OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

High Dimensional Global Pattern Set Mining Based On Genetic Constraint Mechanism

M. Rajasekar[1], S. Vimalathithan[2],

[1, 2] *(PG scholar, Associate Professor, Department of Computer science and Engineering, Indra Ganesan College of Engineering, Trichy-12, Tamil Nadu)*

Abstract: *Pattern set mining troubled with discovery a set of NP interrelated patterns in constraints. The Number of Pattern(NP) set mining crisis is a very universal problem that can be instantiated to a large range of well-known mining responsibilities including concept-learning, rule-learning, redescription mining, abstract clustering and NP-tiling. To resolve this create a large number of constraints for use in NP pattern set mining, both at the narrow level, that is, on individual patterns, and on the universal level for a whole pattern set. The genetic approach used to model by high dimensional data into low dimensional data population for speedy retrieval of frequent itemsets and reducing the storage space of the dataset. This allows exploring a large number of settings within a unified framework and solving multiple mining problems in single step. The experimental results show that there is an improvement of predictive accuracy for classifying data summarized based on Fixed Length Feature Construction With Substitution (FLFCWS) to address the problem by means of optimizing the feature construction for relational data summarization.*

Keywords: *Data mining, Genetic approach, Constraints Constraint programming and pattern set.*

I. Introduction

Problem of native pattern mining will be formalized as that of finding the set of patterns, the set of all patterns that satisfy a constraint p with relation to information. Various approaches to pattern mining are developed to effectively notice the patterns adhering to a collection of constraints; a well known example is that the downside of finding frequent item sets. The generated set of patterns is usually large, that is, the set of patterns is simply too onerous to interpret because the patterns act with each other. This has junction rectifier to a typical step-wise procedure during which pattern mining solely forms associate degree intermediate step within the data discovery method. Within the opening, the patterns adhering to some constraints are thoroughly explored for. These patterns are typically known as native patterns. Within the second step, some patterns are chosen and combined in an exceedingly heuristic thanks to produce a world model.

In distinction to constraint-based native pattern mining, wherever mining issues are usually well formalized and anti-monotonicity is one in every of the guiding principles for the event of algorithms, no unifying principles or algorithms exist for pattern set mining. Pattern set mining tasks are solved mistreatment wide ranges of heuristics and greedy search ways, usually by sorting out patterns in an exceedingly opening and by filtering these patterns in an exceedingly second step. The most downside with these approaches is that it's unclear however they extend toward new and unsolved issues. Indeed, to the most effective of the authors' data, there presently is not any general approach for locating pattern sets that generalizes over an oversized variety of settings. To build a primary step toward formally specifying pattern set mining issues and resolution them by means that of general algorithms. A tendency to develop a framework during which a mess of tasks, as well as concept-learning, abstract bunch, description mining and covering, will be formalized. The most plans during this framework are to formalize mining tasks as issues of finding k patterns that along satisfy constraints. In distinction to earlier approaches, wherever constraints are usually solely formalized on the native level, that is, on individual patterns, among this framework we have a tendency to conjointly formalize constraints on the world level that's on the pattern set as a full. Each level of constraints are formalized at identical time, that is, in an exceedingly single specification; Have a tendency to gift a high-level modeling language, freelance from underlying frameworks, and show a way to use it to formulate several well-known tasks.

A key feature is that has a tendency to open up the chance that mining issues don't seem to be solved in multiple steps, however conjointly in one single step. Our vision is that these declaratively specific issues are solved employing a general convergent thinker in an exceedingly uniform manner. Developing such solvers may be a challenge. Constraint programming may be a generic framework for resolution combinatorial and optimization issues underneath constraints. It's been used with success in various applications, as well as constraint-based mining of individual patterns. The key power of CP lies in its generic approach to downside solving: users model a retardant by specifying constraints, and also the CP convergent thinker can use those constraints to search out the solutions. This has the advantage that new issues will be solved by solely ever-changing the specification in terms of constraints; a replacement convergent thinker isn't required. Use the benefits of Constraint Programming to demonstrate the chances of a declarative approach to data processing.

This paper is organized as follows: Section II Presents Related Works, Section III Proposed technique. The experimental Results of the schemes are presented in section IV and Section V presents conclusion of this paper.

