International Journal of Applied Engineering Research ISSN 0973-4562 Volume 15, Number 11 (2020) pp. 1052-1058
© Research India Publications. http://www.ripublication.com

Soft Set Theory-A Novel Soft Computing Tool for Data Mining

Santhosh Kottam [1*], Dr. Varghese Paul [2]

[1]*Research Scholar, R&D Centre, Bharathiyar University, Coimbatore, Tamilnadu – 641046, India.*

[2]*Professor & HOD, Department of Information Technology, CUSAT, Kochi, Kerala -682022, India.*

Abstract

Presently various soft computing techniques are applied in data mining. They have certain benefits in their workplace. Regularly, the volume of data repositories is changing. More than that, these data are not always precise and crisp in nature. There exists certain ambiguity and vagueness in the data. While the time of knowledge discovery, our computing tools have to consider these limitations. Due to the lack of parameterization tools, conventional soft computing tools cannot deal with these difficulties effectively. To handle these difficulties we tried to bring out the possibilities of a new soft computing tool-soft set theory- in the data mining. First two sections of this paper cover the introduction and preliminaries of soft computing. In the remaining sections, the authors discuss the theoretical aspects of soft set and multi-soft set. Finally, the paper ends with a discussion on possibilities of soft set theory in different data mining functionalities. Soft sets have a lot of use in the fields of business, health, education, agricultural, and many more.

Keywords: Soft Computing, Soft set, Multi-soft set, Fuzzy set, Rough set, Neural network, Genetic algorithms, Classification, Clustering, Frequent itemset.

I. INTRODUCTION

Powerful techniques and methods for mining useful information from huge data repositories have emerged throughout the recent decades. These techniques utilize the capacity of PC to find large volumes of information in a speedy and successful way. Though, the information to search is vague and badly affected with ambiguity. In the case of non homogeneous data repositories such as movie and plain text, the information might be uncertain and inconsistent.

Moreover interesting patterns and relationships are uncertain and not accurate. To make the data mining process more accurate and reliable, there should be effective methods for handling uncertainty, vagueness and errors. These methods should have approximate reasoning power and capacity for handling incomplete data. These methods are collectively known as soft computing. Soft computing aim is to utilize the patience for partial reasoning, uncertainty, vagueness and partial truth in order to achieve amenability, stability, and inexpensive solutions. Common soft computing methodologies - genetic algorithms, fuzzy sets, rough sets and neural networks - are most broadly used in the overall information discovery and data mining procedures.

We introduce a new soft computing tool –soft set theory- for handling ambiguity and uncertainty in various decision making problems. This concept is introduced by famous Russian mathematician D. Molodtsov. Soft set theory provides sufficient parameterization methods for dealing with data vagueness. In the soft set theory, we do not have to bring in the idea of an exact solution and the initial image of the object is approximate in nature [1]. Also we examine the possibilities of soft set theory in different data mining functionalities like classification and association analysis.

II. PRELIMINARIES

A. Soft Computing

One of the major drawbacks of conventional data mining technique is that complicated problems cannot be precisely explained by mathematical models, and therefore it is very difficult to produce good results from such problems. As mentioned in the introduction part, soft computing handles incomplete facts, ambiguity, and approximate reasoning to solve complicated problems. According to founder of fuzzy logic theory Dr Zadeh [2] stated that "the steering principle of soft computing is to make use of the tolerance for vagueness, uncertainty, and incomplete truth to achieve tractability, robustness, low solution cost, better relationship with reality". Zadeh brings up that soft computing is not a single technique, it consists of a number of techniques like- fuzzy logic, rough set, neural networks, and genetic algorithms. All these methods are not standing alone and competitive, but are supportive to each other and can be used collectively to find a solution for a given problem. We can conclude that soft computing objective is to solve complicated problems by using the ambiguity and uncertainty in decision making procedures [3][19]. A comparative picture of the conventional and soft computing

*__Corresponding author:__ Santhosh Kottam, Asst.Professor (Senior Grade), Dept. of Computer Applications, FISAT, Angamaly, Kerala-683577, India. Email: sankottam@gmail.com

based problem solving method is given in fig.1.

