Volume 4, Issue 1, January 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Attribute Reduction using Hybrid Extraction

Mohammed Adam TaheirMohammed,Wan Maseri Binti Wan Mohd, Ruzaini Bin Abdullah Arshah,
Faculty of Computer System & Software Engineering

Liu Yao
Faculty of Technology, Malaysia

Abstract -The major problems facing the decision making is how to handle uncertain data, several theories are dealing with uncertainty, soft set theory also handle this uncertainty problem. Reduction techniques are still an open area to be explored in knowledge management, which focuses on uncertain data removing unlike comparisons. This paper proposes based on rough set theory and soft set theory, it deletes the parameter, then execute Hybrid reduction technique which known as Hybrid Extraction for generating optimal, sub optimal until last optimal result. As part of this analysis, a comparison test with previous reduction techniques. The conclusion part the Hybrid Extraction shows better alternative results compared to Hybrid reduction and normal parameter reduction in terms of efficiency and response time.

Keywords: Boolean-valued information system; Extractions reductions; Parameters reductions; Knowledge Management.

I. Introduction

Handling uncertain data solved by using mathematical principles, and one of them is soft set theory [2]. Soft set is called (binary, basic, elementary) neighborhood systems. As for the standard soft set," it may be redefined as the classification of objects in two distinct classes, thus confirming that soft set can deal with a Boolean-valued information system". Molodtsov [2] advantages it is free from the inadequacy of the parameterization tools, unlike in the theories of fuzzy set, probability and interval mathematics. The main advantage of Knowledge Discovery is to find useful pieces within the data with little or no human relevant. KDD is the process of discovering useful knowledge from data, and data mining for generating extracting focuses on relevant data from a large amount of data which are available in distributed system which has lacked recourses. The additional steps in the KDD process, such as data preparation, data selection, data cleaning, incorporation of appropriate prior knowledge, and proper interpretation of the results of mining, are essential to ensure that useful knowledge is derived from the data [5]. Knowledge Discovery increased interest in learning beside discovery in unstructured semi-structured domains as Text Mining, Web Mining like graphs/networks (Link Analysis) and learning models of relational/first-order of Relational Data Mining and so on.

The rest of this paper is organized as follows. Section III presents an analysis of Hybrid reduction technique, followed by section IV, which is proposed technique, then section V focuses on result and discussions. Finally, the conclusion of this proposal is described in section VI.

II. Related work

Maji et al.[1] introduce definitions in the year 2002 for generating optimal and sub optimal decision because of huge of data which available are available in the distribution systems need to reduce these data amounts and store only significant one which are manage vague data or may has ambiguity. The Maji technique extracting the decisions which based on the maximum weight. It selects the sub decision based on maximum weights partions that any sub parameters have maximum optimal result partions will be a reduction, which has a lot of inconsistency and their sub extractions are not correct. Chen et al. technique [9] overcome the Maji problems in the year 2005 by for removing the inconsistency from it, it focuses only on optimal reductions but their sub extractions are not correct. Kong et al. [8] in the year 2008 defines techniques for solving Chen [9] problem, it introduces the definition of normal parameter reduction in soft set theory to overcome problems in Chen, but not take place at all in our data set. Rose et al. [10] in the year 2010 defines A framework of Decision Making Based on Maximal Supported Sets which reduce columns reduction, it overcomes Maji [1] problems which investigate that for every sub parameters has same optimal result partions, and sup A(u)=sup A(v) it has proper parameter reduction, therefore its problem has not rows (objects) reduction which affect the storage size by low reduction, it focuses only on columns (parameters). All previous techniques which mentioned above focused on column reductions, it ignores the row reduction to this Rose et al. [3] in the year 2011 applies row reduction and column reduction techniques, has parameter reduction based on technique [10] and it selects the object which not in the maximum result partion, it reduces the consistency data by proper data set expansion for sub optimal, next optimal with high quality of informations based on significant object. The key benefit of Hybrid reduction overcomes all techniques mentioned above drawbacks which provide a better result compared to all and overcome their problems which appear in result analysis [6,7].