II. RELATED WORK

Tias Guns et.al.[1] implemented that the framework used to discover the pattern set mining under constraints without underlying any principles. It did not focus on global constraints relationship. L. De Raedt et.al.[2] reported that data mining community has been interested in constraint- based mining, that is, the use of constraints to specify the desired properties of the patterns to be mined. The task of the data mining system is then to generate all patterns satisfying the constraints. A wide variety of constraints for local pattern mining exists and has been implemented in an even wider range of specific data mining systems. On the other hand, the artificial intelligence community has studied several types of constraint-satisfaction problems and contributed many general purpose algorithms and systems for solving them. These approaches are now gathered in the area of constraint programming. In constraint programming, the user specifies the model, that is, the set of constraints to be satisfied, and the constraint solver generates solutions. Thus, the goals of constraint programming and constraint based mining are similar (not to say identical); it is only that constraint programming targets any type of constraint satisfaction problem, whereas constraint- based mining specifically targets data mining applications. Therefore, it is surprising that despite the similarities between these two endeavors, the two fields have evolved in- dependently of one another, and also, that – to the best of the author's knowledge – constraint programming tools and techniques have not yet been applied to pattern mining, and, vice versa, that ideas and challenges from constraint-based mining have not yet been taken up by the constraint programming community. It finds the gap between these two fields by investigating how standard constraint-programming techniques can be applied to a wide range of pattern mining problems. The aim is to formalize the most well-known constraint-based mining problems in terms of constraint programming terminology. This includes constraints such as frequency, closeness and maximalist, and constraints that are monotonic, anti-monotonic and convertible, as well as variations of these constraints, such as δ -closeness. Then incorporate them in off-the-shelf and state-of-the-art constraint programming tools, such as Geode 1 and Eclipse and run experiments. The results are surprising in that 1) using the constraint programming approach, it is natural to combine complex constraints in a flexible manner (for instance, $\delta -$ closeness in combination with monotonic and anti-monotonic constraints); unlike in the existing constraint-based mining systems, this does not require modifications to the underlying solvers; 2) even though the constraint programming methods were not meant to cope with the specifics of data mining (such as coping with large data sets, having 10000s of constraints to solve), and even though the focus of this study is not on the development of efficient algorithms, it turns out that existing constraint programming systems already perform quite well as compared to dedicated data mining solvers in that on a number of benchmark problems their performance is similar.

T. Guns et.al.[3] suggested that the important problems in data mining and machine learning are classification and pattern mining. In recent years an increasing number of publications have studied the combination of these problems. The main idea in these methods is that patterns can be used to define features or can be used as rules; classification models which make use of these features or rules may be more ac curate or more simple to understand. Last, but not least, in structured domains, pattern mining can be considered a propositionalization approach which enables the use of propositional data mining and machine learning algorithms. Despite the large amount of publications devoted to this topic, believe however that an overview of what has been accomplished in this area is missing. It is not uncommon for publications in this area to refer to only a small portion of relevant related work, hence preventing deeper insight or a general theory from evolving. The problems of subgroup discovery, contrast set mining and emerging pattern mining are so similar that their main differences are arguably the terminology used. Believe that this phenomenon is much more wide-spread. For instance, the author pointed out that the independently proposed areas of correlating (or correlated). Other overviews have stressed the fact that there are different types of data, such as graph-based, tree- based and item set- based data. They coupled pattern selection strategies to particular pattern types, and stressed the fact that different pattern mining algorithms are needed to deal with each such data type. Even though this is true, and indeed one often needs to implement a different pattern miner to deal with a pattern type at hand, believe that it is more important in this case to stress the conceptual similarities between these pattern mining algorithms. Doing so leads to the insight that most approaches that have been proposed for complex data types, such as graphs, can easily also be implemented in pattern mining algorithms for simpler data types, such as item sets;

this leads to a large number of additional approaches that item set-mining based approaches could be compared with. Initial approaches which combined pattern mining and classification models took a strict step-wise approach, in which a set of patterns is computed once and these patterns are subsequently used in models. However, in more recent years a large number of methods have been proposed which aim at integrating pattern mining, feature selection and model construction. The use of patterns in predictive models is a topic that has received a lot of attention in recent years. Pattern mining can help to obtain models for structured domains, such as graphs and sequences, and has been proposed as a means to obtain more accurate and more interpretable models. Despite the large amount of publications devoted to this topic, that believe however that an overview of what has been accomplished in this area is missing.