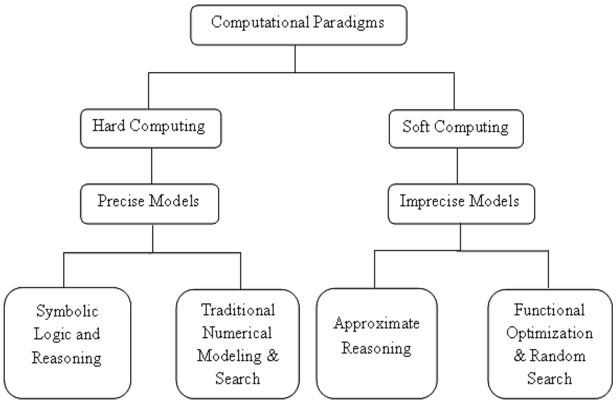

Fig. 1. Conventional and soft computing approaches

The left branch represents the conventional hard computing method where a precise model of the problem under analysis is obtainable and conventional mathematical techniques are used to find the solution. The right branch depicts a soft computing approach where only an approximate model of the problem may be obtainable, and the solution depends upon approximate reasoning methods. Soft computing has many applications like- handwriting recognition, automotive systems and manufacturing, image processing and data compression, architecture, decision-support systems and many more. Soft computing is expected to take a good role in science and engineering, but ultimately its power may extend much beyond. In future soft computing performs an important paradigm shift in the style of computing. A modification, which represents the reality that the human mind, dissimilar to current day PC, retains a noticeable capacity to keep and process data which is commonly uncertain, ambiguity and vagueness. Following techniques are commonly used in soft computing- fuzzy set, rough set, genetic algorithms, neural networks and soft set.

B. Fuzzy Sets

This theory has an alternative name known as possibility theory. In 1965, Dr. Lotfi Zadeh invented fuzzy set as an alternative solution for traditional binary-value logic and probability theory. It allows functioning at a maximum level of abstraction and offers a method for handling data, which possess uncertainty and ambiguity. Fuzzy set is different from traditional set theory. It assigns a membership scale for each object and value lies between 0 to1. It has wide application and use in data mining functionalities [4].

C. Neural Networks

A neural network is a set of linked input/output units in which each link has a weight related to it. It consists of a huge collection of extremely interconnected processing objects called neurons, working together to find solutions for precise problems. Neural networks are competent for extracting meaning from complex or vague data, which can be used to find patterns and perceive trends that are too multifaceted. Neural network widely used in following data mining functionalities like- rule generation, rule assessment, regression, and clustering [5].

D. Genetic Algorithms

In knowledge discovery, genetic algorithms are used to evaluate the strength of other algorithms. These are easy to implement and have been used in the following data mining functionalities- association rule analysis, classification, and other optimization problems. Genetic Algorithms can be considered as an evolutionary process with a collection of possible solutions, from which, solutions with higher degrees of fitness are selected. At each level of process, these chosen solutions undergo- crossover and mutation - to produce a candidate of the next generation. Crossover supports in the exchange of generated knowledge in the form of genes between individuals and mutation supports in restoring lost or unexplored regions in search place [6].

E. *Rough Sets*

Rough set theory can be used for classification to find out structural relationships within vague or noisy data. It applies to discrete-valued attributes and therefore, continuous-valued attributes must be discretized before their use. It evaluates a given model from below and from above, using *lower* and *upper* approximations. A rough set learning method can be used to obtain a set of rules in IF-THEN form, from a *decision table*. It provides an effective tool for extracting knowledge from databases. In rough set method, firstly creates a knowledge base, classifying objects and attributes within the created decision tables. Then, a data pre-processing step is initiated to remove some noisy data. Finally, the data dependency is assessed, in the reduced database, to find the minimal subset of attributes called reduct [5].

