International Journal of Computing and Corporate Research
ISSN (Online) : 2249-054X
Volume 4 Issue 6 November 2014
International Manuscript ID : 2249054XV4I6112014-03

AN EFFECTIVE ARM APPROACH IN VARIOUS ASSORTED APPLICATIONS

SurbhiHuria
School of Computer Sciences
Chitkara University
Rajpura,India

Dr.Jaiteg Singh
School of Computer Sciences
Chitkara University
Rajpura,India

ABSTRACT - **Association rule mining or learning is a prominent and well researched technique for evaluating the interesting relations between variables in large databases. It calculates the strong rules discovered in databases using different measures of interestingness. Using the concept of strong rules, the research introduced the association rules for the extraction of regularities between products in large-scale transaction data recorded by point-of-sale (POS) systems in supermarkets. For example, the rule product1, product2 -> product3 found in the sales data of a supermarket would indicate that if a customer buys products together, and the customer likely to also buy another product. Such type of information can be used as the basis for decision making and about marketing activities such as, e.g., promotional pricing or product placements. In addition to the above example from market basket analysis association rules are employed today in many application areas including Web usage mining, intrusion detection, Continuous production, and bioinformatics. In contrast with sequence mining, association rule learning typically does not consider the order of items either within a transaction or across transactions. In this manuscript, a novel algorithmic approach and technique is proposed that will make use of backtracking for finding out the best results from the given rules using pruning.**

Keywords – Data Mining, Ontology, Association Rule Mining, Associative Classification, Backtracking

I. Introduction

Data mining and knowledge discover, were introduced in 1993 and are used to recognize associations around a set of things in a database. These associations are not independent upon the intrinsic properties of the data themselves (as with viable conditions), yet rather reliant upon the co-occasion of the data things. Affiliation standards are moreover used for diverse orders, for example, desire of frustration in information transfers masterminds by recognizing which events happen before the bafflement. The dominant part of our consideration in this paper will be on bushel market analysis, however in later portions we will look at changed procurements as well. Acquaintanceship rules are the champion amongst the most explored zones of data mining and have started late acknowledged much attention from the bunch of database. They are found of being most extremely advantageous in the advertising and retail gatherings and some other fields. In this paper we give an overview of acquaintanceship rule research. The technique of mining association rules, introduced in [1], is considered as one of the most relevant tasks in Knowledge Discovery in Databases [2]. It aims to discover, among sets of items in transaction databases, thus revealing implicative tendencies from valuable information.

An association rule is described as the implication P → Q where P and Q are sets of items and$P \cap Q = \cent$. The strength of association rule mining lies in its ability to deliver interesting discovered knowledge that exists in data. Unfortunately, due to high dimensions of massive data, its strength becomes its main weakness while analyzing the mining result. The large number of discovered rules makes it very difficult for the decision maker to manually outline the interesting rules. Thus, it is crucial to help the decision maker to efficiently reduce the number of rules.

This is one of the major and critical areas of data mining.

II. Case Analysis - Medical Dataset Rule Mining Using Backtracking

We can apply pruning on the rules by counting the records from the dataset in every category.

