Journal of Computational and Theoretical Nanoscience
Vol. 12, 5928–5939, 2015

RESEARCH ARTICLE

A Soft Set Approach for Clustering Student Assessment Datasets

Suhirman[1,2], Tutut Herawan[3,*], Iwan Tri Riyadi Yanto[4], Jasni Mohamad Zain[2], Qin Hongwu[2], and Zailani Abdullah[5]

[1] *Universitas Teknologi Yogyakarta, Yogyakarta, Indonesia*
[2] *Faculty of Computer System and Software Engineering, Universiti Malaysia Pahang Gambang, Kuantan Pahang, Malaysia*
[3] *AMCS Research Center, Yogyakarta, Indonesia*
[4] *University Ahmad Dahlan Yogyakarta, Yogyakarta, Indonesia*
[5] *Universiti Malaysia Terengganu, Terengganu, Malaysia*

Educational data mining has been studied extensively as it provides useful information for educators to make more accurate decisions concerning their students, and to adapt their teaching strategies accordingly. Data clustering as one of data mining techniques can be considered as an alternative method for educational data mining. In this paper, a data clustering technique based on soft set theory is presented. The Maximum Degree of Domination in soft set theory (MDDS) is proposed and further applied to select the best attribute in educational data clustering. To find meaningful clusters from a dataset, clustering attribute selection is conducted so that attributes within the clusters made will have a high correlation or high interdependence to each other while the attributes in other clusters are less correlated or more independent. The datasets are taken from a survey from a number of courses at the Information Engineering and the Architecture Departments of the University Technology of Yogyakarta, Indonesia. The evaluation criteria uses score range from 0 to 100. Student name, age, race, and attendance are not required in this assessment. In the results, we show how to determine the dominant attributes of a set of attributes of an assessment list by using the proposed technique. The results obtained can potentially contribute to give a recommendation in awarding the final grade of a course more quickly and accurately.

Keywords: Soft Set Theory, Clustering Attributes, Domination Degree, Education, Student.

1. INTRODUCTION

Education is the foundation for achieving sustainable development. Concerning with the importance of this kind of education, the key aspect is needed on the measuring achievement levels in higher environmental education.[1] Higher education institutions are overwhelmed with huge amounts of information regarding student's enrollment, number of courses completed, achievement in each course, performance indicators and other data. This has led to an increasingly complex analysis process of the growing volume of data and to the incapability to take decisions regarding curriculum reform and restructuring. On the other side, educational data mining is a growing field aiming at discovering knowledge from student's data in order to thoroughly understand the learning process and take appropriate actions to improve the student's performance and the quality of the courses delivery.[2]

Educational data mining (EDM) can be applied to wide areas of research including elearning, intelligent tutoring systems, text mining, social network mining, and etc. In education, EDM can function as a replacement for less accurate but more established psychometric techniques. Educational data mining is an interactive cycle of hypothesis formation, testing and refinements that alternates between two complementary types of activities. One type of activity is qualitative analysis, focuses on understanding individual tutorial events. Other type involve, knowledge tracing analyses the growth curve by aggregating over successive opportunities to apply skills.[3] The EDM process converts raw data coming from educational systems into useful information that could potentially have a great impact on educational research and practice. This process does not differ much from other application areas of DM, like business, genetics, medicine, and etc., because it follows the same steps as the general DM process.[4]

*Author to whom correspondence should be addressed.

1546-1955/2015/12/5928/012 doi:10.1166/jctn.2015.4738

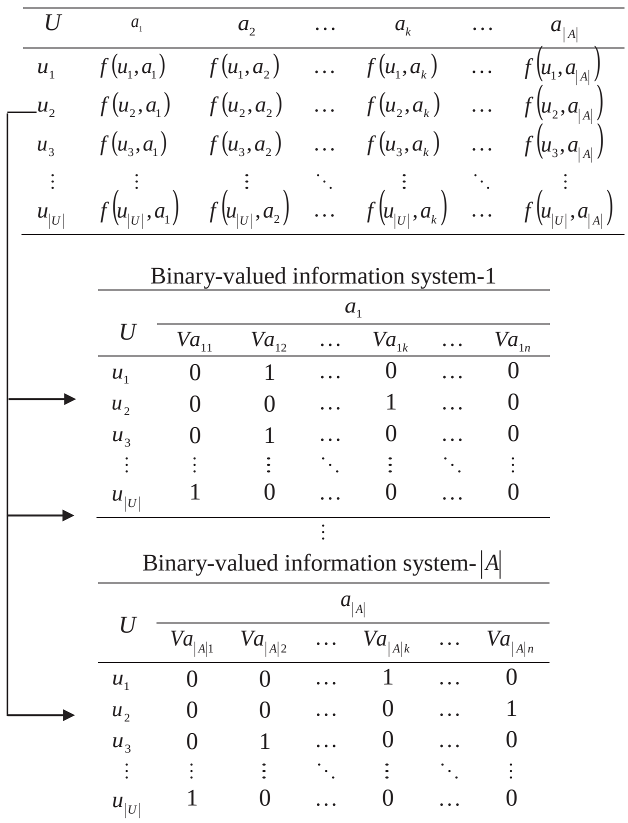

Fig. 1. A decomposition of a multi-valued information system.

$$
(F,E)=\begin{cases}
(F,Education)=\begin{cases}(F,Education_{PhD})=\{1,2,3,4\}\\(F,Education_{Master})=\{5,6,7,8,9\}\end{cases}\\
(F,Attitude)=\begin{cases}(F,Attitude_{Medium}=\{1,2,3,4,5,6,9\}\\(F,Attitude_{Good})=\{7,8\}\end{cases}\\
(F,Math)=\begin{cases}(F,Math_{Good})=\{1,2,9\}\\(F,Math_{Medium})=\{3,4\}\\(F,Math_{Low})=\{5,6,7,8\}\end{cases}\\
(F,Computer)=\begin{cases}(F,Computer_{Good})=\{1,2,8,9\}\\(F,Computer_{Medium})=\{3,4,5,6,7\}\end{cases}\\
(F,Statistic)=\begin{cases}(F,Statistic_{Good})=\{1,2,3,4,9\}\\(F,Statistic_{Medium}=\{5,6,7,8\}\end{cases}\\
(F,English)=\begin{cases}(F,English_{Medium})=\{1,2,5,7,8\}\\(F,English_{Good})=\{3,4\}\\(F,English_{Low})=\{6\}\\(F,English_{No})=\{9\}\end{cases}\\
(F,Comunication)=\begin{cases}(F,Comunication_{Medium})=\{1,2,8,9\}\\(F,Comunication_{Good})=\{3,4\}\\(F,Comunication_{Low})=\{5,7\}\\(F,Comunication_{Less})=\{6\}\end{cases}\\
(F,Management)=\begin{cases}(F,Management_{Good})=\{1,2,3,4\}\\(F,Management_{Medium})=\{5,6,7,8,9\}\end{cases}\\
(F,Experience)=\begin{cases}(F,Experience_{Good})=\{1,2,6,7\}\\(F,Experience_{Medium})=\{3,4,5,8,9\}\end{cases}
\end{cases}
$$

Fig. 2. Multi softsets compotition from dataset in Table I.

