International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 02 Issue: 07 | Oct-2015 www.irjet.net p-ISSN: 2395-0072

ANAFDUKCD: MIXED MINING

Shrinivas D. Gaikawad[1], Ashish M. Pawar[2], Navnath L. Mohite[3], Prashant D. Kamble[4] , Sonali S. Muley[5]

[1] *Student, Department of Computer Engineering, MMIT Pune, Maharashtra, India*
[2] *Student, Department of Computer Engineering, MMIT Pune, Maharashtra, India*
[3] *Student, Department of Computer Engineering, MMIT Pune, Maharashtra, India*
[4] *Student, Department of Computer Engineering, MMIT Pune, Maharashtra, India*
[5] *Professor, Department of Computer Engineering, MMIT Pune, Maharashtra, India*

Abstract - ***Venture data mining applications often involve complex data such as multiple large Heterogeneous data sources, user preferences, and Business Intelligence. In such situations, a single method or one-step mining is often limited in discovering informative knowledge. It is essential to develop effective approaches for mining patterns combining necessary information from multiple relevant business lines. This system builds on our existing works and proposes combined mining as a general approach to mining for instructive patterns combining components from either multiple data sets or multiple features or by multiple methods on demand. Types of combined patterns, such as incremental cluster patterns, can result from such frameworks, which cannot be produced by the existing methods. .***

Key Words: *Combined mining, Complex data, Data mining, Useful Knowledge Discovery, Multiple source data mining, Public service data mining, etc…*

1. INTRODUCTION

1.1 Overview

Combined mining is a 2 to multi step data mining approach, which involves first mining the atomic patterns from each separate data source and then combines those atomic patterns into combined-patterns by pattern-merging method, which is more suitable for a individual problem. In multi-source combined mining approach, we first find the informative patterns from separate data source and then generate the combined patterns, which can't be directly generated by some conventional algorithms like FP-growth etc. In multi-feature combined mining approach, we consider features from multiple data sets while generating the instructive patterns, where it is necessary in order to make the patterns more actionable. In case of cluster patterns, we produce the cluster of patterns with same prefix but the remaining data items in the pattern make the results to be different. The challenges come from many aspects, for instance, the traditional methods usually produce homogeneous features from a single source of data while it is not effective to mine for patterns combining components from various data sources. It is often very costly and sometimes impossible to join multiple data sources into a single data set for pattern mining. The need for developing effective techniques for involving multiple heterogeneous features, data sets, methods in Venture data mining.

We proposed the concepts of combined association rules, mixed rule pairs, and mixed rule clusters to mine for informative patterns in complex data by catering for the comprehensive aspects in various data sets. A combined association rule is composed of multiple heterogeneous item-sets from various data sets while mixed rule pairs and mixed rule clusters are built from combined association rules. Analysis shows that such combined rules cannot be produced by traditional algorithms such as the FP Growth.

The basic ideas of combined mining are as follows.

1. By involving various heterogeneous features, combined patterns are generated which reflect multiple aspects of care and characteristics in businesses.

2. By mining multiple data sources, mixed patterns are generated which reflect various aspects of nature across the business lines.

3. By applying multiple methods in pattern mining, mixed patterns are generated which disclose a deep and comprehensive centre of data by taking advantage of different methods.

4. By applying multiple interesting metrics in pattern mining, patterns are generated which reflect care and significance from multiple perspectives.

As we will study in Section 2.1 about Literature Survey, the existing works in handling the above mentioned challenges can be categorized into the following:
1) Data sampling
2) Joining multiple relational tables
3) Post analysis and mining
4) Involving various methods and
5) Mining multiple data sources.

In real-life data mining, data sampling is frequently not acceptable since it may miss useful data that are filtered out. Table joining may not be able due to the time and space limit such as in dealing with millions of transactions from multiple sources. In addition, techniques for involving various methods and handling various data sources are often specifically developed for particular cases.

