Analysis and Comparative Study of Classifiers for Relational Data Mining

Vimalkumar B. Vaghela
Ph.D. Scholar,
Department of Computer Science & Engineering,
Karpagam University,
Coimbatore,
Tamilnadu, India

Kalpesh H. Vandra
Principal,
C. U. Shah College of Engineering & Technology,
Surendranagar,
Gujarat, India

Nilesh K. Modi
Professor & Head of Computer Science Department,
S V Campus,
Kadi,
Gujarat, India

ABSTRACT
As an important task of relational database, relational classification can directly classify the data that involve multiple relations from a relational database and have more advantages than propositional data mining approaches. The information age has provided us with huge data repositories which cannot longer be analyzed manually. Most available existing data mining algorithms looks for pattern in a single relation. To classify data from relational database need of multi-relational classification arise which is used to analyze relational database and used to predict behavior and unknown pattern automatically which include business data, bioinformatics, pharmacology, web mining, credit card fraud detection, disease diagnosis system, computational biology, online retailers. In this paper, we present the several kinds of multi-relational classification methods including Inductive Logic Programming (ILP) based, Associative based multi-relational classification, Emerging Patterns based, Relational database based classification approaches and discuss each relational classification approaches, their characteristics, their comparisons and challenging issues in detail.

Keywords
Relational data mining, Multi-relational classification, Inductive Logic Programming, Tuple ID Propagation, Selection Graph, Decision Tree.

1. INTRODUCTION
Relational databases [2] [11] are the most popular repository for structured data, and are thus one of the richest sources of knowledge in the world. In a relational database, multiple relations are linked together via entity-relationship links. Unfortunately, most existing data mining approaches can only handle data stored in single relation, and cannot be applied to relational databases. Propositionalization [11] [20] may cause many problems, such as lose information of linkages and relationship, or cause statistical skew. Therefore, it is need to design data mining approaches that can discover knowledge from multi-relational data. Multi-relational databases can often provide much richer information for data mining, and thus multi-relational data mining approaches can often achieve better performance than single relation. Multi-relational data mining faces two major challenges. First, it is much more difficult to model multi-relational data. Unlike records in a single relation which can be modeled by vectors, multi-relational data contains heterogeneous objects and relationships among them, and there has not been widely accepted model for mining such relations. Second, in many data mining approaches (e.g., classification) aim at finding a model (or hypothesis) that fits the data. In a relational database, the number of possible model is much larger than that in single relation. Multi-relational data mining (MRDM) [2] [11] [20] aims to discover useful patterns across multiple relations without joining data of multiple relations into a single relation.

2. MULTI-RELATIONAL CLASSIFICATION
The important task of MRDM is multi-relational classification which aims to build a classification model that utilizes information in multiple relations. Multi-relational classification need not to transform multi-relations into a single relation, which effectively avoid these problems of relational informational loss, statistical skew and efficiency reducing that often happen in propositional or attribute-value classification approaches. A database for multi-relational classification consists of a set of relations, one of which is the target relation R_t, whose tuples are called target tuples and are associated with class labels. The other relations are non-target relations. In relational database, a relation can be defined as $r = (A^r, K^r, FK^r)$ where $A^r = \{A^r_1, \ldots, A^r_n\}$ is set of attributes, $K^r \subseteq A^r$ represents the primary key of the relation and $FK^r = \{FK^r_1, \ldots, FK^r_m\}$ is the set of foreign keys in r. Each foreign key can be defined as $FK^r_i = (F^r_i, s_i)$ where $F^r_i \subseteq A^r$ and s_i is the relation whose primary key K^{si} is referenced by FK^r_i. Based on the available related work multi-relational classification (MRC) divides in main four categories: 1) Inductive Logic Programming based MRC 2) Associative based MRC 3) Relational Database based MRC 4) Emerging Patterns based MRC. These approaches are described briefly in next section.

3. INDUCTIVE LOGIC PROGRAMMING (ILP) BASED RELATIONAL CLASSIFICATION
As its name indicates, ILP [29] [38] [41] [42] is situated at the intersection of two important areas of Computer Science: Induction that is one of the main techniques used in several Machine Learning algorithms to produce models that generalize beyond specific instances and Logic Programming which is the programming paradigm that uses first order logic to represent relations between objects and implements

deductive reasoning. The main representative of this paradigm is Prolog.

ILP is the intersection of machine learning and logic programming and is characterized by the use of logic for the representation [37] [38] of multi-relational data. The core of ILP is the use of logic for representation and the search for syntactically legal hypotheses constructed from predicates provided by the background knowledge. In ILP systems [42] [47], the training examples, background knowledge and induced hypothesis are all expressed in a logic program form. Two measures are used to test the quality of the induced theory. After learning, the theory with background knowledge should cover all positive examples (completeness) and should not cover any negative examples (consistency). Completeness and consistency together form correctness.

In ILP [41], a system often starts with an initial pre-processing phase and ends with a post-processing phase. In pre-processing phase, error (noise) in the given examples can be detected and eliminated. In post-processing phase, redundant clauses in the induced theory are removed in order to improve its efficiency. There are two approaches for the search direction: top-down and bottom-up. ILP is the study of learning methods for data and rules that are represented in first-order predicate logic. Predicate logic allows for quantified variables and relations and can represent concepts that are not expressible using examples described as feature vectors. A relational database can be easily translated into first-order logic and be used as a source of data for ILP. The goal of *inductive logic programming* (ILP) is to infer rules of this sort given a database of background facts and logical definitions of other relations.

3.1 Well-known ILP-based system

3.1.1 Bottom-up System

3.1.1.1 GOLEM

Golem [38] [41] [42] [47] is an inductive logic programming algorithm developed by Stephen Muggleton and Feng. It uses the technique relative least general generalization proposed by Gordon Plotkin. Therefore, only positive examples are used and the search is bottom-up. Negative examples can be used to reduce the size of the hypothesis by deleting useless literals from the body clause. In order to generate a single clause, GOLEM first randomly picks several pairs of positive examples, computes their rlggs and chooses the one with greatest coverage. If the final clause does not cover all positives, the covering approach will be applied. The covered positives are removed from the input and the algorithm will be applied to the remaining positives (Lavrač & Džeroski, 1994; Muggleton & Feng, 1990).

3.1.1.2 CIGOL

CIGOL [29] [37] [41] [42] (logic backwards) is interactive bottom-up relational ILP system based on inverse resolution. CIGOL employs three generalization operators which are relational upgrades of absorption, intra-construction and truncation operators. The basic idea is to invert the resolution rule of deductive inference using the generalization operator based on inverse substitution. CIGOL uses the absorption operator. However, CIGOL also needs oracle knowledge to direct the induction process.