Clustering as one of DM processes is an important data analysis method used to group data with similar characteristics. It has been used in many areas such as gene data processing,[1] transactional data processing,[2] decision support,[3] and radar signals processing.[4] A new way for data clustering is called the soft set theory, which is proposed by Molodtsov.[5,6] Molodtsov also pointed out in Ref. [9] 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 sets,[7] rough sets,[8] intuitionistic fuzzy sets,[9] vague sets,[10] and interval mathematics,[11] and etc. Therefore, it is very convenient and easy to apply soft set theory into practice. It has a rich potential for applications in several directions, few of which had already been demonstrated by Molodtsov, such as the smoothness of functions, game theory, operations research, Riemann integration, Perron integration, probability theory, and measurement theory.[5] Presently, soft set theory has attracted attention of many researchers all over the world, who have contributed essentially to its development and applications. Research on the soft set theory is progressing rapidly. Maji et al. firstly introduced some definitions of the related operations on soft sets.[12] Ali et al. took into account some errors of former studies and put forward some new operations on soft sets.[13] Maji et al. firstly employed soft sets to solve the decision-making problem.[14] Chen et al. pointed out that the conclusion of soft set reduction offered in Ref. [14] was incorrect, and then present a new notion of parameterization reduction in soft sets in comparison with the definition to the related concept of attributes reduction in rough set theory.[15] The concept of normal parameter reduction is introduced in Ref. [16], which overcome the problem of suboptimal choice and added parameter set of soft sets. Ma et al.[17] gave a new efficient normal parameter reduction of soft sets to improve.[16] Zou and Xiao[18] and Hongwu et al.[19] depicted data analysis approaches of soft sets under

Table I. Multi valued information system.

Std	Education	Attitude	Math	Computer	Statistic	English	Comunication	Management	Experience
1	Ph.D	Medium	Good	Good	Good	Medium	Medium	Good	Good
2	Ph.D	Medium	Good	Good	Good	Medium	Medium	Good	Good
3	Ph.D	Medium	Medium	Medium	Good	Good	Good	Good	Medium
4	Ph.D	Medium	Medium	Medium	Good	Good	Good	Good	Medium
5	Master	Medium	Low	Medium	Medium	Medium	Low	Medium	Medium
6	Master	Medium	Low	Medium	Medium	Low	Less	Medium	Good
7	Master	Good	Low	Medium	Medium	Medium	Low	Medium	Good
8	Master	Good	Low	Good	Medium	Medium	Medium	Medium	Medium
9	Master	Medium	Good	Good	Good	No	Medium	Medium	Medium

RESEARCH ARTICLE

Table II. Domination value from data set in Table I.

Soft set	Domination value	Max—domination
$(F, Education_{PhD})$	0.4444,0,0,0,0,0,0,0,0.4444,0,0,0,0,0,0,0,0,0.4444,0,0,0.4444,0,0	0.4444
$(F, Education_{Master})$	0,0.5556,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.5556,0,0,0	0.5556
$(F, Attitude_{Medium})$	0,0.2222,0,0,0.2222,0,0,0.2222,0,0.2222,0,0,0,0,0,0,0,0,0.2222,0.2222,0,0,0	0.2222
$(F, Attitude_{Good})$	0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.7778,0,0	0.7778
$(F, Math_{Good})$	0,0,0.3333,0,0,0.3333,0,0,0.3333,0,0,0,0,0.3333,0,0,0,0,0,0,0.3333,0,0	0.3333
$(F, Math_{Medium})$	0.2222,0,0,0.2222,0,0,0.2222,0,0.2222,0,0.2222,0,0,0,0.2222,0,0,0.2222,0,0,0.2222,0,0.2222	0.2222
$(F, Math_{Low})$	0,0.4444,0,0,0.4444,0,0,0.4444,0,0,0,0,0,0,0,0,0,0,0.4444,0,0,0,0	0.4444
$(F, Computer_{Good})$	0,0,0,0,0,0.4444,0,0,0,0,0,0,0,0.4444,0,0,0, 0,0,0 0,0,0	0.4444
$(F, Computer_{Medium})$	0,0,0,0,0,0,0.5556,0,0,0,0,0,0,0, 0,0,0,0,0,0,0,0,0	0.5556
$(F, Statistic_{Good})$	0,0.4444,0,0,0.4444,0,0,0.4444,0,0,0,0,0,0,0,0,0,0,0.4444,0,0,0,0	0.4444
$(F, Statistic_{Medium})$	0,0,0,0,0,0,0,0,0.5556,0,0,0,0,0,0,0,0,0,0,0,0.5556,0,0	0.5556
$(F, English_{Medium})$	0,0,0,0,0,0,0,0,0,0.5556,0,0,0,0,0,0,0,0,0,0,0,0,0	0.5556
$(F, English_{Good})$	0.2222,0,0,0.2222,0,0,0.2222,0,0.2222,0,0.2222,0,0,0,0.2222,0,0,0.2222,0,0,0.2222,0,0.2222	0.2222
$(F, English_{Low})$	0,0.1111,0,0,0.1111,0,0.1111,0.1111,0,0,0,0.1111,0,0,0,0,0.1111,0,0.1111,0,0.1111,0.1111,0	0.1111
$(F, English_{No})$	0,0.1111,0.1111,0,0,0.1111,0,0,0.1111,0,0,0,0.1111,0.1111,0,0,0,0,0.1111,0,0.1111,0,0.1111	0.1111
$(F, Comunication_{Medium})$	0,0,0,0,0,0.4444,0,0,0,0,0,0,0,0.4444,0,0,0,0,0,0,0,0,0	0.4444
$(F, Comunication_{Good})$	0.2222,0,0,0.2222,0,0,0.2222,0,0.2222,0,0.2222,0,0,0,0.2222,0,0,0.2222,0,0,0.2222,0,0.2222	0.2222
$(F, Comunication_{Low})$	0,0.2222,0,0,0.2222,0,0.2222,0.2222,0,0.2222,0,0,0,0,0,0.2222,0,0,0.2222,0,0,0,0	0.2222
$(F, Comunication_{Less})$	0,0.1111,0,0,0.1111,0,0.1111,0.1111,0,0,0,0.1111,0,0,0,0,0.1111,0,0.1111,0,0.1111,0.1111,0	0.1111
$(F, Management_{Good})$	0.4444,0,0,0,0,0,0,0,0.4444,0,0,0,0,0,0,0,0,0.4444,0,0,0.4444,0,0	0.4444
$(F, Management_{Medium})$	0,0.5556,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.5556,0,0,0,0	0.5556
$(F, Experience_{Good})$	0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.4444,0	0.4444
$(F, Experience_{Medium})$	0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0.5556	0.5556

incomplete information. Xiao et al. described a combined forecasting approach based on fuzzy soft sets.[20] Herawan and Mat Deris proposed the idea of mining association rules under soft set theory.[21] Qin et al. presented the idea of selecting a clustering attribute under soft set theory,[22] the proposed method is based on the notion of multi soft-sets proposed by Herawan et al.[23] Furthermore, the soft set model can also be combined with other mathematical models. Some algebraic concepts based on soft sets were proposed as diverse as soft groups,[24] soft ideals and idealistic soft BCK/BCI-algebras,[25] soft semirings, soft sub semirings, soft ideals and idealistic soft semirings.[26]

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 Ref. [23]. Since a direct proof that every rough set is a soft set have been given in Ref. [27],

MDDS Algorithm

Input: Categorical-valued data-set
Output: A Clustering attribute
Begin
1. Builds the multi-soft set approximation
2. Calculate Domination of Attributes a_i, with respect to all a_j, where $i \neq j$
3. Select the maximum of domination degree of each attributes
4. Select the clustering attribute based on the maximum degree of domination of attributes

End

Fig. 3. MDDS algorithm.

Hongwu et al.[19] 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.

Mamat et al.[28] proposes MAR, an alternative technique to select a clustering attribute using soft set theory. It is based on a concept of Maximum Attribute Relative where the comparison of attributes is made by taking into account the relative of the attribute at the category level. The proposed technique potentially discovers the attributes subsets with better coverage. However, the technique is still facing with high computational time.

