

Combined Mining Approach and Pattern Discovery in “Online Shopping Application”

Kamble Jayshree Raghunath

Information Technology, Siddhant College of Engineering, Sudumbare, Pune, India

Abstract: *Now a day it's mandatory to use data mining tools and services. Business application data involve data from large heterogeneous data sources as well as data such as user preferences and business impacts. Generally the expectation of business person is to view single view of business settings, analytical reports from complex data that help business users to make decisions. Traditional data mining systems has so many drawbacks in order to overcome the drawbacks; a novel idea is combined mining. Combined mining is effective technique for extracting and constructing actionable complex knowledge, patterns. The most common approaches in data mining have been extended to the multi-relational, multi-feature, multi-source and multi-method combined mining. This paper illustrates the concept of combined association rule to find actionable knowledge from that association rule. A traditional association mining often produces large numbers of association rules and it is difficult for users to understand such rule. Combined mining is a post processing method for association rules generated. In this approach, first the association rule are filtered by varying support and confidence levels, then using the interestingness measure rules, association rules are further extracted. This paper briefly illustrates the concept of combined mining, general frameworks, paradigms and basic processes for combined mining.*

Keywords: Actionable Knowledge Discovery, Combined association rule, Combined Mining, Complex Data.

1. Introduction

In this paper, we used the concept of combined mining [1]. At enterprise level when data mining is done, it has to operate on the huge amount of data, businesses produce complex data that is distributed in nature. When mining is done on such data with a single step, it produces results as a particular aspect. However this is not expected by the business person, to take the business decision different aspects, features of data are to be considered. Enterprise needs to perform mining based on multiple features, data sources, and methods. Patterns generated by using combined mining reflect all aspects of the enterprise while the traditional methods discover homogeneous feature from a single source of data.

The combined mining concept is used for easily handling the multi-feature sets, multi-information sources, constraints, multi-method in data mining of “Online Shopping Application” and for analyzing relations between objects or descriptors or between identified patterns. The proposed concept of combined association rule [2] is composed of multiple heterogeneous item sets from different data sets while combined rule pairs and combined rule clusters are built from combined association rules. This paper implements combined mining approach on customer transaction data in online shopping, in order to improve decision making in marketing campaign, pricing strategy and the other key focus areas. In proposed system, we analyze object and pattern relations [5]. Online transaction data was analyzed which had rich repository of transaction data. This paper demonstrates how we can apply combined mining approach to perform customer segmentation analysis and identify the customer target list for the purpose of coming up with recommendations that can help in decision making processes and actionable knowledge discovery [3]. We conducted a customer segmentation study using a comprehensive set of data which contains both transaction records and customer profiles.

The general ideas of combined mining used in this paper are as follows;

- By involving multiple heterogeneous features, multiple aspects of concern and characteristics in transactions are considered.
- By mining multiple data sources, while generating combined patterns multiple aspects of nature across the business lines are used.
- By applying multiple interestingness metrics in pattern mining, Interestingness metrics are nothing but concerns and significance from multiple perspectives.

This paper organized as follows, Section 2 presents related work, Concepts used in combined mining are introduced in Section 3, Section 4 describes proposed mining framework, Section 5 algorithms used in proposed system; Section 6 Experimental works. This paper is concluded in section 7; this section also includes future work.

2. Related Work

Longbing Cao, Yanchang Zhao, Huaifeng Zhang and E. K. Park (2010) presented Flexible Frameworks for Actionable Knowledge Discovery, which gives formal view of the Actionable Knowledge Discovery, to help in decision making, in business.

Karypis and Wang (2005) introduced a direct mining approach using new classifier, HARMONY, which directly mines the resultant set of rules [4] required for classification.

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

W. Fan (2008) built a model based search tree for partitioning the data onto different nodes and at each node, it

directly finds out a discriminative pattern.

B. Liu, W. Hsu, and Y. Ma (1999) proposed a novel technique, this technique removes the insignificant association rules from set of association-rules and prunes the discovered association-rules then finds a subset of the unpruned association-rules; a summary of the discovered association-rules can be formed.

Lent, Swami and J. Widom (1997) introduced the method for clustering two-dimensional associations in large databases. BitOp a geometric-based algorithm is used for clustering, that is embedded within ARCS (Association Rule Clustering System).

J. Han (2006) presented a new approach of CrossMine and CrossMine-Tree. CrossMine and CrossMine-Tree is a set of novel and powerful methods for multi-relational classification including tuple ID propagation.

