A STUDY ON COMPUTATIONAL INTELLIGENCE TECHNIQUES TO DATA MINING

Prof. S. Selvi [1,] R.Priya[2], V.Anitha[3] and V. Divya Bharathi[4]

[1, 2, 3,4]Department of Computer Engineering,
Government college of Engineering, Bargur, India.
[1]sel_raj241@gmail.com,[2]priya231213@gmail.com,
[3]anithacs180294@gmail.com
[4]divyashree0129@gmail.com

ABSTRACT

Nowadays rate of growth of data from various applications of resources is increasing exponentially. The collections of different data sets are formulated into Big Data. The data sets are so complex and large in volume. It is very difficult to handle with the existing Database Management tools. Soft computing is an emerging technique in the field of study of computational intelligence. It includes Fuzzy Logic, Neural Networks, Genetic Algorithm, Machine Learning and Rough Set Theory etc. Rough set theory is a tool which is used to derive knowledge from imprecise, imperfect and incomplete data. This paper presents an evaluation of rough set theory applications to data mining techniques. Some of the rough set based systems developed for data mining such as Generalized Distribution Table and Rough Set System (GDT-RS), Rough Sets with Heuristics (RSH), Rough Sets and Boolean Reasoning (RSBR), Map Reduce technique and Dynamic Data Mining etc. are analyzed. Models proposed and techniques employed in the above methods by the researchers are discussed.

KEYWORDS

Data Mining, GDT-RS, RSH, RSBR, Map Reduce, Dynamic Data Mining, Rough Set Theory.

1. INTRODUCTION

It is a great challenge to deal Big Data. The data sets are characterized in terms of huge volume in quantity, high variety in type or classification, velocity in terms of real time requirements and constant changes in data structure or user interpretation. Basically it is very tedious task to understand the data. Hence big data reflects into revolutionary change in research methodology as well as tools to be employed in various applications. The conventional database management tools which are present now are not suitable to big data processing applications. The challenges include Data Analytics, Data Access, Capture, Curation, Sharing, Storage, Search, Transfer and Visualization etc. Therefore we require Computational Intelligence to solve real world problems. Some of the computational intelligence techniques are Evolutionary Computing, Swarm Intelligence, Fuzzy Logic, Neural Network, Machine Learning, Genetic Algorithm and Rough Set Theory etc.

David C. Wyld et al. (Eds) : COSIT, DMIN, SIGL, CYBI, NMCT, AIAPP - 2014
pp. 247–259, 2014. © CS & IT-CSCP 2014 DOI : 10.5121/csit.2014.4924

Computational Intelligence or Soft Computing techniques are exploiting the tolerance of imprecision, uncertainty, Partial truth information. Due to inconsistencies it is very difficult to mine knowledge. The Rough Set Theory is a mathematical model proposed by Pawlak [1], [2] which deal with vagueness to a great extent. A rapid growth of interest in rough set theory and its applications is being seen now. It is one of the first non-statistical methods of data analysis.

The basic concept of rough set theory is the approximation of spaces. The subset of objects defined by lower approximation is the objects that are definitely part of the interest subset and the subset defined by upper approximation are the objects that will possibly part of the interest subset. The subset defined by the lower and upper approximation [3] is known as Rough Set. Rough set theory has evolved into a valuable tool used for representation of vague knowledge, identification of patterns, knowledge analysis and minimal data set.

In modern decision support systems, data mining is the most prevalent and powerful tool used for extraction of useful, implicit and previously not known information from large data bases. Many of these data mining tasks search for the frequently occurring interesting patterns. This is done by using machine learning techniques. Data mining is part of Knowledge Discovery in Databases (KDD) [4]. Data mining is a specific step in KDD that involve the application of algorithms for extracting hidden patterns from data. It is used to find knowledge in the form of rules that characterizes the property of data or relationships, patterns etc. The data mining systems are mostly designed using traditional machine learning techniques. Rough set theory is powerful data mining tool which is implemented to reduce data sets, to find hidden patterns and to generate decision rules. The main advantage of using rough set theory is that it does not need any preliminary information about data. Recently, Rough set theory finds an important place among the researchers in intelligent information systems.