In this paper, Maximum Degree of Domination in soft set theory (MDDS)—an alternative technique for selecting a clustering attribute based on soft set theory is presented. The MDDS is proposed by applying the concept of dominance relation in multi-soft sets in determining the most dominant attribute. The most dominant attribute will be used as a clustering attribute. The MDDS is further applied to select the best attribute in educational data clustering for the assessment purpose of university students.

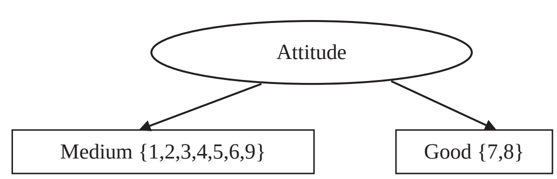

Fig. 4. Result of clustering.

RESEARCH ARTICLE

Table III. Assessment of algorithms courses.

		2011					2012					2013				
Att	Description	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
T1	Task 1	5	0	7	8	10	0	0	1	21	6	1	0	5	23	12
T2	Task 2	0	0	5	25	0	0	1	3	18	6	5	0	9	24	3
T3	Task 3	7	0	0	18	5	0	0	6	20	2	9	0	1	17	14
T4	Task 4	6	0	0	17	7	0	0	4	22	2	12	0	1	24	4
MT	Midterm	1	0	1	22	6	0	0	4	17	7	10	0	2	20	9
FE	Final exam	2	2	3	7	16	0	0	3	18	7	1	0	1	12	27

Table IV. Assessment of software engineering courses.

		2011					2012					2013				
Att	Description	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
T1	Task 1	6	1	20	48	12	51	0	4	14	18	25	3	18	25	15
T2	Task 2	3	0	8	58	18	1	0	9	59	18	3	0	5	58	20
T3	Task 3	6	0	37	44	0	5	0	31	51	0	13	0	36	37	0
T4	Task 4	0	4	3	22	58	0	1	9	27	50	1	2	5	25	53
MT	Midterm	5	17	49	14	2	0	16	55	13	3	0	24	48	10	4
FE	Final exam	5	18	30	25	9	4	5	43	22	13	4	11	37	29	5

Table V. Assessment of system security courses.

		2011					2012					2013				
Att	Description	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
T1	Task 1	1	1	24	64	0	0	0	8	82	0	0	0	15	74	0
T2	Task 2	0	2	10	77	1	0	0	2	88	0	0	0	9	80	0
T3	Task 3	1	2	23	64	0	0	1	10	79	0	0	1	13	75	0
T4	Task 4	1	5	12	45	27	0	0	2	8	80	0	1	5	20	63
MT	Midterm	10	24	32	21	3	4	11	32	34	9	6	22	31	26	4
FE	Final exam	5	3	26	46	10	0	0	11	68	11	2	0	22	56	9

Table VI. Assessment of system security courses.

		2011					2012					2013				
Att	Description	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
T1	Task 1	3	4	15	12	28	5	2	15	8	37	0	2	2	2	55
T2	Task 2	4	0	4	54	0	3	0	7	57	0	0	0	0	61	0
T3	Task 3	1	0	2	49	10	2	0	1	62	2	0	0	0	52	9
MT	Midterm	2	21	22	13	4	5	13	29	15	5	4	7	21	24	5
FE	Final exam	3	7	19	21	12	2	6	32	18	9	3	4	12	20	22

Table VII. Assessment of architectural design courses.

		2011					2012					2013				
Att	Description	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5	#1	#2	#3	#4	#5
T1	Task 1	9	0	2	23	0	2	9	10	12	0	5	0	22	0	0
T2	Task 2	8	0	7	19	0	3	2	16	12	0	13	0	14	0	0
T3	Task 3	9	11	9	5	0	5	0	5	23	0	9	0	18	0	0
T4	Task 4	13	3	15	3	0	6	7	7	13	0	9	0	18	0	0
T5	Task 5	9	4	5	14	2	4	2	6	21	0	14	0	13	0	0
T6	Task 6	19	4	6	5	0	4	9	11	9	0	15	0	12	0	0
T7	Task 7	15	0	11	8	0	9	0	24	0	0	15	0	12	0	0
T8	Task 8	16	0	3	15	0	14	0	19	0	0	3	10	7	6	1
MT	Midterm	3	2	20	9	0	14	0	19	0	0	6	1	4	11	5
FE	Final exam	3	4	22	5	0	5	0	28	0	0	9	0	0	18	0

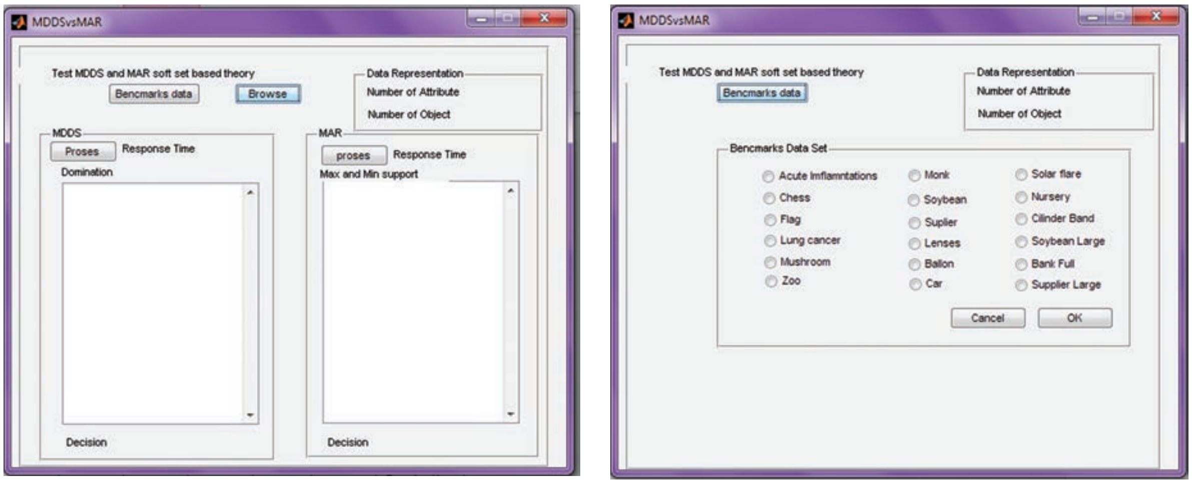

Fig. 5. User interface of the developed MDDS software.

The rest of the paper is organized as follows. In Section 2, we explain some basic knowledge about soft set theory. We present the proposed MDDS technique for selecting a clustering attribute in Section 3. In Section 4, we explain the data used for experiments. Section 5 discusses the experiment results and a series of evaluations on the result. Finally, we conclude this work in Section 6.

