International Journal of Management, IT & Engineering
(ISSN: 2249-0558)

CONTENTS

Chief Patron

Dr. JOSE G. VARGAS-HERNANDEZ
Member of the National System of Researchers, Mexico
Research professor at University Center of Economic and Managerial Sciences,
University of Guadalajara
Director of Mass Media at Ayuntamiento de Cd. Guzman
Ex. director of Centro de Capacitacion y Adiestramiento

Patron

Dr. Mohammad Reza Noruzi
PhD: Public Administration, Public Sector Policy Making Management,
Tarbiat Modarres University, Tehran, Iran
Faculty of Economics and Management, Tarbiat Modarres University, Tehran, Iran
Young Researchers' Club Member, Islamic Azad University, Bonab, Iran

Chief Advisors

Dr. NAGENDRA. S.
Senior Asst. Professor,
Department of MBA, Mangalore Institute of Technology and Engineering, Moodabidri

Dr. SUNIL KUMAR MISHRA
Associate Professor,
Dronacharya College of Engineering, Gurgaon, INDIA

Mr. GARRY TAN WEI HAN
Lecturer and Chairperson (Centre for Business and Management),
Department of Marketing, University Tunku Abdul Rahman, MALAYSIA

MS. R. KAVITHA
Assistant Professor,
Aloysius Institute of Management and Information, Mangalore, INDIA

Dr. A. JUSTIN DIRAVIAM
Assistant Professor,
Dept. of Computer Science and Engineering, Sardar Raja College of Engineering,
Alangulam Tirunelveli, TAMIL NADU, INDIA

Editorial Board

Dr. CRAIG E. REESE
Professor, School of Business, St. Thomas University, Miami Gardens

Dr. S. N. TAKALIKAR
Principal, St. Johns Institute of Engineering, PALGHAR (M.S.)

Dr. RAMPRATAP SINGH
Professor, Bangalore Institute of International Management, KARNATAKA

Dr. P. MALYADRI
Principal, Government Degree College, Osmania University, TANDUR

Dr. Y. LOKESWARA CHOUDARY
Asst. Professor Cum, SRM B-School, SRM University, CHENNAI

Prof. Dr. TEKI SURAYYA
Professor, Adikavi Nannaya University, ANDHRA PRADESH, INDIA

Dr. T. DULABABU
Principal, The Oxford College of Business Management, BANGALORE

Dr. A. ARUL LAWRENCE SELVAKUMAR
Professor, Adhiparasakthi Engineering College, MELMARAVATHUR, TN

Dr. S. D. SURYAWANSHI
Lecturer, College of Engineering Pune, SHIVAJINAGAR

Dr. S. KALIYAMOORTHY
Professor & Director, Alagappa Institute of Management, KARAIKUDI

Prof S. R. BADRINARAYAN
Sinhgad Institute for Management & Computer Applications, PUNE

Mr. GURSEL ILIPINAR
ESADE Business School, Department of Marketing, SPAIN

Mr. ZEESHAN AHMED
Software Research Eng, Department of Bioinformatics, GERMANY

Mr. SANJAY ASATI
Dept of ME, M. Patel Institute of Engg. & Tech., GONDIA(M.S.)

Mr. G. Y. KUDALE
N.M.D. College of Management and Research, GONDIA(M.S.)

Editorial Advisory Board

Dr. MANJIT DAS
Assistant Professor, Deptt. of Economics, M.C.College, ASSAM

Dr. ROLI PRADHAN
Maulana Azad National Institute of Technology, BHOPAL

Dr. N. KAVITHA
Assistant Professor, Department of Management, Mekelle University, ETHIOPIA

Prof C. M. MARAN
Assistant Professor (Senior), VIT Business School, TAMIL NADU

Dr. RAJIV KHOSLA
Associate Professor and Head, Chandigarh Business School, MOHALI

Dr. S. K. SINGH
Asst. Professor, R. D. Foundation Group of Institutions, MODINAGAR

Dr. (Mrs.) MANISHA N. PALIWAL
Associate Professor, Sinhgad Institute of Management, PUNE

Dr. (Mrs.) ARCHANA ARJUN GHATULE
Director, SPSPM, SKN Sinhgad Business School, MAHARASHTRA

