, mms}\}$.

**Example 2:** Table 1 represents the web services accessed sequence. From the concept of support of a service set, we have some supported sets. When $\text{minimum\_support} = 0.2$ and $\text{minimum\_confidence} = 0.5$, we have the following service sets.

<table>
<thead>
<tr>
<th>Service</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>6</td>
</tr>
<tr>
<td>Gmail</td>
<td>9</td>
</tr>
<tr>
<td>WhatsApp</td>
<td>2</td>
</tr>
<tr>
<td>Facebook, Gmail</td>
<td>5</td>
</tr>
<tr>
<td>Facebook, WhatsApp, Gmail</td>
<td>2</td>
</tr>
<tr>
<td>chat</td>
<td>3</td>
</tr>
<tr>
<td>msg</td>
<td>2</td>
</tr>
<tr>
<td>mail</td>
<td>2</td>
</tr>
<tr>
<td>video</td>
<td>3</td>
</tr>
<tr>
<td>photo</td>
<td>1</td>
</tr>
<tr>
<td>news</td>
<td>1</td>
</tr>
<tr>
<td>tag</td>
<td>1</td>
</tr>
<tr>
<td>like</td>
<td>2</td>
</tr>
<tr>
<td>mms</td>
<td>2</td>
</tr>
</tbody>
</table>

From Table 1, let two categories, i.e., $C_1 = \text{Web services}$ and $C_2 = \text{Actions}$, from the concept of $\text{max\_support}$, we have the following maximal web services supported sets.

<table>
<thead>
<tr>
<th>Service</th>
<th>$\text{max_support}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>1</td>
</tr>
<tr>
<td>Gmail</td>
<td>4</td>
</tr>
<tr>
<td>Facebook, Gmail</td>
<td>3</td>
</tr>
<tr>
<td>Facebook, WhatsApp, Gmail</td>
<td>2</td>
</tr>
<tr>
<td>chat, msg</td>
<td>2</td>
</tr>
<tr>
<td>mail</td>
<td>2</td>
</tr>
<tr>
<td>video, photo</td>
<td>2</td>
</tr>
</tbody>
</table>

From the above-mentioned maximal web services supported set if we consider rules as normal association, then for the first rule we have support $= 2$ with confidence $= 22\%$ and $\text{max\_factor} = 1.65$. For the second rule, we have support $= 2$ with confidence $= 40\%$ and $\text{max\_factor} = 1.6$. Similarly for the last rule, we have support $= 2$ with confidence $= 100\%$ and $\text{max\_factor} = 1$. 


Framework for finding maximal association rules in mobile web service

\{Gmail\} \rightarrow \{video, photo\} with sup\textsuperscript{max} = 2 \text{ and } conf\textsuperscript{max} = 2 / 4 = 50%,

\{Facebook, Gmail\} \rightarrow \{mms, like\} with sup\textsuperscript{max} = 2 \text{ and } conf\textsuperscript{max} = 2 / 3 = 66%,

\{Facebook, WhatsApp, Gmail\} \rightarrow \{chat, msg\} with sup\textsuperscript{max} = 2 \text{ and } conf\textsuperscript{max} = 2 / 2 = 100%.

3 Soft set and Boolean value information system

In this section, \(U\) refers to a universe, \(E\) is the set of parameters and \(P(U)\) is the power set of \(U\).

3.1 Information and Boolean value system

An information system is the combination of 4 tuples \(Z = (U, A, V, f)\), where \(U = \{u_1, u_2, u_3, \ldots, u_{|U|}\}\) is a non-empty finite set of objects. A non-empty finite set of attributes is represented as \(A = \{a_1, a_2, a_3, \ldots, a_{|A|}\}\). \(V = \bigcup_{a \in A} V_a\), \(V_a\) is the domain/column value set of attribute \(a\), and \(f: U \times A \rightarrow V\) is an information function such that \(f(u, a) \in V_a\) for every \((u, a) \in U \times A\). An information system is also known as a knowledge representation system of an attribute-valued system and it can be represented in terms of information tables. Table 2 represents attribute-valued system as an information system.

