Discovering in Formative Knowledge using Combined Mining Approach

Aniket A. Yadav
Dept. of Information Technology
Smt. Kashibai Navale College of Engineering,
Pune India

ABSTRACT
Data mining applications involve complex data like multiple heterogeneous data sources, different user preference and create decision making activities. The comprehensive, useful information may not be obtained by using the single data mining method in the form of informative patterns as that would consume more time and space. Combined mining is a hybrid mining approach for mining informative patterns from single or multiple data sources, many features abstraction and merging multiple methods as per the requirements. Some concepts show hybrid or combined mining approach. In this paper, multi method combined mining methodology is designed. Proposed method combine apriori algorithm with Multi-Objective Evolutionary Algorithm. It helps to improve searching for the exact products from complex data. Proposed method combine apriori algorithm with multi-objective evolutionary algorithm to improve searching of documents from complex data. It addressed challenging problems in combined mining and summarized and proposed effective pattern merging and interaction paradigms, combined pattern types, such as pair patterns and cluster patterns, interestingness measures. In the proposed method, by using apriori algorithm, calculate the support and confidence of the frequent item set which improve results of searching by a using multi-objective evolutionary algorithm.

Keywords
Combined mining approach, Multi-objective Evolutionary algorithm, Apriori algorithm, Knowledge Discovery combined mining, complex data, data mining, multiple source data mining.

1. INTRODUCTION
Data mining tools are used to predict behavior businesses, allowing and future trends, to make knowledge-driven decisions and proactive, also data mining tools can help to answer business questions that have traditionally been too time consuming with finding the proper way to resolve. A complex database appears to a user as a single database but is, in fact, a set of databases stored in multiple tables. The data on several tables can be simultaneously accessed, modified and search by user. The goal of the project is to present a fast complex data mining algorithm, which should be based on the idea of capturing mined data with user desired results.

The exponentially increasing amounts of data being generated each year make getting useful information from critical data. The information frequently is stored in a data warehouse which is a repository of data collected from various sources such as corporate databases, data from external sources and summarized information from internal systems. This collected information and data called as complex data. Complex data include variable forms of data [5].

Distributed data mining algorithms [Kargupta and Park (2002)] [7], data sampling is generally not accepted because it may miss useful data that may be filtered out during sampling. Distributed data sets need to join into one large data set but the process may be more time and space consuming. More often such approach of handling multiple data sources can only be developed for specific cases and cannot be applied to all of the domain problems. There are two multistep data mining approach in combined mining. In the first step, it involves mining the atomic patterns from each individual data source and then next step combines atomic patterns into combined-patterns. For particular problem which is more suitable. In multi-source combined mining approach, it generates informative patterns from individual data source and then generates the combined patterns.

Lonfbing Cao [13] proposed a Combined Mining: Analyzing Object and Pattern Relations for Discovering Actionable Complex Patterns. They also briefly illustrate the concepts and discuss how they can be applied to mining complex data for complex knowledge in either a multi-feature, multi-source, or multi-method scenario. Combined pattern and combined mining present a general paradigm with great potential for identifying and producing informative and actionable patterns.

The goal of the proposed system is to present a fast complex data mining algorithm, which should be based on the idea of capturing mined data with the user desired result.

The remaining paper is organized as follows:

Section I Introduction. Section II Related work. Section III Implementation detail. IV Experimental Setup Section V Results and discussions and Section VI conclusion and future scope.

2. RELATED WORK
Dong and J. Li (1999) [8] introduce a new type of Pattern, i.e. emerging patterns (EPs), for discovering knowledge from databases. The Authors define EPs as data item-sets whose support increase more significantly from one to another data set. The Author has used EPs to build very powerful classifiers. W. Fan et al (2008) [9] builds a model based on data search tree, which partitions the data onto different nodes. Each node directly find out a discriminative pattern, this pattern further divide its examples in more pure subsets.

A novel technique was proposed by B. Liu et al. (1999) [16], which first prune the discovered association-rules to remove the insignificant association rules from the entire set

Of association-rules, then finds a subset of the un-pruned association-rules by which a summary of the discovered association-rules can be formed. The algorithm refers it as a subset of association rules as the direction setting (DS) rules because they can be used to set the directions, these directions

are followed by the rest of the association rules. With the help of the summary, the user can have more focus on the important aspects of the particular domain and also can view the relevant details.