III. SOFT SET THEORY

Next we discuss a newly emerged soft computing tool - soft set theory. Theories like rough set, fuzzy set, and genetic algorithms are good mechanisms for dealing uncertainty and ambiguity. These theories have certain limitations caused by the insufficiency of parameterization. This fact is clearly true in the case of fuzzy set theory. Even though a fuzzy set is a good tool for handling uncertainty, there exists a difficulty like how to determine membership value for each object. There is no precise method for setting a membership value. We should not enforce a uniform method to set the membership value. The character of the membership value is particularly individual. For example, consider the statement $\mu_F(x) = 0.7$, everyone understands this statement in his own manner. So, the fuzzy set does not give a realistic approach for fixing a membership function. The basis for these limitations is, due to the insufficiency of the parameterization. Considering these facts, famous Russion Mathematician, Molodtsov contributed an idea of soft theory as a method for handling with uncertainties which is except from the above limitations. Soft set theory has wide applications in different domains.

A. *Definition (Soft Set)*

A pair (G, P) is called a soft set over V, where G is a mapping given by

$$G : P \rightarrow F(V)$$

In other words, a soft set over V is a parameterized family of subsets of the universe V. Every set $G(\varepsilon), \varepsilon \in P$ may be considered as the set of ε - elements of the soft set (G, P). A soft set is not a set.

Consider an example

V is the set of leaders considering an event.

P is the set of qualities of a good leader. Each parameter is a word or a sentence.

P= {Vision, Courage, Integrity, Humility, Strategic planning, Focus, Cooperation}

Here soft set (G, P) describes the "qualities of the leader" which an organization is going to measure. Suppose that there are six leaders in the batch V given by

V={l1,l2,l3,l4,l5,l6} and P={p_1,p_2,p_3,p_4,p_5,p_6,p_7}

where

p1 stands for the parameter 'Vision'

p_2 stands for the parameter 'Courage'

p_3 stands for the parameter 'Integrity'

p_4 stands for the parameter 'Humility'

p_5 stands for the parameter 'Strategic planning'

p_6 stands for the parameter 'Focus'

p_7 stands for the parameter 'Cooperation'

Suppose that

G (p_1) = {l2,l4}

G(p_2) = {l1, l3}

G(p_3) = {l3,l4, l5}

G(p_4) = {l1, l3, l5}

G(p_5) = {l1}

G(p_6) = {l2, l3, l4}

G(p_7) = {l2, l4, l6}

The soft set (G, P) is a parameterized family {G (p_i), i=1, 2, 3……, 7} of subsets of the set V and gives us a collection of approximate descriptions of an object. G (p_1) means "leaders (vision)" whose functional value is the set {l2, l4}. The soft set (G, P) is a collection of approximation as below:

(G, P) = {vision = {l2, l4}, courage= {l1, l3}, Integrity = {l3, l4, l5}, Humility = {l1, l3, l5}, Strategic planning= {l1}, Focus= {l2, l3, l4}, Cooperation= {l2, l4, l6}}

Where each approximation has two parts

(i) a predicate *p* eg: vision

(ii) an approximate value set *v* eg: {l2, l4}

Table I: Tabular representation of a soft set

U	Vision	Courage	Integrity	Humility	Strategic planning	Focus	Cooperation
l1	0	1	0	1	1	0	0
l2	1	0	0	0	0	1	1
l3	0	1	1	1	0	1	0
l4	1	0	1	0	0	1	1
l5	0	0	1	1	0	0	0
l6	0	0	0	0	0	0	1

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 15, Number 11 (2020) pp. 1052-1058
© Research India Publications. http://www.ripublication.com

B. Definition (*Operations with Soft Set)*

Consider a binary operator *, for subsets of the set V. Let (G, A) and (H, B) be soft sets over V. Then, the operation * for soft sets is defined as

$$(G, A) * (H, B) = (I, A \times B),$$

Where $I(\alpha,\beta) = G(\alpha) * H(\beta), \alpha \in A, \beta \in B,$ and $A \times B$ is the Cartesian product of the sets A and B.

C. Definition (Equality of Two Soft Sets)

Two soft sets (G, A) and (H, B) over a common universe V are said to be soft equal if (G, A) is a soft subset of (H, B) and (H, B) is a soft subset of (G, A).