In Hybrid reduction generates their parameter reduction based on grouping which satisfies their conduction to this we define Hybrid Extraction which are quick decision whether the object or parameter is a reduction or not.

III. Analysis of the Previous Works

Let a soft set (F,E) representing studies the communication prowess among selected university student. Let assume that there are eighteen students that has been surveyed in the universe U with $U=\{u_1,u_2,\cdots,u_{18}\}$, and E is a set of parameters representing communication facilities that is been used by the student surveyed, $E=\{p_1,p_2,p_3,p_4,p_5,p_6\}$, where p_1 stands for the parameter for using communication facilities such as "email", p_2 stands for the parameter "facebook", p_3 stands for the parameter "blog", p_4 stands for the parameter "friendsters", p_5 stands for the parameter "yahoo messenger" and lastly p_6 stands for the parameter "sms". Consider the mapping $F:E\to P(U)$ given by "student communication prowess $(\cdot)$", where $(\cdot)$ is to be filled in by one of parameters $p\in E$. Suppose that

As for example, $F(p_2)$ means communication by facebook is been used by and being represented functional value of $\{u_2,u_3,u_4,u_5,u_8,u_9,u_{10},u_{12},u_{13},u_{14},u_{15},u_{16},u_{17}\}$, while $F(p_4)$ means communication through friendsters with its functional value represented by $\{u_2,u_3,u_4,u_8,u_9,u_{10},u_{11},u_{13},u_{15},u_{16},u_{17},u_{18}\}$ as shown in table 2 [13].

Thus, the overall approximation can be represented as the following:

$$(F,E)=\left\{\begin{array}{c}\text{email}=\{u_1,u_2,u_3,u_4,u_8,u_9,u_{10},u_{11},u_{13},u_{14},u_{15},u_{16},u_{17}\},\\ \text{facebook}=\{u_2,u_3,u_4,u_5,u_8,u_9,u_{10},u_{12},u_{13},u_{14},u_{15},u_{16},u_{17}\},\\ \text{blog}=\{u_1,u_2,u_3,u_4,u_6,u_8,u_9,u_{10},u_{11},u_{13},u_{15},u_{16},u_{17},u_{18}\},\\ \text{friendsters}=\{u_2,u_3,u_4,u_8,u_9,u_{10},u_{12},u_{13},u_{15},u_{16},u_{17},u_{18}\},\\ \text{ym}=\{u_2,u_3,u_4,u_8,u_9,u_{10},u_{11},u_{13},u_{15},u_{16},u_{17},u_{18}\},\\ \text{sms}=\{u_2,u_3,u_4,u_8,u_9,u_{10},u_{13},u_{15},u_{16},u_{17}\}\end{array}\right\}$$

Figure 1. The soft set

The previous example can be represented in the following Boolean-valued information system.

Table 1. Tabular representation of a soft set

U/P	p_1	p_2	p_3	p_4	p_5	p_6	$f(.)$
u_1	1	0	1	0	0	0	2
u_2	1	1	1	1	1	1	6
u_3	1	1	1	1	1	1	6
u_4	1	1	1	1	1	1	6
u_5	0	1	0	0	1	0	2
u_6	0	0	1	0	1	0	2
u_7	0	0	0	0	1	0	1
u_8	1	1	1	1	1	1	6
u_9	1	1	1	1	1	1	6
u_{10}	1	1	1	1	1	1	6
u_{11}	1	0	1	0	1	0	3
u_{12}	0	1	0	1	0	0	2
u_{13}	1	1	1	1	1	1	6
u_{14}	1	1	0	0	0	0	2
u_{15}	1	1	1	1	1	1	6
u_{16}	1	1	1	1	1	1	6
u_{17}	1	1	1	1	1	1	6
u_{18}	0	0	1	1	1	0	3

A.Analysis of Hybrid Reduction in Soft Set Decision Making in Rose et al [3]

1-Input soft set (F,E) over universe U.
2- Determine co-occurrences of attributes for every object.
2- Calculate the support co-occurrences.
3-Arranging the support in decreasing order.
4- Rank the result until last result which determine based on support.
5- Determine the U/E clusters partions.
6- Determine any combinations of attributes are satisfaction the optimal decision partions and suppA(u)=sup A(v) for every u,v in U.
7- For any group of attributes satisfies procedure 6 determine the reduction.
8-Determine any row fulfill the definition of ultimate support set.
9- Delete the object partition which in ultimate support.
10-For any ultimate minimum support deletes the partions of inferior object.
11-If there is any ultimate minimum support set, mark the mark the object the inferior object.