By Bing Liu, Wynne Hsu, Yiming Ma [10] in this approach they define association rules are a fundamental class of patterns that exist in the data. The key strength of association rule mining is completely of mining. It finds all associations in the data that satisfy the user specified minimum support and minimum confidence constraints. This strength, however, comes with a major drawback. It often produces a huge number of associations. This is particularly true for data sets whose attributes are highly correlated. The huge number of associations makes it very difficult, if not impossible, for a human user to analyze in order to identify those interesting/useful ones. The technique first prunes the discovered associations to remove those associations which are insignificant, and finds a special subset of the unproved associations to form a summary of the discovered associations.

H. Yu et al. (2003) [14] proposed a new method called as Clustering-Based SVM (CB-SVM), in which, the whole data set can be scanned only once to have an SVM with samples that carry the statistical information of the data by applying a hierarchical micro-clustering algorithm. The Authors also show that CB-SVM is also highly scalable for very large data sets and also generating very high classification accuracy. Longbing Cao (2012) [18] proposed combined mining as an approach from the perspective of object and pattern relation analysis.

3. IMPLEMENTATION DETAILS

In this paper, the system design for improving performance of searching of data in complex database. The Proposed method is combined Multi Objective Evolutionary Algorithm with apriori algorithm which is used for rule mining.

3.1. MOEA

MOEAs have been widely used for classification of complex data. The most commonly studied approach is the use of MOEAs of evolving a good set of classification rules, in the system rules created by association Rule mining. The second approach is to employ MOEAs to Dene the class boundaries (hyper planes) in the training data. The last approach is to use MOEAs for training and to model the construction of well-known classifiers such as neural networks and decision tree classifiers.

In the proposed method admin trained the dataset using the a priori algorithm. In apriori algorithm calculate the candidate item set and minimum support.

3.2. System Overview

The proposed method decomposed problem statements into different modules as admin, user, and query searching using rules created by an apriori algorithm with a Multi Objective Evolutionary algorithm.

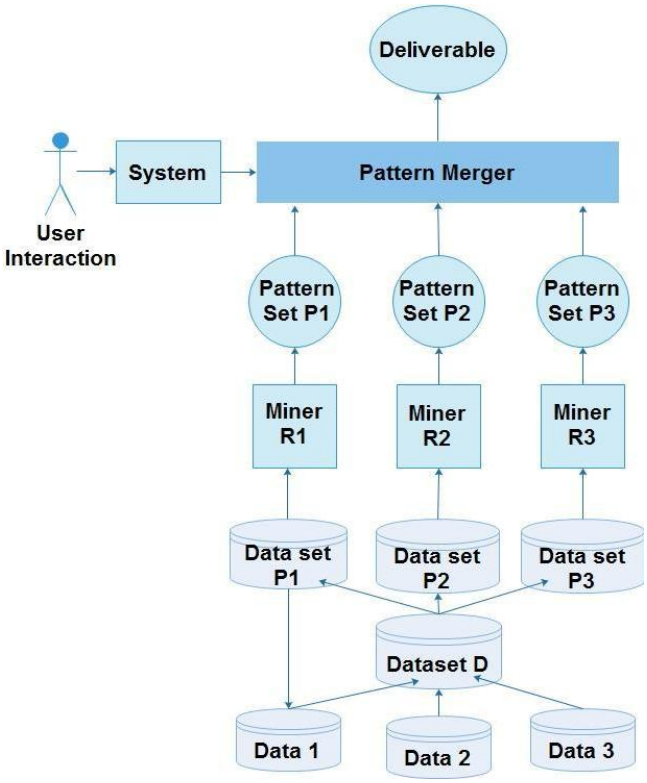

Fig.1. System Architecture

Following modules are explained for Mining of complex data using combined mining approach

a. **Admin login:** In This Module admin maintained various Food, and Electronic product details using relational database with respective attribute. The databases contain attribute like cost, type, name, quantity, and Status of Food and Electronic products.
b. **Query-similarity:** The main purpose of implemented system is reducing searching time required to search proper user desired results. To solve this the index concept used. Indexing actually used to reduce execution time.

When customer login and search the Food and Electronic details with a specific query, then product details will extract from the database using rule mining and COEA. Details are displayed from different types of databases, using join query. Also user can give feedback to that product.

c. **Searching Criteria:** For product searching, Multi-Objective Evolutionary algorithm used. From which User get proper results from complex data. For creating patterns and rules apriori algorithm is used.

