Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list. Note: if you opt to annotate the file with software other than Adobe Reader then please also highlight the appropriate place in the PDF file. To ensure fast publication of your paper please return your corrections within 48 hours.

For correction or revision of any artwork, please consult [http://www.elsevier.com/artworkinstructions](http://www.elsevier.com/artworkinstructions).

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Click on the ‘Q’ link to go to the location in the proof.

<table>
<thead>
<tr>
<th>Location in article</th>
<th>Query / Remark: click on the Q link to go</th>
<th>Please insert your reply or correction at the corresponding line in the proof</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Please confirm that given names and surnames have been identified correctly.</td>
<td></td>
</tr>
<tr>
<td>Q2</td>
<td>Please check whether the identification of corresponding author is okay as typeset.</td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td>As per journal requirements, a minimum of at least five keywords are required. Kindly provide the same.</td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td>Please check whether the insertion of ‘Ref. [2]’ text citation is okay as typeset.</td>
<td></td>
</tr>
<tr>
<td>Q5</td>
<td>Please note that the author name ‘Hongwu et al.’ does not match with Ref. [30]. Kindly check.</td>
<td></td>
</tr>
<tr>
<td>Q6</td>
<td>Please note that ‘Figs. 2–6, 7–10 and 11 and 12’ have identical captions. Kindly check.</td>
<td></td>
</tr>
</tbody>
</table>

Thank you for your assistance.
MAR: Maximum Attribute Relative of soft set for clustering attribute selection

Rabiei Mamat\textsuperscript{a,}\textsuperscript{*}, Tutut Herawan\textsuperscript{b}, Mustafa Mat Deris\textsuperscript{c}

\textsuperscript{a}Department of Computer Science, Universiti Malaysia Terengganu, 21300 Kuala Terengganu, Terengganu, Malaysia
\textsuperscript{b}Department of Information System, University of Malaya, 50603 Pantai Valley, Kuala Lumpur, Malaysia
\textsuperscript{c}Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Parit Raja, 86400 Batu Pahat, Johore, Malaysia

ABSTRACT

Clustering, which is a set of categorical data into a homogenous class, is a fundamental operation in data mining. One of the techniques of data clustering was performed by introducing a clustering attribute. A number of algorithms have been proposed to address the problem of clustering attribute selection. However, the performance of these algorithms is still an issue due to high computational complexity. This paper proposes a new algorithm called Maximum Attribute Relative (MAR) for clustering attribute selection. It is based on a soft set theory by introducing the concept of the attribute relative in information systems. Based on the experiment on fourteen UCI datasets and a supplier dataset, the proposed algorithm achieved a lower computational time than the three rough set-based algorithms, i.e., TR, MMR, and MDA up to 62\%, 64\%, and 40\% respectively and compared to a soft set-based algorithm, i.e., NSS up to 33\%. Furthermore, MAR has a good scalability, i.e., the execution time of the algorithm tends to increase linearly as the number of instances and attributes are increased respectively.

Keywords:
Soft set theory
Clustering attributes
Attribute relative
Complexity

1. Introduction

Data clustering refers to the method of creating a group of objects or cluster. Results of data clustering or so-called data clustering is a collection of data objects where each cluster contains data objects that are similar to each other and their different clusters are quite distinct [1]. Data clustering is also known as cluster analysis, segmentation analysis, taxonomy analysis, and unsupervised classification. Clustering has been used in many areas such as gene data processing [2], transactional data processing [3], decision support [4], and radar signals processing [5]. Recently, greater attentions have been given to categorical data clustering [6,7], where data objects are made up of non-numerical attributes.

The main difference between categorical data and numerical data is the multi-valued attribute that belongs to the categorical data. These properties lead to difficulties in the similarities and dissimilarity measurement in the clustering process, since the normal distance measures cannot be applied directly to the categorical data. Therefore, the best similarity measurement of the categorical data is done by defining the common object for the attribute as well as the common values of the attribute, and the association between the two [8]. A number of algorithms for clustering categorical data have been proposed including works by Ganti et al. [9], Gibson et al. [10], Guha et al. [11] and Zaki [12].

These methods make important contributions to the issue of clustering categorical data. However, these approaches have led to an increment in computational complexity and inefficient system due to the properties of multi-valued attribute in the categorical data, having so many structures and evolved into a very large data set. Another major issue is that they are not designed to handle uncertainties, while in the real world it is very important in many applications since there are no sharp boundaries between clusters.

To this, rough set theory [13–15] has been used by introducing a series of clustering attributes, in which, one of the attributes is selected and used to cluster the objects that are not clustered yet at each time until all objects are clustered. The approaches of finding a clustering attribute had successfully exploited the uncertainties in the multi-valued information system [26,27,29]. But, some scenarios under study have shown that those approaches have their own drawbacks, i.e. there exists some unexpected iteration that leads to an increment in the processing time. In addition, a clustering attribute also leads to the instability to the cluster number in the clustering process.