Dr. NEELAM RANI DHANDA
Associate Professor, Department of Commerce, kuk, HARYANA

Dr. FARAH NAAZ GAURI
Associate Professor, Department of Commerce, Dr. Babasaheb Ambedkar Marathwada University, AURANGABAD

Prof. Dr. BADAR ALAM IQBAL
Associate Professor, Department of Commerce, Aligarh Muslim University, UP

Dr. CH. JAYASANKARAPRASAD
Assistant Professor, Dept. of Business Management, Krishna University, A. P., INDIA

Associate Editors

Dr. SANJAY J. BHAYANI
Associate Professor ,Department of Business Management, RAJKOT (INDIA)

MOID UDDIN AHMAD
Assistant Professor, Jaipuria Institute of Management, NOIDA

Dr. SUNEEL ARORA
Assistant Professor, G D Goenka World Institute, Lancaster University, NEW DELHI

Mr. P. PRABHU
Assistant Professor, Alagappa University, KARAIKUDI

Mr. MANISH KUMAR
Assistant Professor, DBIT, Deptt. Of MBA, DEHRADUN

Mrs. BABITA VERMA
Assistant Professor, Bhilai Institute Of Technology, DURG

Ms. MONIKA BHATNAGAR
Assistant Professor, Technocrat Institute of Technology, BHOPAL

Ms. SUPRIYA RAHEJA
Assistant Professor, CSE Department of ITM University, GURGAON

November 2011 IJMIE Volume 1, Issue 6 ISSN: 2249-0558

Title

HIGHLY QUANTITATIVE MINING ASSOCIATION RULES WITH CLUSTERING

Author(s)

N. Venkatesan

Department of IT,

Bharathiyar College of Engineering and Technology,

Karaikal.

Abstract:

Data mining is a step in the knowledge discovery process consisting of certain data mining algorithms that, under some acceptable computational efficiency limitations, finds patterns or models in data. Association rule analysis starts with transactions containing one or more products or service offerings and some rudimentary information about the transaction. This paper describes the clustering in association rules using quantitative attributes, which are expressive multi-dimensional generalized association rules for university admission. University database which is vast and which has interrelated item sets is chosen for mining. Quantitative attributes can have a very wide range of values defining their domain. 2-Dimensional quantitative association rules predicting the condition on the right hand side, given the quantitative attributes age, status and citizen. In this method age is the main criteria for quantitative process. The Strong association rules are obtained by this prescribed method.

Keywords: Market Basket Analysis, eclat algorithm, Quantitative Association Rule, Clustering, Binning.

Introduction:

Data Mining, also called as data archeology, data dredging, data harvesting, is the process of extracting hidden knowledge from large volumes of raw data and using it to make crucial business decisions. Experimental data in many domains serves as a basis for predicting useful trends.

Association rules [2], first introduced in 1993, are used to identify relationships among a set of items in a database [6]. These relationships are not based on inherent properties of the data themselves (as with functional dependencies), but rather based on co-occurrence of the data items.

Association rules identify collections of data attributes that are statistically related in the database using confidence and support levels. The main problem of association rule induction is that there are so many possible rules. It is obvious that such a vast number of rules cannot be processed by inspecting each one in turn. Efficient methods are needed that restrict the search space and check only a subset of all rules [10], and if possible, without missing important rules. Association rules analyzes how the item purchased by customers. An example of an association rule is as follows.

Bread → Butter (sup 10%, conf 80%)

This rule says that 10% of the customers buy bread and butter together and those who buy butter 80% of the time.

The main problem of association rule induction is that there are so many possible rules. It is obvious that such a vast number of rules cannot be processed by inspecting each one in turn. Efficient algorithms are needed that restrict the search space and check only a subset of all rules, and if possible, without missing important rules. Using a concept hierarchy, quantitative and multiple minimum supports, interesting rules are generated. To select interesting rules from the set of all possible rules, a multiple minimum support [4] and a minimum confidence are fixed.

This paper is organized as follows. In Section 2, Association rule mining through eclat Algorithm as well as description of sample university database with substitution are given. Section 3, proposes the method with multiple minimum support, Concept hierarchy with quantitative concept. In Section 4, Generating Quantitative Association Rules with the help of Binning and Clustering. Section 5 discusses the experimental and performance analysis with single dimensional results. The study is concluded in the section 6, along with a brief on future work.