C. Zhang (2008), proposed new concept of combined pattern for extracting important, actionable and impact oriented information from a large amount of data and to measure interestingness and analyzes redundancy in patterns

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

3. Concepts Used in Combined Mining

3.1 Combined Mining

Combined mining is a multistep data mining process, in which multiple atomic patterns are generated and that atomic patterns are merged using pattern merging method into a combined pattern set for each data set. Combined mining refers to either one or more of the following aspects.

- The Combination of Multiple Data Sources: The pattern identified from multiple data sources are combined together.
- The Combination of Multiple Features: The pattern set involves multiple features of data set or data sets.
- The Combination of Multiple Methods: To generate the pattern multiple data mining methods are applied on different data sets.

The key entities associated in discovering patterns are as follows:

- Data Set D: Data set consists of all data sets which concern with our business problem.
- Feature Set F: Features set includes all features of data set used for pattern mining.
- Method Set R: Method set R includes all methods deployed on the data set.
- Interestingness Set I: Interestingness set I composed of all measures from all methods R
- Impact Set T: Impact set T refers to business impacts or outcomes.
- Pattern Set P: Pattern set P is an atomic pattern set generated after mining process.

3.2 Pattern

Pattern: A pattern [4] is a combination of relevant descriptors or attributes associated with certain relations and constraints.

There are two types of patterns generated which are as follows

- Pair Pattern: Pattern merging method G correlate two pattern P_1 and P_2 with each other into a pair, $P \rightarrow G(P_1, P_2)$, this means resultant pattern P consists combination of two patterns only.
- Cluster Pattern: Pattern merging method G correlate more than two patterns each other into cluster, $P \rightarrow G(P_1, P_2, \dots, P_n)$ ($n > 2$), this means resultant pattern P consists combination of more than two patterns.

3.3 Combined Association Rule

Combined association rule mining is to discover the association among the 'attribute-value' pairs. Suppose item set $D \subseteq I_D$, I_D is the item set with attributes $(S_{D1}, S_{D2}, \dots, S_{Dm})$, another item set $A \subseteq I_A$, I_A is the item set with attributes $(S_{A1}, S_{A2}, \dots, S_{An})$, C is a item set of class attribute, then the combined association rule is represented as

$$D + A_1 \Rightarrow C_1$$

$$D + A_i \Rightarrow C_i$$

4. Proposed Mining Framework

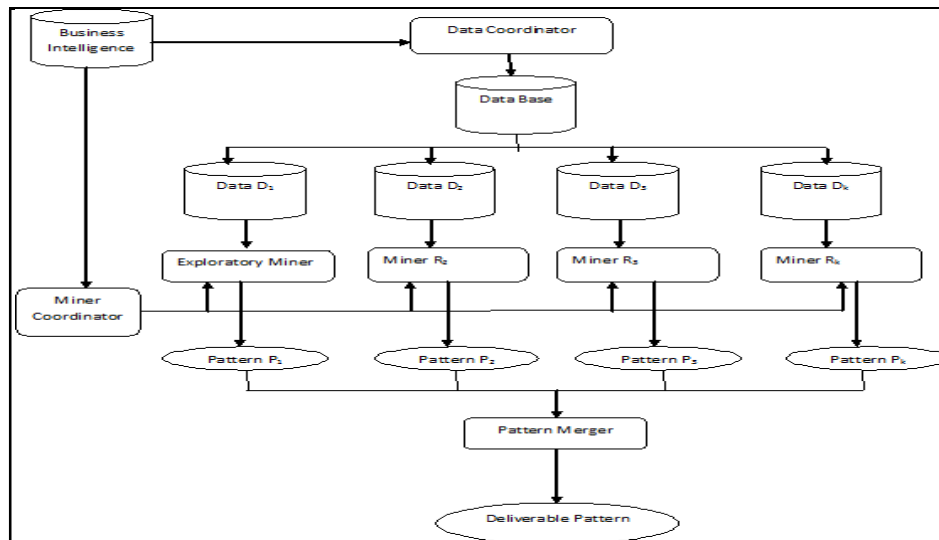


Figure 1: Framework for Combined Mining

Fig. 1 shows a general framework of combined mining system. The common architecture for multi-feature mining, multi method mining and multi source mining are described here.

The proposed framework works as follows.

The data set is partitioned into multiple subsets (D_1, \dots, D_k) through data partitioning. Data coordinator coordinates the data partitioning. One of the data sets or certain data is selected for mining exploration i.e. D_1 . Each data set has separate data miner which are responsible for mining data and generate the pattern (P_1, P_2, \dots, P_k) from that data set. Miner coordinator is present for coordinating miners. Generated patterns are merged by using pattern merging method by the pattern merger. Finally, deliverable pattern is generated.