A new way to manage uncertain data is called the soft set theory, which is proposed by Molodtsov [16] in 1999. According to him, the main advantage of soft set theory is free from the inadequacy of the parameterization tools, unlike in the previous theories.
such as the theory of probability. The soft set theory uses parameterization sets as its main solutions for problem solving, which makes it very convenient and easy to apply in practice. As a result, great progresses of study in soft set theory have been made, including the works of [17–22]. As for standard soft set, it may be redefined as the classification of objects in two distinct classes (yes/1 and no/0), thus confirming that soft set can deal with a Boolean-valued information system.

For a multi-valued information system, the concepts of multi-soft sets have been proposed in [23]. Since a direct proof that every rough set is a soft set have been given in [24], Hongwu et al. [30] proposed a soft set model on the equivalent classes of an information system, which can be easily applied in obtaining approximate sets of rough sets. Furthermore, they use it to select a clustering attribute for categorical datasets and a heuristic algorithm, namely the NSS. Experimental results on fifteen UCI benchmark datasets showed that the proposed NSS algorithm provides a faster decision in selecting a clustering attribute as compared with Maximum Dependency Attributes (MDA) approach. However, since it transfers MDA to a soft set attribute dependency model, the time complexity of the algorithm is still an outstanding issue.

In this paper, by applying the concept of attribute relatives in soft set theory, an alternative technique to select a clustering attribute using soft set theory is proposed. In summary, the contributions of this work are summarized as follows:

a. MAR, an alternative technique to select a clustering attribute using soft set theory is suggested. It is based on a concept of attribute relative where the comparison of attributes is made by taking into account the relative of the attribute at the category level.

b. The proposed technique potentially discovers the attributes subsets with better coverage and achieves a lower computational time. From the experimentation on fifteen UCI benchmark data sets, the MAR achieves a lower computational time than the three rough set-based algorithms, i.e. TR, MMR, MDA up to 62%, 64%, and 40%, respectively and a soft set-based algorithm NSS up to 33%.

The rest of the paper is organized as follows. In Section 2, we explain some basic knowledge about this topic. Then, we give a literature survey and review on rough and soft set-based clustering attribute selection techniques in Sections 3 and 4. In Section 5, we present the MAR’s technique. Section 6 discusses the experiment and a series of evaluations on the result. Finally, we conclude this work in Section 7.

2. Essential rudiments

2.1. Decision system

An information system as defined in [15] is a 4-tuple (quadruple) $S = (U, A, V, f)$, where $U = \{u_0, u_1, \ldots, u_m\}$ is a non-empty finite set of objects, $A = \{a_1, a_2, \ldots, a_n\}$ is a non-empty finite set of attributes, $V$ is the domain (value set) of attributes. $f: U \times A \rightarrow V$ is an information function, such that, $f(x, a) = v$ for every $(x, a) \in U \times A$, is called information (knowledge) function. An information system can be intuitively expressed in terms of an information table such as in Table 1.

Table 1

<table>
<thead>
<tr>
<th>U/B</th>
<th>e_1</th>
<th>e_2</th>
<th>e_3</th>
<th>e_4</th>
<th>e_5</th>
<th>e_6</th>
<th>e_7</th>
</tr>
</thead>
<tbody>
<tr>
<td>c_1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>c_2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>c_3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>c_4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>c_5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>c_6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>c_7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>c_8</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>c_9</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

2.2. Rough set theory

Rough set theory is an approach to aid decision making in the presence of uncertainty [13–15]. Within this theory, a data-set can be formally described using an information system. The notion of information system provides a convenient tool for the representation of objects in terms of their attribute values. Based on the concept of an information system, some basic definitions in the rough set theory are given below.

Let $S = (U, A, V, f)$, an information system as defined in Section 2.1 and $U$ is the set of all objects.

Definition 1 (Indiscernibility Relation $IND(B)$). Given two objects $x_a$, $x_b \in U$, they are indiscernible by the set of attributes $B$ in $A$, if and only if $f(x_a, a) = f(x_b, a)$ for every $a \in B$. That is, $(x_a, x_b) \in IND(B)$, if and only if $\forall a \in B$ where $B \subseteq A$, $f(x_a, a) = f(x_b, a)$.

Obviously, $IND(B)$ is an equivalence relation. It is well-known that an equivalence relation induces a unique partition. The partition of $U$ induced by $IND(B)$ in $S = (U, A, V, f)$ denoted by $U/B$ and equivalence class in the partition $U/B$ containing $x \in U$, denoted by $[x]_{IND(B)}$.

Definition 2 (Equivalence class). Given $IND(B)$, the set of objects in $U$ having the same values for the set of attributes $B$ in $A$ is called an equivalence classes, $[x]_{IND(B)}$. It is also known as an elementary set with respect to $B$.