3.1.2 Top-Down System

3.1.2.1 LINUS

LINUS [38] [41] [42] [47] is one of the most popular attribute-value learning environments in the ILP history. LINUS is a framework that reduces the relational learning problem into a propositional one, employs an attribute-value learning method and transforms the solution hypothesis into relational form. It is a non-interactive ILP system, integrating several ILP attribute-value learning algorithm in a single environment. It can be viewed as a toolkit, in which one or more of the algorithm can be selected in order to find the best solution for the input. The main algorithm behind LINUS consists of three steps. In the first step, the learning problem is transformed from relation to attribute-value form. In the second step, the transformed learning problem is solved by an attribute-value learning method. In the final step, the induced hypothesis is transformed back into relational form.

3.1.2.2 MIS

Model Inference System (MIS) [29] [38] [41] [47] is an interactive top-down relational ILP system, which uses refinement graph in the search process (Shapiro, 1983). In its algorithm, at the beginning the hypothesis is empty (H = Φ).Then it reads the examples (either positive or negative) one by one. If the example is negative and covered by some clauses in the Hypothesis set, then incorrect clauses are removed from the solution set. If the example is positive and it is not covered by any clause in the solution set, with breadth-first search, a clause c, which covers the example [23], is developed and added to solution set. The process will continue until the solution set (H) becomes complete and consistent (Lavrač & Džeroski, 1994).

3.1.2.3 FOIL

First-Order Inductive Learner (FOIL) [20] [29] [37] [41] [42] is a non-interactive top-down relational ILP system, which uses refinement graph in the search process as in MIS. It uses the covering approach for the solution having more than one clause. FOIL is a sequential covering algorithm that builds rules one at a time. After building a rule, all positive target tuples satisfying that rule are removed and FOIL will focus on tuples that have not covered by any rule. When building each rule, predicates are added one by one. At each step, every possible predicate is evaluated, and the best one is appended to the current rule. FOIL chooses the clause according to weighted information gain criteria.

3.1.2.4 PROGOL

PROGOL [38] [42] is a top-down relational ILP system, which is based on inverse entailment (Muggleton, 1995; Muggleton & Tamaddoni-Nezhad, 2008). It performs a search through the refinement graph. Besides a definite program B as background knowledge and a set of ground facts E as examples, PROGOL requires a set of mode declarations for reducing the hypothesis space.

3.1.2.5 WARMR

Design of algorithms for frequent pattern discovery has become a popular topic in data mining. Almost all algorithms have the same of level-wise search known as APRIORI algorithm (Agrawal, Mannila, Srikant, Toivonen, & Verkamo, 1996). The level-wise algorithm is based on a breadth-first search in the lattice spanned by a specialization relation between patterns (Dehaspe & Raedt, 1997; Dehaspe & Toivonen, 2001). The APRIORI method looks at a level of the lattice at a time. It starts from the most general pattern. It iterates between candidate generation and candidate evaluation phases. In candidate generation, the lattice is used for pruning non-frequent patterns from the next level. In candidate evaluation, frequencies of candidates are computed with respect to database. Pruning is based on the voting

system criteria property with respect to frequency: if a pattern is not frequent then none of its specializations are frequent.

3.1.3 Decision trees relational classification approaches

Decision trees are trees that classify instances by sorting them based on feature values. Each node in a tree represents a feature in an instance to be classified, and each branch represents a value that the node can assume. Instances are classified starting at the root of node and sorted based on their feature values. The Decision tree construction does not require any domain knowledge and is appropriate for exploratory knowledge discovery. In general decision tree classifiers have good accuracy. There are two major classification algorithms for inducing relational decision trees (SCART[5] and TILDE[28]) upgraded from the two most famous algorithms for inducting propositional decision trees (CART and C4.5).

3.1.3.1 Top-down induction of first-order logical decision trees (TILDE)

Top-down induction of decision trees (TDIDT) is the best known and most successful machine learning technique. It has been used to solve numerous practical problems. It employs a divide-and-conquer strategy, and in this it differs from its rule based competitors (e.g., AQ, CN2), which are based on covering strategies. Within attribute-value learning (or propositional concept-learning) TDIDT is more popular than the covering approach. Yet, within first-order approaches to concept-learning, where only a few learning systems have made use of decision tree techniques. The main reason why divide-and-conquer approaches are not yet so popular within first-order learning, which lies in the discrepancies between the clausal representations employed within inductive logic programming and the structure underlying a decision tree.

First-order logical decision trees:

A first-order logical decision tree (FOLDT) is a binary decision tree in which (1) the nodes of the tree contain a conjunction of literals, and (2) different nodes may share variables, under the following restriction: a variable that is introduced in a node (which means that it does not occur in higher nodes) must not occur in the right branch of that node.

3.1.3.2 Structural Classification and Regression Trees (SCART)

SCART is capable of inducing first-order trees for both classification and regression problems, i.e., for the prediction of either discrete classes or numerical values. This algorithm is upgraded from a propositional induction algorithm and turns it into a relational learner by devising suitable extensions of the representation language and the associated algorithms. In particular, it is upgraded CART, the classical method for learning classification and regression trees, to handle relational examples and background knowledge. The system constructs a tree containing a literal (an atomic formula or its negation) or a conjunction of literals in each node, and assigns either a discrete class or a numerical value to each leaf. In addition, it is extended the CART methodology by adding linear regression models to the leaves of the trees; this does not have a counterpart in CART, but was inspired by its approach to pruning. The regression variant of SCART [5] is one of the few systems applicable to Relational Regression problems. Experiments in several real-world domains demonstrate that the approach is useful and competitive with existing methods, indicating that the advantage of relatively small and comprehensible models does not come at the expense of predictive accuracy.

3.1.4 Probability relational classification approaches

For dealing with the noise and uncertainty encountered in most real-world domains, probability is introduced into LBRC to integrate the advantages of both logical and probabilistic approaches to knowledge representation and reasoning. At present, the method mainly includes Inductive Logic Programming and Bayesian Networks, ILP and Stochastic Grammars.

Probabilistic relational model (PRM) [15] [27] [39] is an extension of Bayesian networks for handling relational data. A PRM describes a template for a probability distribution over a database. The template includes a relational component, that describes the relational schema for the domain, and a probabilistic component, that describes the probabilistic dependencies that hold in the domain. A PRM, together with a particular universe of objects, define a probability distribution over the attributes of the objects and the relations that hold between them.

Stochastic Logic Programs (SLPs) [34] have been a generalization of Hidden Markov Models, stochastic context-free grammars, and directed Bayes nets. A stochastic logic program consists of a set of labeled clauses p: C, where p is a probability label described the probability information of the corresponding relational pattern and C is a logic clause for extended dependent relationship between data. And by learning the data, the clause set covers each specific example and probabilities record the dependence relationships.