Suppose

C1=200

C2=120

C3=220

C4=80

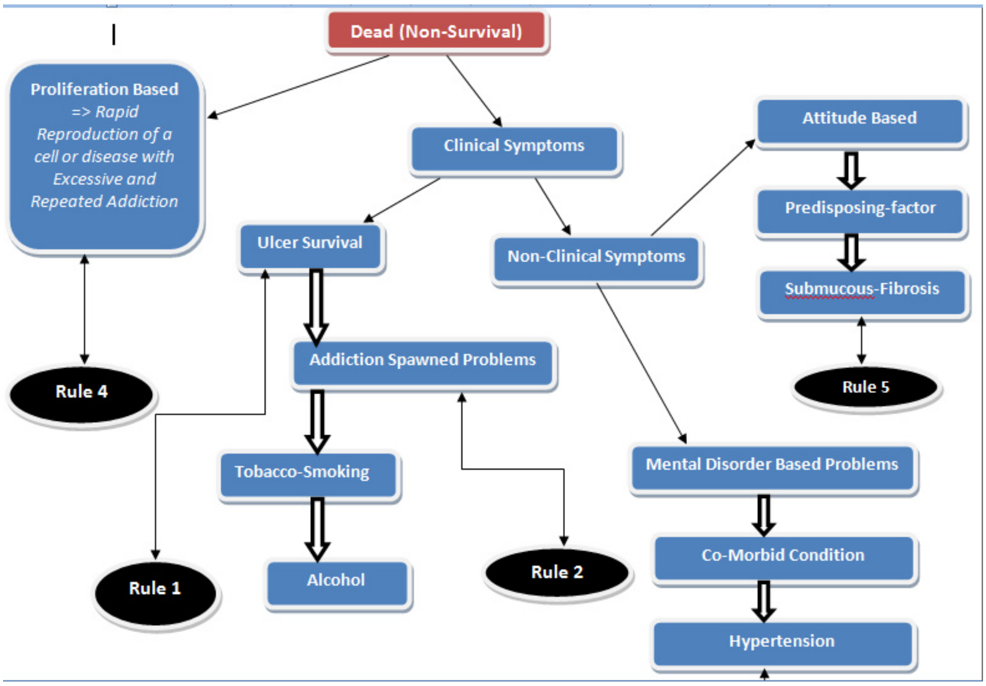
C5=90

We can prune the least 3 counts and finally we will have 2 Best Rules of higher records count from the dataset

C1 (200) ->*Clinical-Symptom (Ulcer)*

C3 (220) ->*Co-Morbid-Condition=Hypertension*

Sr. N.	*Rules*	*State*	*Records Count*
1	*Clinical-Symptom (Ulcer)*	**Survival=Dead**	*C1*
2	*History-of-Addiction(Tobacco-Smoking) && History-of-Addiction1(Alcohol)*		*C2*
3	*Co-Morbid-Condition=Hypertension*		*C3*
4	*Gross-Examination=Ulcero-proliferative*		*C4*
5	*Predisposing-factor=Submucous-Fibrosis*		*C5*



Data mining (the analysis step of the "Knowledge Discovery in Databases" process, or KDD),[1] is an interdisciplinary subfield of computer science, [3][4][5] used for the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems.[3]

To overcome this issue, the post-processing task was proposed to improve the selection of discovery rules. Different complementary post-processing methods can be used, like pruning, summarizing, grouping or visualization. The pruning phase includes the removal uninteresting or redundant rules. In the summarizing phase, summary of rules is being generated. Group of rules is produced in the grouping phase, while the visualization phase is used to have a better presentation.

However, most of the existing post-processing methods isgenerally based on statistical information of the database. Since rule interestingness is strongly dependent upon the user knowledge and goals. For instance, if the user looks for unexpected rules, all the already known rules will be pruned. Or, if the user wants to focus on particular schemas of rules, only the given subset of rules will be selected.

In the terminology of machine learning [6] classification is considered as an instance of supervised learning, i.e. a learning where a training set of correctly identified observations is available.

Association rule learning is a well-known research method for discovering interesting relations of variables in large databases. It is intended to identify strong rules discovered in databases using different measures of interestingness. [9]

Algorithm	**Scan**	**Data structure**	**Comments**
CCPD	m+1	Hash table and tree	Data Parallelism; on shared-memory machine
DD	m+1	Hash table and tree	Task Parallelism; round- robin partition
IDD	m+1	Hash table and tree	Task Parallelism; partition by the first items
HPA	m+1	Hash table and tree	Task Parallelism; partition by hash function
SH	m+1	Hash table and tree	Data Parallelism; candidates generated independently by each processor.
HD	m+1	Hash table and tree	Hybrid data and task parallelism; grid parallel architecture

Table 1 - Comparison of Algorithms on multiple parameters

Algorithm	**Scan**	**Comments**
AIS	m+1	Suitable for low cardinality sparse transaction database; Single consequent
SETM	m+1	SQL compatible
Apriori	m+1	Transaction database with moderate cardinality; Outperforms both AIS and SETM; Base algorithm for parallel algorithms
Apriori-TID	m+1	Very slow with larger number of $\overline{C_k}$; Outperforms Apriori with smaller number of $\overline{C_k}$;
Apriori-Hybrid	m+1	Better than Apriori. However, switching from Apriori to Apriori-TID is expensive; Very crucial to figure out the transition point.
OCD	2	Applicable in large DB with lower support threshold.
Partition	2	Suitable for large DB with high cardinality of data; Favors homogenous data distribution
Sampling	2	Applicable in very large DB with lower support.
DIC	Depends on interval size	Database viewed as intervals of transactions; Candidates of increased size are generated at the end of an interval
CARMA	2	Applicable where transaction sequences are read from a Network; Online, users get continuous feedback and change support and/or confidence any time during process.
CD	m+1	Data Parallelism.