2. SOFT SET THEORY

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

DEFINITION 1 (SEE REFS. [29, 30]). A pair (F, E) is called a soft set over U where F is a mapping given by

$$F: E \to P(U) \quad (1)$$

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

EXAMPLE 1. Let a universe $U = \{s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}\}$ be a set of student candidates and a set of parameters $E = \{e_1, e_2, e_3, e_4, e_5\}$ into a series of assessment which stands for parameters "intelligence," "attitude," "finances," "family support," and "motivation," respectively. Consider F be a mapping of E into the set of all subsets of the set U as $F(e_1) = \{s_1, s_2, s_4, s_5\}$, $F(e_2) = \{s_3, s_8, s_9\}$, $F(e_3) = \{s_6, s_9, s_{10}\}$, $F(e_4) = \{s_2, s_3, s_4, s_5, s_8\}$, and $F(e_5) = \{s_6, s_9, s_{10}\}$. Now consider a soft set (F, E), which describes the "capabilities of the student candidate for hire." In this example, the soft set (F, E) is given by

$$(F, E) = \left\{ \begin{array}{l} \text{intelligence} = \{s_1, s_2, s_4, s_5\} \\ \text{attitude} = \{s_3, s_8, s_9\} \\ \text{finance} = \{s_6, s_9, s_{10}\} \\ \text{family support} = \{s_2, s_3, s_4, s_5, s_8\} \\ \text{motivation} = \{s_2, s_5, s_6, s_7, s_8, s_9, s_{10}\} \end{array} \right\} \quad (2)$$

Obviously, the soft set (F, E) is not a crisp set and (F, E) is a parameterized family $\{F(e_i), i = 1, 2, 3, \ldots, 5\}$ of subsets of the set U that have two parts of approximation: predicate (p) and value (v). For example, for the approximation "attitude $= \{s_3, s_8, s_9\}$," p is attitude and $v = \{s_3, s_8, s_9\}$. In the following definition, we present the notion od value-class of a soft set.

DEFINITION 2 (SEE REF. [14]). The class of all value sets of a soft set (F, E) is called value-class of the soft set and is denoted by $C_{(F,E)}$.

Table VIII. Matric results from algorithm courses 2011.

Attribute (wrt)	Domination degree						Maximum domination
	T1	T2	T3	T4	MT	FE	
T1	0.00	0.00	0.00	0.00	0.07	0.07	0.07
T2	0.60	0.00	0.17	0.23	0.27	0.70	0.70
T3	0.00	0.00	0.00	0.00	0.07	0.13	0.13
T4	0.00	0.00	0.00	0.00	0.07	0.13	0.13
MT	0.17	0.00	0.00	0.00	0.00	0.17	0.17
FE	0.00	0.00	0.00	0.00	0.07	0.00	0.07

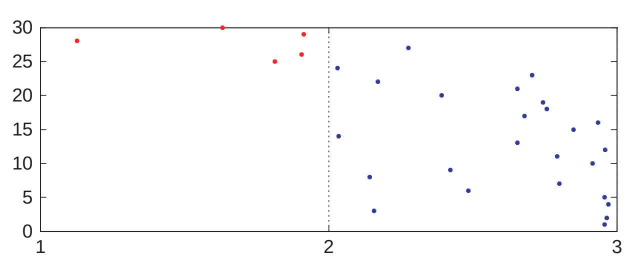

Fig. 6. Clustering visualization students on algorithms courses 2011.

Table IX. Matric results from algorithm courses 2012.

Attribute (wrt)	Domination degree						Maximum domination
	T1	T2	T3	T4	MT	FE	
T1	0.00	0.04	0.00	0.07	0.00	0.11	0.11
T2	0.04	0.00	0.07	0.07	0.00	0.11	0.11
T3	0.04	0.14	0.00	0.07	0.00	0.00	0.14
T4	**0.04**	**0.25**	**0.07**	**0.00**	**0.00**	**0.11**	**0.25**
MT	0.04	0.14	0.00	0.00	0.00	0.11	0.14
FE	0.04	0.14	0.00	0.14	0.00	0.00	0.14

The value-class of a soft set will be used in determining the dominance of soft set given in Section 3 as follow.

3. THE PROPOSED MDDS METHODS

In this section, we present the proposed MDDS for selecting a clustering attribute. Firstly, we recall the notion of multi-soft sets to represent with multi-valued information system. Secondly, we present the notion of domination in multi-soft sets. Finally, MDDS is presented to select the best clustering attribute.

Note that from Definition 1, the "standard" soft set deals with a binary-valued information system. For a multi-valued information system $S = (U, A, V, f)$, where $V = \bigcup_{a \in A} V_a$, V_a is the domain (value set) of attribute a which has multi values, a decomposition can be made from S into $|A|$ number of binary-valued information systems $S = (U, A, V_{\{0,1\}}, f)$. In the following sub-section, we present the construction of *multi-soft sets* representing $S = (U, A, V, f)$ based on such decomposition.

3.1. Multi-Soft Sets

In this sub-section, we will propose an idea of decomposing a multi-valued information system $S = (U, A, V, f)$ into $|A|$ numbers of Boolean-valued information system $S^i = (U, a_i, V_{\{0,1\}}, f)$, where $|A|$ is the cardinality of A. The decomposition of $S = (U, A, V, f)$ is based on decomposition of $A = \{a_1, a_2, \ldots, a_{|A|}\}$ into the disjoint-singleton attribute $\{a_1\}, \{a_2\}, \ldots, \{a_{|A|}\}$. At this stage, only complete information system is given the consideration. Let $S = (U, A, V, f)$ be an information system such that for every $a \in A$, $V_a = f(U, A)$ is a finite non-empty set and for every $u \in U$, $|f(u, a)| = 1$. For every a_i under ith-attribute consideration, $a_i \in A$ and $v \in V_a$, we define the map $a_v^i \colon U \to \{0, 1\}$ such that

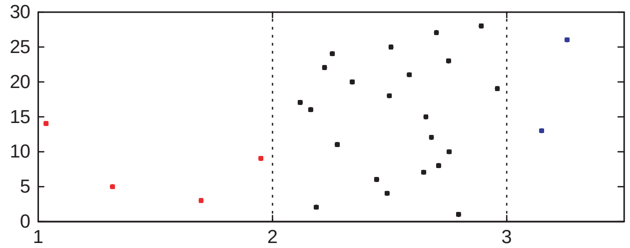

Fig. 7. Clustering visualization students on algorithms courses 2012.

Table X. Matric results from algorithm courses 2013.

Attribute (wrt)	Domination degree						Maximum domination
	T1	T2	T3	T4	MT	FE	
T1	0.00	0.12	0.02	0.02	0.05	0.05	0.12
T2	0.02	0.00	0.02	0.02	0.00	0.05	0.05
T3	0.02	0.00	0.00	0.02	0.05	0.05	0.05
T4	**0.02**	**0.00**	**0.02**	**0.00**	**0.24**	**0.05**	**0.24**
MT	0.02	0.00	0.02	0.12	0.00	0.05	0.12
FE	0.02	0.07	0.02	0.12	0.00	0.00	0.12

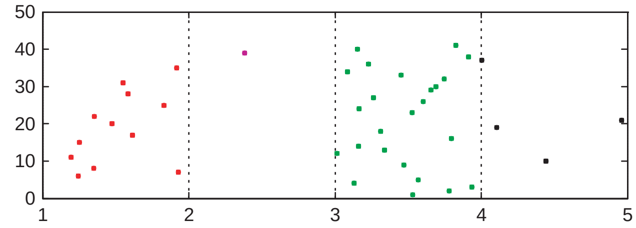

Fig. 8. Clustering visualization students on algorithms courses 2013.

$a_v^i(u) = 1$ if $f(u, a) = v$, otherwise $a_v^i(u) = 0$. The next result, we define a binary-valued information system as a quadruple $S^i = (U, a_i, V_{\{0,1\}}, f)$. The information systems $S^i = (U, a_i, V_{\{0,1\}}, f)$, $i = 1, 2, \ldots, |A|$ is referred to as a decomposition of a multi-valued information system $S = (U, A, V, f)$ into $|A|$ binary-valued information systems, as depicted in Figure 1. Every information system $S^i = (U, a_i, V_{ai}, f)$, $i = 1, 2, \ldots, |A|$ is a deterministic information system since for every $a \in A$ and for every $u \in U$, $|f(u, a)| = 1$ such that the structure of a multi-valued information system and $|A|$ number of binary-valued information systems give the same value of attribute related to objects.