| \(U\) | \(a_1\) | \(a_2\) | \(a_3\) | \(\ldots\) | \(a_{|A|}\) |
|------|-------|-------|-------|-------|-------|
| \(u_1\) | \(f(u_1, a_1)\) | \(f(u_1, a_2)\) | \(f(u_1, a_3)\) | \(\ldots\) | \(f(u_1, a_{|A|})\) |
| \(u_2\) | \(f(u_2, a_1)\) | \(f(u_2, a_2)\) | \(f(u_2, a_3)\) | \(\ldots\) | \(f(u_2, a_{|A|})\) |
| \(u_3\) | \(f(u_3, a_1)\) | \(f(u_3, a_2)\) | \(f(u_3, a_3)\) | \(\ldots\) | \(f(u_3, a_{|A|})\) |
| \(\vdots\) | \(\vdots\) | \(\vdots\) | \(\vdots\) | \(\vdots\) | \(\vdots\) |
| \(u_{|U|}\) | \(f(u_{|U|}, a_1)\) | \(f(u_{|U|}, a_2)\) | \(f(u_{|U|}, a_3)\) | \(\ldots\) | \(f(u_{|U|}, a_{|A|})\) |

In this table, \(f(u, a): U \times A \rightarrow V\) denotes a tuple. Tuple\(_i\) = \((f(u_i, a_1), f(u_i, a_2), \ldots, f(u_i, a_{|A|}))\), where \(i = 1, 2, 3 \ldots |U|\).

It is not necessary that tuple is associated with unique entity. It is possible that two distinct entities could have the same representation of tuple, i.e., duplicate tuples may be in the information table and it is not permissible in a relational DBMS. The information system concept is generalised from relational DBMS (Maji et al., 2003). Information system can be represented in Boolean value information system, \(Z = (U, A, V, f)\), if \(V_a = \{0, 1\}\) for every \(a \in A\).

3.2 Concepts of soft set

Definition 1: A pair \((F, E)\) is called a soft set over \(U\), where \(F\) is a mapping given by \(F: E \rightarrow P(U)\). In other words, a soft set over \(U\) is a parameterised family of subsets of
the universe $U$. For $e \in E$, $F(e)$ may be considered as a set of e-approximate element of the soft set $(F, E)$. Clearly, a soft is not a crisp set (Herawan and Deris, 2011).

**Example 3:** Let us consider a soft set $(F, E)$, which describes the ‘attractiveness of mobiles’ that Mr. John (say) is going to buy. Suppose that there are six mobiles in the universe $U$ under concern

$$U = \{m_1, m_2, m_3, m_4, m_5, m_6\},$$

and

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

where $e_1$ stands for the parameter ‘expensive’, $e_2$ stands for the parameter ‘slim’, $e_3$ stands for the parameter ‘beautiful’, $e_4$ stands for the parameter ‘smartphone’ and $e_5$ stands for the parameter ‘cheap’. The soft set $(F, E)$ is a parameterised family $\{f(e_i), i = 1, 2, 3, \ldots, n\}$ of subsets of set $U$ and gives us a collection of approximate description of the mobile. Consider the mapping function $F : E \rightarrow P(U)$ given by ‘mobile (.)’ where (.) is to be filled by one of the parameters $e \in E$. Suppose that

$$F(e_1) = \{m_2, m_4\},$$

$$F(e_2) = \{m_1, m_3\},$$

$$F(e_3) = \{m_1, m_4, m_5\},$$

$$F(e_4) = \{m_1, m_3, m_5\},$$

$$F(e_5) = \{m_1\}.$$ 

Therefore, $F(e_2)$ means ‘mobile (slim)’, whose functional value in the set $\{m_1, m_3\}$. The approximation collection of soft set $(F, E)$ is defined as

$$(F, E) = \begin{cases} 
\text{expensive mobile} = \{m_2, m_4\} \\
\text{slim mobile} = \{m_1, m_3\} \\
\text{beautiful mobile} = \{m_1, m_4, m_5\} \\
\text{smartphone mobile} = \{m_1, m_3, m_5\} \\
\text{cheap mobile} = \{m_1\} 
\end{cases}.$$ 

Here, each approximation has two parts, a predicate $p$ and an approximate value set $v$.

