American Journal of Applied Sciences 10 (12): 1629-1633, 2013
ISSN: 1546-9239

doi:10.3844/ajassp.2013.1629.1633 Published Online 10 (12) 2013 (http://www.thescipub.com/ajas.toc)

Combined Pattern Mining for D3M Using Fuzzy

S.S. Dhenakaran and S. Maheswari

Department of Computer Science and Engineering, Alagappa University, Karaikudi, TamilNadu, India

Received 2013-09-30, Revised 2013-10-03; Accepted 2013-11-01

ABSTRACT

Many algorithms and models were developed but the findings are not actionable and lack of soft power while solving the complex problems. Domain Driven Data mining is used a major efforts to promote the action ability of the knowledge discovery in the real world smart decision making. Combined mining is one of the common methods for analyzing complex data for identifying complex knowledge. The deliverables of combined mining are combined patterns. The complex environment gives the combined patterns. In this research we process a new technique called Fuzzy Combined Pattern Mining (FCPM) for Domain Driven Data Mining. It was used to find all the rules that satisfy the minimum support and minimum confidence constraints. In FCPM, we first apply the fuzzy concept to find the patterns after that the fuzzy pattern will be merged to combined pattern mining. The proposed algorithm have been implemented and compared with Apriori Its performance was studied on an experimental basis. The main objective is to provide the interesting patterns to the end user. The implementation of fuzzy in combined mining will generate the rules and based on rules we can identify the interesting patterns.

Keywords: Domain Driven Data Mining, Combined Pattern, Fuzzy Combined Pattern Mining (FCPM)

1. INTRODUCTION

Data mining is also called as the knowledge discovery in databases is the non-trivial process of identifying the valid, novel, potentially useful and ultimately understandable knowledge in large scale data. On the other hand the traditional data mining research concentrates more on the developing, demonstrating and pushing the use of the specific algorithms and models. The process of data mining stops at pattern identification. As the fact goes, (1) many algorithms have been designed of which very few are repeatable and executable in the real world, (2) often many patterns are mined but a major proportion of them are either commonsense or of no particular interest to business and (3) end users generally cannot easily understand and take them over for business use. In precise we can notice that the findings are not actionable and lack soft power in solving real-world complex issues. Thorough efforts are essential for promoting the action-ability of knowledge discovery in real world smart decision making. To this end, domain-driven Data Mining (D3M) has been proposed to tackle the above issues.

Combined mining is a technique for analyzing object relations and pattern relations and for extracting and constructing actionable complex knowledge (patterns or exceptions) in complex situations. The combined mining technique is used for handling the complexity of employing multi feature sets, multi information sources, constraints, multi methods and multi models in data mining and for the analyzing complex relations between objects or descriptors (attributes, sources, methods, constraints, labels and impacts) or between identified patterns during the learning process. Combined patterns are formed with the analysis of the internal relations between objects or pattern which constitutes and obtained by a single method on a single dataset. For instance, combined sequential patterns are obtained from

Corresponding Author: S.S. Dhenakaran, Department of Computer Science and Engineering, Alagappa University, Karaikudi, TamilNadu, India

analyzing the relations within a discovered sequential pattern space. Combined Mining, is one of the general methods for directly identifying patterns enclosing constituents from multiple sources or with heterogeneous features such as covering demographics, behavior and business impacts. The deliverables of combined mining are combined patterns such as combined association rules. Combined patterns consist of multiple components, a pair or cluster of atomic patterns, identified in individual sources or based on individual methods.

For the improvement of the efficiency of combined pattern mining we shall propose the new technique called Fuzzy based Combined Pattern Mining (FCPM) for Domain Driven Data Mining. Most real-life data are neither only binary nor only numerical but a combination of both. Quantitative attributes such as age, take values from a partially ordered, numerical scale which is frequently a subset of the real numbers. The general method adopted is to convert numerical attributes into binary attributes using ranges (for example, any numeric value for attribute Age would fit in ranges like "up to 25", "25-60", "60 and above"). This would reduce the pattern to traditional association rule mining with binary values. The best way of solving the problem is to have quality values represented in the interval [0, 1], instead of just 0 and 1 and to have transactions with a given quality represented to a certain extent (in the range [0, 1]). Thus, we need to use fuzzy methods, by which quantitative values for numerical attributes are converted to fuzzy binary values. This would ensure that there is no loss of information whatever may be the value of any numerical attribute. In this study, we would first apply the fuzzy methods to our real time data set and find the patterns. Then we would also merge this with the combined patterns.

