ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

Overview of Association Rule Based on Pattern Mining

K.Dharani, Dr.Antony selvadoss Thanamani

Research scholar, Department of Computer Science, NGM College, Pollachi, India

Associate Professor and Head, Department of Computer Science, NGM College, Pollachi, India

ABSTRACT: Pattern mining is a way of obtaining undetected patterns or facts from massive amount of data in a database. Association rule mining is a major technique in the area of pattern mining. Association rule mining finds frequent item sets from a set of transactional databases. Pattern mining algorithm is one of the newest algorithms of association rule mining. Pattern mining consists of discovering interesting, useful, and unexpected patterns in databases various types of patterns can be discovered in databases such as frequent itemsets, associations, subgraphs, sequential rules, and periodic patterns. Pattern mining is also known as knowledge discovery in databases (KDD). Pattern mining employs an iterative approach known as level wise search. In this article, we have presented a survey of most recent work that has been done by researchers in Association rule based mining using Pattern Mining. In pattern mining various techniques are used. That is decision tree, nearest neighbour, artificial intelligent, genetic algorithm, rule induction. The decision tree algorithm is easy to understand and robust and nearest neighbour is simplicity, but difficult to compute complexity for large system. Genetic algorithm is used to find accurate knowledge. And finally the rule induction algorithm gives high accuracy.

KEYWORDS: Pattern mining, Item Sets, Association Rules, pattern mining algorithm.

I.INTRODUCTION

Pattern mining is more in demand because it helps to reduce cost and increases the revenues [1]. The various applications of pattern mining are customer retention, market analysis, and production control and fraud detection. Pattern mining is designed for different databases such as object-relational databases, relational databases, data warehouses and multimedia databases. Pattern mining methods can be categorized into classification, clustering, association rule mining, sequential pattern discovery, regression etc. Amongst these methods, association rule mining is very important which results in generating strong association rules. Association rules was first proposed by R.Agrawal which aims at finding frequent item sets from a set of transactional databases. The various algorithms in associations rule mining are Pattern mining, FP-Growth, Direct Hashing and Pruning (DHP); it is based on support and confidence values. The definitions of probability are:

Support (A->B) =P (AUB) eq. (1)
Confidence (A->B) =P (B|A)

The rules that meet the condition of minimum support (min_supp) and minimum confidence (min_conf) values are known as strong association rules. The itemsets which appears in the data set frequently are known as frequent itemsets. If the support value of itemsets A is greater than or equal to minimum support threshold value, then itemsets A is called frequent itemsets. If the support value of itemsets A is smaller than the minimum support threshold value, then itemsets A is called infrequent itemsets.

II.BACKGROUND STUDY

Pattern mining algorithm is the fundamental algorithm of association rule mining .Pattern mining employs an iterative approach known as levelwise search. In Pattern mining, (k+1) itemsets are generated from k-itemsets. First,

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

scan the database for count of each candidate and compare candidate support count with minimum support count to generate set of frequent 1-itemsets. The set is denoted as L1. Then, L1 is used to find L2, set of frequent 2-itemsets, which is further used to find L3 and so on, until no more frequent k-itemsets can be found [3]. After finding set of frequent k-itemsets, it is easy to generate strong association rules. The process of finding each Lk requires the database to be scanned completely once. An important Pattern mining property is used to improve efficiency of levelwise generation of frequent itemsets which tells that any subset of frequent itemset must be frequent.

Shuo Yang et al (2012) [2] proposed a theorem to improve the traditional Pattern mining Algorithm. The traditional Pattern mining Algorithm takes more time to scan the database in order to find out the frequent itemsets. This increases the complexity and decreases efficiency. The proposed algorithm decreases the database access on the basis of customer habits. It uses relative theorems to find frequent itemsets. For applying improved Pattern mining algorithm to Ecommerce, there will be a need to develop a shopping site because when customers visit the shopping site the system will automatically find out their next purchasing goods based on goods already available in their shopping basket. So, it will save time and increases the efficiency and provides more benefit. The customers could easily generate association rules with the help of improved Pattern mining algorithm and suggests useful products to customers within a reasonable time. The tool used for developing the site is Macromedia Dreamweaver 8; database management tool is Microsoft SQL Server 2000 and Web server is Tomcat 6.0. According to experimental results, it was shown that improved Pattern mining Algorithm when compared with traditional Pattern mining algorithm is more efficient.