Some of the applications in which the rough set theory is efficiently employed are in the areas of medicine, social networking, aerospace engineering, market analysis etc. This paper presents the rough set theory, basic concepts of data mining, and the techniques of data mining in which rough set theory is used to improve the performance of data mining. Basics of data mining are discussed in section 2. The rough set theory is presented in section 3. The various data mining techniques based on rough set theory presented in various research articles are discussed in section 4. The analysis of the techniques is also consolidated in this section. Finally the survey is concluded in section 5.

2. DATA MINING

Data mining is a process of querying and extracting useful information [5], hidden patterns and unknown data in a database.

Main goals of data mining for many organizations include detecting patterns to improve marketing capabilities, future predictions etc. Decision making is difficult when the size of data base increases a large and obtained from many sources and domains. Thus, consideration is also to be given for the integrity of data.

Partitioning data into groups, associating rule to data and ordering data are the main tasks of data mining. With ubiquitous computing infrastructure, volume of data is also increasing to a larger extent. Hence, it is very difficult to have manual analysis of data. Data mining is specifically uses techniques for extracting features from such database for decision making.

2.1 The knowledge discovery process

Knowledge Discovery process through data mining is divided into four: Selection, Pre-processing, Data Mining and Interpretation [5].

Selection is a process of creating a target data set. It is not that the entire data base is to undergo the data mining process, because of the fact that the data represents a number of different aspects of the unrelated domain. Hence, the very purpose of data mining is to be clearly specified.

Pre-processing is nothing but processing or preparing the data set that could be used for analysis by the data mining software. This further involves activities that resolve undesirable data characteristics like missing data, irrelevant non-variant fields and outlying data points. This pre-processing step results in generating a number of subsets of original set. All the data are converted into a format acceptable for data mining software. The above process of collection and manipulation of data in data mining process is called collection and cleaning.

Data mining is the process involved in analyzing cleaned data by mining software to obtain significant results such as hidden trends and patterns.

Data ⇨ Information⇨ Knowledge ⇨ Wisdom⇨ Intelligence

Fig.1. Activities of KDD Process

Fig.1 shows the activities of KDD process. The rapid growth in IT is because of the KDD process. The purpose of Data mining and knowledge discovery is to develop methodologies and tools for automating the data analysis and thus creating useful information and knowledge. This helps in decision making faster.

2.1.1 Steps of Data Mining:

The steps of data mining are: organizing data, determining desired outcomes, selecting tools, mining the data and pruning the result. The results show that only the useful ones are further considered.

2.1.2 Data Mining Technologies and Techniques:

Data mining process shown in fig.2 is an integration of different technologies such as database management, data warehousing, machine learning and visualization etc. Some of the submitted methods are traditional and established.

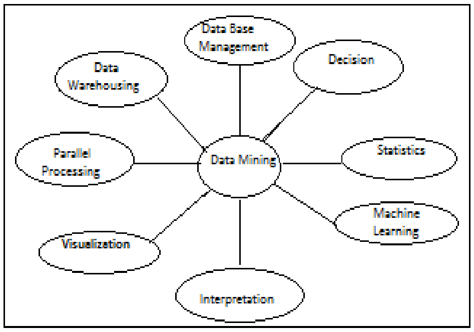

Fig.2. Data Mining Processes

2.1.3 Method Selection:

Some of the technical dimensions for the selection of data mining method, apart from the general considerations such as cost and support are given below:

- Uni-variate vs. multi-variate data.
- Numerical vs. categorical or mixed data.
- Explanation requirements or comprehensibility. Some tools give results, which are implicit to users (black box), while others can give causal and explicit representations.
- Fuzzy or precise patterns. There are methods such as decision trees, which only work with clear-cut definitions.
- Sample independence assumptions. Most methods assume independence of data patterns. If there are dependencies on the data patterns, it is necessary to remove or explore.
- Availability of prior knowledge. Some tools require prior knowledge, which might be not available. On the other hand, some others do not allow input of prior knowledge causing a waste of prior knowledge.

Other challenges come from lack of understanding of the domain problem and assumptions associated with individual techniques. Hence, data mining is not a single step. It requires multiple approaches to use some tools for the preparation of data.