The following are the contribution of this study after converting the categorical data to numerical data:

- Apply Fuzzy Membership function to convert numerical data set to fuzzy data set
- Use Fuzzy Data to find the Fuzzy pattern
- Merge Fuzzy Pattern with combined Pattern
- Finally we show the analytical result, how our FCPM method is efficient

The rest of the study will be organized as under:

In Section 2, we would review several related works.

In Section 3, we would describe the proposed fuzzy based combined mining.

In Section 4, we would evaluate the performance of our work.

In Section 5, the conclusion of the study would be given.

1.1. Related Works

Mining is a process of extracting trends or patterns from historical data. These trends or patterns can provide business intelligence that leads to actionable knowledge. There are many data mining methods or algorithms that exist for mining data to get patterns. However, all the existing algorithms are single-step mining algorithms. This means that they provide business intelligence inadequately. They may not be able to reflect the complex needs of an enterprise to take decisions correctly. When multiple data mining techniques are combined it is possible to get actionable knowledge that can cater to the needs of an enterprise. In this study combining mining algorithms (Cao *et al*., 2011) have been implemented using a prototype application that demonstrates the efficiency of combined mining. The combined knowledge can't be provided by existing algorithms such as sequential pattern Growth (Vijayalakshmi and Mohan, 2010). The existing works on data mining operations on complex data or enterprise generally of different types such as direct mining approaches; post mining of patterns; data sets with extra features; multiple methods integration; and also joining multiple relational tables. Harmony (Wang and Karypis, 2005) proposed an approach to mine for discriminative patterns. Emerging Pattern are useful for mining multifactor interaction (Dong and Li, 2005). These algorithms attempted to use multiple features or mining techniques. Combined mining is best used to provide actionable knowledge in spite of complex data sets and features. Ren and Zhou (2006) Sequential patterns obtained from the prior mining processes.

There are four categories of combined mining approaches in literature. A commonly used approach (Zhao *et al*., 2009) is the post mining or post analysis of obtained patterns. It is best used to prune the rules obtained after mining database or reducing redundancy or even summarizing the patterns obtained (Narmadha *et al*., 2011). Cao *et al*. (2011) combined mining proposed contain direct mining methods. Cao *et al*. (2011) multisource combined mining, multimethod combined mining and multi-feature combined mining were introduced.

Lei and Ren-hou (2007) proposed an algorithm to mine fuzzy association rules. They first transformed quantitative attribute values into linguistic terms and then used the adjusted difference analysis to find interesting

associations among attributes. It had the advantage that the user-specified thresholds were not needed since the statistical analysis was used. In addition, both positive and negative associations could be found. Sankaradass and Arputharaj (2011) Proposed to improve the intelligence assistance in analysis and even the fuzzy will be suitable for multidimentional data analysis.

Martino and Sessa (2012) proposed a fuzzy mining approach to handle numerical data in databases with attributes and derived fuzzy association rules. At nearly the same time, Hong *et al*. (1987) proposed a fuzzy mining algorithm to mine fuzzy rules from quantitative transaction data (Lee *et al*., 2008). Basically, the fuzzy mining algorithms first used membership functions to transform each quantitative value into a fuzzy set in linguistic terms. The algorithm then calculated the scalar cardinality of each linguistic term on all the transaction data. The mining process based on fuzzy counts was then performed to find fuzzy association rules.

2. MARERIALS AND METHODS

Figure 1 shows the architecture of our proposed Fuzzy based Combined Mining Pattern. In this study we first convert the data set into fuzzy set after that we apply the fuzzy pattern mining with combined pattern mining for identifying the patterns in a dataset. The following subsection describes the proposed algorithm indetail.

2.1. Data Conversion

The real time data set is taken for mining pattern. The data set contains both numerical and categorical data so first we have to convert all categorical data to numerical data. The following algorithm is used for numerical data conversion.

Algorithm 1

Input: Column Values-CV;
Distinct Column Values-DCV;

1. Read Distinct Column Values
2. For each A value in CV
3. String s1=””; (Contains binary value)
4. For each B value in DCV
   a. If (A==B)
      s1 = s1+1
      b. Else
      S1 = s1+0;
5. End if
6. Convert s1 to Decimal Value
7. End for