III.FOUNDATION OF PATTERN MINING

Pattern mining techniques are the result of a long process of research and product development. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time. Pattern mining takes this evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery. Pattern mining is ready for application in the business community because it is supported by three technologies that are now sufficiently mature:

- Massive data collection
- Powerful multiprocessor computers
- Pattern mining algorithms

Commercial databases are growing at unprecedented rates. A recent META Group survey of data warehouse projects found that 19% of respondents are beyond the 50 gigabyte level, while 59% expect to be there by second quarter of 1996 in some industries, such as retail, these numbers can be much larger. The accompanying need for improved computational engines can now be met in a cost-effective manner with parallel multiprocessor computer technology. Pattern mining algorithms embody techniques that have existed for at least 10 years, but have only recently been implemented as mature, reliable, understandable tools that consistently outperform older statistical methods.

In the evolution from business data to business information, each new step has built upon the previous one. For example, dynamic data access is critical for drill-through in data navigation applications, and the ability to store large databases is critical to pattern mining.

IV. THE SCOPE OF PATTERN MINING

Pattern mining derives its name from the similarities between searching for valuable business information in a large database — for example, finding linked products in gigabytes of store scanner data — and mining a mountain for a vein of valuable ore. Both processes require either sifting through an immense amount of material, or intelligently

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

probing it to find exactly where the value resides. Given databases of sufficient size and quality, Pattern mining technology can generate new business opportunities by providing these capabilities:

A. ***Automated prediction of trends and behaviours***. Pattern mining automates the process of finding predictive information in large databases. Questions that traditionally required extensive hands-on analysis can now be answered directly from the data — quickly. A typical example of a predictive problem is targeted marketing. Pattern mining uses data on past promotional mailings to identify the targets most likely to maximize return on investment in future mailings. Other predictive problems include forecasting bankruptcy and other forms of default, and identifying segments of a population likely to respond similarly to given events.
B. ***Automated discovery of previously unknown patterns***. Pattern mining tools sweep through databases and identify previously hidden patterns in one step. An example of pattern discovery is the analysis of retail sales data to identify seemingly unrelated products that are often purchased together. Other pattern discovery problems include detecting fraudulent credit card transactions and identifying anomalous data that could represent data entry keying errors.

Pattern mining techniques can yield the benefits of automation on existing software and hardware platforms, and can be implemented on new systems as existing platforms are upgraded and new products developed.

When Pattern mining tools are implemented on high performance parallel processing systems, they can analyze massive databases in minutes. Faster processing means that users can automatically experiment with more models to understand complex data. High speed makes it practical for users to analyze huge quantities of data. Larger databases, in turn, yield improved predictions.

Databases can be larger in both depth and breadth:

- ***More columns***. Analysts must often limit the number of variables they examine when doing hands-on analysis due to time constraints. Yet variables that are discarded because they seem unimportant may carry information about unknown patterns. High performance
- ***More rows***. Larger samples yield lower estimation errors and variance, and allow users to make inferences about small but important segments of a population.

V. THE MOST COMMONLY USED TECHNIQUES IN PATTERN MINING

A. ***Decision trees***: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID). Compared to other data-mining techniques, it is widely applied in various areas since it is robust to data scales [4].They are easily understandable
B. ***Nearest neighbour method:*** A technique that classifies each record in a dataset based on a combination of the classes of the record(s) most similar to it in a historical dataset. Sometimes called the k-nearest neighbour technique. Its prime advantage is its simplicity [5], but its main inconvenience is its computing complexity for large training sets.
C. ***Artificial neural networks***: Non-linear predictive models that learn through training and resemble biological neural networks in structure. It is easy to maintain .but it is over fitting [6].
D. ***Genetic algorithms***: Optimization techniques that use process such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution. It is a search strategy to find accurate and comprehensible knowledge within large database that may be considered as search space [7].
E. ***Rule induction***: The extraction of useful if-then rules from data based on statistical significance. Rule induction gives the high accuracy [8].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

Many of these technologies have been in use for more than a decade in specialized analysis tools that work with relatively small volumes of data. These capabilities are now evolving to integrate directly with industry-standard data warehouse and OLAP platforms.