3. ROUGH SET THEORY (RST)

Rough set theory, a new mathematical model developed by Pawlak in 1980s [1], [2], is an approach for imperfect or vague knowledge. This approach is used in the areas of knowledge acquisition, knowledge discovery, pattern recognition, machine learning and expert systems. Rough set theory provides means of identifying hidden patterns in data, finding minimal set of data, pointing out significant data, generating sets of decision rules from data. Rough set theory assumes that some form of information is associated with every object of universe. Objects are said to be indiscernible, if they are characterized by similar information. The mathematical basis of rough set theory is the indiscernibility relation exhibited by the above information.

The following are the listing of the concepts and research ideas of rough set theory:

Information System

Data for rough set based analysis are usually formatted into an *information system* IS= (U,A), where the set U is the universe of *objects* and the set A consists of *attributes*; any attribute $a \in A$ is a mapping from U into a value set V_a. Subsets of U are *concepts*.

Indiscernibility Relation

It is the main concept in rough set theory, and is considered as a similarity relation between two objects or more, where all the values are identical in relation to a subset of considered attributes. Indiscernibility relation IND (A) is an equivalence relation [3].

Elementary Set

A set of indiscernible objects is known as elementary set. A union of elementary sets is referred to as crisp set otherwise the set is rough set

.

Lower and Upper Approximation

Vague concepts, in contrast to precise concepts, cannot be characterized in terms of information about their elements. Therefore, in the proposed approach, we assume that any vague concept is replaced by a pair of precise concepts, called the lower and the upper approximation of the vague concept [6].

Reducts

For a set B of attributes, one can look after an inclusion-minimal set C ⊆ B with the property that IND(C) =IND (B), i.e., C is the minimal subset of attributes in B that provides the same classification of concepts as B. Such C is said to be a B-*reduct*.

Functional Dependence

For given A= (U,A), C,D ⊆ A, by C→D is denoted the *functional dependence* of D on C in A that holds iff IND(C) ⊆ IND(D). In particular, any B-reduct C determines functionally D. Also dependencies to a degree are considered [2].

3.1 Algorithms used

To obtain the decision rules from the decision table, the algorithms LEM2 [7], [8], Explore [9] and MODLEM [8] are utilized. LEM2, Explore and MODLEM algorithms for rule induction are defined briefly as follows. These algorithms are strong for both complete and incomplete decision table induction.

3.1.1 LEM2 Algorithm

LERS [7] (LEarning from examples using Rough Set) is a rule induction algorithm that uses rough set theory to handle inconsistent data set, LERS computes the lower approximation and the upper approximation for each decision concept. LEM2 algorithm of LERS induces a set of certain rules from the lower approximation, and a set of possible rules from the upper approximation. The procedure for inducing the rules is the same in both cases [10]. This algorithm covers all examples from the given approximation using a minimal set of rules [11].

3.1.2 MODLEM Algorithm

Preliminary discretization of numerical attributes is not required by MODLEM. The algorithm MODLEM handles these attributes during rule induction, when elementary conditions of a rule are created. MODLEM algorithm has two version called MODLEM-Entropy and MODLEM – Laplace. In general, MODLEM algorithm is analogous to LEM2. MODLEM also uses rough set theory to handle inconsistent examples and computes a single local covering for each approximation of the concept [10]. The search space for MODLEM is bigger than the search space for original LEM2. Consequently, rule sets induced by MODLEM are much simpler and stronger.

3.1.3 Explore Algorithm

Explore is a procedure that extracts from data all decision rules that satisfy requirements such as strength, level of discrimination, length of rules and conditions on the syntax of rules. It may also be adapted to handle inconsistent examples either by using rough set approach or by tuning a proper value of the discrimination level. Induction of rules is performed by exploring the rule space imposing restrictions reflecting these requirements. The main part of the algorithm is based on a breadth-first exploration which amounts to generating rules of increasing size, starting from one-condition rules. Exploration of a specific branch is stopped as soon as a rule satisfying the requirements is obtained or a stopping condition, reflecting the impossibility to fulfill the requirements, is met [11].

4. DATA MINING TECHNIQUES USING ROUGH SET THEORY

This section presents the various data mining techniques proposed by many researchers using the rough set theory.

4.1 Generalized distribution table and rough set system (GDT-RS)

GDT-RS is a soft hybrid induction system which helps to discover classification rules from databases with uncertain and incomplete data [12], [13]. The system is based on a hybridization of the Generalization Distribution Table (GDT) and the Rough Set (RS) methodology. The GDT-RS system can generate, from noisy and incomplete training data, a set of rules with the minimal (semi-minimal) description length, having large strength and covering all instances.

