SOFT SET AND ROUGH SET APPROACH IN DECISION MAKING WITH UNCERTAINTIES

A.Jenneth[1], Dr .K. Thangavel [2]

ABSTRACT

The first part of this paper gives a general view of Knowledge Discovery in Database. The information system contains redundant attributes which will not be useful for Knowledge Discovery and may sometimes be misleading. The tools -Rough Set and the Soft Set, deal with uncertainty of the information system in data mining. In the second part, the basic definitions in Rough and Soft Sets are included and in the last, a decision making problem is addressed. The problem is presented in tabular form containing the information. Then, Reduct-soft-set is defined by applying reduction of the knowledge representation system in Rough set theory. An algorithm is applied to select the object with optimal choice. It uses a small number of parameters to determine the optimal objects for the problem of making decisions.

Keywords **:** Data mining, Soft Set, Reduct Soft Set, Rough Set, Knowledge Discovery in Databases

1. INTRODUCTION

Data mining refers to digging out or obtaining knowledge or concealed information from a huge amount of data. Data mining is treated as synonym for another popularly used term, Knowledge Discovery in Databases or KDD [1]. KDD is the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data. The objective of KDD is to get meaningful knowledge from the database. The key steps in Knowledge Discovery cycle are Data cleaning, Data integration, Data selection, Data transformation, Data mining, Pattern Evaluation and Knowledge presentation. Apart from Data mining there are some other commonly used techniques such as statistical methods, case-based reasoning (CBR), Bayesian Belief Networks (BBN), etc.,

1.1. SOFT SET

In a real life scenario, huge data encompassed in fields such as Economics, Medicine or engineering are not clear and understandable and contain uncertainties in several situations. But the tools for modeling, cognition and computation deal with problems that are certain in Economics, Medicine and so on. But three theories are there to deal with uncertainty: Probability theory, Fuzzy set theory and Theory of interval mathematics[2].

These theories have their own insufficiencies. The Theory of probability requires doing a huge lot of experiments in order to achieve check samples based on full samples. In reality such experiments on economy are difficult to carry out. The theory of interval mathematics considers calculation discrepancies and makes an interval estimation to achieve accurate results. But this theory is not capable of describing the

[1] Asst.Prof, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore.

[2] Professor, Department of Computer Science, Periyar University, Salem

smooth variation, uncertainty and conflict of the data. Theory of fuzzy set is very suitable for handling fuzzy data, which was proposed by L.A. Zadeh. The difficulty with fuzzy sets is establishing proper subject function, though it can be applied successfully in many fields. A uniform method is not there to make subject function, as the property of the subject function is utmost individuation.

The lack of the concept of uttering parameters is the major reason behind the said deficiencies. The uncertainty of parameters is the obstruction of applying these theories, since the tools for creating certain parameters are so deprived. D.Molodtstov formulated (1999) the theory of soft sets to deal with vivid, fuzzy and uncertain data [2].

1.2. ROUGH SET

Rough set which was introduced by Zdzislaw Pawlak in 1982 is used to analyze uncertain data. Rough set algorithm makes use of only internal information. It does not depend on any added model suppositions whereas Fuzzy set or probabilistic methods rely on these added assumptions. The rough step algorithm furnishes connotations and techniques to identify the attributes which distinguishes one class from the other. Since inconsistencies are permitted and affiliation in a set needs not to be absolute, the potential for treating noise elegantly is immense. The rough set approach provides results that shall be a set of rhetoric rules which could be simple for a human perception syntactically and semantically. Even though the complications of dynamic databases, time and memory constrictions are handled uniquely for each system, the time intricacy shall be large and possibly polynomial.

The common pattern in case of uncertain and incomplete data can be identified by the rough set methodology, based on the mathematical theory on the set approximation of the categorizing space [7].

Rough set theory was initially developed for a finite universe of discourse in which the knowledge base is a partition, which is obtained by any equivalence relation defined on the universe of discourse [3]. The data are formulated as a decision table with rows related to objects and columns to attributes. Class label in the data set is to identify the class of a row and is the decision attribute and the remaining attributes are condition attributes.