PRELIMINERY WORKS:

ECLAT ALGORITHM:

In Eclat algorithm [18][20] implementation the set of transactions as a (sparse) bit matrix and intersects rows to determine the support of item sets. The search space of Eclat algorithm is based on depth first traversal of a prefix tree [19].

Éclat principle:-

A convenient way to represent the transactions for the Eclat Algorithm is a bit matrix, in which each row corresponds to an item, each column to a transaction. A bit is set in this matrix if the item corresponding to the row is contained in the transaction corresponding to the column, otherwise it is cleared. Eclat searches a prefix tree. The transition of a node to its first child consists in constructing a new bit matrix by intersecting the first row with all following rows. For the second child, the second row is intersected with all following rows and so on.

The item corresponding to the row is intersected with the following rows to form the common prefix of the item sets, processed in the corresponding child node. Of course, rows corresponding to infrequent item sets should be discarded from the constructed matrix, which can be done most conveniently if it stores with each row the corresponding item identifier rather than relying on an implicit coding of this item identifier in the row index.

University Database:

Experimental data in many domains serves as a basis for predicting useful trends. I opted to generate association rule in one such university database records of 1000 students [12, 13]. The database which is chosen depicts the student choice of getting degree in the university level. All the students were in the age group of 15 to 35 years, and the nationality is the addition category. All of them choose either undergraduate or graduate degrees. They were sub-categorized as science degree and arts degree. Further classification based on the applied science and science (computer science and engineering).

Table 2.1 Sample University Student Database

Degree	Decipline	Major	Status	Age	Natio nality	Parent Income (in lac)
MA	Graduate	History	Junior	21..25	Indian	2.2
MSc	Graduate	Maths	Senior	26..30	Indian	1.2
BSc	UnderGra duate	Physics	Junior	16..20	French	4.4

Eclat with University Database

For manual substitution purpose only 10 sample records are selected. Table 2.2 represents the confidence level as well as rules obtained by the general algorithm.

Table 2.2 Support &Confidence with some association rules

X→Y	S	Confidence
IndianStudent → UnderGraduate	30%	3/5*100=60%
IndianStudent → Graduate	30%	3/5*100=60%
ForeignStudent → UnderGraduate	30%	3/7*100=42.8%
ForeignStudent → Graduate	40%	4/7*100=57.1%

Using frequent item set [10], Support levels are obtained for each and every process. Table 2.2 describes various support levels of the sample records. Based on Apriori, Confidence levels are obtained.

The following are the rules obtained by the Apriori Algorithm.

IndianStudent → UnderGraduate (Support 30% Confidence 60%)

IndianStudent → Graduate (Support 30% Confidence 60%)

ForeignStudent → UnderGraduate (Support 30% Confidence 2.8%)

ForeignStudent → Graduate (Support 40% Confidence 57.1%)

TERMINOLOGY:

Multi Level Quantitative Approaches:

A concept hierarchy [4, 16] defines a sequence of mappings from a set of low level concepts to higher level more general concepts. Concept hierarchies may also be defined by grouping values for a given dimension or attribute resulting in a set grouping hierarchy. Concept hierarchies [17] allow data to be handled at varying levels of abstraction. A top-down progressive deepening method is developed for efficient mining of multiple-level Quantitative association rules from large transaction databases.