S. Nijssen et.al.[4] presented that the correlated pattern mining is concerned with finding the highest scoring patterns a correlation measure (such as information gain). By reinterpreting correlation measures in ROC space and formulating correlated item set mining as a constraint programming problem, that obtain new theoretical insights with practical benefits. More specifically, contributed 1) an improved bound for correlated item set miners, 2) a novel iterative pruning algorithm to exploit the bound, and 3) an adaptation of this algorithm to mine all item sets on the convex hull in ROC space. The algorithm does not depend on a minimal frequency threshold and is shown to outperform several alter- native approaches by orders of magnitude, both in runtime and in memory requirements. Correlated pattern mining is amongst the most popular data mining tasks. As opposed to frequent item set mining, correlated pattern mining involves transactions that belong to two different classes and the task is to find those patterns that are correlated with the class attribute, that is, those patterns that are indicators of one of the two classes. In this paper, that formalizes this task as that of finding the patterns that score high. A large number of publications are concerned with correlated pattern mining, but the problem is also known under the names of interesting item set mining, contrast set mining, emerging item set mining, sub- group discovery and discriminative item set mining. Further publications have extended these settings to structured domains, such as graphs, trees and sequences. The popularity of correlated pat- terns is in part due to their use for classification, where they has been used both as classification rules and as features for building classifiers. The key difference between traditional rule learning and correlated pattern mining approaches is that the former are typically heuristic and the latter provide guarantees concerning completeness and optimality of the computed solutions.

H. Cheng et.al.[5] described that the DDPMine often adopts a two-step approach: frequent pattern (or classification rule) mining followed by feature selection (or rule ranking). However, this two-step process could be computationally expensive, especially when the problem scale is large or the minimum support is low. It was observed that frequent pattern mining usually produces a huge number of “patterns” that could not only slow down the mining process but also make feature selection hard to complete. The proposed direct discriminative pattern mining approach, DDPMine, to tackle the efficiency issue arising from the two-step approach. DDPMine performs a branch-and bound search for directly mining discriminative patterns without generating the complete pattern set. Instead of selecting best patterns in a batch, that introduces a “feature-centered” mining approach that generates discriminative patterns sequentially on a progressively shrinking FP-tree by incrementally eliminating training instances. The instance elimination effectively reduces the problem size iteratively and expedites the mining process. Empirical results show that DDPMine achieves orders of magnitude speedup without any downgrade of classification accuracy. It outperforms the state-of-the-art associative classification methods in terms of both accuracy and efficiency.M. Khiari et.al.[6] introduced that there are lot of works to discover local patterns under constraints but there are not so many methods to combine local patterns. Compare patterns between them; they are mainly based on the reduction of the redundancy. But not implemented globally. S. Ruggieri et.al.[7] detailed that the techniques of disjunction-free sets, non-derivable itemsets, closed non-derivable itemsets are maintained the disjunctive or the negative support of itemsets, or that refer to measures other than support. W. Li et.al.[8] discovered that building accurate and efficient classifiers for large databases is one of the essential tasks of data mining and machine learning research. Given a set of cases with class labels as a training set, classification is to build a model (called classifier) to predict future data objects for which the class label is unknown. Previous studies have developed heuristic/greedy search techniques for building classifiers, such as decision trees, rule learning, natıve-Bayes classification, and statistical approaches. These techniques induce a representative subset of rules (e.g., a decision tree or a set of rules) from training data sets for quality prediction. Recent studies propose the extraction of a set of high quality association rules from the training data set which satisfies certain user specified frequency and confidence thresholds. Effective and efficient classifiers have been built by careful selection of rules. Such a method takes the most effective rule(s) from among all the rules mined for classification. Since association rules explore highly confident associations among multiple variables, it may overcome some constraints introduced by a decision-tree induction method which examines one variable at a time. Extensive performance studies show that association- based classification may have better accuracy in general. However, this approach may also suffer some weakness as shown below. On one hand, it is not easy to identify the most effective rule at classifying a new case. Some method, such as simply selects a rule with a maximal user -defined measure, such as confidence. As a selection may not always be the right choice in many cases. Such a simple pick may affect the classification accuracy.