5. Algorithm

The proposed system takes data set as input that is collected according to the problem defined and generates pattern as per the requirements of business person depending on the interestingness metric. Output patterns are either pair patterns or cluster patterns.

- Step 1:** Identify a suitable data for initial mining exploration
- Step 2:** Import those data into the Database.
- Step 3:** Partition the database into k datasets.
- Step 4:** Select the features using FPGrowth algorithm.
- Step 5:** Extract frequent item set.
- Step 6:** Find unwanted data.
- Step 7:** Generate the Combined Association Rules.
- Step 8:** Merge the atomic pattern into combined pattern
- Step 9:** Generate the conceptual pattern.
- Step 10:** Invoke Business Intelligence on the generated pattern.
- Step 11:** Output the compound pattern either pair or cluster

6. Experimental Work

The environment used to build this application is JDK 1.6/1.7, Net beans, MYSQL, and Windows XP. Hardware used is a PC with RAM: 128Mb SD RAM, Processor: Pentium III, 1.13GHz, Mouse: 2-Buttoned mouse, Hard Drive: 40GB HDD. The data set used for the experiment is pertaining to online shopping are Transaction, and Products list.

Modules:

- Loading multisource datasets
- Combining feature set
- Mining frequent pattern
- Evaluation of impact set
- Result of patterns

Modules Description is as follows:

6.1 Loading multisource datasets:

The multisource dataset is loaded into our project. There are two types of dataset is to be used. Like;

- Demographic dataset
- Transaction dataset

Demographic data consists of population, Transaction data describes the event or transactions occurred in a system.

6.2 Combining feature set

This module organizes related feature classes into common data set. Feature set used to integrate related feature classes. Sometimes, users organize data access privileges using feature datasets.

6.3 Mining frequent pattern

Frequent pattern is a pattern that occurs frequently in a data set. The FPgrowth is an influential algorithm for mining frequent itemsets for Boolean association rules. Scan data and find the support for each item. Discard infrequent items.

6.4 Evaluation of impact set

The concatenation of multiple databases that usually contain some amount of missing data along with a variable percentage of inaccurate data, pollution, outliers, and noise. A missing value can signify a number of different things. Perhaps the field was not applicable, the event did not happen, or the data was not available. The impact set is used to evaluate the missing value and also be ignored.

6.5 Result of patterns

After completing the previous techniques, we obtain the atomic patterns. The atomic patterns are merged according to pattern merging methods.

7. Conclusion and Future Work

Proposed model for “Combined mining in Online Shopping Application” is developed using combined mining approach of data mining. This presented a comprehensive and general approach named *combined mining* for discovering informative knowledge in complex data of “Online Shopping Application”.

We focus on discussing the frameworks for handling multi-feature, multi-source related issues. We 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 the future we can work on a deep understanding of pattern similarity, dissimilarity and dependence between patterns, to develop more complex but actionable knowledge. We create a pattern descriptive language which can be used for pattern representation, inference and ontology

References

- [1] H. Zhang, Y. Zhao, L. Cao, and C. Zhang, 2012, “Combined mining: Discovering Informative Knowledge in Complex Data”, IEEE Transaction.
- [2] H. Zhang, Y. Zhao, L. Cao, and C. Zhang, 2008 “Combined association rule mining”, IEEE Transaction.
- [3] H. Zhang, Y. Zhao, L. Cao, and C. Zhang, 2010, “Flexible Framework for Actionable Knowledge Discovery”, IEEE Transaction.
- [4] J. Wang, G. Karypis, 2005 “Harmony: Efficiently mining the best rules for classification”, SDM
- [5] L. Cao, 2012, “Combined Mining: Analyzing Object and Pattern Relations for Discovering Actionable Complex Pattern”, IEEE Transaction.

- [6] L.Cao, 2008, “Mining Impact- Targeted Activity Patterns in Imbalanced Data”, IEEE Transactions.
- [7] Y. Ma, B.Liu, W.Hsu, 1999, “ Pruning and Summarizing the Discovered Associations”, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
- [8] Dr.Ela Kumar, 2010, “A Combined Mining Approach and Application in Tax Administration”, IJERT.
- [9] Suvarna R. Bhagwat, 2013, “Combined and Actionable Pattern Discovery using DDID-PD Framework:A Review”, IJERT.
- [10] T. Ramanjaneyulu, B. Naresh Reddy, Jogi Suresh, 2013, “Informative Pattern Discovery Using combined Mining”, IJERT.

Author Profile



Ms. Kamble Jayshree R. received B.E. (Computer) degree from Bharti Vidyapeeth College of Engg. For Women, Pune University in 2008, Now pursuing M.E.(Information Technology) degree from Siddhant College of Engg., Pune University.