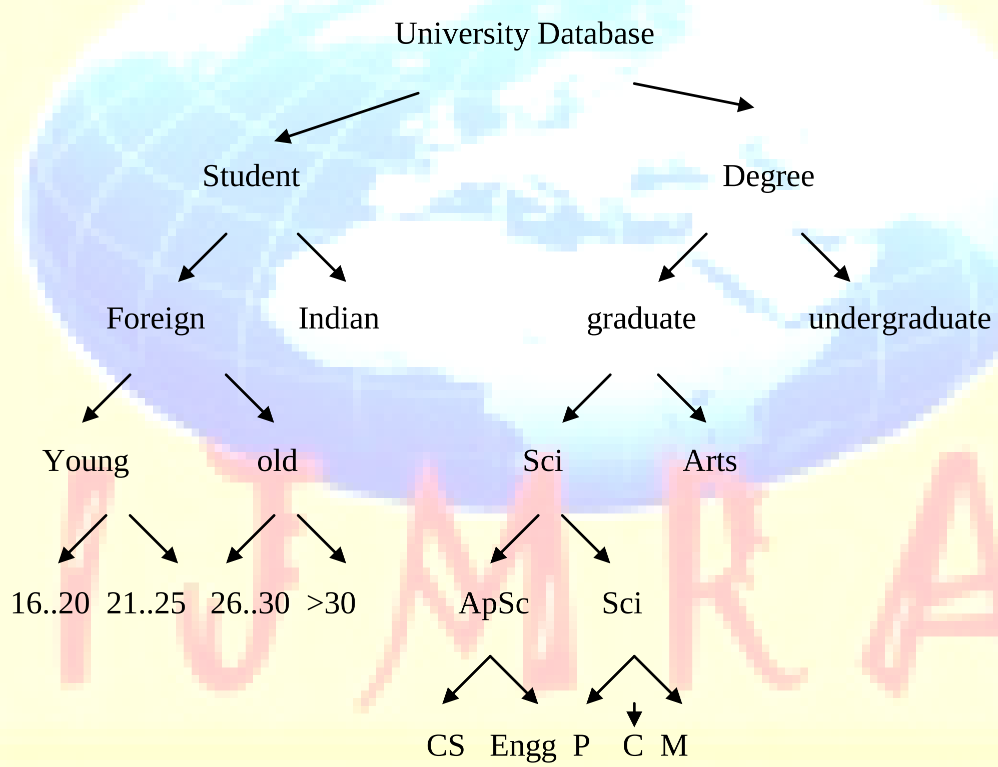

Figure 3.1: Concept hierarchy of University Database

Quantitative attributes, in this case, are discretised prior to mining using predefined concept hierarchies, where numeric values are replaced by ranges. Categorical attributes may also be generalized to higher conceptual levels if desired.

Reduced Minimum Support at Lower Levels:

Each level of abstraction has its own minimum support threshold. The lower the abstraction level, the smaller the corresponding threshold. Using multiple minimum support, rules should be generated. The transformed task relevant data may be stored in a data cube [4].

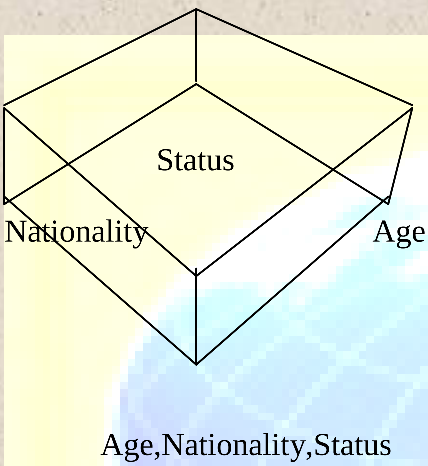

Figure 3.2. Data cube consists of a lattice of cuboids

Data cube consists of a lattice of cuboids that multidimensional data structures. These structures can hold the given task relevant data as well as aggregate, group by information. Figure 3.1shows the lattice of cuboids defining a a data cube for the dimensions of age, nationality and status. The cells of n-dimensional cuboids are used to store the support counts of the corresponding predicate sets. Cuboids contain the total number of transactions in the task relevant data, and so on.

QUANTITATIVE CLUSTERING:

Binning:

Quantitative attributes can have a wide range of values defining their domain. To keep grids down to a manageable size, instead partition the ranges of quantitative attributes into intervals. These intervals are dynamic in that they may later be further combined during the mining process. The partitioning process is referred to as binning. In the below example, partitioning the attribute is done with respect to age.

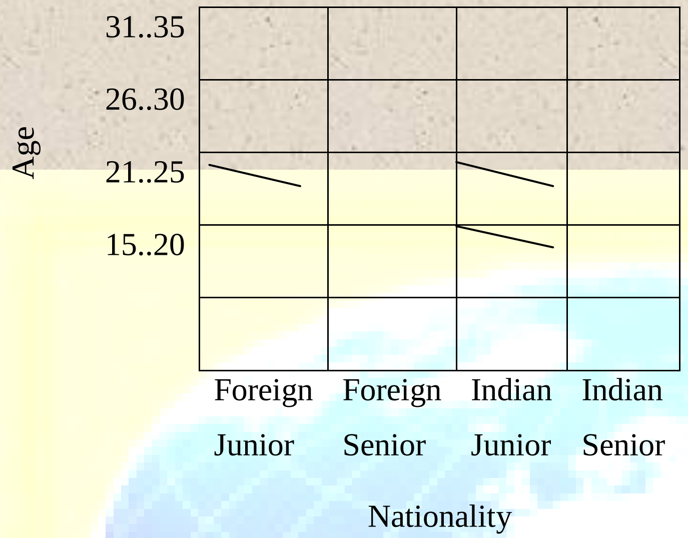

Figure 4.1. A 2-D gird for tuples

Clustering:

The strong association rules are obtained by this mapping of 2-Dimensional grid [4]. 2-D Quantitative association rules predicting the condition on the rule right hand side given the quantitative attributes age. Notice that these rules are quite "close" to one another, forming a rule cluster on the grid. Rules can be combined or "clustered" together to form the simpler rule which produces the good rules.

Using the clustering system process scans the grid, search for rectangular clusters of rules. Bins the quantitative attributes occurring within a rule cluster may be further combined and further dynamic discretization of the quantitative attributes occur. This grid based technique describer here assumed that the initial association rules can be clustered into rectangular region. Rectangular clusters may oversimplify the data. Quantitative attributes are dynamically partitioned using equidepth binning and the partitions are combined based on a measure of partial completeness which quantifies the information lost due to partitioning.

Clustering in Quantitative Process:

Algorithm: ClusQuantitative

Input:

Database

Output:

Grouping the data for Quantitative Association Rules

Step 1: Find all the elements of the database

Step 2: Partition the database elements to form a concept hierarchy structure. (See figure 3.1)

Step 3: Find the Frequent item set value for higher level of the concept hierarchy. With the help of minimum threshold condition, support levels are calculated for each itemsets.

Step 4: Draw the data cube for the elements with are related with student particulars. (See figure 3.2)

Step 5: Using Binning concept to fix the 2-D grid layout for the given data.

Step 6: Find the range field (Age variation of students)

Step 7: Using the Clustering technique to combine the related data items with respect to 2-D grid. (See figure 4.1)

Step 8: Find the confidence level for the frequent itemset. Generate Rule with the help of Apriori_Gen algorithm.

EXPERIMENTAL RESULTS & PERFORMANCE ANALYSIS:

In this section, the university dataset is implemented in eclat based algorithm and the performance evaluation is also carried out to prove the efficiency of the new approach.

EXPERIMENTAL ANALYSIS:

For experimental Intel Pentium 2.5 GHz processor, Windows XP with 256 MB RAM was used. The results for these data sets are discussed

In Level 1, support level value is very high (by getting minimum threshold support value). By applying the reduced support concept, values of support are minimum one. Different levels of the concept hierarchy will yield rules differently. Level 1 rule has the low confidence values. But in the lower level, some rules have either less or higher value than level 1.

In the least level, rules have the full confidence values. Because of the reduced support this will yield interesting and useful rules.

Age(x,16..20),status(x,young), citizen(x,Indian) → Degree(x, CompSci ApSci Undergraduate)

Age(x,21..25),status(x,young),citizen(x,French) → Degree(x,CompSciApSciUndergraduate)

The clustering the related data item as follows:

16..20 age group young Indian Citizen → CompSciApSciUndergraduate (confidence 98%)

21..25 age group young French Citizen → CompSciApSciUndergraduate (confidence 86%)

In the above example, IndianStudent → Undergraduate has confidence value 60% in the higher level 1. But at the level 5, the same rule has the confidence value as 98%

The overall substitution of the whole database will yield more number of useful rules.

PERFORMANCE ANALYSIS:

With the help of Clus-Quantitative algorithmic approach, quantitative clustering process actually creates sets of group of data. For mining multiple level association rules, concept hierarchy and multiple minimum support should be provided for generalizing interesting rules [14, 15]. This may lead to the generation of many uninteresting associations such as "IndianStudent → Undergraduate" before the discovery of some interesting ones, such as **"16..20 age group young Indian Citizen → CompSciApSciUndergraduate"**, because the former may occur more frequently and thus have larger support than the latter.

This observation lead us to examine the methods for mining association rules at multiple minimum support using concept hierarchy, which may only discover rules at Quantitative basis with clustering technique have high potential to find nontrivial informative association rules because of its flexibility at focusing the attention to various data and applying different threshold at different levels.