D. *Definition* (*Not Set of a Set of Parameters)*

Let P = {p_1, pe_2, pe_3, ...pe_n} be a set of parameters. The Not set P denoted by ˥ P is defined by ˥ P = {˥p_1, ˥p_2, ˥p_3, ……., ˥p_n} where ˥p_i = not p_i, $\forall i$

E. Definition (*Complement Soft Set of a Soft Set*)

The complement of' a soft set (G: A) is denoted by $(G, A)^c$ and is defined by $(G, A)^c = (G^c$, ˥A) where G^c: ˥A $\rightarrow$ F(U) is a mapping given by $G^c(\alpha) = V - G$ (˥α), $\forall \alpha \in$ ˥A.

F. Definition 4.6 (Null Soft Set)

A soft set (*G*, *A*) over *V* is said to be a NULL soft set denoted by ϕ, if $\forall \varepsilon \in A, G(\varepsilon) = \phi$, *(null-set).*

G. Definition (Absolute Soft Set)

A soft set (G, A) over V is said to be absolute soft set denoted by $\tilde{A}, if \forall \varepsilon \in A, G(\varepsilon) = V$

H. Definition (And Operation on Two Soft Sets)

If (G, *A*) and (H, *B*) be two soft sets then (G, *A*) AND (H, *B*) denoted by (G, *A*) ∧ (H, *B*) and is defined by (G, *A*)(H, *B*) = (I, *A* × *B*) where I(*α*, *β*) = G(*α*) ∩ H(*β*) for all (*α*, *β*) ∈ *A* × *B*.

I. Definition (Or Operation on Two Soft Sets)

If (G, *A*) and (H, *B*) be two soft sets then (G, *A*) OR (H, *B*) denoted by (G, *A*) ∨ (H, *B*) is defined by (G, *A*) ∨ (H, *B*) = (I, *A* × *B*) where I(*α*, *β*) = G(*α*) ∪ H(*β*) for all (*α*, *β*) ∈ *A* × *B*.

IV. MULTI-SOFT SET THEORY

A. Decomposition of Multi-valued Information Systems

We decompose a multi-valued information system T=(V, B, W, g) into |*B*| number of binary-valued information systems. The decomposition of T=(V, B, W, g) is based on decomposition of B={$b_1, b_2, b_3, \ldots, b_{|B|}$} into the disjoint-singleton attribute {b_1 },{b_2 },… {$b_{|B|}$ } . Let T = (V, B,*W*,g) be an information system such that for every b∈B , W_b = *g(V,B*), is a finite non-empty set and for every *v* ∈*V* , |*f* (*v*, b)| =1. For every b_i under i^{th} -attribute consideration, b_i ∈B and w∈W_b , we define the map $b^i_w : V \rightarrow \{0,1\}$ such that $b^i_w(v) = 1$ if *g*(v, b) = w , otherwise $b^i_w(v) = 0$. The next result, we define a binary-valued information system as a quadruple T*'=(V,* b_i, $W_{\{0,1\}}$, *g).*The information systems T*'=(V,* b_i, $W_{\{0,1\}}$*,g), i=1,2,….*|*B*| is referred to as a decomposition of multi-valued information system *T* = (V, B,W, g) into |B| binary-valued information systems. Every information system T*'=(V,*b_i, W_{bi}*,g), i=1,2,….,*|*B*| is a deterministic information system since for every b∈ B and for every v ∈V ,|*g (v*,b)| = 1 such that the structure of a multi-valued information system and |B| number of binary-valued information systems give the same value of attribute related to objects[7].

Proposition 1. *If* (H, P) *is a soft set over the universe S, then* (H, P) *is a binary valued information system R=* (S, *A*,$V_{\{0,1\}}$, *h*).

Proof. Let (H, P) be a soft set over the universe *S*, we define a mapping

H ={h1 , h2 ,….. , hn} , where

$$h_i : S \rightarrow V_i \text{ and } h_i(x) = \begin{cases} 1, & x \in H(y_i) \\ & , \text{ for } 1 \leq i \leq n \\ 0, & x \notin H(y_i), \end{cases}$$

Hence, if A=P, V= $\bigcup_{e_i \in A} V_{e_i}$ where V_{e_i} = {0,1}, then a soft set (H, P) can be considered s a binary-valued information system R= (S, *A*,$V_{\{0,1\}}$, *h*).