For the approximation ‘smartphone mobile = $\{m_1, m_3, m_5\}$’, we have the predicate name as smartphone mobile and the approximate value set as $\{m_1, m_3, m_5\}$. Thus, a soft set $(F, E)$ can be represented as the group of approximation

$$(F, E) = \{p_1 = v_1, p_2 = v_2, p_3 = v_3, p_4 = v_4, \ldots, p_n = v_n\}.$$
Framework for finding maximal association rules in mobile web service

**Proposition 1:** If \((E, F)\) is a soft set over universe set \(U\), then \((E, F)\) can be represented in Boolean value system as

\[ Z = (U, A, V_{(0,1)}, f) \]

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

\[ F = \{ f_1, f_2, f_3, \ldots, f_n \} \]

where

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

\[ f_2 : U \rightarrow V_2 \quad \text{and} \quad f_2(x) = \begin{cases} 1, & x \in F(e_2), \\ 0, & x \notin F(e_2), \end{cases} \]

\[ f_3 : U \rightarrow V_3 \quad \text{and} \quad f_3(x) = \begin{cases} 1, & x \in F(e_3), \\ 0, & x \notin F(e_3), \end{cases} \]

\[ \vdots \]

\[ f_n : U \rightarrow V_n \quad \text{and} \quad f_n(x) = \begin{cases} 1, & x \in F(e_n), \\ 0, & x \notin F(e_n). \end{cases} \]

So that if \(A = E\), \(V = \bigcup_{e_i \in E} V_{e_i}\), where \(V_{e_i} = \{0,1\}\), then soft set \((F, E)\) can be thought as a Boolean value information system \(Z = (U, A, V_{(0,1)}, f)\) (Herawan and Deris, 2011).

It is also possible that a Boolean value in formation system can be represented as a soft set. There are lots of researches going on the soft set. By a soft set, we can easily classify objects into two classes, i.e., 0 (No) or 1 (yes). It is very simple to make one to one association between a Boolean value system and soft set. It is represented in Table 3.

<table>
<thead>
<tr>
<th>(U)</th>
<th>Expensive ((e_1))</th>
<th>Slim ((e_2))</th>
<th>Beautiful ((e_3))</th>
<th>Smartphone ((e_4))</th>
<th>Cheap ((e_5))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(m_1)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(m_2)</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(m_3)</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(m_4)</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(m_5)</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(m_6)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The complete framework for finding maximal association rule using a soft set is presented in Figure 2. This figure shows that mobile user accessed various web services using the internet, and then the sequences of accessed web services are stored in databases. The user may access these services by different techniques like WiFi, landline internet or mobile data, etc. Now, soft set strategy can be applied to the stored web services sequences to find maximal rules using the Boolean-valued system.
4 Conversion of web service sequences into a soft set

In this section, we discuss about the conversion of web service sequences into soft set. This conversion is performed using a Boolean value information system. For example, we have $W = \{ws_1, ws_2, ws_3, \ldots, ws_n\}$ be a set of web services and $D = \{s_1, s_2, s_3, \ldots, s_{|U|}\}$ is a sequence of service accessed. For a Boolean value information system, we have $Z = (U, A, V_{\{0,1\}}, f)$. For every $a \in A$ and $u \in U$, we proposed a mapping $f: U \times A \rightarrow \{0,1\}$, such that $f(u, a) = 1$ if $a$ is present in $s$, otherwise $f(u, a) = 0$. This representation is referred to as a conversion of web services sequences into the Boolean value information system. If every $a \in A$ and $u \in U$, $|f(u, a)| = 1$, then Boolean value information system $Z = (U, A, V_{\{0,1\}}, f)$ is said to be deterministic. According to soft set theory, Boolean value representation is an efficient data structure for data storage. Therefore, web service sequences are easily transferred into bitmap table or Boolean value information system. The first column of web services sequence is the sequence number and remaining columns are the web service sequences. This structure is competent because the column address of a Boolean value attribute does not change. In this transformation, searching of any accessed service is fast, by accessing column only. If searched service have 1 value in column means it is present otherwise not. By Proposition 1, it is clear that every Boolean value information system can be
represented as a soft set. The following are the conversion of web services sequences into the Boolean value information system.