M. Thoma et.al.[9] implemented that the classification of graphs is an increasingly important step in numerous application domains, such as function prediction of molecules and proteins, computerized scene analysis, and anomaly detection in program flows. Among the various approaches proposed in the literature, graph classification based on frequent sub graphs is a popular branch: Graphs are represented as (usually binary) vectors, with components indicating whether a graph contains a particular sub graph that is frequent across the dataset. On large graphs, however, one faces the enormous problem that the number of these frequent sub graphs may grow exponentially with the size of the graphs, but only few of them possess enough discriminative power to make them useful for graph classification. Very frequent sub graphs are not useful since they are not discriminative between classes. Therefore, frequent sub graph based classification usually sets up a pretty low frequency threshold, resulting in thousands or even millions of features. Given such a tremendous number of features, a complicated feature select ion mechanism is likely to fail. Efficient and discriminative feature selection among frequent sub graphs is hence a key challenge for graph mining. The author proposed an approach to feature selection on frequent sub graphs, called CORK that combines two central advantages. First, it optimizes a sub modular quality criterion, which means that can yield a near-optimal solution using greedy feature selection. Second, our sub modular quality function criterion can be integrated into g Span, the state-of-the-art tool for frequent sub graph mining, and help to prune the search space for discriminative frequent sub graphs even during frequent sub graph mining. Consequently, should need an efficient algorithm to select discriminative features among a large number of frequent sub graphs. In earlier work, we adopted a heuristic approach and demonstrated that it could outperform methods using low dimensional features. The goal is to define an efficient near-optimal approach to feature selection among frequent sub graphs generated by g Span. The key idea is to pick frequent sub graphs that greedily maximize a sub modular quality criterion, thereby guaranteeing that the greedy solution to the feature selection problem is close to the global optimal solution. To make this approach efficient, that integrate it into g Span, the state-of-the-art tool for frequent sub graph mining, and derive pruning criteria that allow us to narrow down the search space when looking for discriminative sub graphs.

L. De Raedt et.al.[10] presented that the local pattern mining algorithms generate sets of patterns, which are typically not directly useful and have to be further processed before actual application or interpretation. A key contribution is that show how well known properties from local pattern mining, such as monotonicity and anti-monotonicity, can be adapted for use in pattern set mining. Finally, do the similarities between local pattern and global model mining also raise some general questions and not solved globally.

X. Yan et.al.[11] suggested that the major challenge of frequent-pattern mining is not at the efficiency but at the interpretability, the huge number of frequent patterns makes the patterns themselves difficult to explore, thus hampering the individual and global analysis of discovered patterns. How to summarize a collection of itemset patterns using only K representatives, a small number of patterns that a user can handle easily.

A. Potential Difficulties

The following difficulties found in the related work.

1. It didn't focus on global constraints relationship.
2. Negative support of itemsets or that refers to measures other than support.
3. A large number of discriminative itemset can be generated during the mining step
4. Limited constraints.
5. The number of overlapping clustering found in different data sets, for varying patterns.

III. PROPOSED TECHNIQUE

Literatures in Section II have some potential difficulties. Proposed FLFCWS framework presented in this paper is used to overcome those difficulties. In this proposed work, High Dimensional Global Pattern Set Mining (HDGPS) Genetic is classified using the two following approaches:

1. HDGPS Genetic from high dimensional data into low dimensional data. Genetic algorithm is a search heuristic that mimics the process of natural selection. This heuristic is normally used to generate useful solutions to optimization and search problem of high dimensional data. The constructed features will be used to generate relevant patterns that characterize non- target records associated to the target record as an input representation for data summarization process. Numerous feature scoring measures are used as fitness function to find the best set of constructed features. Extract the high dimensional data into low dimensional data for constructing global pattern easily.

2. Data summarization algorithm for FLFCWS. In data summarization, the representation of the multi-instances stored in non-target tables that have many-to-one relationship with record stored in target table influences the descriptive accuracy of the summarized data.