Based on the notion of a decomposition of a multi-valued information system in the previous sub-section,

Table XI. Matric results from software engineering courses 2011.

Attribute (wrt)	Domination degree						Maximum domination
	T1	T2	T3	T4	MT	FE	
T1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
T2	0.01	0.00	0.00	0.00	0.00	0.00	0.01
T3	0.01	0.00	0.00	0.00	0.00	0.00	0.01
T4	0.01	0.00	0.00	0.00	0.02	0.00	0.02
MT	**0.01**	**0.03**	**0.00**	**0.03**	**0.00**	**0.00**	**0.03**
FE	0.01	0.00	0.00	0.00	0.02	0.00	0.02

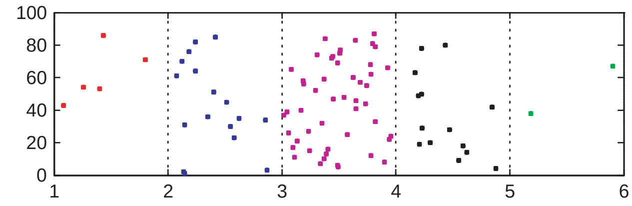

Fig. 9. Clustering visualization students on software engineering courses 2011.

RESEARCH ARTICLE

Table XII. Matric results from software engineering courses 2012.

Attribute (wrt)	Domination degree T1	T2	T3	T4	MT	FE	Maximum domination
T1	**0.00**	**0.01**	**0.06**	**0.01**	**0.00**	**0.00**	**0.06**
T2	0.00	0.00	0.00	0.01	0.00	0.06	0.06
T3	0.00	0.01	0.00	0.01	0.03	0.00	0.03
T4	0.00	0.01	0.00	0.00	0.00	0.00	0.01
MT	0.00	0.01	0.06	0.01	0.00	0.00	0.06
FE	0.00	0.01	0.00	0.01	0.00	0.00	0.01

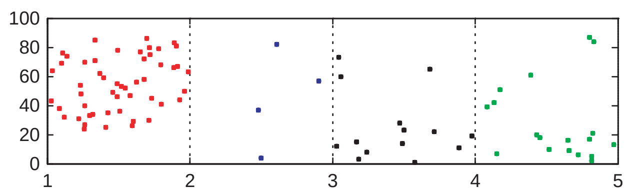

Fig. 10. Clustering visualization students on software engineering courses 2012.

in this sub-section we present the notion of multi-soft set representing multi-valued information systems. Let $S = (U, A, V, f)$ be a multi-valued information system and $S^i = (U, a_i, V_{ai}, f)$, $i = 1, 2, \ldots, |A|$ be the $|A|$ binary-valued information systems. From Proposition 1, we have

$$S = (U, A, V, f)$$
$$= \begin{cases} S^1 = (U, a_1, V_{\{0,1\}}, f) & \Leftrightarrow & (F, a_1) \\ S^2 = (U, a_2, V_{\{0,1\}}, f) & \Leftrightarrow & (F, a_2) \\ \vdots & \vdots & \vdots \\ S^{|A|} = (U, a_{|A|}, V_{\{0,1\}}, f) & \Leftrightarrow & (F, a_{|A|}) \end{cases}$$
$$= ((F, a_1), (F, a_2), \ldots, (F, a_{|A|}))$$

Table XIII. Matric results from software engineering courses 2013.

Attribute (wrt)	Domination degree T1	T2	T3	T4	MT	FE	Maximum domination
T1	0.00	0.00	0.00	0.01	0.00	0.00	0.01
T2	0.00	0.00	0.00	0.07	0.00	0.00	0.07
T3	0.00	0.06	0.00	0.03	0.00	0.00	0.06
T4	**0.00**	**0.00**	**0.00**	**0.00**	**0.16**	**0.00**	**0.16**
MT	0.00	0.00	0.00	0.09	0.00	0.00	0.09
FE	0.00	0.00	0.00	0.03	0.00	0.00	0.03

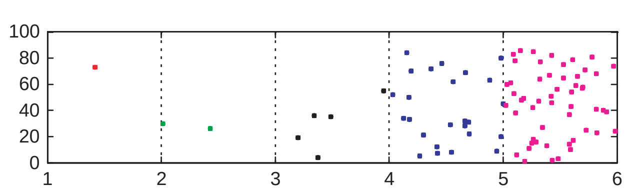

Fig. 11. Clustering visualization students on software engineering courses 2013.

Table XIV. Matric results from system security courses 2011.

Attribute (wrt)	Domination degree T1	T2	T3	T4	MT	FE	Maximum domination
T1	0.00	0.12	0.01	0.44	0.03	0.14	0.44
T2	**0.02**	**0.00**	**0.01**	**0.51**	**0.03**	**0.00**	**0.51**
T3	0.02	0.01	0.00	0.31	0.03	0.11	0.31
T4	0.02	0.01	0.03	0.00	0.00	0.00	0.03
MT	0.02	0.01	0.03	0.01	0.00	0.00	0.03
FE	0.02	0.01	0.01	0.01	0.03	0.00	0.03

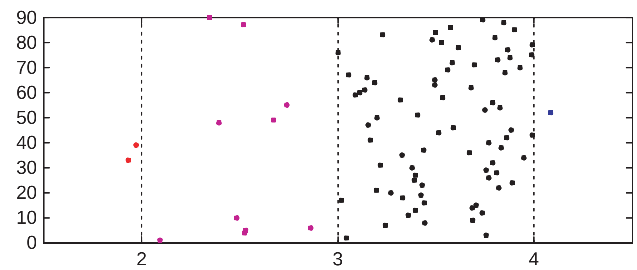

Fig. 12. Clustering visualization students on system security courses 2011.

We define $(F, E) = ((F, a_1), (F, a_2), \ldots, (F, a_{|A|}))$ as a multi-soft set over universe U representing a multi-valued information system $S = (U, A, V, f)$.

Furthermore, from Table I, the composition of multi-soft set is made, as shown in Figure 2 as following.

3.2. Domination in Multi-Soft Sets

The notion of soft set-based domination is presented as follow.

Table XV. Matric results from system security courses 2012.

Attribute (wrt)	Domination degree T1	T2	T3	T4	MT	FE	Maximum domination
T1	0.00	0.02	0.89	0.91	0.10	0.12	0.91
T2	**0.91**	**0.00**	**0.89**	**1.00**	**0.27**	**0.24**	**1.00**
T3	0.00	0.00	0.00	0.09	0.10	0.12	0.12
T4	0.00	0.02	0.89	0.00	0.10	0.12	0.89
MT	0.00	0.00	0.01	0.00	0.00	0.00	0.01
FE	0.00	0.02	0.01	0.02	0.10	0.00	0.10

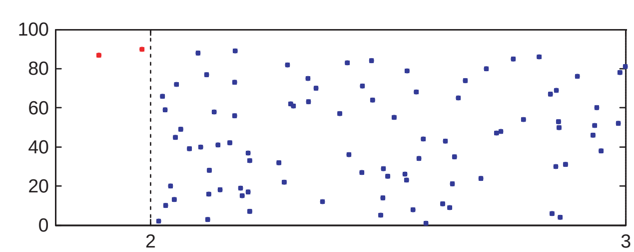

Fig. 13. Clustering visualization students on system security courses 2012.

Table XVI. Matric results from system security courses 2013.