\[ \text{ws}_1 \to a_1, \]
\[ \text{ws}_2 \to a_2, \]
\[ \text{ws}_{|S|} \to a_{|S|} \]

or

\[ W = \{w_1, w_2, w_3, \ldots, w_{|A|}\} \to A = \{a_1, a_2, a_3, \ldots, a_{|A|}\} \]

and

\[ s_1 \to u_1, \]
\[ s_2 \to u_2, \]
\[ s_{|U|} \to u_{|U|} \]

or

\[ D = \{s_1, s_2, s_3, \ldots, s_{|U|}\} \to U = \{u_1, u_2, u_3, \ldots, u_{|U|}\} \]

5 Association rule using soft set

In this section, we have used a soft set for finding association rule and maximal association rule in web service accessed sequences. For this purpose, we need to convert web service sequences into soft set. This conversion is shown here

\[ W = \{w_1, w_2, w_3, \ldots, w_{|A|}\} \to A = \{a_1, a_2, a_3, \ldots, a_{|A|}\} \to E = \{e_1, e_2, e_3, \ldots, e_{|E|}\} \]

\[ D = \{s_1, s_2, s_3, \ldots, s_{|U|}\} \to U = \{u_1, u_2, u_3, \ldots, u_{|U|}\} \]

and

\[ F_j : U \to V_j \text{ and } f_j(x) = \begin{cases} 1, & x \in F(e_j) \\ 0, & x \notin F(e_j) \end{cases}. \]

Thus

\[ D = \{s_1, s_2, s_3, \ldots, s_{|U|}\} \to (F, E). \]

5.1 Association rule mining using soft set

In this proposed method, we use the concept of coexistence of parameters for generating association rules (Bi et al., 2003). Here, \((F, E)\) represent to the soft set over the universe \(U\) by the Boolean value information system. The following are the soft set representation of Table 1 web services sequences.
Definition 2 (Coexist set): Let \((F, E)\) be a soft set over \(U\) and \(u \in U\). A service coexist set in a web service sequence \(u\) can be defined as \(\text{coex}(u) = \{e \in E : f(u, e) = 1\}\). We have the following coexist of each web service:

\[
\begin{align*}
\text{coex}(u_1) &= \{\text{Facebook, WhatsApp, Gmail, chat, msg}\}, \\
\text{coex}(u_2) &= \{\text{Facebook, WhatsApp, Gmail, chat, msg}\}, \\
\text{coex}(u_3) &= \{\text{Gmail, mail}\}, \\
\text{coex}(u_4) &= \{\text{Gmail, video, photo}\}, \\
\text{coex}(u_5) &= \{\text{Gmail, video, photo}\}, \\
\text{coex}(u_6) &= \{\text{Facebook, news, tag}\}, \\
\text{coex}(u_7) &= \{\text{Facebook, Gmail, like, mms}\}, \\
\text{coex}(u_8) &= \{\text{Facebook, Gmail, like, mms}\}, \\
\text{coex}(u_9) &= \{\text{Facebook, Gmail, mail}\}. 
\end{align*}
\]

Definition 3: Suppose \((F, E)\) be a soft set over the universe \(U\) and \(X \subseteq E\). A set of parameters \(X\) is said to be supported by a sequence \(u \in U\) if \(X \subseteq \text{coex}(u)\) (Herawan and Deris, 2011).

Definition 4: Suppose \((F, E)\) be a soft set over the universe \(U\) and \(X \subseteq E\). The support of a set of attributes \(X\), denoted by \(\text{support}(X)\), is defined by the number of sequences, \(U\) supporting \(X\), i.e., \(\text{support}(X) = |\{u : X \subseteq \text{coex}(u)\}|\), where \(|X|\) is the cardinality of \(X\). According to Definitions 3 and 4, the supported set obtained in Section 5.1 can be represented as:

\[
\begin{align*}
\text{support}\{\text{Facebook}\} &= |\{u_1, u_2, u_3, u_4, u_5, u_6, u_9, u_{10}\}| = 6, \\
\text{support}\{\text{Gmail}\} &= |\{u_1, u_2, u_3, u_4, u_5, u_6, u_9, u_{10}\}| = 9, \\
\text{support}\{\text{WhatsApp}\} &= |\{u_1, u_2\}| = 2, \\
\text{support}\{\text{Facebook, Gmail}\} &= |\{u_1, u_2, u_9, u_{10}\}| = 5, \\
\text{support}\{\text{Facebook, WhatsApp}\} &= |\{u_1, u_2\}| = 2, \\
\text{support}\{\text{WhatsApp, Gmail}\} &= |\{u_1, u_2\}| = 2, \\
\end{align*}
\]
Framework for finding maximal association rules in mobile web service

support \{\text{Facebook, WhatsApp, Gmail}\} = \left\lceil \left\lceil u_1, u_2 \right\rceil \right\rceil = 2,

support \{\text{chat}\} = \left\lceil \left\lceil u_1, u_2 \right\rceil \right\rceil = 2,  \text{support} \{\text{msg}\} = \left\lceil \left\lceil u_1, u_2 \right\rceil \right\rceil = 2,

support \{\text{mail}\} = \left\lceil \left\lceil u_1, u_6, u_{10} \right\rceil \right\rceil = 3,  \text{support} \{\text{video}\} = \left\lceil \left\lceil u_1, u_5 \right\rceil \right\rceil = 2,

support \{\text{photo}\} = \left\lceil \left\lceil u_1, u_5, u_6 \right\rceil \right\rceil = 3,  \text{support} \{\text{news}\} = \left\lceil \left\lceil u_5 \right\rceil \right\rceil = 1,

support \{\text{tag}\} = \left\lceil \left\lceil u_5 \right\rceil \right\rceil = 1,  \text{support} \{\text{like}\} = \left\lceil \left\lceil u_3, u_5 \right\rceil \right\rceil = 2,

support \{\text{mms}\} = \left\lceil \left\lceil u_3, u_5 \right\rceil \right\rceil = 2,  \text{support} \{\text{chat, msg}\} = \left\lceil \left\lceil u_1, u_5 \right\rceil \right\rceil = 2,

support \{\text{video, photo}\} = \left\lceil \left\lceil u_1, u_5 \right\rceil \right\rceil = 2,  \text{support} \{\text{mail, photo}\} = \left\lceil \left\lceil u_5 \right\rceil \right\rceil = 1,

support \{\text{news, tag}\} = \left\lceil \left\lceil u_5 \right\rceil \right\rceil = 1,  \text{support} \{\text{like, mms}\} = \left\lceil \left\lceil u_3, u_5 \right\rceil \right\rceil = 2.

Definition 5: Suppose \((F, E)\) be a soft set over the universe \(U\) and \(X, Y \subseteq E\), where \(X \cap Y = \emptyset\). Here, \(X \rightarrow Y\) form inference to the association rules between \(X\) and \(Y\). The web services set \(X\) and \(Y\) is called antecedent and consequent, respectively.

Definition 6 (Support): Suppose \((F, E)\) be a soft set over the universe \(U\) and \(X, Y \subseteq E\), where \(X \cap Y = \emptyset\). The support of an association rule \(X \rightarrow Y\), denoted by \(\text{support}(X \Rightarrow Y)\), is defined by

\[
\text{support}(X \Rightarrow Y) = \text{support}(X \cup Y) = \left\lceil \left\lceil u : X \cup Y \subseteq \text{coex}(u) \right\rceil \right\rceil
\]

Definition 7 (Confidence): Suppose \((F, E)\) be a soft set over the universe \(U\) and \(X, Y \subseteq E\), where \(X \cap Y = \emptyset\). Here, \(\text{confidence}(X \Rightarrow Y)\) represent the confidence of association rule \((X \Rightarrow Y)\). This confidence can be defined by

\[
\text{confidence}(X \Rightarrow Y) = \frac{\text{support}(X \cup Y)}{\text{support}(X)} = \frac{\left\lceil \left\lceil u : X \cup Y \subseteq \text{coex}(u) \right\rceil \right\rceil}{\left\lceil \left\lceil u : X \subseteq \text{coex}(u) \right\rceil \right\rceil}
\]

5.2 Soft-set-based categorisation and Taxonomy

Suppose \((X, E)\) be a soft set over the universe \(U\). The taxonomy \(T_d\) of \(E\) is partition of \(E\) into disjoint subsets like \(T_d = \{E_1, E_2, E_3, \ldots, E_n\}\). Here, each member of the \(T_d\) is called a category. For a web service \(w\), we denote \(T_d(w)\) the category that contains \(w\). If \(Z\) is a web services set and indicate to a single category, then it can be denoted as a \(T_d(Z)\).