A. NP-Pattern set Mining

Each resulting item set corresponds to a set of local patterns and is therefore a pattern set. The constraints implement the desired size NP and correctness. The constraint c specifies both limited and universal constraints at the general pattern set level. The NP-pattern set consists of NP individual itemset patterns. Every pattern is represented by its item set and transaction set. The k-pattern setting is configured by selected pattern to set the heuristic value and find out the specific constraints value in mining values. Pattern mining is concerned with finding all patterns that adhere to some constraint p. The patterns are distinct by a pattern language L. The constraint c is usually a combination of multiple constraints. All patterns that stick on to some chosen local constraints are mined fully. Then, these patterns are combined beneath a set of global constraints, often including an optimization function f. Because of the size of the local pattern set, normally heuristic techniques are used when penetrating for the approximately best pattern set. The constraint p specifies both local and global constraints at the overall pattern set level. In addition, as the number of pattern sets can become extremely large, then study how to find the best pattern set with respect to an optimization criterion.

B. Instantiation Mining

One of the contributions of this paper is that shows how it can be instantiated to address several well-known data mining problems as given below.

B1. Abstract Clustering

Each cluster is described by a pattern covered by the pattern are part of the cluster. Then formalize the abstract clustering as resulting pattern sets that do not overlap and cover all sets. The minimum cluster size would be maximal if all clusters have the same size; hence this formulation will prefer more fair solutions.

B2. NP-tiling

The main aspire of tiling is to cover up as many 1s in a binary matrix with a given number of patterns or tiles. A tiling can be measured useful as the patterns in a tiling are in some way most characteristic for the data. The closeness constraint is not strictly necessary, but closed item sets cover a larger area than their non closed counterparts.

B3. Redescription Mining

The main endeavor of redescription mining is to find sets of syntactically special formulas that all envelop the same set of transactions; such sets of formulas are of interest as they point toward equivalences in the attributes in the data. The frequency of one pattern needs to be maximized as all patterns have to cover exactly the same transactions.

C. Constraint Families

A constraint is localized when it is characterized on one individual pattern and international when it is characterized on many patterns. This definition of a global constraint disagrees from the usual definition in constraint programming, where a global constraint indicates a constraint that connects a number of variables in a nontrivial way. Throughout the study of the effectiveness of each constraint, that recognizes two new and special categories. Within the class of localized constraints we recognize localized look-ahead constraints, and inside the class of international constraints, the pairwise international constraints.

C1. Entity Pattern Constraints

Entity pattern constraints are by definition local constraints. It is not distinguishing local constraints by being monotonic or anti monotonic with respect to set inclusion, as frequent in constraint-based mining. In the constraint programming structure, always calculate bounds on the domains of variables, which are not tied to set inclusion/exclusion specifically. Note down from a constraint programming perspective, every constraint is monotonic with respect to the domain of the variables, as a propagator can only remove values from it.

$\Pi=(I=\psi(T)=\{i\varepsilon L \mid t\varepsilon T(t,i)\varepsilon D\})$

I=closure operator

C2. Idleness Constraints

In CP the cardinality of a set can in standard be designed by summing 0/1 variables on behalf of the fundamentals in the set. However, to deal with idleness, often require to analyze the cardinality of a set which is the outcome of comparing two other sets. Instead of the intermediary set with further variables would construct our details weighty. This constraint would not broadcast very well as a change in one pattern will not openly power the other patterns. Naturally a lot of the distances would have to be identified before the constraint can circulate a change on the other distances. An improved solution would be to constrain every distance separately.