Definition 3 (Lower Approximation). Given a set of attributes $B$ in $A$ and set of objects $X$ in $U$. The lower approximation of $X$ is defined as the union of all elementary set which is contained in $X$. That is $\underline{X}_B = \{x \in U | [x]_{IND(B)} \subseteq X\}$

Definition 4 (Upper Approximation). Given a set of attributes $B$ in $A$ and set of objects $X$ in $U$. The upper approximation of $X$ is defined as the union of the elementary sets which has a non-empty intersection with $X$. That is $\overline{X}_B = \{x \in U | [x]_{IND(B)} \cap X \neq \emptyset\}$
Definition 5 (Roughness). The accuracy of approximation (roughness) of any subset \( X \subseteq U \) with respect to \( B \subseteq A \) denoted by \( \sigma_B(X) \) is measured by

\[
\sigma_B(X) = \frac{X_B}{X^B} \quad (3)
\]

2.3. Soft set theory

Definition 6 (See [16,25]). A pair \((F, A)\) is called a soft set over \( U \) where \( F \) is a mapping given by

\[
F : A \rightarrow P(U) \quad (4)
\]

In other words, a soft set \((F, A)\) over \( U \) is a parameterized family of subsets of \( U \). For \( \alpha \in A, F(\alpha) \) may be considered as the set of \( \alpha \)-elements of the soft set \( F(A) \) or the set \( \alpha \)-approximate elements of the soft set \( F(A) \). Clearly, a soft set is not a (crisp) set.

Example 1. Let a universe \( U = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}\} \) be a set of candidates and a set of parameters \( E = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7\} \) be a set of soft skills which stand for the parameters "communicative", "critical thinking", "team work", "information management", "entrepreneurship", "leadership" and "moral", respectively. Consider \( F \) be a mapping of \( E \) into the set of all subsets of the set \( U \) as \( F(e_1) = \{c_1, c_2, c_4, c_5\}, F(e_2) = \{c_6, c_9, c_{10}\}, F(e_3) = \{c_2, c_3, c_4, c_5, c_8, c_{10}\}, F(e_4) = \{c_2, c_3, c_4, c_5, c_7, c_9, c_{10}\}, F(e_5) = \{c_2, c_3, c_{10}\} \) and \( F(e_6) = \{c_6, c_9, c_{10}\} \). Now consider a soft set \((F, E)\), which describes the "capacities of the candidate for hire". According to the data collected, the soft set \((F, E)\) is given by

\[
(F, E) = \begin{cases} 
\text{communicative} = \{c_1, c_2, c_4, c_5\}, \\
\text{critical thinking} = \{c_3, c_9, c_{10}\}, \\
\text{team work} = \{c_6, c_9, c_{10}\}, \\
\text{information management} = \{c_2, c_3, c_4, c_5, c_8\}, \\
\text{entrepreneurship} = \{c_2, c_3, c_4, c_5, c_7, c_9, c_{10}\}, \\
\text{leadership} = \{c_6, c_9, c_{10}\}, \\
\text{moral} = \{c_6, c_9, c_{10}\} 
\end{cases}
\]

Definition 7. Let \( S = (U, A, V_{(0,1)}) \) be an information system. If \( V_{a} = [0,1] \) for every \( a \in A \), then \( S = (U, A, V_{(0,1)}) \) is called a Boolean-valued information system.

Proposition 1. Each soft set can be considered as a Boolean-valued information system.

Proof. Let \((F, E)\) be a soft set over the universe \( U \), \( S = (U, A, V, f) \) be an information system. Obviously, the universe \( U \) in \((F, E)\) can be considered as the universe \( U \) in \((F, E)\), the parameter set \( E \) may be considered as the attributes \( A \). Then, the information function \( f \) is defined by

\[
f = \begin{cases} 
1, & h \in F(e) \\
0, & h \not\in F(e) 
\end{cases}
\]

That is, when \( h \in f(e) \), where \( h \in U \) and \( e \in E \), then \( f(h, e) = 1 \), otherwise \( f(h, e) = 0 \). To this, we have \( V(h, e) = [0,1] \).

3. Rough set-based clustering attribute selection techniques

Currently, there are two measurement approaches based on the theory of rough set in clustering attribute selection. The first approach is based on the roughness of the attribute, and the second one is based on the attribute dependencies. In this section, two techniques of first approach, i.e. Total Roughness (TR) and Min–Min Roughness (MMR), and a technique of second approach called Maximum Dependency of Attribute (MDA) will be reviewed.

We define \( F(E) = \{f(a_1), (F, a_2), \ldots, (F, a_n)\} \) as a multi-soft sets over universe \( U \) representing a multi-valued information system \( S = (U, A, V, f) \). As an example, by using an information system in Table 2 which describes the student’s enrollment qualification, the concept of multi-soft set is illustrated in Fig. 1.