Rough Sets theory delineates three regions based on the equivalent classes induced by the attribute values: lower approximation, upper approximation and boundary [3]. While lower approximation region comprises the objects that can be categorized certainly, the upper approximation region comprises the objects that are categorized probably and the boundary is the variance between the other two regions [8].

Classification is a term which designates the sub-categorization of the universal set, which has all possible categories, into a number of distinct categories referred to as elementary sets [4].

Elementary sets can be supposed as controlling guidelines elucidating the objects of the classification.

Every object can be classified using the elementary set of parameters that cannot be split any more, even if other elementary sets of parameters present.

2. PRELIMINARIES

Definition of Soft set

Let us take U as the universal set and V as parameters set.

Definition: A pair (S, V) is called a soft set over U if S is a charting of V to the set of all subsets of U

$\longrightarrow$ P(U) S: V , where P(U) is the power set of U.

Soft set can be defined as a parameterized clique of subsets of the universal set. Each set S( ), from the clique may be treated as the set of elements of the soft set (S,V) , or as the set of approximate components of the soft set [5].

For more explanation the below example is taken.

(1) A soft set (S, V) pronounces the features of the computer to be bought whereas,

U = the set of computer models;

V = the set of parameters; each can be a descriptive text

V = {{expensive; cheap; high speed processor; normal speed processor; multimedia keyboard; LCD ; CRT; free annual maintenance contract}, and in this case defining a soft set means point out cheap, high speed processor and so on.

It can be noted that S( )may be blank for some V [5]

There is a fundamental difference between classical mathematics and soft set theory in formulating or describing any object. The mathematical model of an object would be constructed and the conception of exact solution is defined in case of classical mathematics. In many cases the exact solution cannot be arrived at, as this model is very intricate. Hence, to overcome this difficulty, the concept of approximate solution is introduced in the next step to arrive at the solution.

There is no restriction on the suitable description which makes the soft set theory comfortable and easy to use. The parameters can be formulated with any word, sentence or real numbers as preferred.

3. AN ILLUSTRATION OF THE TOOLS DEALING WITH UNCERTAINTY

Molodstov [9] presented some applications of the soft set theory in several directions viz., study of smoothness of functions game theory, operations research, Riemann-integration, Perron integration, probability, theory of measurement etc., [12].

Here, an application is presented which makes use of soft set theory with rough approach to solve a decision making problem.

Let set U = $\{c_1, c_2, c_3, c_4, c_5\}$,

Whereas c_1, c_2 … c_6 are the different models of computer system, and P is the set of parameters as given below,

P = {expensive; low-cost; high speed processor; normal speed processor; multimedia keyboard; LCD ; CRT; free AMC}.

We consider the soft set (S,P) which describes the features of the models, given by

(S,P)= {expensive =$\{c_1, c_4\}$, low-cost = $\{c_2, c_3, c_5, c_6\}$, high speed processor =$\{c_1, c_4, c_5, c_6\}$, normal speed processor = $\{c_2, c_3,\}$. multimedia key board= $\{c_1, c_4, c_5, c_6\}$, LCD= {c1,c2,,c4,c5,c6 }, CRT= {c3,,c6 }, free AMC = $\{c_1, c_2,, c_4, c_6\}$}.

Incase someone (Customer X) wants to buy a computer based on the parameters 'low-cost; high speed processor; multimedia keyboard; LCD; free AMC' etc., they formulate the subset

P1 = {low-cost; high speed processor; multimedia keyboard; LCD; free AMC} of the set P. Thus one needs to select the computer that has the most number of parameters of P, from the availabilities designated in U.

Let us consider another two persons (Customer Y and Customer Z), who want to get a computer with a model on the basis of their choice parameters P_2 P and P_3 P respectively.

The objective of the problem here is to find out the most suitable computer model with choice parameters of Customer X. The most suitable model for one customer may not be suitable for another as the final selection depends on the choice parameters of individual customers [5].

Some hypothetical depictions are done on soft set theory to unravel the problem as elucidated as follows.

3.1 Soft sets represented in tabular form[5]

Soft sets can be represented in an easily accessible manner in a tabular form. This was introduced by Lin[9] and Yao[10] previously. This is more or less an analogous representation. The (S,P) mentioned above is represented in tabular form as shown in Table 1.