3.1.5 Distance relational classification approaches

RIBL (Relational Instance-Based Learning) [12] is to learn through right distance definition between the objects in multi-relational environment. The basic idea is as follows. To calculate the distance between two objects/examples, their properties are taken into account first (at depth 0). Next, objects immediately related to the two original objects are taken into account (at depth 1), or more precisely, the distances between the corresponding related objects. At depth 2, objects related to those at depth 1 are taken into account, and so on, until a user-specified depth limit is reached. It uses the k-nearest neighbor (kNN) method in conjunction with the RIBL distance measure to solve the prediction problem. RIBL2 upgrades the RIBL [12] [26] distance measure by considering lists and terms as elementary types, much like discrete and numeric values. RIBL was successfully applied to the practical problem of diterpene structure elucidation. RIBL2 has been used to predict mRNA signal structure and to automatically discover previously uncharacterized mRNA signal structure classes.

Kernel functions can project data in non-linear space into high dimensional feature spaces permitted linear hyper sphere to classify the data according to distances. So the key of the method is to construct a kernel for learning from relational data.

4. ASSOCIATE BASED RELATIONAL CLASSIFICATION

Associative classification uses association mining techniques that search for frequently occurring patterns in large databases. The patterns may generate rules, which can be analyzed for use in classification. Several algorithms have

been proposed for associative classification such as Classification based on Multiple Association Rule (CMAR), Classification based on Predictive Association Rules (CPAR).

CMAR determines the class label by a set of rules. To improve both accuracy and efficiency, it employs a data structure called Classification Rule-tree, to compactly store and retrieve a large number of rules for classification. To speed up the mining of complete set of rules, it adopts a variant of Frequent-Pattern growth method.

CPAR combine the advantages of both associative classification and traditional rule-based classification. It adopts a greedy algorithm to generate rules directly from training data. All the above algorithms only focus on processing data in a single table and applying these algorithms in multi relational environment will result in many problems.

The paper [10] extends Apriori to mine the association rules in multiple relations. The paper [35] is also based on deductive databases. These two approaches cannot be applied in relational databases directly. They have high computational complexity, and the pattern they find is hard to understand.

A Multi-relational classification algorithm based on association rules is proposed in MrCAR [17]. It uses class frequent closed item-sets. It reflects the association between class labels and other item-sets, and used to generate classification rules. MrCAR have higher accuracies comparing with the existing multi relational algorithm. The rules discovered by MrCAR have more comprehensive characterization of databases.

4.1 CMAR: Accurate and Efficient Classification Based on Multiple Class Association Rules

Associative classification has high classification accuracy and strong flexibility at handling unstructured data. However, it still suffers from the huge set of mined rules and sometimes biased classification or over-fitting since the classification is based on only single high-confidence rule. CMAR, i.e., Classification based on Multiple Association Rules. The method extends an efficient frequent pattern mining method, FP-growth, constructs a class distribution-associated FP-tree, and mines large database efficiently. Moreover, it applies a CR-tree structure to store and retrieve mined association rules efficiently, and prunes rules effectively based on confidence, correlation and database coverage. The classification is performed based on a weighted analysis using multiple strong association rules. CMAR is consistent, highly effective at classification of various kinds of databases and has better average classification accuracy in comparison with CBA and C4.5.

4.2 CPAR: Classification based on Predictive Association Rules

Recent studies in data mining have proposed a new classification approach, called associative classification, which, according to several reports, such as, achieves higher classification accuracy than traditional classification approaches such as C4.5. However, the approach also suffers from two major deficiencies: (1) it generates a very large number of association rules, which leads to high processing overhead; and (2) its confidence-based rule evaluation measure may lead to over-fitting.

In comparison with associative classification, traditional rule-based classifiers, such as C4.5, FOIL and RIPPER, are substantially faster but their accuracy, in most cases, may not be as high.

CPAR (Classification based on Predictive Association Rules), which combines the advantages of both associative classification and traditional rule-based classification. Instead of generating a large number of candidate rules as in associative classification, CPAR adopts a greedy algorithm to generate rules directly from training data.

Moreover, CPAR generates and tests more rules than traditional rule-based classifiers to avoid missing important rules. To avoid over-fitting, CPAR uses expected accuracy to evaluate each rule and uses the best k rules in prediction.

4.3 Faster Association Rules for Multiple Relations

The formalism of association rules was introduced by Agrawal [1996] for the purpose of basket analysis. An important step in the discovery of such rules is the construction of frequent item sets. These are, for instance, sets of items that are frequently bought together in one supermarket transaction. As this discovery step is time critical, it is obligatory that it is performed reasonably fast. Much research has been done in order to develop efficient algorithms. A well known algorithm resulting from this research is APRIORI, of which many variants have been developed, such as APRIORITID [Agrawal *et al.*, 1996] and a breadth-first algorithm introduced by Pijls and Bioch [1999].

On the other hand, efforts have been done to extend the usability of association rules beyond the basic case of basket analysis. Dehaspe and De Raedt [1997] use the notion of atom sets as a first order logic extension of item sets. The incorporation of techniques from Inductive Logic Programming allows for more complex rules to be found which also take into account background knowledge. Consequently, this also allows data mining of data which is spread over tables which cannot reasonably be merged into one table. An algorithm was implemented based on this notion, which was called WARMR. The usefulness of this algorithm was demonstrated in several real-world situations (see, for example, [Dehaspe *et al.*, 1998]). These experiments, however, also showed the major shortcoming of the algorithm: its efficiency proved to be very low, some experiments even taking several days.

4.4 MrCAR: A Multi-relational Classification Algorithm based on Association Rules

Classification based on association rules is one of the most effective classification method, whose accuracy is higher and discovered rules are easier to understand comparing with classical classification methods. However, current algorithms for classification based on association rules is single table oriented, which means they can only apply to the data stored in a single relational table. Directly applying these algorithms in multi-relational data environment will result in many problems. MrCAR mines relevant features in each table to predict the class label. Close item-sets technique and Tuple ID Propagation method are used to improve the performance of the algorithm. MrCAR has higher accuracy and better understandability comparing with a typical existing multi-relational classification algorithm.

Bing Liu et al. (1998) presented the first associative classification algorithm CBA (Classification Based on Associations). After that, associative classification has been

studied extensively and has a great progress. Many associative classification algorithms have been proposed successively such as: CAEP, ADT, CMAR, CPAR, etc. While all the above algorithms can only process data organized in single table, applying these algorithms in multi-relational data environment will result in many problems.

The general steps of the existing associative classification algorithms are: (1) by taking advantage of the traditional association rules algorithms such as: Apriori, FP-growth, etc, generate each class's frequent item-sets from the training sample set; (2) construct classification rules based on the frequent item-sets (3) Use these rules to classify unseen objects. Based on the above steps, it seems that if we want to extend associative classification algorithms to multi-relational environment, we can (1) mine frequent item-sets utilizing multi-relational association rules algorithms; (2) construct multi-relational classification rules and (3) predicting class labels bases on these rules.

However, the existing multi-relational association rules algorithms have more or less shortcomings. The existing algorithm classified in two categories: algorithms based on ILP (Inductive Logic Programming) and algorithms focusing on data organized in star schema. While the first kind of approaches, such as WARMR and FARMER are based on deductive databases and cannot be applied in relational databases directly. Besides, they have high computational complexity, and the pattern they find is hard to understand. The second kind of approaches, such as JSApriori, masl, masb and MultiClose, can only be applied in databases which are organized in star schema, and they may have statistical skew problem. From above, we cannot utilize existing multi-relational association rules directly.