12- Remove every row or columns which have empty objects (zero significant).

The parameter p6 is removed because having no object as shown in table 2.
The result of reduction of Hybrid reductions as shown in table 3 which reduced table data size up to 37%.

Table 2

UF	p_1	p_2	p_3	p_4	p_5	p_6	$f(.)$
u_1	1	0	1	0	0	0	2
u_5	0	1	0	0	1	0	2
u_6	0	0	1	0	1	0	2
u_7	0	0	0	0	1	0	1
u_{11}	1	0	1	0	1	0	3
u_{12}	0	1	0	1	0	0	2
u_{14}	1	1	0	0	0	0	2
u_{18}	0	0	1	1	1	0	3

Table 3

U / P	p_1	p_2	p_3	p_4	p_5	$f(.)$
u_1	1	0	1	0	0	2
u_5	0	1	0	0	1	2
u_6	0	0	1	0	1	2
u_7	0	0	0	0	1	1
u_{11}	1	0	1	0	1	3
u_{12}	0	1	0	1	0	2
u_{14}	1	1	0	0	0	2
u_{18}	0	0	1	1	1	3

IV. Propose technique

This technique, known as Hybrid Extraction which deletes every object has all parameters after that removed every parameter has empty object [3] or has all object [4] or their value are complemented [8] then delete every parameter if satisfied Hybrid reduction technique, then forwarded to delete parameter set, finally select every parameter not in deleted parameter set then execute Hybrid reduction technique.

A.The propose technique procedure:

1-Input soft set (F, E) over universe U.
2- Determine co-occurrences of attributes for every object.
2- Calculate the support co-occurrences.
3-Determine the order of supports by arranging in decreasing order.
4- Rank the result until the inferior object which determine based on support.
5- Determine the any object has all parameters (maximum ultimate support cluster).
6-Dlete the maximum ultimate support cluster.
7-Determine any parameter has empty object.
8- Determine any parameter has all object.
9- Delete every parameter has empty object.
10- Delete every parameter if has all objects.
11- Delete every parameter if satisfied Hybrid reduction technique, then forwarded to delete parameter set.
12-Repeat for second, third parameter until last parameter.
13-If any combination is deleted are satisfied the Hybrid reduction techniques, then remove the combination to delete parameter set.
14- Select every parameter not in deleted parameter set and execute Hybrid reduction for generating optimal, sub optimal result until the inferior object.
15-For any ultimate minimum support deletes the partions of inferior object.
16-If there is any ultimate minimum support set, mark the object the inferior object.
17-This techniques known as Hybrid Extraction for special reduction.

B. Propose analysis

First, it deletes the partion of the object $\{u_2,u_3,u_4,u_8,u_9,u_{10},u_{11},u_{13},u_{15},u_{16},u_{17},u_{18}\}$ as shown in table 4.

Second, it the parameter p6 because it has an empty object as shown in table 5.

Table 4

U/P	p_1	p_2	p_3	p_4	p_5	p_6	$f(.)$
u_1	1	0	1	0	0	0	2
u_5	0	1	0	0	1	0	2
u_6	0	0	1	0	1	0	2
u_7	0	0	0	0	1	0	1
u_{11}	1	0	1	0	1	0	3
u_{12}	0	1	0	1	0	0	2
u_{14}	1	1	0	0	0	0	2
u_{15}	0	0	1	1	1	0	3

Table 5

U/P	p_1	p_2	p_3	p_4	p_5	p_6	$f(.)$
u_1	1	0	1	0	0	0	2
u_5	0	1	0	0	1	0	2
u_6	0	0	1	0	1	0	2
u_7	0	0	0	0	1	0	1
u_{11}	1	0	1	0	1	0	3
u_{12}	0	1	0	1	0	0	2
u_{14}	1	1	0	0	0	0	2
u_{15}	0	0	1	1	1	0	3

V. Result and Discussions

In Hybrid reduction [3] based on Table 2, it reduces data set size up to 37%, it generates extraction parameters as well as next sub optimal object which reduces a number of parameters and objects in Boolean databases drastically but still being able to maintain consistency in decision making as shown Figure 2. Special extraction has an algorithm for removing every parameters and calculated its support then determine its support cluster if equal to the original data set support cluster then remove this parameter from expanding.