5.3 Soft-set-based maximal association rules

We have defined the concept of association rules between two sets of parameters in Section 5.1. We have also defined support and confidence of web services in a sequence using soft sets. Here, we have found maximal association rules using soft set. Soft-set-based maximal associated rules are identical and faster when compared with (Amir et al., 2005; Feldman et al., 1997) and rough-set-based rules (Bi et al., 2003; Guan et al., 2003, 2005).
**Definition 8:** Suppose \((F, E)\) be a soft set over the universe \(U\) and \(X, Y \subseteq E\), The maximal support of a set of \(X\) attribute denoted by \(\text{support}(X)\) and defined by the number of sequences \(U\) maximal supporting \(\max \text{supp}(X) = \left\{ u : X = \text{coex}(u) \cap E \right\}\) where \(|X|\) is the cardinality of \(X\). Definitely, \(\max \text{supp}(X) = |e : e \in X \cap F(e) = 1|\).

**Example 4:** Suppose \(C_1 = \text{Web services and } C_2 = \text{Action}\), from the concept \(\max \text{supp}\), we have the following maximal supported sets.

\[
\begin{align*}
\text{max supp}\{\text{Facebook}\} &= \{u_1\} = 1, \\
\text{max supp}\{\text{Gmail}\} &= \{u_3, u_4, u_5, u_6\} = 4, \\
\text{max supp}\{\text{Facebook, Gmail}\} &= \{u_3, u_4, u_6\} = 3, \\
\text{max supp}\{\text{Facebook, WhatsApp, Gmail}\} &= \{u_1, u_2\} = 2, \\
\text{max supp}\{\text{chat, msg}\} &= \{u_1, u_2\} = 2, \\
\text{max supp}\{\text{mail}\} &= \{u_3, u_6\} = 2, \\
\text{max supp}\{\text{video, photo}\} &= \{u_3, u_5\} = 2, \\
\text{max supp}\{\text{mail, photo}\} &= \{u_6\} = 1, \\
\text{max supp}\{\text{news, tag}\} &= \{u_2\} = 1, \\
\text{max supp}\{\text{like, mms}\} &= \{u_3, u_4\} = 2.
\end{align*}
\]

**Definition 9:** Suppose \((F, E)\) be a soft set over the universe \(U\) and two maximal web service sequences \(X, Y \subseteq E\), where \(X \cap Y = \emptyset\). A maximal association rule between \(X\) and \(Y\) is represented as \(max\). The web services sequence \(X\) and \(Y\) are called maximal antecedent and maximal consequent, respectively.

**Definition 10:** Suppose \((F, E)\) be a soft set over the universe \(U\) and two maximal web service sequences \(X, Y \subseteq E\), where \(X \cap Y = \emptyset\). The maximal support of the rule \((X \Rightarrow Y)\) is denoted as \(\text{max supp}(X \Rightarrow Y)\) and is defined by

\[
\max \text{supp}(X \Rightarrow Y) = \max \text{supp}(X \cup Y) = \left\{ u : X \cup Y = \text{coex}(u) \cap E \right\}
\]

**Definition 11:** Suppose \((F, E)\) be a soft set over the universe \(U\) and two maximal web service sequences \(X, Y \subseteq E\), where \(X \cap Y = \emptyset\). The maximal confidence of the association rule \(X \Rightarrow Y\), denoted by \(\text{max confidence}(X \Rightarrow Y)\) and defined as

\[
\max \text{confidence}(X \Rightarrow Y) = \frac{\text{support}(X \cup Y)}{\text{support}(X)} = \frac{\left\{ u : X \cup Y = \text{coex}(u) \cap E \right\}}{\left\{ u : X = \text{coex}(u) \cap E \right\}}
\]