There are attributes, namely *condition* attributes and *decision* attributes (sometimes called class attributes) in a database. The condition attributes are used to describe possible instances in GDT, while the decision attributes correspond to concepts (classes) described in a rule. The GDT consists of three components: *possible instances*, *possible generalizations* of instances, and *probabilistic relationships* between possible instances and possible generalizations.

Possible instances are defined by all possible combinations of attribute values from a database. *Possible generalizations* of instances are all possible cases of generalization for all possible instances. The *probabilistic relationships* between possible instances and possible generalizations are defined by means of a probabilistic distribution describing the strength of the relationship between any possible instance and any possible generalization.

4.1.1 Simplification of the Decision Table by GDT-RS:

The process of rule discovery consists of the decision table preprocessing, including selection and extraction of the relevant attributes (features), and the appropriate decision rule generation. The relevant decision rules can be induced from the minimal rules (i.e. with the minimal length of their left-hand sides with respect to the discernibility between decisions) by tuning them (e.g. dropping some conditions to obtain more general rules which are better predisposed to classify new objects even if they do not classify properly some objects from the training set). The relevant rules can be induced from the set of all minimal rules, or from its subset covering the set of objects of a given decision table [14], [15]. A representative approach to the problem of generation of the so called local relative reducts of condition attributes is the one to represent knowledge to be preserved about the discernibility between objects by means of the discernibility functions.

It is obvious that by using the GDT one instance can be matched by several possible generalizations, and several instances can be generalized into one possible generalization. Simplifying a decision table by means of the GDT-RS system leads to a minimal (or sub-minimal) set of generalizations covering all instances. The main goal is to find a relevant (i.e. minimal or semi-minimal with respect to the description size) covering of instances still allowing us to resolve conflicts between different decision rules recognizing new objects. The first step in the GDT-RS system for decision rule generation is based on computing local relative reducts of condition attributes by means of the discernibility matrix method.

Relevant attributes are searched using bottom-up method instead of searching for dispensible attributes. Any generalization matching instances with different decisions should be checked by means of noise rate. If the noise level is smaller than a threshold value, such a generalization is regarded as a reasonable one. Otherwise, the generalization is contradictory.

Furthermore, a rule in the GDT-RS is selected according to its priority. The priority can be defined by the number of instances covered (matched) by a rule (i.e. the more instances are covered, the higher the priority is), by the number of attributes occurring on the left-hand side of the rule (i.e. the fewer attributes, the higher the priority is), or by the rule strength [12].

4.1.2 Searching Algorithm for an Optimal Set of Rules:

Searching algorithm developed by Dong *et al.* [13] for a set of rules and based on the GDT-RS methodology is outlined below.

Step 1. Create the GDT.
Step 2. Calculate the probabilities of generalizations.
Step 3. For any compound instance u' (such as the instance u'_1 in the above table), let $d\ (u')$ be the set of the decision classes to which the instances in u' belong.
Step 4. Using the idea of the discernibility matrix, create a discernibility vector (i.e. the row or the column with respect to u in the discernibility matrix) for u.
Step 5. Compute the entire local relative reducts for instance u by using the discernibility function.
Step 6. Construct rules from the local reducts for instance u, and revise the strength of each rule using (4).
Step 7. Select the best rules from the rules (for u) obtained in *Step 6* according to its priority [12].
Step 8. $U' = U' - \{u\}$. If $U' \neq \emptyset$;, then go back to *Step 4*. Otherwise, go to *Step 9*.
Step 9. If any rule selected in *Step 7* covers exactly one instance, then Stop, otherwise, repeat the above steps to select a minimal set of rules covering all instances in the decision table.

4.2 Rough sets with heuristics (RSH)

Rough set with heuristics (RSH) is a system that helps to select attribute subset [16]. The development of the RSH is based on the following observations: (i) a database always contains a lot of attributes that are redundant and not necessary for rule discovery; (ii) if these redundant attributes are not removed, not only does the time complexity of the rule discovery increase, but also the quality of the discovered rules can be significantly decreased. The goal of attribute selection is to find an optimal subset of attributes according to some criterion so that a classifier with the highest possible accuracy can be induced by an inductive learning algorithm using information about data available only from the subset of attributes.