This system provides following search criteria

i. Multi Method: User can search products by keyword with the index or keywords without an index.

Proposed system uses covering index. No any constraints are enforced during the index creation phase. Hence, the result extraction can be performed by only index, without accessing the original database.

ii. Multi feature: User can select multiple features of product for searching.

iii. Display result: Results are displayed with support and confidence count.

3.3. Constraints

1. The Database should be loaded.
2. Provide login and password for each user.
3. Only registered user able to view the product.
4. User search product by inserting query.

3.4. Mathematical Model

- Let, S is system={R, L, CD, A, U, S, O, I}
- R is registration of user= {U1, U2....}
- User U= {u1, u2 ....}
- L is login process
- A is admin module={U, P, I, D}
- I is input for product detail= {I1, I2...}
- D is loaded Database tables= {F, E}
- E is an electronic product detail tables= {ED1, ED2 ...}
- F is an electronic product detail tables= {F D1, F D2...}
- ED contains attributes
  {Name, description, cost, type, brand}
- FD contains attributes {Name, cost, type, description}

Apply apriori algorithm to calculate confidence and support using following algorithm

- L1 =(Frequet itemset of cardinality 1); for(k=1;lk !=0;k++)do begin
- Nk+1 = apriori − gen(Lk);//new candidates for all transactions t N _do n.count++;
- End
- Lk+1 =candidates in Nk+1 with min support end
- end
- return Uk Lk
- IQ is the input query entered by user= {Q1, Q2...}
- O ⊂ E or F

4. EXPERIMENTAL SETUP

For the experimentation work JAVA (Net Beans IDE 8.0) is used with Processor Pentium IV, RAM 1 GB. User can execute system on Windows operating systems like XP/7/Vista. SQL Server is used to store electronic products detail tables.

5. RESULT AND DISCUSSION

5.1 Results

In the proposed method, for product sourcing, multi-objective combined mining approaches are used. For creating patterns and rules the traditional apriori algorithm applied.

The performance of the proposed system is calculated by using Support and confidence in search result. Following graph shows system performance. The designed algorithm provides scalability and reduces execution time.

Fig. 5 shows system performance of old and new support and confidence. Fig.6 shows the difference between old and new support and confidence. Fig.4 shows the overall system performance of support and confidence.

Figure 6 and figure 7 Shows support verses confidence graph of both new and old system.

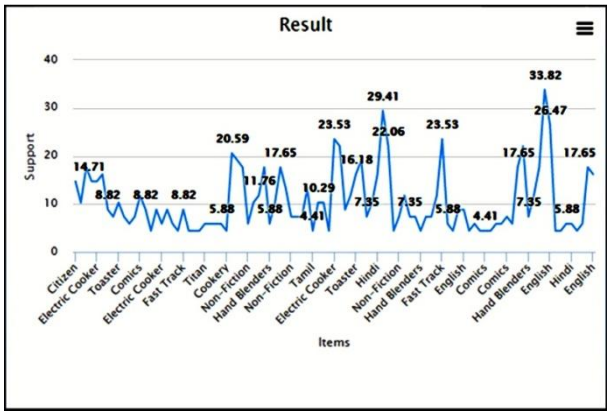

Figure 2: Proposed Support Graph.

Figure 2 shows proposed support graph. Here y axis shows support of items and x axis shows items. Graph numbers gives support level of items.

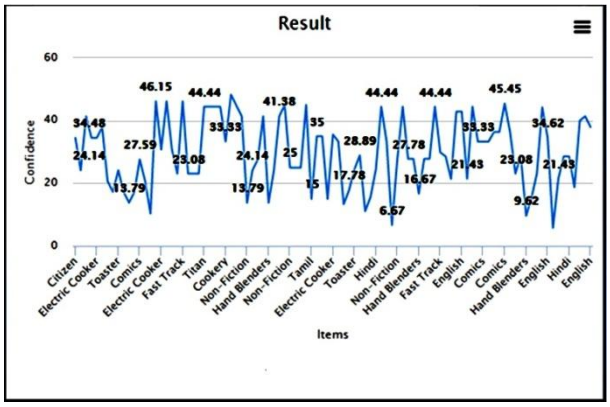

Figure 3: Proposed Confidence Graph

Figure 3 shows proposed confidence graph. Here y axis shows confidence of items and x axis shows items. Graph numbers gives confidence level of items.