Fig. 1. Multi-soft set composition from dataset in Table 2.
3.1. Total Roughness (TR)

Total roughness technique, which is proposed by Mazlack et al. [26], has three main steps. The first one is to determine the roughness of each attribute by exploiting the relationship between lower and upper approximation which is defined in Eqs. (1) and (2), respectively. Let \( S = (U, A, V, f) \), an information system as described in Section 2.1, that \( a_i \in A \) has \( k \)-different values, say \( \beta_k \), \( k = 1, 2, \ldots, n \). Let \( X(a_i = \beta_k) \), \( k = 1, 2, \ldots, n \) be a subset of the objects having \( k \)-different values of attribute \( a_i \). The roughness of the set \( X(a_i = \beta_k) \), \( k = 1, 2, \ldots, n \) with respect to \( a_i \), where \( i \neq j \), denoted by \( \text{TR}_{Rough\_a_i}(X(a_i = \beta_k)) \), is defined by

\[
\text{TR}_{Rough\_a_i}(X(a_i = \beta_k)) = \frac{|X(a_i = \beta_k) \setminus X(a_i = \beta_j)|}{|X(a_i = \beta_k)|}, \quad k = 1, 2, \ldots, n
\]

where \( X(a_i = \beta_k) \) and \( X(a_i = \beta_j) \) are, respectively, the cardinality of lower and upper approximations of a subset of the objects having \( k \)-different values of attribute \( a_i \) with respect to attribute \( a_j \), where \( i \neq j \).

The next step, the mean roughness of attribute \( a_i \in A \) with respect to attribute \( a_j \in A \), where \( i \neq j \), denoted by \( \text{TR\_MeanRough\_a_i}(a_j) \), is evaluated as follows:

\[
\text{TR\_MeanRough\_a_i}(a_j) = \frac{\sum_{k=1}^{n} |X(a_i = \beta_k) \setminus X(a_i = \beta_j)|}{|V(a_i)|}
\]

where \( V(a_i) \) is the set of values of attribute \( a_i \in A \). Finally, the total roughness of attribute \( a_i \in A \) with respect to attribute \( a_j \in A \), where \( i \neq j \) denoted as \( \text{TR}(a_i) \), is obtained by the following formula:

\[
\text{TR}(a_i) = \sum_{a_j \in A} \text{TR\_MeanRough\_a_i}(a_j)
\]

Mazlack et al. suggested that the highest of TR is the best selection for a clustering attribute.

3.2. Min–Min Roughness (MMR)

MMR, proposed by Parmar et al. [27], is another rough set-based technique. The value of roughness of MMR technique is given below:

\[
\text{MMR\_Rough\_a_i}(X(a_i = \beta_k)) = 1 - \text{TR\_Rough\_a_i}(X(a_i = \beta_k)) = 1 - \frac{|X(a_i = \beta_k) \setminus X(a_i = \beta_j)|}{|X(a_i = \beta_k)|}, \quad k = 1, 2, \ldots, n
\]

It is clear that MMR technique uses Marczewski–Steinhaus metric [28] to measure the roughness of the set \( X(a_i = \beta_k) \), \( k = 1, 2, \ldots, n \), with respect to \( a_i \), where \( i \neq j \).

The value of the mean roughness of MMR technique is given as follows:

\[
\text{MMR\_MeanRough\_a_i}(a_j) = \frac{\sum_{k=1}^{n} |X(a_i = \beta_k) \setminus X(a_i = \beta_j)|}{|V(a_i)|}
\]

From Eqs. (9) and (10), it is easily seen that the value of the mean roughness of MMR technique is an opposite of the value of the mean roughness of TR.

\[
\text{MMR\_MeanRough\_a_i}(a_j) = \frac{\sum_{k=1}^{n} (1 - \text{TR\_Rough\_a_i}(X(a_i = \beta_k)))}{|V(a_i)|} = \frac{|V(a_i)| - \sum_{k=1}^{n} \text{TR\_Rough\_a_i}(X(a_i = \beta_k))}{|V(a_i)|} = 1 - \text{TR\_MeanRough\_a_i}(a_j)
\]

for \( i \neq j \).

Let \( m \) be the attributes, Min-Roughness (MR) of attribute \( a_i \) with respect to all \( a_j \), where \( i \neq j \) is defined as

\[
\text{MinR}(a_i) = \min \{\text{MMR\_MeanRough\_a_i}(a_j), \ldots, \text{MMR\_MeanRough\_a_i}(a_j)\}
\]

where \( 1 \leq i, j \leq m \).

The minimum of the Min-Roughness or known as Min–Min Roughness, which is denoted by MMR, is formulated as follows:

\[
\text{MMR} = \min(\text{MinR}(a_1), \ldots, \text{MinR}(a_j), \ldots, \text{MinR}(a_m))
\]

where \( A \) is the number of attribute in the information system, \( a_i \in A \). Parmar et al. suggested that the least mean roughness is the best selection of a clustering attribute.