Attribute (wrt)	Domination degree T1	T2	T3	T4	MT	FE	Maximum domination
T1	0.00	0.10	0.01	0.78	0.04	0.12	0.78
T2	**0.83**	**0.00**	**0.85**	**0.78**	**0.04**	**0.12**	**0.85**
T3	0.00	0.00	0.00	0.72	0.04	0.10	0.72
T4	0.00	0.00	0.01	0.00	0.04	0.10	0.10
MT	0.00	0.00	0.01	0.01	0.00	0.00	0.01
FE	0.00	0.00	0.01	0.01	0.04	0.00	0.04

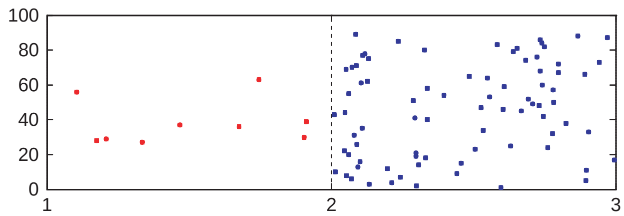

Fig. 14. Clustering visualization students on system security courses 2013.

Definition 3. Let (F, A) be multi-soft sets over U representing $S = (U, A, V, f)$ and $(F, a_i), (F, a_j) \in (F, A)$. Soft set (F, a_i) is said to be dominated by (F, a_j), denoted $(F, a_i) \leq (F, a_j)$ if for every $X \in C_{(F, a_i)}$, there exist $Y \in C_{(F, a_j)}$, such that $X \subseteq Y$.

The generalized soft set-based domination based on its degree is presented as follow.

Definition 4. Let (F, A) be multi-soft sets over U representing $S = (U, A, V, f)$ and $(F, a_i), (F, a_j) \in (F, A)$. (F, a_j) is said to be dominated in degree k by (F, a_i), denoted $(F, a_i) \leq_k (F, a_j)$, where

$$k = \left|\bigcup X : X \subseteq Y\right| / |U| \tag{3}$$

and, $X \in C_{(F, a_i)}$ and $Y \in C_{(F, a_j)}$.

Table XVII. Matric results from file system courses 2011.

Attribute (wrt)	Domination degree T1	T2	T3	MT	FE	Maximum domination
T1	0.00	0.06	0.02	0.06	0.00	0.06
T2	**0.76**	**0.00**	**0.05**	**0.10**	**0.19**	**0.76**
T3	0.06	0.00	0.00	0.03	0.00	0.06
MT	0.00	0.00	0.05	0.00	0.00	0.05
FE	0.00	0.00	0.05	0.00	0.00	0.05

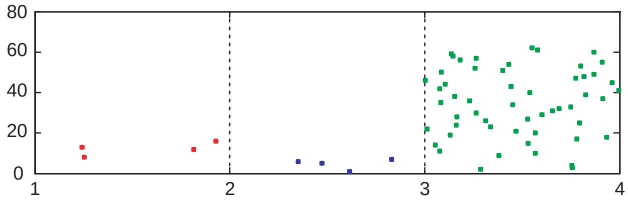

Fig. 15. Clustering visualization students on file system courses 2011.

Table XVIII. Matric results from file system courses 2012.

Attribute (wrt)	Domination degree T1	T2	T3	MT	FE	Maximum domination
T1	0.00	0.10	0.04	0.07	0.00	0.10
T2	**0.15**	**0.00**	**0.04**	**0.07**	**0.36**	**0.36**
T3	0.22	0.00	0.00	0.27	0.22	0.27
MT	0.00	0.00	0.04	0.00	0.00	0.04
FE	0.03	0.10	0.01	0.00	0.00	0.10

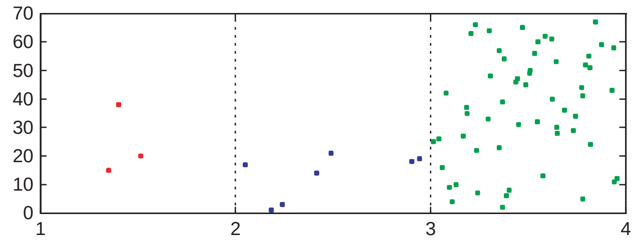

Fig. 16. Clustering visualization students on file system courses 2012.

Obviously $0 \leq k \leq 1$. If $k = 1$, then (F, a_i) is dominated totally by (F, a_j). Otherwise, (F, a_i) is dominated partially by (F, a_j).

3.3. MDDS

Let (F, A) be multi-soft sets over U representing $S = (U, A, V, f)$, based on Definition 5, the soft set (F, a_i) with maximum degree of domination will be selected as a clustering attribute i.e.,

$$\max\{k_1, k_2, \ldots, k_n\} \tag{4}$$

Proposition 2. *MAR and MDDS select the same clustering attribute.*

Table XIX. Matric results from file system courses 2013.

Attribute (wrt)	Domination degree T1	T2	T3	MT	FE	Maximum domination
T1	0.00	0.00	0.00	0.26	0.48	0.48
T2	**1.00**	**0.00**	**1.00**	**1.00**	**1.00**	**1.00**
T3	0.07	0.00	0.00	0.18	0.31	0.31
MT	0.03	0.00	0.00	0.00	0.00	0.03
FE	0.07	0.00	0.00	0.00	0.00	0.07

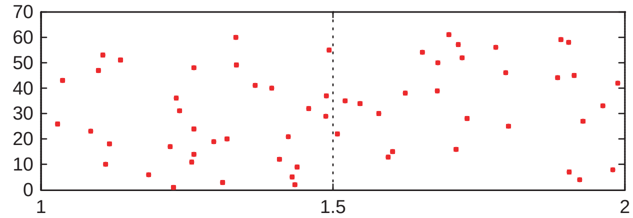

Fig. 17. Clustering visualization students on file system courses 2013.

RESEARCH ARTICLE

Table XX. Matric results from architectural design courses 2011.

Attribute (wrt)	Domination degree										Max. domination
	T1	T2	T3	T4	T5	T6	T7	T8	MT	FE	
T1	0.00	0.00	0.00	0.09	0.06	0.15	0.00	0.00	0.26	0.12	0.26
T2	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.12
T3	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.06
T4	0.06	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.15	0.00	0.15
T5	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.09
T6	0.00	0.00	0.00	0.18	0.26	0.00	0.00	0.00	0.00	0.12	0.26
T7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.12
T8	**0.00**	**0.00**	**0.00**	**0.18**	**0.44**	**0.26**	**0.00**	**0.00**	**0.00**	**0.00**	**0.44**
MT	0.06	0.00	0.00	0.18	0.06	0.00	0.00	0.09	0.00	0.12	0.18
FE	0.00	0.00	0.15	0.09	0.00	0.15	0.24	0.00	0.09	0.00	0.24

Proof. Let (F, A) be a multi-soft sets over U representing an information system $S = (U, A, V, f)$. Given $(F, a_h), (F, a_i), (F, a_j) \subseteq (F, A)$, where $h \neq i \neq j$.

Fact: MAR and MDDS select the same clustering attribute, say a_h. Logically, U/a_h will be the coarsest and un-balanced partition. To prove above proposition, we use indirect proof as follow.

Suppose MDDS selects another attribute say a_j, where $j \neq h$.

Thus, the value of k_j achieves the highest value, for $1 \leq j \leq |A|$. Let say the value achieved on (F, a_j), where $j \neq l$. Thus from (3), the cardinality $|\bigcup X: X \subseteq Y|$ i.e., the domination degree of (F, a_j) by (F, a_l) must achieve maximum (the highest). To this, from (4) U/a_j must be the coarser and un-balanced partition. This contradict with the fact that the coarsest un-balanced partition is on attribute a_h. The proof complete.