Consider a multi valued information system R= (S, A, *V*, *h*) and R^i = (S, a_i, V_{ai}, *h*), ι=1,2,....|A| be the |A| b*inary*-valued information systems.

From *proposition 1.*

$$Q = (S, A, V, h) = \begin{cases} R^1 = (S, a1, V_{\{0,1\}}, h) \Longleftrightarrow (H, a_1) \\ R^2 = (S, a2, V_{\{0,1\}}, h) \Longleftrightarrow (H, a_2) \\ \quad . \qquad\qquad\qquad . \\ \quad . \qquad\qquad\qquad . \\ R^{|A|} = (S, a_{|A|}, V_{\{0,1\}}, h) \Longleftrightarrow (H, a_{|B|}) \end{cases}$$

$$= ((H,a_1),(H,a_2),\ldots..,(H,a_{|A|}))$$

We can conclude (H,P) = $((H,a_1),(H,a_2),\ldots..,(H,a_{|A|}))$ as a multi soft set over universe S representing a multi-valued information system R = (*S, A,V, h*).

V. APPLICATION OF SOFT SET THEORY IN DATA MINING

A. Decision Making

All kinds of decision making problems consist of following processes- formal modelling, reasoning and computing. Traditionally, these methods have been deterministic, crisp, and exact in character. In conventional mathematics, we build a mathematical replica of an item and define the concept of the precise solution of this replica. There are many complex problems existing in the real world that involve data which are not always all crisp. Usually the mathematical replica is too complex and we cannot find the precise solution. Nowadays, we apply different soft computing methods to overcome different challenges posed by data mining. The main techniques of soft computing include - fuzzy set, rough sets, neural networks, and genetic algorithms. These methods face challenges from uncertain data handling. Soft set theory contributes sufficient parameterization tools for handling data uncertainty. Here, the initial picture of the object has an approximate nature, and we do not need to set up the idea of precise solution [9-13].

B. Classification

Classification is one of the important data mining functions and used in many decision making problems. Simply, it is a process of finding the class label of an unknown object. Classification is a two-step process. In the first phase, we build up a model from the training data set and in the second step; the derived model is used for finding the class label of unknown object. In the first phase, we use efficient and scalable algorithms.

Generate a multi soft set from the given data set. Apply operations and properties of multi-soft set on generated soft set until all records are classified [17-18].

C. Frequent itemset mining

Association analysis is a widely used functionality in knowledge discovery. Nowadays, it has a good number of applications in academia, industry, and research scholars. For making decisions in all areas, an interesting pattern among the items is very important. Usually, we find out frequent item sets from huge data sets. Different algorithms and methods exist for finding frequent item sets[20-25]. Among them, Apriori and FP growth algorithms are commonly used. Apriori algorithm is well known for its simplicity. Since it is an iterative method, it will consume a huge amount of time to finish its execution. Similarly, FP-Growth algorithm emerged for eliminating the drawbacks of apriori algorithm. The generation of FP-growth algorithm requires a high volume of memory. Scanning the dataset twice reduces the full efficiency of FP- growth algorithm and produces a large number of FP-trees. Two overcome these limitations soft set theory can play a vital role [14-16]

D. Clustering

Clustering is one of the most prominent methods in data mining. Clustering is a process for identifying class and finding interesting correlations and patterns in the underlying data, which is needed in a number of data analysis tasks like - unsupervised classification and data summation, segmentation of large homogeneous data sets into smaller homogeneous, that can be separately modeled and analyzed. Rough set theory, introduced by Z. Pawlak in 1982, is a mathematical tool to deal with vagueness and uncertainty. One of the popular techniques in data clustering is based on rough set theory. The main approach of rough set-based data clustering is the clustering repository which is mapped as the decision table and this method can be executed by introducing a clustering attribute. Hence, from a database, we select only one attribute, which is the best way to partition the elements, is of primary importance for this approach. Currently, there have been researches in the area of implementing rough set theory in the selection process of clustering attribute. Relationship between the rough set and soft set theories are interesting. Combination of these two theories is used for determining a clustering attribute from a given set of attributes [8].