3.3. Maximum Attribute Dependency (MDA)

Maximum Dependency of Attributes, which is proposed by Herawan et al. [29], disseminate the attributes of information system by using the attribute dependencies. MDA has a 4-step algorithm which starts with a search for indiscernibility relation. Subsequently, attribute dependency for each attribute in the information system is calculated. If given two attributes, \( a_i, a_j \in A \), the dependency degree in \( k \) of attribute \( a_i \) on attribute \( a_j \) denoted by \( d_{k\_a_i \_a_j} \) is calculated using the following equation:

\[
d_{k\_a_i \_a_j} = \frac{\sum_{i \neq j} |X(a_i = a_j)|}{|U|}
\]

The highest dependency degree (Max-Dependency), which is denoted as MD for each \( a_i \in A \), \( 1 \leq i \leq |A| \), is selected in the following step. If given \( m \) attributes, then, MD of attribute \( a_1, a_2 \in A \) is defined as

\[
\text{MD}(a_i) = \max(k_{a_i \_a_j}(a_i), \ldots, k_{a_i \_a_j}(a_i))
\]

where \( a_i \neq a_j \) and \( 1 \leq i, j \leq m \). Lastly, MDA technique suggested that the highest MD is selected as a clustering attribute using Eq. (14). If there are more than one attribute shared the highest \( k \) value, then, the next highest \( k \) in each attribute will be used and compared until the tie is broken.

\[
\text{MDA} = \max(\text{MD}(a_1), \ldots, \text{MD}(a_j), \ldots, \text{MD}(a_m))
\]

In reviewing the above three techniques to handle large data-sets, even with increasing computing capabilities, the complexity of computation is still an outstanding issue. From the analysis, we have found that two techniques, i.e, MMR and MDA, are suffered with ‘unpredicted iteration’. If there are \( n \) objects in the data-set, by taking the worst case scenarios, this situation can trigger \( n^2 \) of processing time which can increase the computational complexity when involving large data-sets. Although TR does not face the same problems as discussed above, TR initially has insufficient computational time as discussed in [29]. Therefore, there is a need for a technique in clustering attribute selection to further improve the computational complexity.

4. Soft set-based clustering attribute selection techniques

A novel soft-set (NSS) approach has been proposed by Hongwu et al. in [30]. In this approach, the re-definition of attribute dependency in the rough set theory is given in the context of soft set theory. In NSS, the degree of dependency is defined as:

\[
k_{a_i}(a_j) = \frac{|a_i(a_j)|}{|U|}
\]

The cardinality of \( a_i(a_j) \) can be calculated as

\[
|a_i(a_j)| = \sum_{k=1}^{D-1} |a_i| |a_j|
\]
where lower approximation of attribute \( a_i \) with respect to attribute \( a_j \) is defined as

\[
\tilde{a}_i(a_j) = \{x \in a_i(a_j), \ k = D + 1, \ldots, D + |U/a_j|\}
\]

(17)

where \( D = \sum_{i=1}^{n} |U/a_i| \).

As another alternative approach of the rough set theory, the NSS

Definition 8. Let \((F, A)\) be a multi soft-sets over the universe \( U \), where \((F, a_i), \ldots, (F, a_n) \subseteq (F, A)\) and \((F, a_i), \ldots, (F, a_m) \subseteq (F, A)\).

Support of \((F, a_i)\) by \((F, a_j)\) denoted \( \sup_{F,a_j}(F, a_i) \) is defined as

\[
\sup_{F,a_j}(F, a_i) = \frac{|(F, a_i) \cap (F, a_j)|}{|F, a_j|}
\]

(18)

Definition 9. Maximum support is a summation of all support with value equals to 1. For each soft set \((F, a_i)\), the maximum support \( \max \sup_{F,a_j}(F, a_i) \) is defined as

\[
\max \sup_{F,a_j}(F, a_i) = \sum (\sup_{F,a_j}(F, a_i) = 1)
\]

(19)

Definition 10. Minimum support is a summation of all support with value less than 1. For each soft set \((F, a_i)\), the minimum support is denoted by \( \min \sup_{F,a_j}(F, a_i) \) is defined as

\[
\min \sup_{F,a_j}(F, a_i) = \sum (\sup_{F,a_j}(F, a_i) \neq 1)
\]

(20)

Definition 11. Mode refers to the value that is most frequently occurred in the probability distribution.

Definition 12. Max refers to the value that is the highest in the probability distribution.

Proposition 2. If

\[
\text{Mode}(\max \sup_{F,a_j}(F, a_i)), \ldots, \max \sup_{F,a_j}(F, a_m) = 1 \]

then \((F, a_i)\) is a clustering attribute.

Proof. Let \((F, a_i)\) and \((F, a_j)\) be two soft sets over the universe \( U \), if \( a_i = a_j \) then the support value of \((F, a_i)\) is equal to 1, therefore it is said that \( a_i \) is relative to \( a_j \). Therefore, \( a_i \) can be used to describe \( a_j \) and vice versa. If \( a_i = a_j \) then there exists \((F, a_k)\) where \( a_k = a_k \) and \( a_i \neq a_j \) then, the support value of \((F, a_k)\) is greater than 1. Based on Definition 5, it is clear that \((F, a_i)\) is selected as a clustering attribute.