Distinct $(\pi^1, \pi 2) = (\mid T1 Ụ T2) \backslash (T1 \Omega T2)$

T=Transaction set

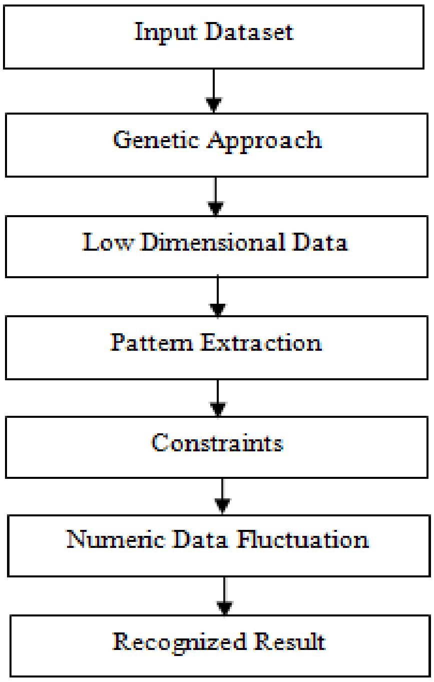
Π=pattern set

C3. Coverage constraints and Discriminative constraints

The wrap of the whole pattern set depends on the cover of each individual pattern. If a transaction is enclosed by one pattern, it is enclosed by the pattern set. Therefore, a transaction is together with this. Coverage and discriminative constraints, which are principally defined on the transaction groups of individual patterns, can also be characterized on such provisional variables. Because of the indirect relative between patterns through these provisional variables, coverage, and discriminative constraints are categorized as usual global constraints.

Coverage(P)▬$(T=\pi(I)=\{t€T \mid ▼i€I:(t,i)€D\})$
T=transaction
I=Itemset
Pattern(π)=(I,T)
Discriminative(p)=freq$^{+}(\pi)$=$|$cover$^{+}(\pi)|=|T^{II}\Omega T^{+}|$
Discriminative(n)=freq$^{-}(\pi)$=$|$cover$^{-}(\pi)|=|T^{II}\Omega T^{-}|$
p=positive example
n=negative example
π=pattern set
T^{+}=total no. of positive
T^{-}=total no. of negative

D. *System Architecture*



E. Algorithm

1. Read an Input Data.
2. Pre-processing stage- Quality Data.
3. Generate random population of n chromosomes (suitable solutions for the problem) range of the candidate road sign color.
4. Evaluate the fitness f(x) of each chromosome x in the population.
5. Create a new population by repeating following steps until the new population is complete.
   - Select two parent chromosomes from a population according to their fitness (the better fitness, the bigger chance to be selected).
   - With a crossover probability cross over the parents to form a new offspring (children). If no crossover was performed, offspring is an exact copy of parents.
   - With a mutation probability mutate new offspring at each locus (position in chromosome).
   - [Accepting] Place new offspring in a new population.
6. [Replace] Use new generated population for a further run of algorithm.
7. [Test] If the end condition is satisfied, stop, and return the best solution in current population.

8. [Loop] Go to step 4.
9. Get the low dimensional data from the output of genetic.
10. Create pattern and apply constraints.
11. Recognize the result

IV. Results And Analysis

The datasets are from the UCI Machine Learning repository and the FIMI repository [Voting]. This framework has been implemented in JAVA and experiments have been run on a Pentium IV with 1GB of memory.Fig.1 shown the performance result of HDGPS mining which is used in less compile time and fast retrievals of itemset. i.e Whenever the iterations are more/complexity it will take less compile time.

Scalability

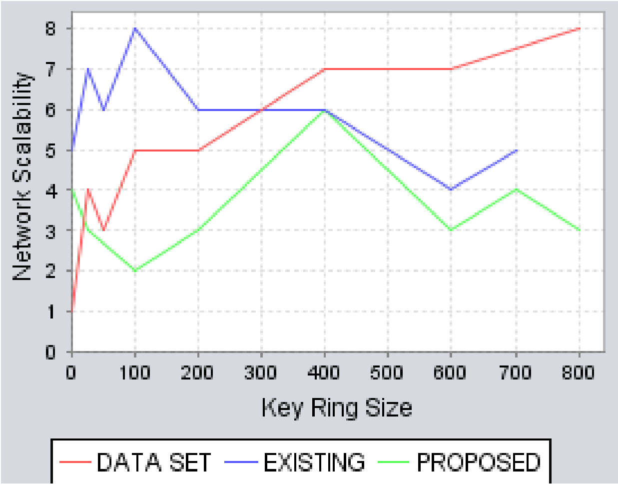


Fig.1 Performance Graph of HDGPS mining

V. Conclusion

The presented HDGPS Genetic framework implements the FLFCWS for data mining tools. The advantage of proposed system is that the framework is to formalize mining tasks as problems of finding NP patterns that together satisfy constraints. Constraint programming is a standard framework for solving combinatorial and optimization difficulty beneath constraints. The new troubles can be solved by only altering the specification in terms of constraints. To search values of exact patterns can be easily derived from the database used a numeric values. It does not partly cover of the data values. Extract the high dimensional data into low dimensional data for constructing low clustering pattern size. It allows build guidelines in how to model new troubles successfully and how to mold existing problems more powerfully. It also untie up the technique for other solver technologies.