3.4. Algorithm

The proposed algorithm for the proposed technique is described in the following figure.

From Table II shows that the algorithms give the result, namely Attitude attribute was selected as the most dominant attribute.

In this example, there are two attributes, the splitting is on the attribute value which has the overall maximum domination versus the other attributes. The partition at this stage can be represented as a tree-spliting model and is shown in Figure 4.

4. DATASETS

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. The Data is taken from the judgment of the Department of information engineering. Algorithm courses consist of 99 students with 5 attributes, Software engineering courses consist of 260 students with 6 attributes, System Security courses consists of 269 students with 6 attributes, and the attribute file system consists of 190 students with 5 attributes. Other assessment Data are also taken from the Department of architecture, namely architectural design course which consists of 94 students with 10 attributes. All the data taken from University Technology Yogyakarta Indonesia for three years from 2011 until 2013. The assessment consists of several components (attribute) are different. Each of the majors and courses have not the same assessment criteria, all of it in the form of assignments, midterm and final exams. Midterms done in the middle of the semester is done in writing, the final exams is given at the end of the semester. Both are done on a scheduled basis. Students' name, age, race, and the force were not necessary in this assessment. Evaluation criteria used range from [0–100]. Nevertheless the data transformation in the data category. A value of 20 or less to 1, the value of 21–40 to 2, the value of 41–60 to 3, the values 61–80 into 4 and 81 more to 5. From these data later in the process to give weight or a certain percentage to get the final value in the form of A for the highest value to the E to the lowest value. The soft set theory is to classify and determine the most dominant attributes.

4.1. Algorithm Course

Assessment of algorithm courses has six components, namely task 1 to task 4, mid-term and final exams as shown in Table III.

4.2. Software Engineering Course

Assessment of Software Engineering courses has six components, namely task 1 to task 4, mid-term and final exams as shown in Table IV.

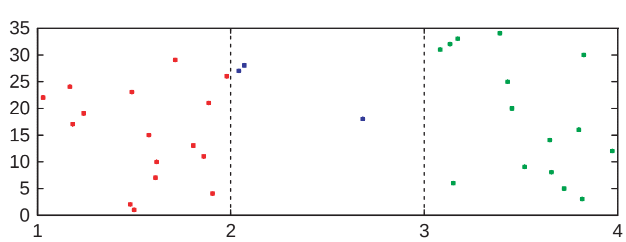

Fig. 18. Clustering visualization students on architectural design courses 2011.

Table XXI. Matric results from architectural design courses 2012.

Attribute (wrt)	Domination degree										Max. domination
	T1	T2	T3	T4	T5	T6	T7	T8	MT	FE	
T1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
T2	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.06
T3	0.00	0.00	0.00	0.39	0.00	0.00	0.00	0.00	0.00	0.00	0.39
T4	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15
T5	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06
T6	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06
T7	0.06	0.06	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.06
T8	0.00	0.09	0.00	0.00	0.06	0.12	0.00	0.00	0.00	0.15	0.15
MT	0.00	0.09	0.00	0.00	0.06	0.12	0.00	0.00	0.00	0.15	0.15
FE	**0.00**	**0.06**	**0.00**	**0.00**	**0.06**	**0.27**	**0.00**	**0.58**	**0.58**	**0.00**	**0.58**

4.3. System Security Course

Assessment of System Security courses has six components, namely task 1 to task 4, mid-term and final exams as shown in Table V.

4.4. File System Course

Assessment of File System courses has five components, namely task 1 to task 3, mid-term and final exams as shown in Table VI.

4.5. Architectural Design Course

Assessment of Architectural Design courses has ten components, namely task 1 to task 10, mid-term and final exams as shown in Table VII.

5. RESULTS AND DISCUSSION

The MDDS techniques will be implemented using Matlab programming language version R2009a under Windows 7 Home Edition operating system powered by Intel i3 processor with 4 GB memory. The user interface of the software is shown in Figure 5. The data used can use benchmarks data or other data. Software will provide the results of the calculation of the domination degree in the form of a matrix and clustering visualization.

Table VIII is a matrix that indicates the degree of dominance attribute value algorithm courses in 2011. The MDDS algorithm provides a T2 as the most dominant attributes compared to other attributes, whereas visualization is divided in two clusters as shown in Figure 6.

Table IX is a matrix that indicates the degree of dominance attribute value algorithm courses in 2012. The MDDS algorithm provides a T4 as the most dominant attributes compared to other attributes, whereas visualization is divided in three clusters as shown in Figure 7.

Table X is a matrix that indicates the degree of dominance attribute value algorithm courses in 2013. The MDDS algorithm provides a T4 as the most dominant attributes compared to other attributes, whereas visualization is divided in four clusters as shown in Figure 8.

Table XI is a matrix that indicates the degree of dominance attribute value software engineering courses in 2011. The MDDS algorithm provides a MT as the most dominant attributes compared to other attributes, whereas visualization is divided in five clusters as shown in Figure 9.

Table XII is a matrix that indicates the degree of dominance attribute value software engineering courses in 2012. The MDDS algorithm provides a T1 as the most dominant attributes compared to other attributes, whereas visualization is divided in four clusters as shown in Figure 10.

Table XIII is a matrix that indicates the degree of dominance attribute value software engineering courses in 2013. The MDDS algorithm provides a T1 as the most dominant attributes compared to other attributes, whereas visualization is divided in five clusters as shown in Figure 11.

Table XIV is a matrix that indicates the degree of dominance attribute value system security courses in 2011. The MDDS algorithm provides a T2 as the most dominant attributes compared to other attributes, whereas visualization is divided in four clusters as shown in Figure 12.

Table XV is a matrix that indicates the degree of dominance attribute value system security courses in 2012. The MDDS algorithm provides a T2 as the most dominant attributes compared to other attributes, whereas visualization is divided in two clusters as shown in Figure 13.

Table XVI is a matrix that indicates the degree of dominance attribute value system security courses in 2013. The MDDS algorithm provides a T2 as the most dominant attributes compared to other attributes, whereas visualization is divided in two clusters as shown in Figure 14.

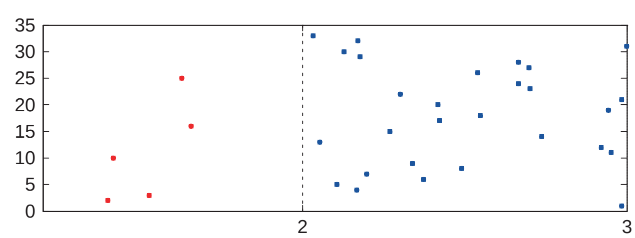

Fig. 19. Clustering visualization students on architectural design courses 2012.

RESEARCH ARTICLE

Table XXII. Matric results from architectural design courses 2013.

Attribute (wrt)	Domination degree										Max. domination
	T1	T2	T3	T4	T5	T6	T7	T8	MT	FE	
T1	**0.00**	**0.00**	**0.00**	**0.00**	**0.48**	**0.44**	**0.44**	**0.26**	**0.63**	**0.00**	**0.63**
T2	0.00	0.00	0.33	0.33	0.00	0.00	0.00	0.04	0.04	0.00	0.33
T3	0.00	0.52	0.00	0.00	0.00	0.00	0.00	0.04	0.04	0.00	0.52
T4	0.00	0.52	0.00	0.00	0.48	0.00	0.44	0.04	0.04	0.00	0.52
T5	0.19	0.00	0.00	0.33	0.00	0.00	0.00	0.15	0.41	0.00	0.41
T6	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.37	0.00	0.37
T7	0.19	0.00	0.00	0.33	0.00	0.00	0.00	0.15	0.26	0.00	0.33
T8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.04
MT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.00	0.00	0.15
FE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.44	0.00	0.44

Table XVII is a matrix that indicates the degree of dominance attribute value file system courses in 2011. The MDDS algorithm provides a T2 as the most dominant attributes compared to other attributes, whereas visualization is divided in three clusters as shown in Figure 15.