Technique Name	Application Used	Performance	Accuracy
Decision Trees [4]	Common Applications	70 %	83 %
Nearest Neighbour[5]	General Database	72 %	84 %
Artificial Neural Network[6]	Advanced Decision Making	76 %	86 %
Genetic Algorithms[7]	Medical Applications	81 %	89 %
Rule Induction[8]	Business Rules	83 %	90 %

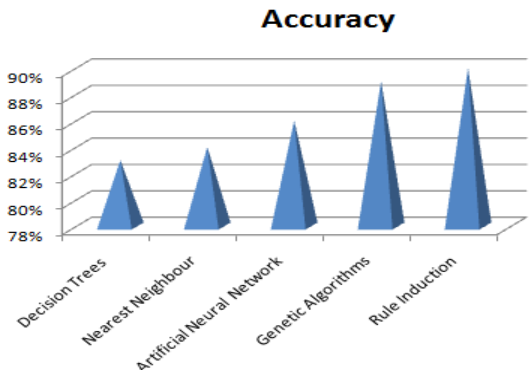

Fig 1. Accuracy chart

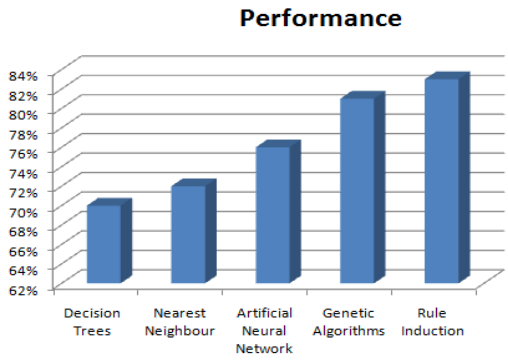

Fig 2. Performance chart

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

VI.GLIMPSE OF LITERATURE

SNO	AUTHORS	YEAR & TITLE	PROS	CONS
1	crows corporations	Third edition Introduction to pattern mining and knowledge discovery ISBN: 1-892095-02-5	Reduce cost Increase revenues	-
2	Shuo yang et al	2012, To improve the traditional pattern mining algorithm 978-0-7695-4818-0/12 $26.00 © 2012 IEEE DOI 10.1109/DCABES.2012.51	Complexity	Efficiency
3	Xece , Gang hu, de-xing wang,xlao-ping liu,jun guo,hao wang	2004, The analysis on method of association rules mining based on concept, third international conference 0-7603-8403- 2/04/$20.00 @2W4 IEEE.	Improve efficiency	-
4	k.m.osei-bryson	2007 decision tree induction using multi performance 34: 3331-3345	Easy understandable Robust	-
5	T.m.cover and P.e.hart	2004 Jan Nearest neighbour pattern classification Vol. IT-13, pp 21-27	Simplicity	Computing complexity for large training sets.
6	M.charles arokiaraj	Application of Neural network in data mining Vol.3, Issue 1 (May 2013), PP 08-11	Easy Maintenance	Over fitting
7	Kamble,atul	2010, International journal of computer theory using genetic algorithm data mining , Vol. 2, No. 3, June, 2010	Accurate knowledge	-
8	w.chung and h.chen	2009 discovering business intelligence from online product reviews: A rule induction frame work	High accuracy	-

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

VII.CONCLUSION

Comprehensive data warehouses in pattern mining with that integrate operational data with customer, supplier, and market information have resulted in an explosion of information. Competition requires timely and sophisticated analysis on an integrated view of the data. However, there is a growing gap between more powerful storage and retrieval systems and the user's ability to effectively analyze and act on the information they contain. In this article the rule induction gives the high accuracy while comparing to other pattern mining techniques.

REFERENCES

1. "Introduction to Pattern mining and Knowledge Discovery" Third Edition by Two Crows Corporation ISBN: 1-892095-02-5
2. Shuo Yang, "Research and Application of Improved the traditional pattern mining algorithm 978-0-7695-4818-0/12 $26.00 © 2012 IEEE DOI 10.1109/DCABES.2012.51
3. xue-gang hu, de-xing wang, xiao-ping liu, jun guo, hao wang, "the analysis on model of association rules mining based on concept lattice and apriori algorithm" Proceedings of the Third International Conference on Machine Learning and Cybernetics, Shanghai, 26-29 August 2004, 0-7603-8403- 2/04/$20.00 @2W4 IEEE.
4. K-M. Osei-Bryson, 2007, "Post-pruningin decision tree induction using multiple performance measures", Computers &Operations Research, 34: 3331-3345.
5. T. M. Cover and P. E. Hart, "Nearest Neighbor Pattern Classification", IEEE Trans. Inform. Theory, Vol. IT-13, pp 21-27, Jan 2004.
6. M.Charles Arockiaraj , "Applications of Neural Networks In Data Mining", International Journal Of Engineering And Science, Vol.3, Issue 1 (May 2013), PP 08-11
7. Kamble, Atul (2010). "Incremental Clustering in Data Mining using Genetic Algorithm". International Journal of Computer Theory and Engineering, Vol. 2, No. 3, June, 2010.
8. Discovering business intelligence from online product reviews: a rule-induction frame work w.chung and h.chen, web-based intelligence system.