VI. CONCLUSION AND FUTURE WORK

Soft computing techniques have a good role in data mining. Many research works are going in this direction and produced various techniques which have dependable accuracy and computation performance. In data mining following soft computing techniques are frequently used for mining knowledge huge repositories – Fuzzy set, rough set, and soft set. Soft set theory is eliminating deficiencies of the above theories. Except for soft set theory, others face the problem of data uncertainty. Reason for these problems is due to the lack of parameterization tools. Soft set theory has sufficient parameterization techniques. This property automatically increased the popularity of soft sets. It has a good number of applications in data mining. We studied these applications and presented the possibilities of soft set theory in data mining.

In our research work, firstly, we presented soft computing and it's important in computation modelling. Our study went through following soft computing methods-fuzzy set, genetic algorithms, neural network and rough set. We tried to bring out limitations of these techniques. Next, we presented theoretical aspects of soft set and one of its branch multi-soft set theory. We discussed different definitions, propositions and properties of these theories. At the end, our paper concluded with a discussion on possibilities of soft set theory in data mining.

In the future, we would like to extend our study to find more opportunities for soft set theory in data mining.

REFERENCES

[1] D. Molodtsov," Soft set theory- First results", An International Journal of Computers & mathematics with applications", Elsevier, Volume 37, Issue 4-5, Page 19-31, February-March1999.

[2] Zadeh LA. Fuzzy logic, neural networks and soft computing. One-page course announcement of CS 294-4. Spring 1993. University of California at Berkeley; Nov. 1992.

[3] Zadeh LA. Fuzzy logic, neural networks, and soft computing. Communications of the ACM 1994; vol. 37. no. 3. pp. 77-84.

[4] L A Zadeh, Fuzzy sets, Inor. and Control 8, 338-353, 1965.

[5] Sushmita Mitra, Senior Member, IEEE, Sankar K. Pal, Fellow, *IEEE*, and Pabitra Mitra, Data Mining in Soft Computing Framework: A Survey, *IEEE* transactions on neural networks, vol. 13, no. 1, january 2002.

[6] Jiawei Han, Micheline Kamber, Data Mining: Concept and Techniques, Elsevier, second edition, 2006.

[7] Tutut Herawan, Mustafa Mat Deris “On Multi-soft Sets Construction in Information Systems”, 5th International Conference on Intelligent Computing, ICIC 2009, Ulsan, South Korea, September 16-19, 2009, Proceedings (pp.101-110).

[8] Hongwu Qin, Xiuqin Ma, Jasni Mohamad Zain, Tutut Herawan “A novel soft set approach in selecting clustering attribute”, Knowledge-Based Systems, 36 (2012) 139–145, Elsevier.

[9] P.K. Maji, A.R. Roy, R. Biswas, An application of soft sets in a decision making problem, Computers and Mathematics with Applications 44 (2002) 1077– 1083.

[10] D. Chen, E.C.C. Tsang, D.S. Yeung, X. Wang, The parameterization reduction of soft sets and its applications, Computers and Mathematics with Applications 49 (2005) 757–763.

[11] A.R. Roy, P.K. Maji, A fuzzy soft set theoretic approach to decision making problems, Journal of Computational and Applied Mathematics 203 (2007) 412–418.

[12] Y. Zou, Z. Xiao, Data analysis approaches of soft sets under incomplete information, Knowledge Based Systems 21 (2008) 941–945.

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

[14] Tutut Herawan, Mustafa Mat Deris,"A soft set approach for association rules mining",Elsevier,Knowledge-Based Systems 24 (2011) 186–195.

[15] R. Agrawal, R. Srikant, Fast algorithms for mining association rules, in: Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), 1994, pp. 487–499.

[16] R. Agrawal, T. Imielinski, A. Swami, Mining association rules between sets of items in large databases, in: Proceedings of the ACM SIGMOD International Conference on the Management of Data, 1993, pp. 207–216.