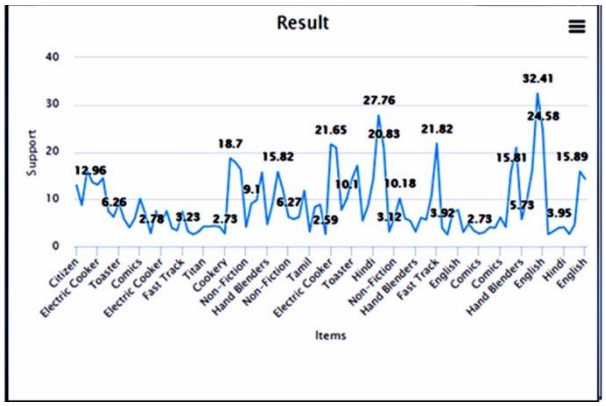

Figure 4: Existing Support Graph

Figure 4 shows existing support graph. Here y axis shows support of items and x axis shows items. Graph numbers gives support level of items.

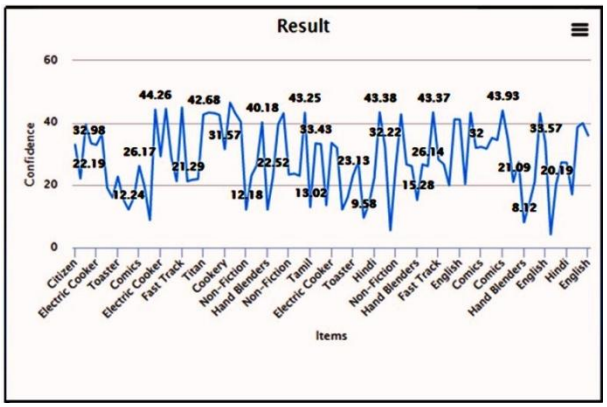

Figure 5: Existing Confidence graph

Figure 5 shows existing confidence graph. Here y axis shows confidence of items and x axis shows items. Graph numbers gives confidence level of items.

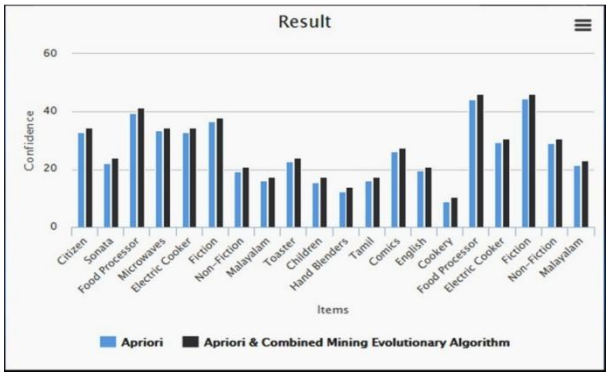

Figure 6: Confidence existing v/s Proposed

Above figure shows proposed confidence results and existing confidence results. Blue bar shows existing Apriori algorithm results and black bar shows the proposed results.

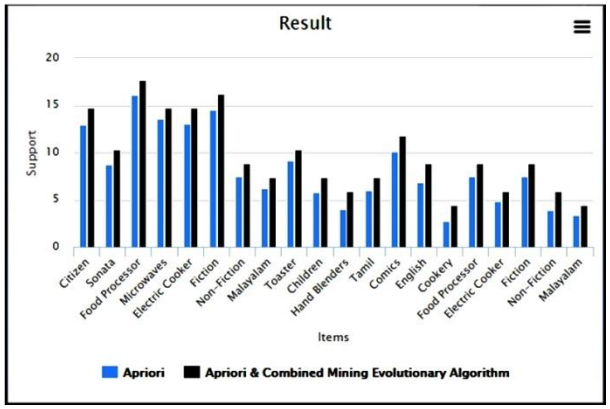

Figure 7: Support existing v/s Proposed

Above figure shows proposed support results and existing support results. Blue bar shows existing Apriori algorithm results and black bar shows the proposed results.

6. CONCLUSION AND FUTURE SCOPE

To design MOEA algorithm studied different existing system related to combine mining approach in data mining. This system comprehensive and general approach named combined mining for discovering informative knowledge in complex data of for any kind of data mining applications. Here I focus on discussing the frameworks for handling multi feature, multi-source related issues. I have addressed challenging problems in combined mining and summarized and proposed effective pattern merging and interaction paradigms, combined pattern types, such as pair patterns and cluster patterns, interestingness measures. In proposed method, by using apriori algorithm, calculate the support and confidence of frequent item set which improve results of searching by using multi objective evolutionary algorithm.