Corollary 1. If

\[
\text{Mode}(\max \sup_{F,a_j}(F, a_i)), \ldots, \max \sup_{F,a_j}(F, a_m) > 1
\]

then \( \min \sup_{F,a_j}(F, a_i), \ldots, \min \sup_{F,a_j}(F, a_m) \) is a clustering attribute.

Proof. The proof is clear from Definition 10 and Proposition 2.
The highest support index. This is followed by a clustering attribute is selected using Definition 7. From Table 4, the clustering attribute without any prejudice. Thus, it is clear that TR and MMR belong to different attributes. In contrast to MDA, the MMR technique is the reversal form of TR technique. But, MMR technique needs to compare two attributes, i.e., Exp and IT, that shared minimum roughness value 0, before Exp attribute is finally selected with the second minimum roughness value 0.75.

As a comparison, the result of the implementations using TR, MMR, MDA and NSS is given in Tables 5–8, respectively. From Table 5, the highest total roughness is 0.3155 which belongs to Exp attribute. Thus, it is clear that TR and MMR techniques also select the Exp attribute as the clustering attribute. The result is reasonable since MMR technique is the reversal form of TR technique. But, MMR technique needs to compare two attributes, i.e., Exp and IT, that shared minimum roughness value 0, before Exp attribute is finally selected with the second minimum roughness value 0.75.

It is also clear that, MDA technique also selects Exp attribute as the clustering attribute. However, the MDA needs 3 iterations before attribute “Exp” is selected as the clustering attribute. From Table 6, after the first iteration, the MDA selects Exp, IT, Math and Prog as candidate attributes. Further, it selects Exp, Math and Prog in the second iteration. The decision to select attribute Exp can only be made in the third iteration since attribute Exp has the highest MDA as compared to attributes Math and Prog. In contrast to MDA, NSS does not face the same problem as MDA even though they are using the same approach. It is clearly shown in Table 8 that NSS can directly choose Exp as the clustering attribute without any prejudice.
6. Experiment results

This section explains and discusses the experimental results of the proposed technique. The main focus of the experiments is on the performance measurement of the proposed technique in which execution time is used as a parameter. For comparisons, four techniques which have been discussed earlier will be used with fourteen datasets obtained from the benchmark UCI machine learning Repository [31] and a Supplier Dataset from [32] as described in Table 9 below.

All the selected data sets are different from one another in terms of size, either horizontally or vertically aimed to analyze the performance of the proposed technique when involving a high number of records as well as the high number of attributes. Some datasets have been modified by removing instances that have incomplete data and removing an attribute that only have one categorical value. All techniques (TR, MMR, MDA, NSS and MAR) are implemented using Ruby programming language version 1.9 under Windows 7 Home Edition operating system powered by Intel i5 processor with 4 GB memory.

The results of the experiments were summarized in five (5) figures. Fig. 2 shows the execution time results between 0.001 s and

<table>
<thead>
<tr>
<th>No</th>
<th>Data sets</th>
<th>Number of instances</th>
<th>Number of attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Acute Inflammation</td>
<td>120</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>Balloons</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>Bank Marketing</td>
<td>45,211</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>Car Evaluation</td>
<td>1728</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>Chess</td>
<td>3196</td>
<td>36</td>
</tr>
<tr>
<td>6</td>
<td>Cylinder Bands</td>
<td>512</td>
<td>15</td>
</tr>
<tr>
<td>7</td>
<td>Lenses</td>
<td>24</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>Lung Cancer</td>
<td>32</td>
<td>56</td>
</tr>
<tr>
<td>9</td>
<td>Mushroom</td>
<td>8124</td>
<td>22</td>
</tr>
<tr>
<td>10</td>
<td>Nursery</td>
<td>12,960</td>
<td>7</td>
</tr>
<tr>
<td>11</td>
<td>Solar Flare</td>
<td>1389</td>
<td>10</td>
</tr>
<tr>
<td>12</td>
<td>Soybean (Large)</td>
<td>307</td>
<td>35</td>
</tr>
<tr>
<td>13</td>
<td>Soybean (Small)</td>
<td>47</td>
<td>35</td>
</tr>
<tr>
<td>14</td>
<td>Supplier Dataset</td>
<td>27</td>
<td>7</td>
</tr>
<tr>
<td>15</td>
<td>Zoo</td>
<td>101</td>
<td>17</td>
</tr>
</tbody>
</table>

Fig. 2. The execution times (in seconds) of TR, MMR, MDA, NSS and MAR.

Fig. 3. The execution times (in seconds) of TR, MMR, MDA, NSS and MAR.

Fig. 4. The execution times (in seconds) of TR, MMR, MDA, NSS and MAR.

Fig. 5. The execution times (in seconds) of TR, MMR, MDA, NSS and MAR.

Fig. 6. The execution times (in seconds) of TR, MMR, MDA, NSS and MAR.

Please cite this article in press as: R. Mamat et al., MAR: Maximum Attribute Relative of soft set for clustering attribute selection, Knowl. Based Syst. (2013), http://dx.doi.org/10.1016/j.knosys.2013.05.009
Table 10
Summary of comparisons between TR, MMR, MDA, NSS and MAR in term of execution time.