International Journal of Computing and Corporate Research
ISSN (Online) : 2249-054X
Volume 4 Issue 6 November 2014
International Manuscript ID : 2249054XV4I6112014-03

PDM	m+1	Data Parallelism; with early candidate pruning
DMA	m+1	Data Parallelism; with candidate pruning

Table 2 - Comparison of Algorithms on multiple parameters

III. Association Of Ontologies And Data Mining

Ontologies [7],introduced in data mining for the first time in early 2000, can be used in several ways [14]: Domain and Background Knowledge Ontologies, Ontologies for Data Mining Process, or Metadata Ontologies. Background Knowledge Ontologies organize domain knowledge and play important roles at several levels of knowledge discovery process. Ontologies [8] for Data Mining Process signifies the mining process description and choose the most appropriate task according to the given problem, while Metadata Ontologies describe the construction process of items. Related to Generalized Association Rules, the notion of raising was introduced. Raising is the operation of generalizing rules in simpler terms making rules more abstract in order to increase support in keeping confidence high enough. This allows the discovery of strong rules and evaluatesufficient support for the given rules. Data mining (DM) domain deals with the analyzing of the different types of data. [18] Actions, time, physical objects and beliefs abstract concepts of large-scale representation are the examples of ontological engineering. [19]

Ontological engineering is an emerging field to study the concerning ontology development process, ontology life cycle, methods and methodologies for building ontologies [20] [21]

Ontology engineering focuses in making the knowledge explicit within the software applications, enterprises and business procedures for a given domain. Ontology engineering provides a direction while solving the inter-operability problems brought by semantic obstacles, which means the obstacles that are related to the business terms and software classesdefiniations. Ontology engineering is a set of tasks related with the development of ontologies for a particular domain - Line Pouchard, NenadIvezic and Craig Schlenoff, Ontology Engineering for Distributed Collaboration in Manufacturing [22]

IV. Proposed Effective Approach

In the proposed methodology, as the case study, the information set of handloom records will be entered in the back – end database. In the event that, any indication is looked, the proposed methodology looks from the back end database and makes the metaphysics dynamic in execution. It overcomes the unsupervised methodology of philosophy era and is executed so that the unprejudiced effects can be accomplished.

The proposed work is based on a well-known algorithmic paradigm known as backtracking in which all possible solutions are analyzed and the best result is selected based on the optimization criteria.

Backtracking refers to the general algorithm for finding all (or some) solutions to the given computational problem that incrementally builds candidates to the solutions, and abandons each partial candidate cn ("backtracks") as soon as it determines that cn cannot possibly be completed to a valid solution.

The new approach defines a new formal environment to prune and group discovered associations integrating knowledge into specific mining process of association rules. Firstly, a basic mining process is applied over data extracting a set of association rules. Secondly, the knowledge base allows formalizing user knowledge and goals. Domain knowledge allows a general view over user knowledge in database domain, and user expectations express user already knowledge over the discovered rules. Finally, the post-processing step consists in applying several operators (i.e. pruning) over user expectations in order to extract the interesting rules. The proposed algorithm is as follows:

- Step 1: Read dataset(di)
- Step 2: Extract Data Items(Ij)
- Step 3: Generate association(a) and relationship(r) of every occurrence of Ij
- Step 4: Discard initial generation of permutations less than 3 support in dataset
- Step 5: Calculate support and reply dataitems connections

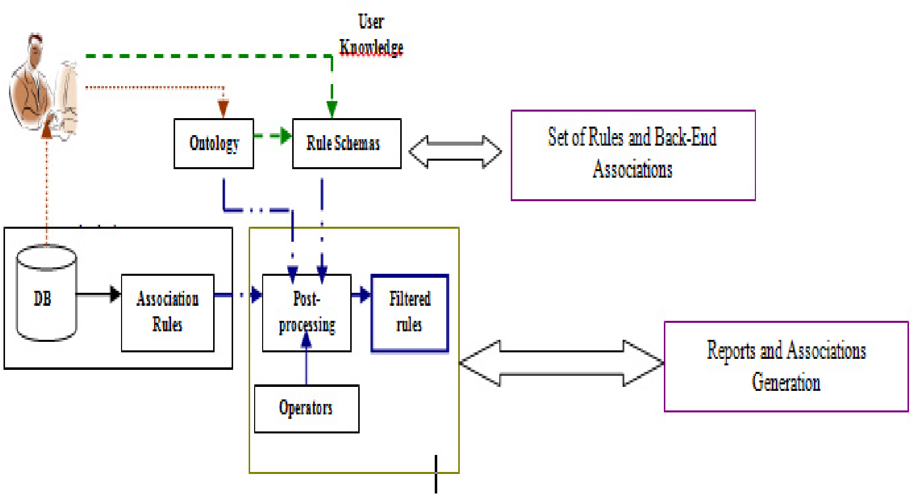


Figure 1 - Framework Description

The novelty of this approach resides in supervising the knowledge discovery process using different conceptual structures for user knowledge representation: one or several ontologies and several rule schemas.

V. Database And Association Rule Mining

The association rules mining techniques are applied over databases described as D = {I, T}. Let I = { I1, I2, ... ,Ip } be the set of attributes (called items) and T = { t1, t2, … tn } be the transaction set. Each transaction ti = { I1, I2, ... ,Imi } is a set of items, such as $t_i \subset I$ and each subset of items, X, is called itemset.