Table XVIII is a matrix that indicates the degree of dominance attribute value file system courses in 2012. The MDDS algorithm provides a T2 as the most dominant attributes compared to other attributes, whereas visualization is divided in three clusters as shown in Figure 16.

Table XIX is a matrix that indicates the degree of dominance attribute value file system courses in 2013. The MDDS algorithm provides a T2 as the most dominant attributes compared to other attributes, whereas visualization is divided in a clusters as shown in Figure 17.

Table XX is a matrix that indicates the degree of dominance attribute value architectural design courses in 2011. The MDDS algorithm provides a T8 as the most dominant attributes compared to other attributes, whereas visualization is divided in three clusters as shown in Figure 18.

Table XXI is a matrix that indicates the degree of dominance attribute value architectural design courses in 2012. The MDDS algorithm provides a FE as the most dominant attributes compared to other attributes, whereas visualization is divided in two clusters as shown in Figure 19.

Table XXII is a matrix that indicates the degree of dominance attribute value architectural design courses in 2013. The MDDS algorithm provides a T1 as the most dominant attributes compared to other attributes, whereas visualization is divided in two clusters as shown in Figure 20.

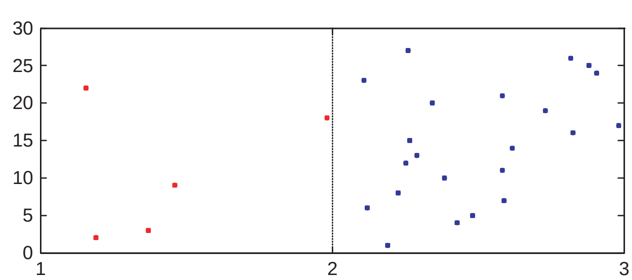

Fig. 20. Clustering visualization students on architectural design courses 2013.

6. CONCLUSION

Data clustering under soft set theory can be considered as a technique for data mining. In this paper, soft set theory has been used as an alternative technique for clustering attribute selection of a college student assessment data sets. The technique described in this paper is Maximum Degree of Domination in Soft set theory (MDDS). The proposed MDDS has been applied to select the best clustering attribute among all candidates in databases. To find meaningful clusters from a dataset, clustering attribute is conducted so that attributes within the clusters made will have a high correlation or high interdependence to each other while the attributes in other clusters are less correlated or more independent. In the experiments, datasets are taken from a survey on a few courses at the Information Engineering and the Architecture Departments of the University Technology Yogyakarta Indonesia during the last 3 years. In the experiments, we show how to determine the dominant attributes of a set of attributes of an assessment list data by using the MDDA technique. The results obtained has potentially contributed to give a recommendation in awarding the final grade of a course more quickly and accurately.

Acknowledgment: This work was supported under the research grant No Vote GRS 130326, Universiti Malaysia Pahang, Malaysia.

References

1. N. Petrovic, S. Isljamovic, and V. Jeremic, *Energy Education Science and Technology Part B: Social and Educational Studies* 5, 809 (**2013**).
2. E. Trandafili, et al., Discovery and evaluation of student's profiles with machine learning, *5th Balkan Conference in Informatics*, Novi Sad (**2012**).
3. B. R. Sachin and S. M. Vijay, A survey and future vision of data mining in educational field, *Second International Conference on Advanced Computing and Communication Technologies*, IEEE Computer Society (**2012**).

4. C. Romero, S. Ventura, and P. D. Bra, Knowledge Discovery with Genetic Programming for Providing Feedback to Courseware Author, Kluwer Academic Publisher **(2004)**.
5. D. Molodtsov, *Comput. Math. Appl.* 37, 19 **(1999)**.
6. D. Molodtsov, The Theory of Soft Sets, URSS Publishers, Moscow **(2004)**, (in Russian).
7. L. A. Zadeh, *Information and Control* 8, 338 **(1965)**.
8. Z. Pawlak, *International Journal Information and Computer Science* 11, 341 **(1982)**.
9. K. Atanassov, *Fuzzy Sets and Systems* 20, 87 **(1986)**.
10. W. L. Gau and D. J. Buehrer, *IEEE Transactions on System Man Cybernetic* 23, 610 **(1993)**.
11. M. B. Gorzalzany, *Fuzzy Sets and Systems* 21, 1 **(1987)**.
12. P. K. Maji, R. Biswas, and A. R. Roy, *Computers and Mathematics with Applications* 45, 555 **(2003)**.
13. M. I. Ali, F. Feng, X. Liu, W. K. Min, and M. Shabira, *Computers and Mathematics with Applications* 57, 1547 **(2009).**.
14. P. K. Maji and A. R. Roy, *Computers and Mathematics with Applications* 44, 1077 **(2002)**.
15. D. Chen, E. C. C. Tsang, D. S. Yeung, and X. Wang, *Computers and Mathematics with Applications* 49, 757 **(2005)**.
16. Z. Kong, L. Gao, L. Wang, and S. Li, *Computers and Mathematics with Applications* 56, 3029 **(2008)**.
17. X. Ma, S. Norrozila, H. Qin, T. Herawan, and J. M. Zain, *Computers and Mathematics with Applications* 62, 588 **(2011)**.
18. Y. Zou and Z. Xiao, *Knowledge Based System* 21, 941 **(2008)**.
19. H. Qin, X. Ma, J. M. Zain, and T. Herawan, *Knowledge Based System* 36, 139 **(2012)**.
20. Z. Xiao, K. Gong, and Y. Zou, *Journal of Computational and Applied Mathematics* 228, 326 **(2009)**.
21. T. Herawan and M. Mat Deris, *Knowledge Based Systems* 24, 186 **(2011)**.
22. H. Qin, X. Ma, T. Herawan, and J. M. Zain, *International Journal of Applied Mathematics and Computer Science* 22, 817 **(2012)**.
23. T. Herawan, M. M. Deris, and J. H. Abawajy, Matrices Representation of Multi Soft-Sets and Its Application, Lecture Notes in Computer Science, eidted by D. Taniar, et al., ICCSA **(2010)**, Vol. 6018, pp. 201–214.
24. H. Aktas and N. Cagman, *Inf. Sci.* 177, 2726 **(2007)**.
25. Y. B. Jun and C. H. Park, *Inf. Sci.* 178, 2466 **(2008)**.
26. F. Feng, Y. B. Jun, and X. Zhao, *Computers and Mathematics with Applications* 56, 2621 **(2008)**.
27. T. Herawan and M. Mat Deris, A direct proof of every rough set is a soft set, *Proceeding of the Third Asia International Conference on Modeling and Simulation, AMS'09*, IEEE Press **(2009)**, pp. 119–124.
28. R. Mamat, T. Herawan, and M. M. Deris, *To Appear in Knowledge-Based Systems* 1 **(2013)**.
29. Z. He, X. Xu, and S. Deng, A link clustering based approach for clustering categorical data, *Proceedings of the WAIM Conference* **(2004)**.
30. Z. Pawlak and A. Skowron, *International Journal of Computer and Information Science* 177, 3 **(2007)**.

Received: 8 July 2015. Accepted: 26 July 2015.

RESEARCH ARTICLE