REFERENCES

[1] Tias Guns, Siegfried Nijssen, and Luc De Raedt ”k-Pattern Set Mining under Constraints” IEEE Transactions on knowledge and data engineering, Vol. 25, No. 2, February 2013.

[2] L. De Raedt, T. Guns, and S. Nijssen, “Constraint Programming for Itemset Mining,” Proc. 14th ACM SIGKDD Int’l Conf. Knowledge Discovery and Data Mining (KDD), pp. 204-212, 2008.

[3] T. Guns, S. Nijssen, and L. De Raedt, “Itemset Mining: A Constraint Programming Perspective,” Artificial Intelligence,vol. 175, nos. 12/13, pp. 1951-1983, 2011

[4] S. Nijssen, T. Guns, and L. De Raedt, “Correlated Itemset Mining in ROC Space: A Constraint Programming Approach,” Proc. 15th ACM SIGKDD Int’l Conf. Knowledge Discovery and Data Mining (KDD), pp. 647-656, 2009.

[5] H. Cheng, X. Yan, J. Han, and P.S. Yu, “Direct Discriminative Pattern Mining for Effective Classification,” Proc. IEEE 24th Int’l Conf. Data Eng. (ICDE), pp. 169-178, 2008.

[6] M. Khiari, P. Boizumault, and B. Cre´milleux, “Constraint Programming for Mining N-Ary Patterns,” Proc. 16th Int’l Conf. Principles and Practice of Constraint Programming (CP), pp. 552-567,2010.

[7] S. Ruggieri, “Frequent Regular Itemset Mining,” Proc. 16th ACM Int’l Conf. Knowledge Discovery and Data Mining (KDD), pp. 263-272, 2010.

[8] W. Li, J. Han, and J. Pei, “CMAR: Accurate and Efficient Classification Based on Multiple Class-Association Rules,” Proc. IEEE Int’l Conf. Data Mining (ICDM), pp. 369-376, 2001.

[9] M. Thoma, H. Cheng, A. Gretton, J. Han, H.-P. Kriegel, A.J. Smola,L. Song, P.S. Yu, X. Yan, and K.M. Borgwardt, "Near-Optimal Supervised Feature Selection among Frequent Subgraphs," Proc. SIAM Int'l Conf. Data Mining (SDM), pp. 1075-1086, 2009

[10] L. De Raedt and A. Zimmermann, "Constraint-Based Pattern Set Mining," Proc. SIAM Int'l Conf. Data Mining (SDM), pp. 1-12, 2007.

[11] X. Yan, H. Cheng, J. Han, and D. Xin, "Summarizing Itemset Patterns: A Profile-Based Approach," Proc. Int'l Conf. Knowledge Discovery and Data Mining (KDD), pp. 314-323, 2005.

[12] Alan Frisch, Warwick Harvey, Christopher Jefferson, Bernadette M. Hern´andez, and Ian Miguel.Essence: A constraint language for specifying combinatorial problems. Constraints, 13(3):268–306, 2008.

[13] Matti J¨arvisalo. Itemset mining as a challenge application for answer set enumeration. In Logic Programming and Nonmonotonic Reasoning, volume 6645 of LNCS, pages 304–310. Springer, 2011

[14] Pascal Van Hentenryck and Laurent Michel. Constraint-Based Local Search. MIT Press, 2005.

[15] F. Bonchi, F. Giannotti, C. Lucchese, S. Orlando, R. Perego, and R.Trasarti, "A Constraint-Based Querying System for Exploratory Pattern Discovery," Information Systems, vol. 34, pp. 3-27, 2009.

[16] F. Rossi, P. van Beek, and T. Walsh, Handbook of Constraint Programming. Elsevier, 2006.