<table>
<thead>
<tr>
<th>No</th>
<th>Data sets</th>
<th>Techniques name</th>
<th>Improvement (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Acute Inflammation</td>
<td>TR 0.0362</td>
<td>0.0345 0.0300</td>
</tr>
<tr>
<td>2</td>
<td>Balloons</td>
<td>MMR 0.0015</td>
<td>0.0017 0.0016</td>
</tr>
<tr>
<td>3</td>
<td>Bank Marketing</td>
<td>MDA 0.2337</td>
<td>0.1640 0.1093</td>
</tr>
<tr>
<td>4</td>
<td>Car Evaluation</td>
<td>NSS 118.55</td>
<td>69.27 95.36</td>
</tr>
<tr>
<td>5</td>
<td>Chess</td>
<td>MAR 0.7900</td>
<td>0.4160 0.3860</td>
</tr>
<tr>
<td>6</td>
<td>Cylinder Bands</td>
<td></td>
<td>0.1500 0.1364</td>
</tr>
<tr>
<td>7</td>
<td>Lenses</td>
<td></td>
<td>112.91 128.39</td>
</tr>
<tr>
<td>8</td>
<td>Lung Cancer</td>
<td></td>
<td>0.2700 0.2500</td>
</tr>
<tr>
<td>9</td>
<td>Mushroom</td>
<td></td>
<td>0.2685 0.2691</td>
</tr>
<tr>
<td>10</td>
<td>Nursery</td>
<td></td>
<td>0.9905 0.9891</td>
</tr>
<tr>
<td>11</td>
<td>Solar Flare</td>
<td></td>
<td>0.8597 0.9284</td>
</tr>
<tr>
<td>12</td>
<td>Soybean (Large)</td>
<td></td>
<td>0.0575 0.0425</td>
</tr>
<tr>
<td>13</td>
<td>Soybean (Small)</td>
<td></td>
<td>0.0028 0.0039</td>
</tr>
<tr>
<td>14</td>
<td>Supplier Dataset</td>
<td></td>
<td>0.0593 0.0406</td>
</tr>
<tr>
<td>15</td>
<td>Zoo</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Average improvement 33.45

Fig. 7. The scalability of TR, MMR, MDA, NSS and MAR to the number of instances.

Fig. 8. The scalability of TR, MMR, MDA, NSS and MAR to the number of instances.

Fig. 9. The scalability of TR, MMR, MDA, NSS and MAR to the number of attributes.

Fig. 10. The scalability of TR, MMR, MDA, NSS and MAR to the number of attributes.

0.004 s which are involving the datasets of Ballons, Lenses, and Supplier.

Fig. 3 illustrates the execution times results between 0.017 s and 0.06 s which are involving the datasets of Acute Inflammation, Soybean (Small), and Zoo.

The results of the execution time between 0.03 s and 0.3 s shown in Fig. 4, which are involving the datasets of Car Evaluation, Lung Cancer, and Nursery.

Fig. 5 illustrates the results of the execution time between 0.1 s and 1 s, which are involving the datasets of Cylinders Bands, Solar Flare, and Soybean (Large).

Table 10 summarizes the comparison results between TR, MMR, MDA, NSS and MAR in term of execution time. To compute the relative improvement of MAR with respect to TR, MMR, MDA, and NSS i.e. how much faster is MAR with respect to them, we use the following formula,

\[
\text{Improvement} = \frac{\text{TR} - \text{MAR}}{\text{TR}} \times 100\%
\]

Please cite this article in press as: R. Mamat et al., MAR: Maximum Attribute Relative of soft set for clustering attribute selection, Knowl. Based Syst. (2013), http://dx.doi.org/10.1016/j.knosys.2013.05.009
In reviewing the scalability of TR, MMR, MDA, NSS and MAR algorithms on fifteen data sets in Table 9. The first is their scalability to the number of instances and the second is their scalability to the number of attributes.

Figs. 7 and 8 show the scalability to the number of instances of using these algorithms in fifteen datasets. It can be observed that the execution time of these algorithms tends to increase. However, for a contrast example that the execution time to select a clustering attribute of TR, MMR, MDA, NSS and MAR algorithms on Chess data set with the number of instances 3196 is 118.55, 156.36, 69.27, 59.57, and 3.3663 s, respectively. Meanwhile, in Nursery data set with the number of instances 12,960 (about four times of Chess's instances), these algorithms achieve 0.2685, 0.2691, 0.2395, 0.2056, and 0.0362 s, respectively. Hence, in general that the five algorithms increases non-linearly as the number of instances in the fifteen datasets is increased.

Figs. 9 and 10 show scalability to the number of attributes of using these algorithms in fifteen datasets. From Fig. 9, it can be observed that the execution time of these algorithms tends to increase linearly to the dataset with attribute size ranging from 4 to 10.