1.2 Problem Definition

The goal of this system is to design an accurate and efficient dataset classifier with good scalability, which should be able to overcome the problems of both the conventional dataset and the recently proposed association-based classifiers. The single mining is frequently limited in discovering informative knowledge. It would be very time and space consuming. Venture data mining applications frequently involve complex data such as various large heterogeneous data sources, user preferences, and business impact. In existing system only single mining is used i.e. single method mining is limited to produce informative knowledge in complex data. mixed rules cannot be produced by basic algorithms such as the FP Growth. It is often costly and sometimes impossible to join various data sources into a single data set for pattern mining. The need for developing effective techniques for involving various heterogeneous features, data sets and methods in venture data mining. We proposed the concepts of mixed association rules, mixed rule pairs, and mixed rule clusters to mine for useful patterns in complex data by humor for the comprehensive aspects in various data sets. A mixed association rule is composed of various heterogeneous item-sets from various data sets while combined rule pairs and mixed rule clusters are built from mixed association rules. By applying various methods in pattern mining, mixed patterns are generated which disclose a deep and comprehensive centre of data by taking advantage of different methods.

2. LITERATURE SURVEY

It is challenging to mine for comprehensive and useful knowledge in such complex data suited to real-life decision needs by using the available methods. The challenges come from many aspects, for instance, the traditional methods normally discover homogeneous features from a single source of data while it is not effective to mine for patterns combining components from various data sources. It is often very costly and sometimes impossible to join multiple data sources into a single data set for pattern mining.

2.1 Existing System

In 2005, Wang etal. proposed the HARMONY algorithm, which improved the efficiency much, specially when min-sup is not very low. In their algorithm, only the rules with highest confidence covering every sample are kept while the others are pruned. However it still follows the two stage framework but employs pruning in mining process. When the min sup is set a low value, the searching space cannot be pruned much.

Advantages:
Well for accuracy and computational efficiency with respect to database size.

Disadvantages:
HARMONY algorithm is not suitable for large database.

In 1999, G. Dong and J. Li defined a new concept of knowledge discovery by using new type of patterns that is emerging patterns (EPs); EPs are used to build very powerful classifiers.

Advantages:
Build very much powerful classifiers which are more accurate than C4.5 and CBA.

Disadvantages:
1. EP algorithm manupulates EPs patterns using Borders only.
2. Without borders it is time consuming to mine Eps.

In 2011, C. Zhang proposed combined mining as a basic approach to mine the informative patterns. They introduced various types of combined mining, basic process of combined mining.

In 2004, Jian Pie and Jiawei Han promote a divide-and-conquer approach, called pattern-growth approach, which is an extension of FP-growth, an efficient pattern-growth algorithm for mining frequent patterns without candidate generation. An easy pattern growth method, Prefix Span, is proposed.

Advantages:
It allows sequential pattern mining.

Disadvantages:
FP-Growth algorithm can't generate combined pattern directly.

First the most of existing single-handed data mining methods don't target the discovery of informative patterns in complex data. Second, approaches to mining for more informative and important knowledge in complex data can be generally categorized as follows:

1. Direct mining by finding effective approaches;
2. Post analysis and post mining of learned patterns;
3. Involving extra features from other data sets;
4. Integrating various methods;
5. Joining multiple relational tables.

2.2 Proposed System

This system builds on our available works and proposes mixed mining as a general approach to mining for informative patterns combining components from either various data sets or multiple features or by multiple methods on demand.