An association rule is an implication $X \rightarrow Y$, where X and Y are two item sets and $X \cap Y = \phi$. This rule holds on D with the confidence c if c% of transactions in T that contain X, also contain Y. The rule has support s in transaction set T if s% of transactions contain $X \cup Y$.

Since their early definition, association rules are mined using Apriori algorithm proposed for the first time in Agrawal et al., 1993.

VI. Operations In Post-Processing Step

The post-processing task that we design is based on operators applied over rule schemas allowing to user to perform several actions over the discovered rules. We propose two important operators: pruning and filtering association rules. The filtering operator is composed by three operators: conforming, unexpectedness and exception.

VII. Conclusion And Future Work

This paper proposes a compelling system for the philosophy plan and talks about the issues of Association Rule Mining using Apriori Algorithm and developing a Reverse Apriori Algorithm using backtrack method. The proposed work could be executed in any space including handloom database, medicinal database, market wicker bin examination, web server log documents and numerous others. Various strategies will be utilized for the enhancement of the results identified with order and guideline mining. The methodology will be upgraded utilizing superior parallel calculations and figuring ideal models. The future extent of work incorporates execution of the algorithmic approach on half and half methodologies making utilization of lattices and parallel frameworks.

VIII. References

[1] Agrawal, R., T. Imielinski, and A. Swami. (1993). Mining Association Rules between Sets of Items in Large Databases. Proceedings of the 12th ACM SIGMOD International Conference on Management of Data, pages 207 - 216.

[2] Fayyad, U. M., G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy. (1996) Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press.

[3] Data Mining Curriculum". ACM SIGKDD. 2006-04-30.

[4] Clifton, Christopher (2010). "Encyclopædia Britannica: Definition of Data Mining". Retrieved 2010-12-09.

[5] Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome (2009). "The Elements of Statistical Learning: Data Mining, Inference, and Prediction". Retrieved 2012-08-07

[6] Anderson,T.W. (1958) An Introduction to Multivariate Statistical Analysis, Wiley.

[7] Gruber, Thomas R. (June 1993). "A translation approach to portable ontology specifications" (PDF). Knowledge Acquisition5 (2): 199–220. doi:10.1006/knac.1993.1008.

[8] Arvidsson, F.; Flycht-Eriksson, A. "Ontologies I".

[9] Piatetsky-Shapiro, Gregory (1991), Discovery, analysis, and presentation of strong rules, in Piatetsky-Shapiro, Gregory; and Frawley, William J.; eds., Knowledge Discovery in Databases, AAAI/MIT Press, Cambridge, MA.

[10] A. Colorni, M. Dorigo et V. Maniezzo, Distributed Optimization by Ant Colonies, actes de la première conférenceeuropéennesur la vie artificielle, Paris, France, Elsevier Publishing, 134-142, 1991.

[11] M. Dorigo, Optimization, Learning and Natural Algorithms, PhD thesis, Politecnico di Milano, Italy, 1992.

[12] de Castro, Leandro N.; Timmis, Jonathan (2002).Artificial Immune Systems: A New Computational Intelligence Approach. Springer. pp. 57–58. ISBN 978-1-85233-594-6.

[13] Comparison and Analysis of Algorithms for Association Rules, IEEE Database Technology and Applications, 2009 First International Workshop, Comput. Sci. Dept., Dalian Nat. Univ., Dalian, China

[14] Rakesh Agrawal and RamakrishnanSrikant Fast algorithms for mining association rules in large databases. Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pages 487-499, Santiago, Chile, September 1994.

[15] A high efficient AprioriTid algorithm for mining association rule, IEEE Sch. of Electron. Inf. Eng., Tianjin Univ., China, Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference

[16] Background for Association Rules and Cost Estimate of Selected Mining Algorithms, Jia, Ashley, ACM 2013

[17] C. Hidber Online Association Rule Mining. In Proceedings of the ACM SIGMOD International Conference on Management of Data, 2014

[18] The OntoDM ontology, URL : http://www.ontodm.com/doku.php 2013

[19] URL :http://ontology.buffalo.edu/bfo/BeyondConcepts.pdf

[20] Asunción Gómez-Pérez, Mariano Fernández-López, Oscar Corcho (2004). Ontological Engineering: With Examples from the Areas of Knowledge Management, E-commerce and the Semantic Web. Springer, 2004.

[21] Denicola, A; Missikoff, M; Navigli, R (2009). "A software engineering approach to ontology building". Information Systems34 (2): 258. doi:10.1016/j.is.2008.07.002.

[22] Line Pouchard, NenadIvezic and Craig Schlenoff (2000) "Ontology Engineering for Distributed Collaboration in Manufacturing". In Proceedings of the AIS2000 conference, March 2000.