In [5] any combination of parameters has support cluster same as the original data set is a reduction to this technique not focuses on the object which has all parameters because if the number of parameters is large then it take time to decide whether these groups has an optimal result partion and has supA(u)=sup A(v). By using the propose technique which has a quick decision for parameter selection compared to Hybrid reduction. It shows the same result as hybrid reduction as shown in Figure 3, but has more efficiency when applied to large data set.

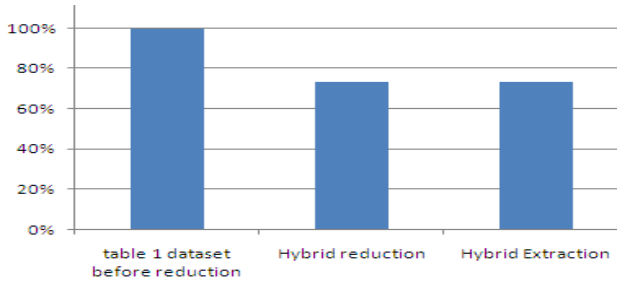

Figure 2

VI. Conclusion

This paper reduced data set size without loss of any original information using soft set theory. The propose techniques known as Hybrid Extraction which used for parameter reduction, it makes quick decision making based on the object which has all parameters after that we can decide for any parameter whether it belong to expanding result or not then the combination. It has efficient response time for large data set compared to Hybrid reduction, which has proper reduction unlike comparisons in Boolean databases. Hybrid Extraction reduce data set, but still being able to maintain consistency in decision making.

References

[1] P.K.Maji, A.R.Roy, and R Biswas 2002. An application of soft sets in a decision making problem. Computer and Mathematics with Application 44, 1077–1083.

[2] D.Molodtsov 1999. Soft set theory-first results. Computers and Mathematics with Applications 37, 19–31.

[3] A. N.M. Rose, M.I.Awang, H.Hassan, A.H.Zakaria, T.Herawan, and M.M.Deris 2011 , Hybrid Reduction in Soft Set Decision Making, Springer, pp. 108–115.

[4] D. A.Kumar and R. Rengasamy 2013. Parameterization Reduction Using Soft Set Theory for Better Decision Making. Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, February 21-22. 978-1-4673-5845-3/13©2013 IEEE.

[5] U.Fayyad, G.P-Shapiro, and P.Smyth 1996. From Data Mining to Knowledge Discovery in Databases. American ficial Intelligence. 0738-4602-1996.

[6] M.A.T.Mohammed,W.M.B.W.Mohd, R.B.A.Arshah, L.Yao 2013. Parameter Reduction Comparisons. Asian Academic Research Associates, vol.1 issue 19 January 2014.

[7] M.A.T.Mohammed,W.M.B.W.Mohd, R.B.A.Arshah, L.Yao 2013. Parameter Reduction using Special Order. Asian Academic Research Associates, vol.1 issue 19 January 2014.

[8] Z.Kong, L.Gao, L.Wang, and S.Li 2008. The normal parameter reduction of soft sets and its algorithm. Computers and Mathematics with Applications 56, 3029–3037.

[9]D. Chen, E.C.C Tsang, D. S Yeung, and X Wang 2005. The Parameterization Reduction of Soft Sets and its Applications. Computers and Mathematics with Applications . vol.49, pp.757–763.

[10] A.N.M.Rose, T.Herawan and M.M.Deris 2010. A framework of Decision Making Based on Maximal Supported Sets. ISNN 2010, Part I, LNCS 6063, pp. 473–482, 2010. © Springer-Verlag Berlin Heidelberg 2010.