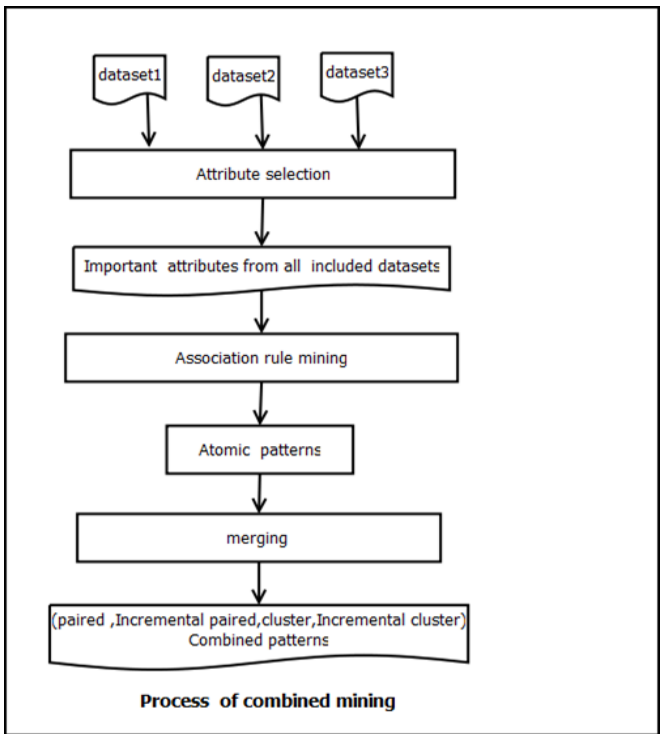

Fig -1: Process of Combined Mining

Combined mining is a 2 to multi step data mining approach, which involves first mining the atomic patterns from each individual data source and then combines those individual patterns into mixed patterns by pattern-merging method, which is more suitable for a separate problem. In multi-source mixed mining approach, we first find the useful patterns from separate data source and then generate the mixed patterns, which cant be generated by some basic algorithms like FP-growth etc. In multi-feature mixed mining approach, we consider features from various data sets while generating the useful patterns, where it is required in order to make the patterns more useful. In case of cluster patterns, we made the cluster of patterns with equal prefix but the remaining data items in the pattern make the results to be variant. In our system ,we are going to implement the concepts of mixed association rules, mixed rule pairs, and mixed rule clusters to mine for useful patterns in complex data by catering for the comprehensive aspects in multiple data sets. A combined association rule is composed of multiple heterogeneous item sets from various data sets while mixed rule pairs and mixed rule clusters are built from mixed association rules.

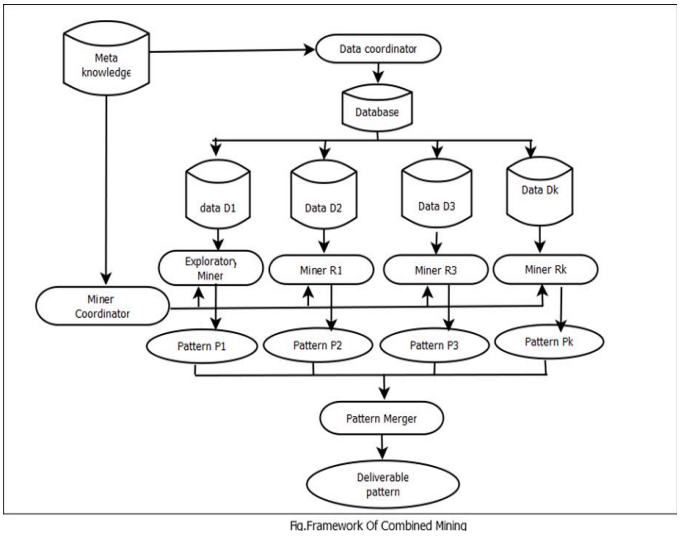

Fig -2: Framework of Combined Mining

It supports the discovery of mixed patterns either in multiple data sets or subsets (D1, . . . ,DK) through data partitioning mixed mining for actionable patterns. the following manner:

1. Based on domain knowledge, business understanding, and goal definition, one of the data sets (say D1) are selected for mining observation (R1);

2. The findings are used to guide either data partition through the data coordinator and to design policies for managing and conducting serial or parallel pattern mining on applicable data sets or subsets or mining respective patterns on related remaining data sets; the deployment of method Rk (k = 2, . . . , L), which could be either in parallel or through combination, is knowledgeable by the understanding of the data and objectives, and if required,

another step of pattern mining is conducted on data set Dk with the supervision of the results from step k 1;

3. After finishing the mining of all data sets, patterns (PRn) identified from particular data sets are merged (GPn) with the participation of domain knowledge and further extracted into final deliverables (P). In this way the system architecture is explained.