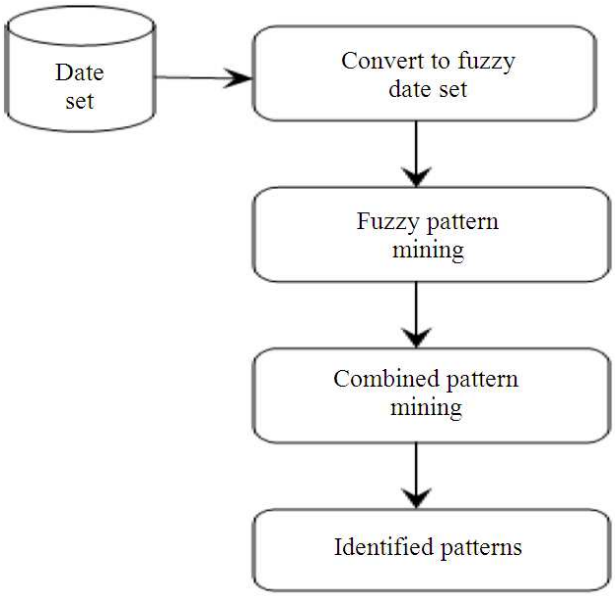

Fig. 1. Architecture of our proposed work

After the numerical data conversion we have to apply the fuzzy technique for creating the fuzzy data set. Here we use Trapezoidal or triangular fuzzifier technique.

Given a data set X = $\{x_1, x_2, x_3, \ldots\ldots x_n\}$ find minimum (a) and maximum (b) value of X. The value of the membership function presents the possibility value of x, as denoted by F(x):

$$F(x)=\begin{cases}2^{*}\left[(x-a)/(b-a)\right]^{2}, a\le x\le (a+b)/2\\ 1-2^{*}\left[(x-b)/(b-a)\right]^{2}, ((a+b)/2)\le x\le b\\ 1, \quad x\ge b\\ 0, \quad \text{otherwise}\end{cases}$$

After conversion of data into fuzzy set we have apply fuzzy pattern mining for finds the frequent pattern. The following section describes the fuzzy combined pattern mining.

2.2. Fuzzy Combined Pattern Mining

For Fuzzy based combined pattern mining, first we have to find the fuzzy association rule. After that we merge the fuzzy association rule with combined pattern mining.

The fuzzy association rules can be discovered in two steps:

- Mining frequent item sets

- Generating fuzzy association rules from the discovered set of frequent item sets

A set of fuzzy transactions may be represented by a table again. Columns and rows are labeled with identifiers of items and transactions respectively. The cell for item ik abd transaction Tj contains a [0 1] value.

Algorithm 2

FCPM
Input: Data Set, minSup, minConf
Output: Fuzzy Combined Pattern

1. Convert Categorical Data (CD) to Numerical Data (ND)
2. Convert Data Set (DS) to Fuzzy Data Set (FDS)
3. Compute sup, conf, lift
4. Extract Rules (FR) from Fuzzy Data Set using minSup and minConf
5. Apply Combined Pattern Mining with FR
6. Extract Fuzzy Combined Patterns

2.3. Performance of Algorithms

Figure 2 Graph represents the comparisions of assoication rules of Apriori algorithm and our proposed FCPM algorthim. With respect to all the minimum support values and generation time of our proposed FCPM algorithm shows the better performance than other existing work. The comparition of Apriori and FCPM algorithm is done with the no of rules generated. In FCPM generate the more no of rules than apriori. So with that rules it is very easy to identify the patterns available in complex data.

3. RESULTS

The sample of heart attack dataset is used in our proposed algorithm. The existing algorithm apriori is also compared with the FCPM. The Fuzzy Membership function is used to convert numerical data set to fuzzy data set and the fuzzy pattern can be identified from the fuzzy data. The fuzzy pattern and the combined pattern is merged. The algorithm 2 shows the convertion of categorical data into numerical data. The thershold value of support, confidence and lift are identified. The rules are extracted from fuzzy Dataset using minimum support and minimum confidence. Thus the result of FCPM shows the outstanding performance of generating rules.

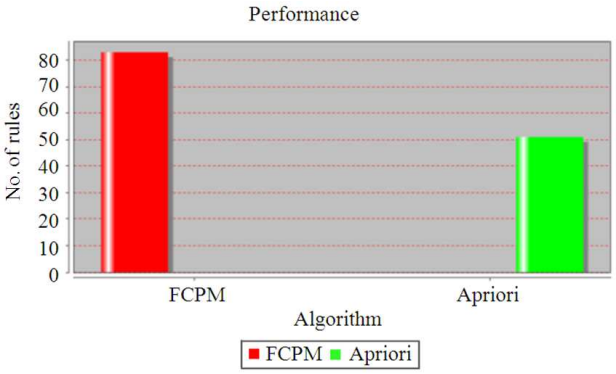

Fig. 2. Performance of FCPM

Table 1. Sample heart data set

Age	Sex	chest pain Type	BP	Cholesterol	Blood
63	Male	asymptomatic	145	233	True
67	Male	asymptomatic	160	286	False
67	Male	asymptomatic	120	229	False
37	Male	non-anginal	130	250	False
41	Female	atypical	130	204	False
56	Male	atypical	120	236	False
62	Female	asymptomatic	140	268	False
57	Female	asymptomatic	120	354	False
63	Male	asymptomatic	130	254	False
53	Male	asymptomatic	140	203	True