Let us take the soft set (S,P) defined based on the choice parameters of customer X as mentioned above. It is represented as shown in Table 1.

If c_i S then $c_{ij} = 1$. otherwise $_{cij} = 0$ where $_{cij}$ are the entries in Table 1[5]

U	p_1	p_2	p_3	p_4	p_5
c_1	0	1	1	1	1
c_2	1	0	0	1	1
c_3	1	0	0	0	0
c_4	0	1	1	1	1
c_5	1	1	1	1	1

Thus it becomes a knowledge representation system, by replacing the attributes by parameters [5].

3.2 Creating Reduct-Table [5].

for any P1 P. (S, P1) is a soft subset of (S,P) which is mentioned above.

Reduct-soft -set is defined below:

If there is a set P11 which is a reduct of set P1, and (S,P1) is the soft set, (S,P11) is referred to as the reduct-soft-set of (S,P1) [5].

3.3 Algorithm for Selection of the Computer

1. Formulate the soft set (S,P) and give it as input
2. Frame the set of choice parameters (P1) of Customer X whereas P1 P
3. Make a list of reduct-soft-sets[5] of (S,P1).
4. Take one reduct-soft-set among all the reduct-soft-sets : (S,P11) of (S,P1).
5. Find out K, where Chk= max of chi, then ck is the optimal choice object and if k has more than one value , Customer X can choose any one of them .

This algorithm can be applied to the original problem as illustrated below.

It can be observed that $(p_1,p_3,p_4,p_5\}$, $\{ p_1,p_2,p_4,p_5\}$, $\{p_1,p_2,p_3,,p_5\}$ are the 3 reducts of

P1 = $\{ p_1,p_2,p_3,p_4,p_5 \}$. Choose any one say P11 = $\{ p_1,p_2,,p_4,p_5\}$.

By taking into account the choice values, the reduct -soft-set is given as in Table 2 shown below.

U	p_1	p_2	p_4	p_5	Choice value
c_1	0	1	1	1	C1 = 3
c_2	1	0	1	1	C2 = 3
c_3	1	0	0	0	C3 = 1
c_4	0	1	1	1	C4 = 3
c_5	1	1	1	0	C5 = 3
c_6	1	1	1	1	C6 = 4

Here max Ci = C6.

Decision: It is better for Customer X to buy the model m6.

4. CONCLUSION

In this paper, we studied and analyzed the Soft Set theory of Molodtsov [9] and Rough Technique of Pawlak [8] which propose a broad tool for dealing with uncertain data. We also presented an application of Soft Set theory using the rough technique of Pawlak [8] to handle a decision making problem.

REFERENCES

1. J. Han and M. Kamber, Data MINING: Concepts and techniques. Morgan Kaufman Publishers, 1995
2. Zhi Xiao, Ling Chen, Bo Zhong, Shijie Ye, "Recognition for soft information based on the theory of soft sets", Proceedings of ICSSSM '05. 2005 International Conference on Services Systems and Services Management, 2005.
3. P. Jaganathan. "Mammography Feature Selection using Rough set Theory", 2006 International Conference on Advanced Computing and Communications, 12/2006
4. www2.cs.uregina.ca
5. P.K. Maji, A.R. Roy, R. Biswas. "An application of soft sets in a decision making problem", Computers & Mathematics with Applications, 2002
6. [6] Sun, Xiao Lei. "A New Method of Fault Diagnosis with Soft Set Theory", Advanced Materials Research, 2011.
7. Z. Pawlak, Rough Sets, Theoretical aspects of reasoning about data, Kluwer Academic Publishers, 1991.
8. Z. Pawlak, "Rough Sets", International Journal of Computer and Information Sciences, vol. 11, no. 5, 1982, pp 341-356.
9. T. Y. Lin, Granular computing on binary relations 11. Rough set representations and belief functions, In Rough Sets in Knowledge Discovery,(Edited by A. Skoworn and L. Polkowski), pp 121 - 140, Springer Verlag, 1998.
10. Y. Y. Yao. Relational interpretations of neighbourhood operators and rough set approximation operators, Information Sciences 111(1-4), 239 -259, 1998.11. D . Molodtsov, Soft Set theory - first results, Computers Math. Applic., 37 (4/5), 19 - 31, 1999.