5. EMERGING PATTERN BASED CLASSIFICATION

Emerging patterns (EPs) is namely item-sets whose supports change significantly from one class to another, capture discriminating features that sharply contrast instances between the classes. The discovery of emerging patterns (EPs) is a descriptive data mining task defined for pre-classified data. Emerging patterns are classes of regularities whose support significantly changes from one class to another. Classification by Aggregating Jumping Emerging Patterns is proposed in (JEP-Classifier), Classification by aggregating emerging patterns (CAEP), are eager-learning based approaches.

JEP-Classifier uses Jumping EPs (JEPs) whose support increases from zero in one dataset to non-zero in the other dataset whereas CAEP uses general EPs. For datasets with more than two classes CAEP uses the classes in a symmetric way, whereas JEP-Classifier uses them in an ordered way.

5.1 CAEP: Classification by aggregating emerging patterns

CAEP is based on the following two main new ideas: 1) use a new type of knowledge, the so-called emerging patterns (EPs). EPs are those item-sets whose supports increases significantly from one class of data to another. 2) An individual EP is usually sharp in telling the class of only a very small fraction of all instances, and thus it will have very poor overall classification accuracy if it is used by itself on all instances. To build an accurate classifier, we first find, each class C, all the EPs meeting some support and growth rate thresholds, from the (opponent) set of all none-C instances to the set of all C instances. Then we aggregate the power of the discovered EPs for classifying in instance s. We derive aggregate the power of the discovered EPs for classifying an instance. We derive an aggregate differentiating score for each class C, by summing the differentiating power of all EPs of C that occur in s; the score for C is then normalized by dividing it by some base score (e.g. median) of the training instances of C. finally, we let the largest normalized score determine the winning class. Normalization is done to reduce the effect of unbalanced distribution of EPs among the classes. CAEP achieves very good predictive accuracy.

EP/JEP (jumping emerging patterns) - based classifiers such as CAEP and JEP-classifier have good overall predictive accuracy. But they suffer from the huge number of mined EPs/JEPs, which makes the classifiers complex. Paper [13] propose a special type of EP, essential jumping emerging patterns (eJEPs), which are believed to be high quality patterns with the most differentiating power and thus are sufficient for building accurate classifiers. Existing algorithms such as border-based algorithms and consEPMiner [44] cannot directly mine such eJEPs. Paper [13] proposed a new single-scan algorithm to effectively mine eJEPs of both data classes (both directions) and results show that the classifier based exclusively on eJEPs, which uses much fewer JEPs than JEP-classifier, achieves the same or higher testing accuracy and is often also superior to other state-of-the-art classification systems such as C4.5 and CBA.

To achieve much better accuracy and efficiency than the previously EP-based classifiers, an instance based classifier using EPs (DeEPs) is proposed in [31], [32]. This approach achieves high accuracy, because the instance-based approach enables DeEPs to pinpoint all EPs relevant to a test instance, some of which are missed by the eager-learning approaches. It also achieves high efficiency by using a series of data reduction and concise data-representation techniques. CAEP, JEP Classifier, are the two relatives to DeEPs. DeEPs have considerable advantages on speed, and dimensional scalability over CAEP and the JEP-Classifier, because of its efficient ways to select the sharp and relevant EPs and to aggregate the discriminating power of individual EPs. Another advantage is that DeEPs can handle new training data without the need to retrain the classifier which is, commonly required by the eager learning based classifiers. This feature is extremely useful for practical applications where the training data must be frequently updated.

The paper [25] proposed in which first mine as many EPs as possible (called eager-learning) from the training data and then aggregate the discriminating power of the mined EPs for classifying new instances. In their propose a new, instance-based classifier using EPs, called DeEPs, to achieve much better accuracy and efficiency than the previously proposed EP-based classifiers. High accuracy is achieved because the instance-based approach enables DeEPs to pinpoint all EPs relevant to a test instance, some of which are missed by the eager-learning approaches. High efficiency is obtained using a series of data reduction and concise data-representation techniques.

ConsEPMiner [1], which adopts a level wise, generate and test approach to discover EPs, which satisfy several constraints. All these methods assume that data to be mined are stored in a single table. Mr.-EP [48], which discovers EPs from data scattered in multiple tables of a relational database. Generated EPs can capture the differences between objects of two classes which involve properties possibly spanned in separate data tables. In [8], two EPs- based relational classifiers Multi-Relational Classification based on Aggregating Emerging Patterns (Mr-CAEP) and Multi

Relational Probabilistic Emerging Patterns Based Classifier (Mr-PEPC) are proposed. Mr-CAEP upgrades the EP-based classifier CAEP from the propositional setting to the relational setting. It computes the membership score of an object to each class. The score is computed by aggregating a growth rate based function of the relational EPs covered by the object to be classified. In Mr-PEPC, relational emerging patterns are used to build a naïve Bayesian classifier which classifies any object by maximizing the posterior probability.

6. RELATIONAL DATABASE BASED CLASSIFICATION

Relational database based classification (RDBC) includes mainly 1) Selection Graph based (SGB) relational classification 2) Tuple ID propagation based relational classification. Selection graph based approach uses database language SQL to directly deal with multiple relations. While Tuple ID propagation is a technique for performing virtual join among the tables, which greatly improve efficiency of relational classification. Selection graph based RC, from a multi-relational data mining frame, get out of ILP approaches and transform the relationship between the tables into intuitive selection graph that is easy to be represented by SQL. That is to say, the query by SQL can complete RC. Multi- relational decision tree learning algorithm (MRDTL) [30] constructs a decision tree whose nodes are selection graphs is an extension of logical decision tree induction algorithm Top down Induction of Logical Decision Trees. It adds decision nodes to the tree through a process of successive refinement until some termination criterion is met. By using suitable impurity measure e.g. information gain, the choice of decision node to be added at each step is determined. MRDTL -2 [3] which improved the calculation efficiency and information loss of MRDTL.

Tuple ID propagation is a method for transferring information among different relations by virtually joining them. It is a convenient method that enables flexible search in relational databases and is much less costly than physical joins in both time and space. Multi-relational naïve bayes classifier Mr-SBC [9] is an integrated approach of first-order classification rules with naive Bayesian classification, in order to separate the computation of probabilities of shared literals from the computation of probabilities for the remaining literals. However, while searching first–order rules, only tables in a foreign key path can be considered and other join paths are neglected. It handles categorical as well as numerical data through a discretization method. CrossMine [43] and Graph-NB [22] was proposed by Jiawei Han and Xiaoxin Yin. CrossMine [43] is a divide and conquer algorithm, which uses rules for classification. It searches for the best way to split the target relation into partitions, and then recursively works on each partition. It also employs selective sampling method, which makes it highly scalable with respect to the number of relations. CrossMine is a sequential covering algorithm that builds rules one by one with the same FOIL. At every step, the foil gain of each of these predicate is evaluated and the best one is added to the current rule. The comprehensive experiments demonstrate the high scalability and accuracy of CrossMine. Graph-NB [22] which upgrades Naïve Bayesian classifier, and use the semantic relationship graph (SRG) to describe the relationship and to avoid unnecessary joins among tables. To improve the accuracy, a pruning strategy named "cutting off" strategy is used to simplify the graph to avoid examining too many weakly linked tables.