4. DISCUSSION

For experiments we take the real time dataset from UCI Machine Learning Repository. **Table 1** shows the sample data of heart dataset. This heart data set contains both numerical and categorical types of data. For applying fuzzy concept we have to convert all the categorical data into numerical data using our data conversion algorithm. After that we apply fuzzy membership function for getting the fuzzy data. Find the fuzzy rules from fuzzy data. For Fuzzy Rules (FR) we set the threshold value for min_Sup = 0.25 and min_Conf = 0.15. Finally Apply Combined Pattern Mining with proposed a mining of dataset with apriori algorithm the mining process of n-dimensional intertransaction rules are data preparation, Frequent-itemset discovery (Nandagopal *et al*., 2012).

5. CONCLUSION

The most challenging problem in the data mining research and development is the mining complex data

for complex knowledge. Typical enterprise applications involve multiple distributed and heterogeneous features and data sources with large quantities, catering for user demographics, preferences, behavior, business appearance, service usage and business impact. This study has presented the most comprehensive and a general approach called the combined mining using fuzzy, for discovering informative knowledge in complex data. A general framework for arriving at more informative knowledge in complex date is provided in the combined mining. Typical challenges such as mining heterogeneous data sources can benefit from combined mining. The performance of the FCPM is improved with the help of the fuzzy approach. More efforts are taken to develop the effective paradigms, combined pattern types, combined mining methods, pattern merging methods and interestingness measures for large and multiple sources of data.

6. RFERENCES

Cao, L., S. Member, H. Zhang, Y. Zhao and D. Luo *et al*., 2011. Combined mining: Discovering informative knowledge in complex data. IEEE Trans. Syst. Man. Cybern. B Cybern., 41: 699-712. DOI: 10.1109/TSMCB.2010.2086060

Dong, G. and J. Li, 2005. Mining border descriptions of emerging patterns from dataset pairs. Knowl. Inform. Syst., 8: 178-202. DOI: 10.1007/s10115-004-0178-1

Hong, C.K., Z.Y. Ou and L. Mandel, 1987. Measurement of subpicosecond time intervals between two photons by interference. Phys. Rev. Lett., 59: 2044-2046. PMID: 10035403

Lee, Y.C., T.P. Hong, T.C. Wang, 2008. Multi-level fuzzy mining with multiple minimum supports. Expert Syst. Applic., 34: 459-468. DOI: 10.1016/j.eswa.2006.09.011

Lei, Z. and L. Ren-hou, 2007. An algorithm for mining fuzzy association rules based on immune principles. Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering, Oct. 14-17, IEEE Xplore Press, Boston, MA., pp: 1285-1289. DOI: 10.1109/BIBE.2007.4375732

Martino, F.D. and S. Sessa, 2012. Detection of fuzzy association rules by fuzzy transforms. Adv. Fuzzy Syst., 2012: 258476-258487. DOI: 10.1155/2012/258476

Nandagopal, S., V.P. Arunachalam and S. Karthik, 2012. Mining of datasets with an enhanced apriori algorithm. J. Comput. Sci., 8: 599-605. DOI: 10.3844/jcssp.2012.599.605

Narmadha, D., G. NaveenSundar and S. Geetha, 2011. An efficient approach for mining association rules in large databases. Int. J. Comput. Sci., 8: 409-415.

Ren, J.D. and X.L. Zhou, 2006. A new incremental updating algorithm for mining sequential patterns. J. Comput. Sci., 2: 318-321. DOI: 10.3844/jcssp.2006.318.321

Sankaradass, V. and K. Arputharaj, 2011. A descriptive framework for the multidimensional medical data mining and representation. J. Comput. Sci., 7: 519-525. DOI: 10.3844/jcssp.2011.519.525

Vijayalakshmi, S. and V. Mohan, 2010. Mining sequential access pattern with low support from large pre-processed web logs. J. Comput. Sci., 6: 1293-1300. DOI: 10.3844/jcssp.2010.1293.1300

Wang, J. and G. Karypis, 2005. Harmony: Efficiently mining the best rules for classification. Proceedings of the SIAM International Conference on Data Mining, (CDM' 05), pp: 205-216.

Zhao, Y., C. Zhang and L. Cao, 2009. Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction. 1st Edn., IGI Global Snippet, Hershey, PA., ISBN-10: 1605664057, pp: 372.