4.2.1 Heuristic Algorithm for Feature Selection:

The attributes from database set called *CORE are* as an initial attribute subset. Next, attributes were selected one by one among the unselected ones using some strategies, and add them to the attribute subset until a reduct approximation is obtained.

Algorithm

Let R be a set of selected condition attributes, P a set of unselected condition attributes, U a set of all instances, and *EXPECT* an accuracy threshold. In the initial state, let $R = CORE(C)$, $P = C - CORE(C)$ and $k = 0$.

```
Step 1. Remove all consistent instances:
            $U = U - POS_R(D)$.
Step 2. If $k$ _ $EXPECT$ then STOP
            else if $POS_R(D) = POS_C(D)$,
            return `only k is available' and
            STOP
Step 3. Calculate $v_p$ , $m_p$.
Step 4. Choose the best attribute $p$, i.e. that with the largest $v_p \times m_p$, and set $R = R \cup \{p\}$, $P = P - \{p\}$
Step 5. Go back to Step 2.
```

4.3 Rough sets and boolean reasoning (RSBR)

RSBR is a system for discretization of real-valued attributes. Discretization of real valued attributes is an important preprocessing step in the rule discovery process. The development of RSBR is based on the following observations: (i) real-life data sets often contain mixed types of data such as real-valued, symbolic data, etc. (ii) real-valued attributes should be discretized in preprocessing (iii) the choice of the discretization method depends on the analyzed data.

The main module in the rule discovery process is the GDT-RS. In the GDT-RS, the probabilistic distribution between possible instances and possible generalizations depends on the number of the values of attributes. The rules induced without discretization are of low quality because they will usually not recognize new objects.

4.3.1 Discretization based on RSBR*:*

In order to solve the discretization problems, a discretization system called the RSBR was developed which is based on hybridization of rough sets and Boolean reasoning. [17], [18]

A great effort has been made [19], [20] to find effective methods of discretization of real-valued attributes. Different results may be obtained by using different discretization methods. The results of discretization affect directly the quality of the discovered rules. Some of discretization methods totally ignore the effect of the discretized attribute values on the performance of the induction algorithm. The RSBR combines discretization of real-valued attributes and classification. In the process of the discretization of real-valued attributes it should also take into account the effect of the discretization on the performance of the induction system GDT-RS.

Roughly speaking, the basic concepts of the discretization based on the RSBR can be summarized as follows: (i) discretization of a decision table, where $Vc = (v_{c,}w_c)$ is an interval of real values taken by attribute *c,* is a searching process for a partition P_c of V_c for any $c \in C$ satisfying some optimization criteria (like a minimal partition) while preserving some discernibility constraints [17, 18] (ii) any partition of V_c is defined by a sequence of the so-called *cuts* $v_1 < v_2 <$ ---$< v_k$ from V_c (iii) any family of partitions $\{P_c\}$ $c \in C$ can be identified with a set of cuts.

4.4 Map reduce method

Map Reduce Method is a programming model [21]. It is capable of processing large data sets called Big Data. It is well suited to handle in a distributed computing environment of real world tasks. The Map reduce computation is described in the fig.3 by two function as follows,

Map Function

It takes input pair (attribute/value) and produces set of intermediate attribute/value pairs.

Reduce Function

It accepts intermediate attribute/value pairs, merges to form smaller sets of values. Typically 0 or 1 output value produced per reduce invocation.

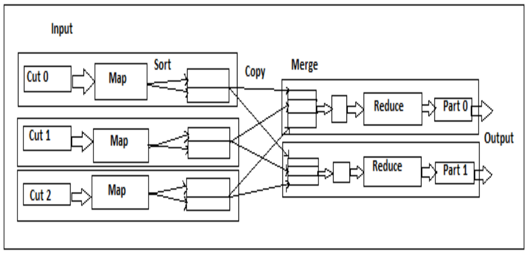

Fig.3. Map Reduce Programming Model

4.4.1 Applying Map Reduce Model to Compute Rough Set Approximation:

From the given Information System, the Universal set is first partitioned into a number of subsets [22]. From the subsets Equivalence Classes are obtained in a single step using Map Function. Now these equivalence classes are combined if it derives the same information set with respect to their conditional attributes. Similarly Equivalence classes of different sub decision classes / tables can be combined together if their information set is same. The above steps are executed in parallel.