[17] https://medium.com/@dhiraj8899/top-5-advantages-and-disadvantages-of-decision-tree-algorithm-428ebd199d9a

[18] A.S.Aneeshkumara , Dr.C.Jothi Venkateswaranb, “Reverse sequential covering algorithm for medical Data mining”, Procedia Computer Science, Elseiver, December 2015.

[19] Santhosh Kottam, Varghese Paul “Uncertain Data Handling in Data Mining using Soft Set Theory” International Journal of Applied Engineering Research, ISSN 0973-4562 vol. 10 no.73 2015.

[20] Saravanan Suba, T. Christopher, “An Efficient Data Mining Method to Find Frequent Item Sets in Large Database Using TR- FCTM”, ictact journal on soft computing, January 2016, Volume: 06, Issue: 02.

[21] Prasanta Gogoi, Ranjan Das, B Borah, D K Bhattacharyya, “Efficient Rule Set Generation using Rough Set Theory for Classification of High Dimensional Data”, International Journal of Smart Sensors and Ad Hoc Networks (IJSSAN) ISSN No. 2248-9738 (Print) Volume-1, Issue-2, 2011, pp 13-20.

[22] Wai-Ho Au, Keith C C Chan, “FARM: A Data Mining System for Discovering Fuzzy Association Rules”, Conference: Fuzzy Systems Conference Proceedings, 1999. FUZZ-IEEE '99. 1999 IEEE International Volume:3.

[23] Amit Bhagat, Sanjay Sharma, K.R. Pardasani,”Ontological Frequent Patterns Mining by potential use of Neural Network”, International Journal of Computer Applications (0975–8887), Volume 36- No. 10, December 2011, pp 44-53.

[24] R. Vijaya Prakash, Govardhan, S. S.V.N. Sharma, “Mining Frequent Itemsets from Large Data Sets using Genetic Algorithms”, IJCA Special Issue on “Artificial Intelligence Techniques - Novel Approaches & Practical Applications”, AIT, 2011, pp 38-43.

[25] Jiye Li, Nick Cercone, “A Rough Set Based Model to Rank the Importance of Association Rules”, In Proceedings of 10th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, RSFDGrC 2005, LNAI Volume 3642, pp 109-118.

AUTHORS PROFILE

Mr. Santhosh Kottam completed his Master of Computer Applications (MCA) from Madras University, Tamilnadu and B.Sc mathematics degree from M G University, Kerala. He is currently pursuing PhD in computer science at Bharatiyar University, Coimbatore. His research area is data mining and has more than 19 years of teaching experience, which includes UG and PG. He has been serving Federal Institute of Science and

Technology (FISAT), Angamaly, Kerala as HOD & Assistant Professor (Senior Grade) in the Department of Computer Applications since May 2008. During this period of time, he has taught many subjects including Programming Languages, System Analysis and Design, Operating Systems, Object Oriented Modeling and design and Data Mining. He has published research papers in the International Journals, National and International Conferences. Mr. Santhosh received the best teacher award of the FISAT College during the year 2017 for his overall contribution.

Dr. Varghese Paul is completed B.Sc (Engg) in Electrical Engineering from Kerala University, M.Tech in Electronics and Ph.D in Computer Science from Cochin University of Science and Technology. Research Supervisor of Cochin University of Science and Technology, M G University Kottayam, Anna Technical University Chennai, Bharathiar University Coimbatore, Bharathidasan University Trichy and Karpagam University Coimbatore. Under the guidance, 29 research scholars had already completed research studies and degree awarded. Research areas are Data Security using Cryptography, Data Compression, Data Mining, Image Processing and E_Governance. Developed TDMRC Coding System for character representation in computer systems and encryption system using this unique coding system. Published many research papers in international as well as national journals and a text book also. Earlier worked as Industrial Engineer with O/E/N India Ltd Cochin, Communication Engineer with KSE Board, SCADA Engineer in Saudi Electricity Department, Head of IT Department CUSAT and Dean (CS, IT and Research) in Toc H Institute of Science and Technology. Certified Software Test Manager, Ministry of Information Technology, Govt of India. Life Member, Information System Audit and Control Association, USA (ISACA), Indian Society for Technical Education, India (ISTE) amd National Geographic Society, USA.