However, for the number of attributes in data sets increased ranging from 11 to 56 as shown in Fig. 10, the five algorithms increases non-linearly. As an example that the execution time to select a clustering attribute of TR, MMR, MDA, NSS and MAR algorithms on Bank Marketing data set with the number of attributes 11 is 297.46, 297.61, 262.72, 210.28, and 18.6556 s, respectively. Meanwhile, in Lung Cancer data set with the number of attributes 56 (about five times of Bank Marketing's attributes), these algorithms achieve 0.2685, 0.2691, 0.2395, 0.2056, and 0.0362 s, respectively. Hence, in general, the five algorithms increases non-linearly as the number of instances in the fifteen datasets is increased.

Figs. 9 and 10 show scalability to the number of attributes of using these algorithms in fifteen datasets. From Fig. 9, it can be observed that the execution time of these algorithms tends to increase linearly to the dataset with attribute size ranging from 4 to 10.

However, for the number of attributes in data sets increased ranging from 11 to 56 as shown in Fig. 10, the five algorithms increases non-linearly. As an example that the execution time to select a clustering attribute of TR, MMR, MDA, NSS and MAR algorithms on Bank Marketing data set with the number of attributes 11 is 297.46, 297.61, 262.72, 210.28, and 18.6556 s, respectively. Meanwhile, in Lung Cancer data set with the number of attributes 56 (about five times of Bank Marketing's attributes), these algorithms achieve 0.2685, 0.2691, 0.2395, 0.2056, and 0.0362 s, respectively. Hence, in general, the five algorithms increases non-linearly as the number of instances in the fifteen datasets is increased.

In reviewing the scalability of TR, MMR, MDA, NSS and MAR algorithms on the fifteen data sets, the value varies in several numbers of instances and attributes. However, on the whole, these algorithms have a good scalability to data size. Let \( S = (U, A, V, J) \) be an information system as described in Section 2.1, the data size here is based-on the complexity of computing \( S = (U, A, V, J) \). Since there in are \(|U|\) number of instances and \(|A|\) number of attributes, thus the complexity will be a \( O(|U| \times |A|) \). From the fifteen data sets in Table 9, Figs. 11 and 12 describe the scalability of these algorithms on the fifteen data sets.

**Other Algorithms – MAR**

In summary, based on the experiment on fourteen UCI datasets and a supplier dataset, the proposed algorithm achieved a lower computational time than the three rough set-based algorithms, i.e. TR, MMR, and MDA up to 62%, 64%, and 40% respectively and compared to a soft set–based algorithm, i.e. NSS up to 33%. Finally, on the average that MAR outperforms the four algorithms up to 33.45%.

Furthermore, we present two types of scalability of TR, MMR, MDA, NSS and MAR algorithms. The first is their scalability to the number of instances and the second is their scalability to the number of attributes.

Figs. 7 and 8 show the scalability to the number of instances of using these algorithms in fifteen datasets. It can be observed that the execution time of these algorithms tends to increase. However, for a contrast example that the execution time to select a clustering attribute of TR, MMR, MDA, NSS and MAR algorithms on Chess data set with the number of instances 3196 is 118.55, 156.36, 69.27, 59.57, and 3.3663 s, respectively. Meanwhile, in Nursery data set with the number of instances 12,960 (about four times of Chess's instances), these algorithms achieve 0.2685, 0.2691, 0.2395, 0.2056, and 0.0362 s, respectively. Hence, in general, the five algorithms increases non-linearly as the number of instances in the fifteen datasets is increased.

Figs. 9 and 10 show scalability to the number of attributes of using these algorithms in fifteen datasets. From Fig. 9, it can be observed that the execution time of these algorithms tends to increase linearly to the dataset with attribute size ranging from 4 to 10.

However, for the number of attributes in data sets increased ranging from 11 to 56 as shown in Fig. 10, the five algorithms increases non-linearly. As an example that the execution time to select a clustering attribute of TR, MMR, MDA, NSS and MAR algorithms on Bank Marketing data set with the number of attributes 11 is 297.46, 297.61, 262.72, 210.28, and 18.6556 s, respectively. Meanwhile, in Lung Cancer data set with the number of attributes 56 (about five times of Bank Marketing's attributes), these algorithms achieve 0.2685, 0.2691, 0.2395, 0.2056, and 0.0362 s, respectively. Hence, in general, the five algorithms increases non-linearly as the number of instances in the fifteen datasets is increased.

In reviewing the scalability of TR, MMR, MDA, NSS and MAR algorithms on the fifteen data sets, the value varies in several numbers of instances and attributes. However, on the whole, these algorithms have a good scalability to data size. Let \( S = (U, A, V, J) \) be an information system as described in Section 2.1, the data size here is based-on the complexity of computing \( S = (U, A, V, J) \). Since there in are \(|U|\) number of instances and \(|A|\) number of attributes, thus the complexity will be a \( O(|U| \times |A|) \). From the fifteen data sets in Table 9, Figs. 11 and 12 describe the scalability of these algorithms on the fifteen data sets.