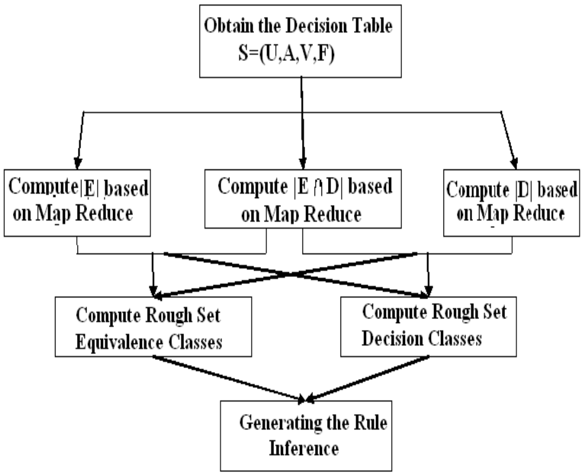
Obtain the Decision Table S=(U,A,V,F)
Compute |E| based on Map Reduce
Compute |E ∩ D| based on Map Reduce
Compute |D| based on Map Reduce
Compute Rough Set Equivalence Classes
Compute Rough Set Decision Classes
Generating the Rule Inference

Fig.4. An example of calculating Rough Set Approximations based on Map Reduce

As Equivalence (E) and Decision classes (D) are computed, construction of association between these two classes can be done in parallel as shown in fig. 4. If there is a relation between the two classes, then association exist, otherwise no. Then lower and upper approximation indexes are computed directly, leading to the calculation of lower and upper approximation.

These parallel methods run on different clusters namely Hadoop, Phoenix, and Twister [23]. Among these Twister is faster than other two systems and Hadoop is slower than other two. Users can decide which runtime system to be used in their application.

4.4.2 Dynamic Data Mining:

Variation in the Data plays a vital role in recent years. Hence it is necessary to dynamically update knowledge as given in fig.5. The knowledge updating takes place under the variation of object set, attribute set and attribute value.

Knowledge Updating due to Variation of Object Set

Here two situations arise, there may be single or multi object enters in/gets out of the information system [24]. Former is called Immigration and latter called Emigration of object. These processes reflect the refining of knowledge in neighborhood decision table.

The following steps help to understand the above said cases. Due to arrival of new single/multi objects, there may or may not be new decision classes generated.

Step 1: The decision classes are updated.
Step 2: Neighborhoods of Immigration object is computed.
Step 3: Neighborhoods of the universe are updated.
Step 4: Finally lower, upper approximation of the decision classes are updated.
Similar to Immigration, in the Emigration of Single/Multi objects also, there may or may not be deletion of existing decision classes. The steps involved are same as above.

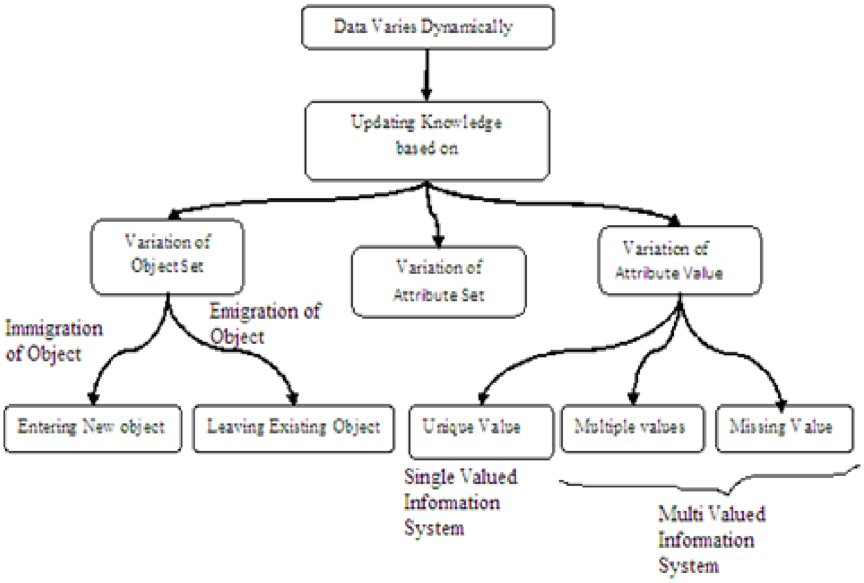

Fig.5. Dynamic Data Interpretation

Knowledge Updating due to Variation in Attribute Value-Matrix Based Approaches

Attribute value in the information system may be represented either by single value or multi value even sometimes may be missed [25]. If an attribute has unique value, then the information system is called single valued Information system. If an attribute has multi value, then the Information system is called multi valued Information system. If there is missing value, it can also be treated as multi valued Information system referred in fig.5.