The paper [43] proposed two methods for classification: CrossMine-Rule is a rule-based classifier and CrossMine-Tree, is decision tree based classifier. The comprehensive experiments demonstrate the high scalability and accuracy of CrossMine. The Relational decision tree (RDC) [16] is an extension of MRDTL algorithm with the usage of tuple ID propagation. For dealing with the missing attribute, a naïve bayes model for each attribute in a table is built based on the other attributes excluding the class attribute. The missing values are filled with the most likely predicted value by the naïve bayes predictor. It achieves higher efficiency and is more efficient in running time than MRDTL-2.

Classification with aggregation of Multiple Features (CLAMF) method is proposed in [14], which is an adaptation of the sequential covering algorithm and classifies the multi relational data using aggregation involving single and multiple features. In temporal databases, classification with multi feature aggregation could provide very interesting rules that are much more meaningful to the end–user by allowing temporal trends**.** The paper [36] is based on two pruning strategy. Firstly, get rid of some attributes based on the foil gain, and make use of relationship between the accuracy of the attribute to give them the second pruning. In the second step, the remaining attributes are used to classify the data. This method guarantees the accuracy and also saves much time. In [19], novel approach proposed to conduct both Feature and Relation Selection for efficient multi-relational classification. In this approach symmetrical uncertainty is used to measure correlation between attributes in a table or cross tables. It also measures the correlation between a table and a class attribute. Based on the correlations, it selects relevant attributes and tables from the database. A multiple view strategy is proposed in [18], which enable us to classify relational objects by applying conventional data mining methods, while there is no need to flatten multiple relations to a universal one. It employs multiple view learners to separately capture essential information embedded in individual relation. The acquired knowledge is incorporated into a Meta learning mechanism to construct the final model.

7. COMPARATIVES STUDIES OF MULTI-RELATIONAL CLASSIFIER

Based on the study approaches related to multi-relational classification, we perform the comparative studies of various multi-relational classification approaches.

7.1 Mutagenesis Database

This database is widely used in the area of ILP. It contains 4 relations and 15,218 tuples. The target relation contains 188 elements of Mutagenesis. Table 1 describes the specification of Mutagenesis dataset and Table 2 demonstrates the performance study on Mutagenesis dataset which represents accuracy of different multi-relational classification approaches. Figure 1 shows the foreign-key primary-key relationship diagram of a Mutagenesis database. Relation Mole is the target relation and attribute label is the target attribute. One can see from figure 1 that there is a cycle between relation Atom and Bond. In figure 2, it is done by duplication of Atom once. This kind of duplication can be done virtually and automatically by doing some mapping.

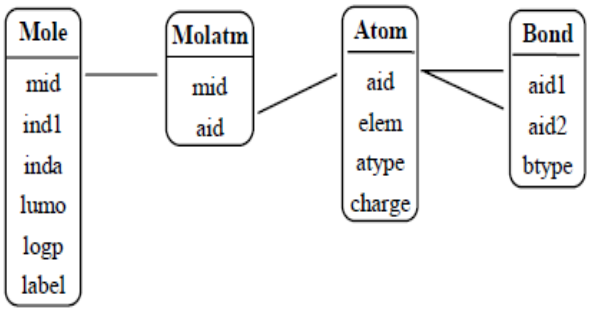

Fig. 1. Relationship of Mutagenesis dataset

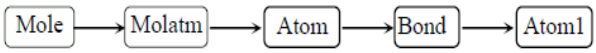

Fig. 2. SRG for Mutagenesis dataset

Table 1. Specification of Mutagenesis dataset

Relation Name	Molecule	Atom	Molecule-Atom	Bond
Tuples	188	4893	4893	5244
Attributes	5	4	2	3

Table 2. Performance Comparison on Mutagenesis dataset

Approach	Accuracy (%)	Reference
FOIL	68.3	[6]
TILDE	79.7	[6]
Mr-SBC	82.4	[33]
Graph-NB	82.3	[22]
MVC	86.7	[23]
MRDTL	80.6	[21]
MRDTL-2	87.5	[21]
CrossMine	87.7	[46]
MrCAR	89.3	[45]

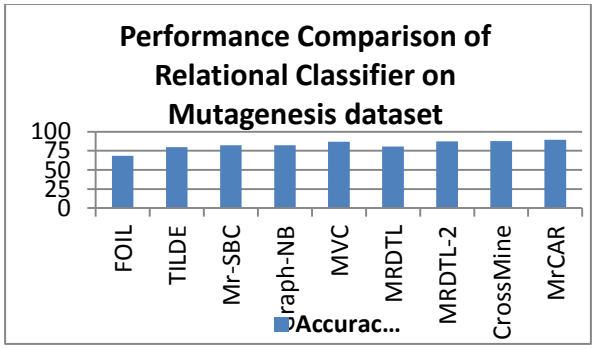

Fig. 3. Performance Comparison of Relational Classifier on Mutagenesis dataset

7.2 Financial database (PKDD CUP 1999)

This database is the financial database used in PKDD CUP 1999. This datasets has eight tables and 75982 tuples totally. The target table Loan contains 324 positive tuples and 76 negative tuples. Figure 4 shows the foreign-key primary-key relationship diagram of a financial database. Figure 5 shows the SRG for this database. Table 3 describes the specification of financial dataset and Table 4 demonstrates the performance study on financial dataset which represents accuracy of different multi-relational classification approaches.