7. ACKNOWLEDGEMENT

We would like to thank the authors, researchers and publishers for making their resources. We would also thank the college authority and teachers for providing the required infrastructure and support. Finally, we would like to extend a heartfelt gratitude to friends and family members.

8. REFERENCES

[1] C. Zhang, D. Luo, H. Zhang, L. Cao and Y. Zhao (2011), "Combined Mining: Discovering Informative Knowledge in Complex Data", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS PART B: CYBERNETICS, VOL. 41, NO. 3, pp. 699-712.

[2] Han and Kamber (2006), "Data Mining Concepts and Techniques", 2nd Ed., United State of America.

[3] C. Zhang, F. Figueiredo, H. Zhang, L. Cao and Y. Zhao (2007), "Mining for combined association rules on multiple datasets", in Proc. DDDM, pp. 18–23.

[4] Longbing Cao (2012), "Combined Mining: Analyzing Object and Pattern Relations for Discovering Actionable Complex Patterns" sponsored by Australian Research Council Discovery Grants (DP1096218 and DP130102691) and an ARC Linkage Grant(LP100200774).

[5] Kavitha, K. And Ramaraj, E. (2012), "Mining Actionable Patterns using Combined Association Rules" International Journal of Current Research, Vol. 4, Issue, 03, pp.117-120, March, 2012.

[6] Jaturon Chattratichat, John Darlington, et al (1999), "An Architecture for Distributed Enterprise Data Mining". Proceedings of the 7th International Conference on High Performance Computing and Networking.

[7] B. Park, D. Hershbereger, E. Johnson and H. Kargupta (1999), Collective data mining: A new perspective toward distributed data mining". Accepted in the Advances in Distributed Data Mining, Eds: Hillol Kargupta and Philip Chan, AAAI/MIT Press (1999).

[8] G. Dong and J. Li (1999), "Efficient mining of emerging patterns: Discovering trends and differences" in Proc. KDD, pp. 43–52.

[9] H. Cheng, J. Han, J. Gao, K. Zhang, O. Verscheure, P. Yu, W. Fan and X. Yan (2008), "Direct mining of discriminative and essential graphical and item set features via model-based search tree" in Proc. KDD, pp. 230–238.

[10] B. Liu, W. Hsu and Y. Ma (1999), "Pruning and summarizing the discovered associations" in Proc. KDD, pp. 125–134.

[11] A. N. Swami, B. Lent and J. Widom (1997), "Clustering association rules" in Proc. ICDE, pp. 220– 231.

[12] J. Han, J. Yang, P. S. Yu and X. Yin (2006), "Efficient Classification across multiple database relations: A CrossMine approach" IEEE Trans. Know. Data Eng., vol. 18, no. 6, pp. 770–783.

[13] C. Zhang, H. Bohlscheid, H. Zhang, L. Cao and Y. Zhao (2008), "Combined pattern mining: From learned rules to actionable knowledge" in Proc. AI, pp. 393–403

[14] H. Yu, J. Han and J. Yang (2003), "Classifying large data sets using SVM with hierarchical clusters" in Proc. KDD, pp. 306–315.

[15] Frank, A. & Asuncion, A. (2010), "UCI Machine Learning Repository" [http://archive. ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[16] David A. Bell, Jie Cheng and Weiru Liu (1997), "An Algorithm for Bayesian Belief Network Construction for Data," In Proceedings of AI & STAT'97, pp. 83-90.

[17] Dr. E. Kumar and A. Solanki (2010), "A Combined Mining Approach and Application in Tax Administration" International Journal of Engineering and Technology Vol.2(2), 2010, 38-44.

[18] C. Zhang, H. Bohlscheid, H. Zhang, L. Cao and Y. Zhao (2008), "Combined pattern mining: From learned rules to actionable knowledge" in Proc. AI, pp. 393–403.

[19] L. Cao, Y. Zhao, H. Zhang, D. Luo, and C. Zhang, "Flexible frameworks for actionable knowledge discovery" IEEE Trans. Know. Data Eng., Vol. 22, no. 9, pp.1299– 1312, Sep. 2010.

[20] M. Antonelli, P. Ducange, and F. Marcelloni, "Multi objective evolutionary rule and condition selection for designing fuzzy rule-based classifiers" in *Proc. FUZZ-IEEE*, Jun. 2012, pp. 1–7.