Static / Dynamic Algorithm for updating approximations under adding / deleting objects steps as follows.

When the object enters in / gets out of the decision table,

Step 1: Relation matrix is updated.
Step 2: Induced Diagonal matrix is updated.
Step 3: Decision matrix is updated.
Step 4: Intermediate matrices or cut matrices are also updated.
Step 5: Lower and Upper approximations are generated.
Step 6: Finally probabilistic positive, boundary and negative approximations are updated.

Incremental Algorithm-Matrix Based Approaches

Static / Dynamic Algorithm [26] for updating approximations under adding / deleting objects is given below.
When the object enters in / gets out of the decision table,
Step 1: Relation matrix is updated.
Step 2: Induced Diagonal matrix is updated.
Step 3: Decision matrix is updated.
Step 4: Intermediate matrices or cut matrices are also updated.
Step 5: Finally lower and upper approximations are generated.

5. CONCLUSION

This paper attempts to bring out the concepts of rough set theory applied to data mining. Basic concepts of the data mining and the various steps normally employed were discussed. In traditional methods of decision making, scientific expertise in combination with some statistical methods is used to support the management. But these cannot be used to handle big data. This paper discusses the various techniques employed by the researchers in identifying the vagueness in data bases using rough set theory. The potential applications of rough set theory in data mining are reviewed in this paper. The concepts of many recent data mining techniques using rough set theory such as GDT-RS, RSH, RSBR, Map Reduce, Dynamic Data mining are also consolidated and presented in the above sections.

REFERENCES

[1] Pawlak, Z., “Rough Sets”, International Journal of Information and Computer Sciences, Vol. 11, pp. 341- 356, 1982.

[2] Pawlak, Z., “Rough Sets: Theoretical Aspects of Reasoning about Data”, Kluwer Academic Publishers, ISBN 0-79231472, Norwell-USA, 1991.

[3] Pawlak Z, Grzymala-Busse J, Slowinski R, and Ziarko W, “Rough Sets”, Communications of the ACM, Vol. 38, No. 11,pp 89-95,Nov 1995.

[4] Zarandi M.H.F, Kazemi A, “Application of Rough Set Theory in Data Mining for Decision Support Systems (DSSs)”, Journal of Industrial Engineering, Vol. 1, pp 25 – 34, 2008.

[5] Mert Bal, “Rough Sets Theory as Symbolic Data Mining Method”, An Application on Complete Decision Table Information Science Letters, Vol. 2, pp. 35-47, NSP Natural Sciences Publishing, 2013.

[6] Pawlak Z, “Rough Classification”, International Journal of Man-Machine Studies, Vol. 20, No. 5, pp. 469-483, 1984.

[7] Grzymala-Busse, J.W., “LERS-A System for Learning from Examples Based on Rough Sets”, Slowinski, R., (Ed.) Intelligent Decision Support Handbook of Application and Advances of the Rough Sets Theory, Kluwer Avademic Publishers, 1992.

[8] Stefanowski, J., “The Rough Set Based Rule Induction Technique for Classification Problems”, Proceedings of 6th European Conference on Intelligent Techniques and Soft Computing, EUFIT 98, Aachen, Germany, pp.109-113, 1998.

[9] Mienko R., Stefanowski, J., Taumi, K.& Vanderpooten, D., “Discovery-Oriented Induction of Decision Rules”, Cahier du Lamsade, No.141, Université Paris Dauphine,1996.

[10] Grzymala-Busse, J.W., Stefanowski, J., “Three Discretization Methods for Rule Induction”, International Journal of Intelligent Systems, Vol. 16, pp. 29-38, 2001.

[11] Stefanowski, J., Vanderpooten, D., “Induction of Decision Rules in Classification and Discovery-Oriented Perspectives”, International Journal of Intelligent Systems, Vol. 16, pp. 13-27, 2001.