TABLE 3. Specification of Financial dataset

Relation Name	Account	Client	Disposition	Order	Loan	Card	District	Transaction
Tuples	4500	5369	5369	6471	400	892	77	52904
Attributes	4	4	4	6	6	4	16	8

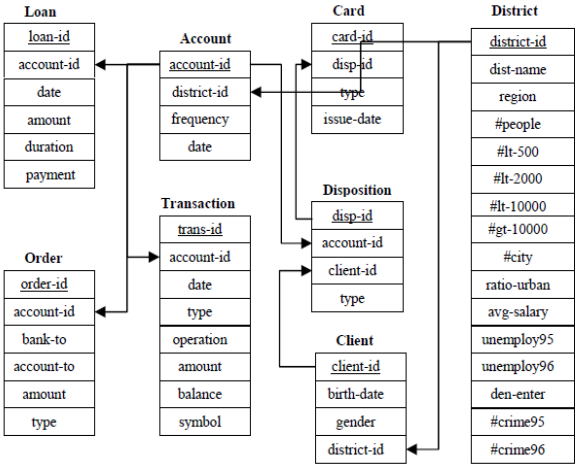

Fig.4. Relationship of Financial dataset (PKDD CUP 1999)

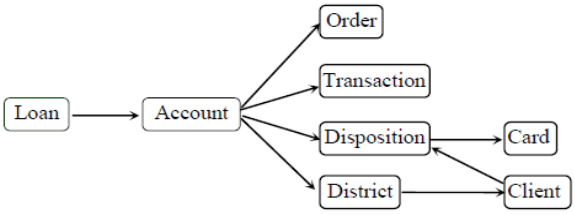

Fig.5. SRG for Financial dataset

TABLE 4. Performance Study on Financial dataset

Approach	Accuracy (%)	Reference
FOIL	71.5	[6]
TILDE	81.3	[6]
Graph-NB	85.25	[22]
CrossMine	89.8	[46]
RDC	83.2	[24]

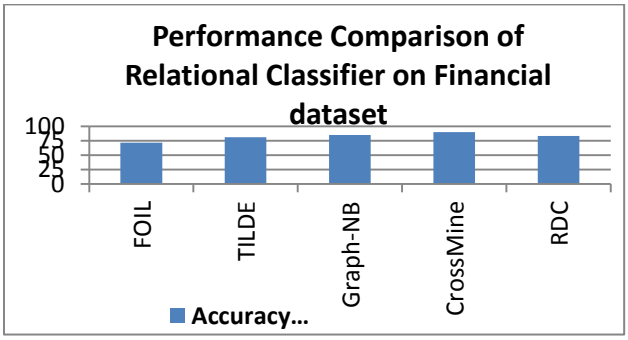

Fig. 6: Performance Comparison of Relational Classifier on Financial dataset

The popular first-order inductive learner (FOIL) method (Quinlan and Cameron-Jones, 1993) upgrades the well-known CN2 algorithm (Clark and Niblett, 1989) to deal with first-order representations. This approach employs a general-to-specific search to build rules to explain many positive examples and cover few negative examples. The top-down induction of logical decision trees (TILDE) method (Blockeel and Raedt, 1998) extends the popular C4.5 propositional learner to tackle relational representations. The TILDE algorithm applies logical queries in nodes of the decision tree instead of testing attributes values. The divide-and-conquer nature embedded in the decision tree construction makes the TILDE method an efficient one.

Mr-SBC is an extension of the navie bayes classification method to the multi-relational setting. In this setting, training data are stored in several tables related by foreign key constrains and each example is represented by a set of related tuples rather than a single row as in the classical data mining setting. It is characterized by three aspects. First, an integrated approach in the computation of the posterior probabilities for each class that makes use of first order classification rules. Second, the applicability to both discrete and continuous attributes by means a supervised discretization. Third, the consideration of knowledge on the data model embedded in the database schema during the generation of classification rules.

Graph-NB is upgraded Navie Bayesian Classifier to deal with multiple tables directly which used the concept of semantic relationship graph in relational environment.

MRDTL-2 includes techniques for significantly speeding up, often by a couple of orders of magnitude, some of the most time consuming components of multi-relational data mining algorithms like MRDTL that rely on the use of selection graphs. MRDTL-2 includes a simple and computationally efficient technique which uses Naive Bayes classifiers for 'filling in' missing attributes values.

Multiple view (MVC) approach can be characterized using different representations (view), and that learning from these representations separately can lead to better gains than merging them into a single dataset. Using a relational database as input, multi-view relational classification strategy learns from multiple views (feature set) of a relational database, and then information acquired by view learners are integrated to construct a final classification model.

MRDTL is augmented with principled methods for handling missing attribute values, is likely to be competitive with the state-of-the-art algorithms for learning classifiers from multiple relations on real-world data sets.

CrossMine is an efficient and scalable approach for multi-relational classification. Is uses a novel method tuple ID propagation to perform virtual joins, so as to achieve high classification accuracy and high efficiency on databases with complex schemas.

MrCAR for classification based on association rules in multi-relational data environment, which mines relevant features in each table to predict the class label. Close item-sets technique and Tuple ID Propagation method are used to improve the performance of the algorithm.

RDC is a new approach for multi-relational decision tree classification. It integrates the multi-relational decision tree and tuple ID propagation. So it can achieve much higher efficiency. These features make it appropriate for multi-relational classification in real world databases. Its main innovation is updating Multi-relational Decision Tree Learning algorithm with the usage of ID propagation, which propagate the tuple IDs (together with their associated class labels) in the target relation to other relations.

8. RESEARCH CHALLENGES

- **Learning from networks of Examples:**

Most ILP (Inductive Logic Programming) Research has focused on problems where individual examples have Relational Structure, but examples are still independent of each other. For instance, an example might be a molecule, with the bonds between atoms as the relational structure, and the task being to predict whether the molecule is a Carcinogen.

However, arguable the most interesting and challenging problem in MRDM is that of dependencies between examples. For example, molecules do not act independently in the cell; rather they participate in complex chains of reactions whose outcomes we are interested in.

- **Learning from time-changing Relational Data:**

Many relational phenomena of interest take place over time(e.g., a shopper cruising through an e-commerce site, a terrorist group preparing an attack, the response of an organism's immune system to an infection). Temporal phenomena pose some of the most complex problems of MRDM.

- **Integration with traditional KDD(Knowledge Discovery from database):**

The emphasis in the MRDM literature has often been on the contrasts between relational and propositional approaches, which are often understandable in a new field, but ultimately counterproductive. Emphasizing the continuity between the two will make the widespread adoption of MRDM easier. Further, MRDM should build on the extensive research already done on Propositional methods, by making them easy to plug them into MRDM algorithms.

- **Integration with databases:**

Traditionally, a single table is extracted manually from a Multi relational database. Finding the best way to do this is often quite difficult, and can consume a large fraction of a KDD Project's time. One of the great potential benefits of MRDM is the ability to automate this process to a significant extent. Fulfilling this potential requires solving the significant efficiency problems that arise when attempting to do data mining directly from a relational database, as opposed to from a single pre-extracted flat file.

- **Learning from multiple sources of information:**

Data from MRDM often comes from multiple sources. They can be of many different types (e.g., Relational databases, plain text, XML, audio, video, sensors, etc.). Even when they are of the same type, different sources often use different representations for the same entities, and can be of widely varying quality. We would like MRDM systems to be able to range autonomously over all sources relevant to their task, perhaps even discovering them on their own.

- **Using domain knowledge:**

If MRDM can take advantage of uncertain reasoning to absorb domain knowledge in less polished forms, this may greatly increase its range of applicability, and its success in the applications it tackles.

- **Making the results of research accessible:**

One of the bottlenecks preventing the wider use of MRDM is the fact that relational algorithms tend to be much more involved and difficult to understand than propositional ones.