**Example 5:** Suppose we have minimum \(\max \text{supp} = 2\) and minimum \(\max \text{confidence} = 0.5\), then we have the following maximal association rules
Framework for finding maximal association rules in mobile web service

\{\text{Gmail}\} \Rightarrow \{\text{video, photo}\} \text{ with max\_support = 2 and max\_confidence = 50\%},
\{\text{Facebook, Gmail}\} \Rightarrow \{\text{mms, like}\} \text{ with max\_support = 2 and max\_confidence = 66\%},
\{\text{Facebook, WhatsApp, Gmail}\} \Rightarrow \{\text{chat, msg}\} \text{ with max\_support = 2 and max\_confidence = 100\%}

6 Experimental analysis

We evaluate the performance of a proposed soft-set-based maximal association rule generation approach in this section. The experiments were performed on a 3.0 GHz processor with 2GB of memory and the operating system Microsoft Windows 7. The proposed approach is implemented in MATLAB 8.3.0.532 (R2014a). The comparison is performed with different maximal association generation and rough set approaches (Bi et al., 2003; Feldman et al., 1997; Guan et al., 2005). The next subsections provide the details about dataset used, experimental results in terms of memory consumption and execution time efficiency.

6.1 Dataset description

We evaluate the performance of a proposed soft-set-based maximal association rule using WS-DREAM web service QoS dataset (Zheng and Lyu, 2010). The WS-DREAM dataset includes QoS performance of about 1.5 million real-world web service invocations of 100 publicly available web services observed by 150 distributed users. From the QoS values of the 100 web services observed by the 150 service users, we have generated 1k, 5k, 10k and 20k web services sequences, respectively. We compared execution time and memory utilisation of a proposed soft-set-based maximal association rule generation approach with algorithms of Bi et al. (2003), Feldman et al. (1997) and Guan et al. (2005). The comparison results are shown in Figures 3 and 4, respectively. The proposed framework is compared with other approaches in terms of execution time and memory consumptions.

Figure 3 shows the response time to generate maximal association rules. It indicated that when the web service sequence increases, computational time also increased. The proposed framework takes less time to generate maximal association rules when compared with rough set approach and other maximal association rule generation approaches.

Figure 4 represents the response time to generate maximal association with respect to different minimum support values. When the minimum support value increases like from 0.5\% to 2\%, the response time takes less time.

The memory consumption of the proposed approach is compared with the different association rule mining approaches. This comparison is shown in Figure 5. It shows that while the data size is increased, memory comparison is also gradually increased. The proposed framework takes less memory to generate maximal association rules when compared with other maximal association rule generation approaches. On the basis of the above-mentioned experiment, it is clear that the soft-set-based association mining techniques improved the result in different scenario. We can generate various kinds of
association and maximal association relationships between different web services. The prediction has also become easier for web services. Web service companies can generate new kind of knowledge for their growth, because they know which web services are mostly accessed together. The proposed work is useful in different areas of e-commerce. With the help of this approach, web services maintenance has become more simple and easy.

**Figure 3** Time comparison with different sequences (see online version for colours)

![Graph showing time comparison with different sequences](image)

**Figure 4** Time comparison with different minimum support (see online version for colours)

![Graph showing time comparison with different minimum support](image)
7 Conclusion

In this paper, we have proposed a framework, to extract maximal association rules from regular association rules of web services accessed sequences. The framework uses soft set concepts for extracting maximal association rules. In this framework, we have converted web services sequences onto soft set via a Boolean value information system, and then used coexistence between attributes. The performance of the proposed framework is measured in terms of memory and execution time and compared with other association rule generation approaches. The proposed framework is faster than rough-set-based approach, and other maximal association rules based approaches. The proposed approach is useful for business companies for web services prediction for users and launching new web services. It is also useful in e-commerce for decision-making and business enhancing. It is used to extract associated web services from huge sequences. In the future, we will attempt to enhance the framework with the use of various constraints and preferences of web services. We will further develop the framework using fuzziness concept.

Acknowledgements

We thank the referees for their valuable suggestions and the Editor-in-Chief for the publication.
References


Framework for finding maximal association rules in mobile web service