[12] Zhong N., Dong J.Z. and Ohsuga S, “Data mining: A probabilistic rough set approach”, Rough Sets in Knowledge Discovery, Vol.2, Heidelberg, Physica-Verlag, pp.127-146, 1998.

[13] Dong J.Z., Zhong N. and Ohsuga S, “Probabilistic rough induction: The GDT-RS methodology and algorithms”, Foundations of Intelligent Systems, Berlin, Springer, pp.621-629, 1999.

[14] Komorowski J., Pawlak Z., Polkowski L. and Skowron A, “Rough sets: A tutorial, In Rough Fuzzy Hybridization”, A New Trend in Decision Making, Singapore, Springer, pp.3-98, 1999.

[15] Pawlak Z. and Skowron A, “A rough set approasch for decision rules generation”, Proceedings Workshop W12: The Management of Uncertainty in AI at 13th IJCAI, pp.1-19, 1993.

[16] Dong J.Z., Zhong N. and Ohsuga S, “Using rough sets with heuristics to feature selection”, New Directions in Rough Sets, Data Mining, Granular-Soft Computing, Berlin, Springer, pp.178-187, 1999.

[17] Nguyen H. Son and Skowron A, “Quantization of real value attributes”, Proceedings International Workshop Rough Sets and Soft Computing, 2nd Joint Conference on Information Sciences (JCIS'95), Durham, NC, pp.34-37, 1995.

[18] Nguyen H. Son and Skowron A, “Boolean reasoning for feature extraction problems”, Foundations of Intelligent Systems, Berlin, Springer, pp.117-126, 1997.

[19] Fayyad U.M. and Irani K.B. (1992) “On the handling of real-valued attributes in decision tree generation”, Machine Learning, Vol.8, pp.87-102. 1992.

[20] Nguyen H. Son and Nguyen S. Hoa, “Discretization methods in data mining”, Rough Sets in Knowledge Discovery, Heidelberg, Physica-Verlag, pp.451-482, 1998.

[21] Zhang J, Li T, Pan Y, “Parallel Rough Set Based Knowledge Acquisition Using Map Reduce from Big Data”, ACM, Beijing, China, August 12, 2012.

[22] Zhang J, Li T, Ruan D, Gao Z, Zhao C, “A parallel method for computing rough set approximations”, Information Sciences, Elsevier Inc, Vol.194, pp 209–223, 2012.

[23] Zhang J, Wong J.S, Lia T, Pan Y, “A comparison of parallel large-scale knowledge acquisition using rough set theory on different map reduce runtime systems”, International Journal of Approximate Reasoning, Elsevier, 2013.

[24] Zhang J, Li T, Ruan D, Liu D, “Neighborhood Rough Sets for Dynamic Data Mining”, International Journal of Intelligent Systems, Wiley Periodicals, Inc.,Vol. 27, pp 317–342, 2012.

[25] Zhanga J, Li T, Ruan D, Liud D, “Rough sets based matrix approaches with dynamic attribute variation in set-valued information systems”, International Journal of Approximate Reasoning, Elsevier Inc.,Vol. 53, pp 620–635, 2012.

[26] Zhang J, Li T, Chen H, “Composite rough sets for dynamic data mining”, Information Sciences, Elsevier Inc., Vol. 257, pp 81–100, 2013

AUTHORS

Selvi received her B.E., degree from Madras University, Chennai, India in 1998. She also received her M.E., degree from Anna University, Chennai, India in 2007. She is currently a PhD candidate at the Faculty of Information and Communication Engineering, Anna University Chennai, India. She has got 14 Years of Teaching Experience. Now she is working as Assistant Professor at the department of Computer Science and Engineering, Government College of Engineering, Bargur , Tamilnadu, India from 2013. She is interested in Data Mining, Cloud Computing, Soft Computing, Network Security and Wireless Sensor Network.

R.Priya,
B.E Final Year,
Department of Computer Science and Engineering,Government college of Engineering-Bargur,
Tamilnadu,
India.

V.Anitha,
B.E Final Year,
Department of Computer Science and Engineering,
Government college of Engineering-Bargur,
Tamilnadu,
India.

V.Divya Bharathi,
B.E Final Year,
Department of Computer Science and Engineering,
Government college of Engi neering-Bargur,
Tamilnadu,
India.
.