9. CONCLUSION

Multi –relational data mining has become popular due to the limitations of propositional problem definition in structured domains and the tendency of storing data in relational databases. In these paper we presents the several kind of classification approaches across multiple database relations including ILP based, Emerging Pattern based, Associative based, Relational database based approaches. We have presented the comparative studies of different multi-relational classifiers on the mutagenesis and financial datasets. The Relational Classification challenges are relational classification approaches are mainly from inductive logic programming technology, which is developed from propositional classification and also how to extend other proposition methods to logic based relational classification. Relational based relational classification opens up a new way for relational classification research. At present, the focus of the selection graph based relational classification is on MRDM with decision tree inductive methods.

10. ACKNOWLEDGMENT

We are thankful to the great GOD for making us able to do something.

This research is the part of Ph.D. programme in Computer Science & Engineering, Karpagam University, India.

11. REFERENCES

[1] Appice, A., Ceci M., Malgieri C., Maleraba D. “Discovering relational emerging patters”, AI*AI 2007, LNCS (LNAI), Vol. 4733, 206-217, Springer, Heidelberg, 2007.

[2] Arno J. Knobbe, Arno Siebes, Hendrik Blockeel, Daniël van der Wallen, “Multi-Relational Data Mining, using UML for ILP”, PKDD '00 Proceedings of the 4th European Conference on Principles of Data Mining and Knowledge Discovery Springer-Verlag London, UK ©2000.

[3] Atramentov, A., Leiva, H., and Honavar, V. “A Multirelational Decision Tree Learning Algorithm-Implementation and Experiments”, ILP LNCS, Vol.2835, pp.38-56, 2003.

[4] Atramentov, A., Leiva, H., and Honavar, V. “Experiments with MRDTL -- A Multi-relational Decision Tree Learning Algorithm”, ILP LNCS, Vol.2835, pp.38-56, 2003.

[5] Blockeel, H. 1998. “Top-down induction of first order logical decision trees”, Artificial Intelligence Journal, vol.101,pp. 285-297.

[6] Blockeel H., De Raedt L., and Ramon J. “Top-down induction of logical decision trees”. In Proc. 1998 Int. Conf. Machine Learning (ICML'98), Madison, WI, Aug. 1998.

[7] C. L. Curotto ,N. F. F. Ebecken, H. Blockeel. “Multi-relational data mining in Microsoft SQL Server”, In Seventh International Conference on Data, Text and Web Mining and their Business Applications, Prague, Tsjechie, July, 2006.

[8] Ceci, M., Appice, A., Maleraba, D. “Emerging Pattern Based Classification in Relational Data Mining”, DEXA 2008, LNCS, vol.5181, pp.283-296.

[9] Ceci M., Appice A., and Malerba D. “Mr-SBC: a Multi-Relational Naive Bayes Classifier”, in N. Lavrac, D. Gamberger, L. Todorovski & H. Blockeel (Eds.), Knowledge Discovery in Databases PKDD 2003, Lecture Notes in Artificial Intelligence, 2838, 95-106, Springer, Berlin, Germany.

[10] Dehaspe, L., Raedt, “Mining Association Rules in Multiple Relations”, In Proceedings of the ILP, Springer-Verlang, London UK, pp.125-132, 1997.

[11] Dzeroski, S., Lavtac, N. 2001. eds, “Relational data mining”, Berlin: Springer.

[12] Emde, W., Wettschereck, “Relational instance based learning”, In Proceedings of the 13th Int. Conference on Machine Learning, Morgan Kaufmann, San Mateo, CA, 122-130, 1996.

[13] Fan, H., Ramamonanarao, K. “An efficient single scan algorithm for mining essential jumping emerging patterns for classification”, In Pacific-Asia Conference on Knowledge Discovery and Data Mining , pp.456-462, 2002.

[14] Frank, R., Moser, F., Ester, M. “A Method for Multi-Relational Classification Using Single and Multi-Feature Aggregation Functions”, In Proceedings of 11th European Conf. on PKDD, Springer, Verlag Berlin Heidelberg, 2007.

[15] Getoor, L., Friedman, N., Koller, D., and Pfeffer, A. 2001. “Learning Probabilistic Relational Models”, pp.307-355, Springer Verlage, New York.

[16] Guo, JF., Li, J., Bian, WF. “An Efficient Relational Decision Tree Classification Algorithm”, In proceedings of 3rd ICNC, vol.3, 2007.

[17] Gu,Y., Liu, H., He, J. “MrCAR: A Multi relational Classification Algorithm based on Association Rules”, Int. Conf. on Web Information Systems and Mining, pp.256- 260, 2009.

[18] Guo, H., Herna, L., Viktor. “Multirelational classification: a multiple view approach”, Knowl. Inf. Systems, vol.17, pp.287–312, Springer-Verlag London, 2008.

[19] H, J. Liu,H.,et at, “Selecting Effective Features and Relations For EfficientMulti-Relational Classification”, Computational Intelligence, Vol 26, No.3, 2010.

[20] Han, J., Kamber, M. 2007. Data Mining: Concepts and Techniques”, 2nd Edition, Morgan Kaufmann.

[21] Héctor Ariel Leiva , Shashi Gadia , Drena Dobbs, “MRDTL: A multi-relational decision tree learning algorithm (2002)” Proceedings of the 13th International Conference on Inductive Logic Programming (ILP 2003).

[22] Hongyan Liu, Xiaoxin Yin, Jiawei Han, “An Efficient Multi-relational Naïve Bayesian Classifier Based on Semantic Relationship Graph”, Proceeding MRDM '05 Proceedings of the 4th international workshop on Multi-relational mining, Pages 39 – 48, ACM New York, NY, USA ©2005.

[23] Hongyu Guo, Herna L. Viktor, “Mining relational databases with multi-view learning”, ACM, DOI: 10.1145/1090193.1090197, 2005.

[24] Jing-Feng Guo, Jing Li, Wei-Feng Bian, "An Efficient Relational Decision Tree Classification Algorithm", IEEE 2007, Natural Computation, ICNC 2007.

[25] Jinyan Li1, Guozhu Dong, Kotagiri Ramamohanarao, "Instance-Based Classification by Emerging Patterns, Principles of Data Mining and Knowledge Discovery", Lecture Notes in Computer Science, 2000, Volume 1910/2000, 191-200, DOI: 10.1007/3-540-45372-5_19.

[26] Kirsten, M., Wrobel, S., Horvath, "Distance Based Approaches to Relational Learning and Clustering: Relational Data Mining", Morgan Kaufmann (2005) 6, pp.213-232, springer, Heidelberg.

[27] Koller, Pfeffer, A. 1998. "Probabilistic frame-based systems", In Proceedings of the 15th National Conference on Artificial Intelligence, pp. 580–587, Madison, WI.

[28] Kramer, S., Widmer, G. 2001. "Inducing Classification and Regression Tress in First Order Logic: Relational Data Mining", pp.140-159, Springer.