3. ALGORITHM

Venture data mining applications often include complex data such as multiple large heterogeneous data sources, user preferences, and business impact. Then there are several steps for getting predicted results from the all process of mining. All these algorithms are compared according to various parts like type of data set, support counting, rule generation, candidate generation and some other element .The compared algorithms are presented together with some examples that lead to the final conclusions. Association rules are widely used in various areas such as telecommunication, market and risk management, inventory control etc.

Following are the algorithms for used in these Systems:

1] Analysis for Multisource Combined Mining

For input datasets 'n' requires 'm' no. of transactions and 'p' no of patterns are evaluated. 'k' no of pattern merging methods are applied so it requires O(k) iterations.

Time Complexity:O(nm)+O(Pk)

2] Analysis for Apriori algorithm

For calculating support value for all the transaction requires O(n) time complexity.

Time Complexity: O(mo)+O(n)

3] **Analysis For Multifeature Combined Mining**

For input dataset requires 'm' no. of transactions and 'p' no of patterns are evaluated 'k' no of pattern merging methods are applied so it requires O(k) iterations.

Time Complexity: O(m)+O(Pk)

4] Analysis For Multimethod Combined Mining

For input datasets requires 'm' no. of transactions and 'p' no of patterns are evaluated using multiple methods, 'k' no of pattern merging methods are applied so it requires O(k) iterations.

Time Complexity: O(m2)+O(Pk)

Where,
n = No. of datasets.
m = No. of transaction.
p = No. of Patterns.
k = No. of patterns merging method

4. CONCLUSIONS

Combined mining to discover the informative knowledge in complex data , in which mining algorithms are applied on multiple input datasets which are relevant to each other. Combined mining as a general approach to mining for informative patterns combining components from either multiple data sets or multiple features or by multiple methods on demand. Multiple merging algorithms are used to obtain the combined pattern, which gives more informative knowledge. Combined patterns gives the more accurate and actionable prediction. The identified combined patterns are more informative and actionable than any single patterns identified in the traditional way.

ACKNOWLEDGEMENT

I am deeply indebted to my project Guide **Prof. Muley S. S.** and **Prof. Patil S. K.** from the Computer Department for contributing to the completion of project and helped me with valuable suggestions for improvement.

I would like to express my gratitude to Principal **Dr. Chavan D. K.** and HOD **Prof. Daflapurkar P. M .**with best facilities and atmosphere for providing me creative work guidance and encouragement. My all friends, lab assistants and family members supported me in my project work. I thank to all above people for extending their cooperation during the project.

REFERENCES

[1] C. Zhang, D. Luo, H. Zhang, L. Cao and Y. Zhao, IEEE "Combined Mining: Discovering Informative Knowledge in Complex Data," *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS* VOL. 41, NO. 3, YEAR 2012.

[2] C. Zhang, D. Luo, H. Zhang, L. Cao and Y. Zhao (2011), IEEE "Combined Mining: Discovering Informative Knowledge in Complex Data," *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS* 2011.

[3] L. Cao, Y. Zhao, H. Zhang, D. Luo, and C. Zhang, "Flexible frameworks for actionable knowledge discovery," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 9, pp. 12991312, Sep. 2010.
[4] Dong and J. Li, "Efficient mining of emerging patterns: Discovering trends and differences," *in Proc. KDD,* 1999, pp. 4352.
[5] J. Wang and G. Karypis, "HARMONY: Efficiently mining the best rules for classification," *in Proc. SDM,* 2005, pp. 205216.

BIOGRAPHIES

Shrinivas Dasharath Gaikawad, Student, Department of Computer Engineering, Marathwada Mitra Mandal's Institute of Technology, Lohagaon, Pune. Maharashtra (INDIA).

Ashish Mohan Pawar, Student, Department of Computer Engineering, Marathwada Mitra Mandal's Institute of Technology, Lohagaon, Pune. Maharashtra (INDIA).

Navnath Laxman Mohite, Student, Department of Computer Engineering, Marathwada Mitra Mandal's Institute of Technology, Lohagaon, Pune. Maharashtra (INDIA).

Prashant Devidas Kamble, Student, Department of Computer Engineering, Marathwada Mitra Mandal's Institute of Technology, Lohagaon, Pune. Maharashtra (INDIA).