[29] Lappoon R. Tang, Raymond J. Mooney, and Prem Melville, "Scaling Up ILP to Large Examples: Results on Link Discovery for Counter-Terrorism", KDD-2003 Workshop on Multi-Relational Data Mining (MRDM-2003), pp.107-121, Washington DC, August, 2003.

[30] Leiva, HA. "A multi-relational decision tree learning algorithm", ISU-CS-TR, Lowa State University, pp.02-12, 2002.

[31] Li, J., Dong, G., Ramamohanarao, K., Wong, L. "A new instance-based lazy discovery and classification system", Machine Learning, vol.54, No.2, pp0. 99-124, 2004.

[32] Li, J., Dong, Ramamohanarao, K. "DeEPs: Instancebased classification by emerging patterns", Technical Report, Dept of CSSE, University of Melbourne, 2000.

[33] Michelangelo Ceci , Donato Malerba, "Mr-SBC: a Multi-Relational Naive Bayes Classifier (2003)" Knowledge Discovery in Databases PKDD 2003, Lecture Notes in Artificial Intelligence.

[34] Muggleton, "Learning Stochastic Logic Programs", In Proceedings of the AAAI-2000 Workshop on Learning Statistical Models from Relational Data, Technical Report WS-00-06, pp. 36-41.

[35] Nijssen S., Kok, J. "Faster Association Rules for Multiple Relations", In Proceedings of the IJCAI, pp.891-896, 2001.

[36] Pan Cao, Wang Hong-yuan. "Multi-relational classification on the basis of the attribute reduction twice", Communication and Computer,Vol. 6, No.11. pp: 49-52, 2009.

[37] Raymond J. Mooney, Prem Melville, Lappoon Rupert Tang, "Relational Data Mining with Inductive Logic Programming for Link Discovery",National Science Foundation Workshop on Next Generation Data Mining, Nov. 2002, Baltimore, MD.

[38] Seda Daglar Toprak,Pinar Senkul, "A New ILP-based Concept Discovery Method for Business Intelligence", 2007, IEEE.

[39] Taskar B, Segal E, Koller D, "Probabilistic Classification and Clustering in Relational Data", In Proceedings of International Conf. Artificial Intelligence, vol.2, 2001.

[40] V.B. Vaghela, A. Ganatra, and A. Thakkar. "Boost a weak learner to a strong learner using ensemble system approach". IEEE International Advance Computing Conference, 3:1432{1436, 2009.

[41] Vimalkumar B. Vaghela, Dr. Kalpesh H. Vandra, Dr. Nilesh K. Modi, "Multi-Relational Classification Using Inductive Logic Programming", International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 3, May - 2012 ISSN: 2278-0181.

[42] Wrobel S, "Inductive Logic Programming for Knowledge Discovery in Databases: Relational Data Mining", Berlin: Springer, pp.74-101, 2001.

[43] Xiaoxin Yin, Jiawei Han, Jiong Yang,et al. "Efficient Classification across Multiple Database Relations:A CrossMine Approach". IEEE Transactions on Knowledge and Data Engneering, 2006, 18 (6):770-783.

[44] Xiuzhen Zhang , Guozu Dong , Ramamohanarao Kotagiri, "Exploring constraints to efficiently mine emerging patterns from large high-dimensional datasets", Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, p.310-314, August 20-23, 2000, Boston, Massachusetts, United States.

[45] Yingqin Gu, Hongyan Liu, Jun He, Bo Hu and Xiaoyong Du, "MrCAR: A Multi-relational Classification Algorithm based on Association Rules", IEEE, Web Information Systems and Mining, 2009.

[46] Yin X, Han J, and Yu PS, "CrossMine: Efficient Classification across Multiple Database Relations". In Proceedings of 20th Int. Conf. on Data Engineering (ICDE'04), 2004.

[47] Yusuf Kavurucu, Pinar Senkul, Ismail Hakki Toroslu, "AGGREGATION IN CONFIDENCE-BASED CONCEPT DISCOVERY FOR MULTI-RELATIONAL DATA MINING", IADIS European Conference Data Mining 2008.

[48] Zhang, X., Dong, G., Ramamohanarao, K. "Exploring constraints to efficiently mine emerging patterns from large high-dimensional datasets", In Proceedings of 6th SIGKDD international conference on Knowledge Discovery and Data Mining, pp. 310-314, 2000.

12. AUTHOR'S PROFILE

Prof. Vimalkumar B Vaghela, is currently doing Ph.D. in Computer Science & Engineering at Karpagam University, India. This author is Young Scientist awarded from Who's Whos Science & Engineering 2010-2011 & also his biography is included in American Biographical Institute in 2011. His publication is also available in ieeeexplorer and also in spocus online database. He is currently working as Assistant Professor in Computer Engineering Department at L.D. College of Engineering, Ahmedabad, Gujarat, India. He received the B.E. degree in Computer Engineering from C. U. Shah College of Engineering and Technology, in 2002 & M.E. degree in Computer Engineering from Dharmsinh Desai University, in 2009. He has published book titled "Ensemble Classifier in Data Mining" in LABERT Academic Publisher, Germany, 2012. His research areas are Relational Data Mining, Ensemble Classifier, Pattern Mining. Author has published / presented more than 8 international papers and 5 national papers.

Dr. Kalpesh H Vandra received the B.E. degree in Electronics & Communication form North Gujarat University, Patan, in 1995, the M.E degree in Microprocessor System Applications from M.S. University, Vadodara, in 1999 and PhD from Saurashtra University, Rajkot, in 2009. He is working as Principal of C. U. Shah College of Engineering & Technology, Wadhwan City, Gujarat, INDIA. Author has more than 15 years of UG & 6 years of PG (MCA and M.E) Teaching Experience. Author had written more than 10 Books related to Computer & IT related area. He has published / presented 10 International and 7 national Research Papers & Guided more than 10 PG Students and more than 155 UG Students Dissertation work. Dr. K H.Wandra is a Chairman of Board of Studies Computer Engineering at Saurashtra University, Rajkot, Core Committee member For Syllabus of UG & PG at Gujarat Technology University, Ahmedabad. Section Managing Committee Member of ISTE Gujarat Section, Life member of ISTE, member of IEEE and CSI, Interested for working in the area of Wireless communication, Networks, Advanced Microprocessors, Data Mining.

Dr. Nilesh K Modi received the MCA degree from A.M.P. Institute of Computer Studies, Kherva, Gujarat, India and PhD from Bhavnagar University in 2006. He is working as a professor & head of MCA department in Sarva Vidyalaya's Institute of Computer Studied, S V Campus, Kadi, Gujarat, India. He is Associate Life Member in Computer Society of India (CSI) Mumbai, Senior Associate Member in International Association of Computer Science and Information Technology (IACSIT) Singapore, Senior Member in International Association of Engineers (IAEng) Hong Kong, Senior Member in Computer Security Institute New York, Member in Data Security Council of India (DSCI) a NASSCOM initiative New Delhi. Author has published / presented more than 18 international papers and more than 25 national papers. His areas of research interest are Data mining, Computer network, information security.