In this ClusQuantitative has knowledge about what are possible combinations of multiple minimum support itemsets with the grouping of related data items. But ClusQuantitative generate multilevel interesting rules.

The following are the interesting information retrieved from this new approach.

To find who is most likely to join the course Computer Science Undergraduate or Students willingness in the age group of 16 to 20 for Indian students or 21 to 25 Foreign Students.

Figure 5.1 shows the Rules which are having rules between Single dimension and Clus-Quantitative university datasets. From this figure, a large number of rules are generated with higher confidence as 100%. Nearly 60% of the Clus-quantitative association rules are having the higher confidence 100% while comparing with single dimensional dataset having 5% of rules only. So, the Clus-quantitative data set produces the highly informative association rules. For any database each and every combination of dataset is essential to predict interesting relationships Hence, more hidden information are retrieved from this new approach.

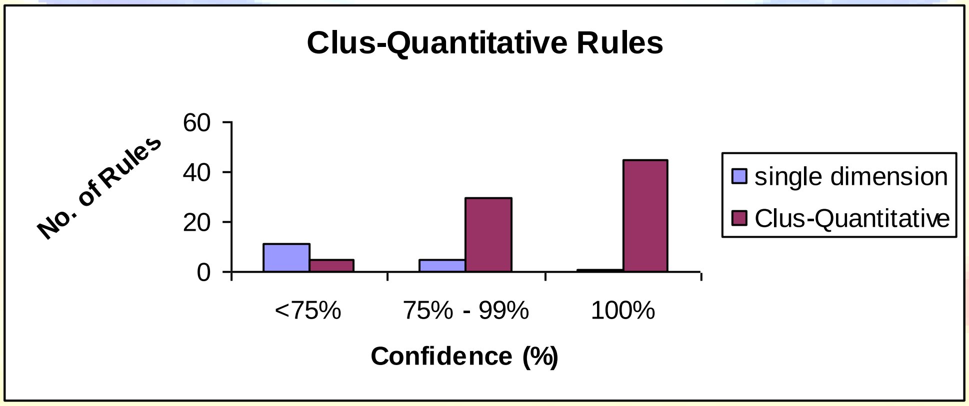

Figure 5.1: Single Vs Quantitative Rules for University Dataset

CONCLUSION AND FUTURE WORK:

Clus-Quantitative which mines Quantitative association rules may lead to progressive mining of refined knowledge from the data and have interesting applications for knowledge discovery in transaction based university admission database as well as other database. With the help of multiple minimum support, binning and clustering of concept hierarchy finding interesting frequent item set at various levels of itemsets is made easy. With the help of Clus-Quantitative substitution, execution time reduction and interesting mining rules are all powerful. Based on the clustering technique, a set of rules can be obtained. This paper mined association rules (did not leave out any data including the extent of status, nationality, etc.) in a step wise manner, so that none of the important data is missed.

Extension of methods for mining knowledge rules poses many interesting issues for further investigation.

REFERENCE:

- R..Agarwal & J.C..Shafer Parallel mining of Association rules IEEE transactions on knowledge and Data Engineering Pages 962-969, December 1996
- Jean-mare Adamo Data Mining for association rules and Sequential patterns
- Fayyad U.M. et al Advance in knowledge Discovery and Data Mining
- J. Han and M. Kamber Data Mining: Concepts and Techniques
- Teutsch SM. Churchill Principles and Practice of public health surveillance
- R. Agarwal, T. Imielinski and Arun N. Swami Mining association rules between sets of items in large databases. Proceedings of the ACM International Conference on Management of Data Pages 207-216, 1993
- R. Agarwal and R. Srikant Fast Algorithms for Mining Association Rules, Proceedomgs of 1994 Int'l Conf on Very Large Databases page. 487-499, Santiago, Chile, Sept. 1994
- Jiawei Han and Yongiiam Fu Discovery of Multiple level association rules from large databases Proceedings of the Int'l conf on Very Large Databases conference, Pages 420-431, 1995
- Bing Liu, W.Hsu, &Y. Ma. Mining Association rules with Multiple supports Proceedings of KDD, Pages 337-341, 1999

- Ke Wang, Yu He, and Jiawel Pushing Support Constraints into Association rules Mining IEEE transaction on knowledge and data engineering May 2003
- C.F. Simmons Introduction to Topology and Modern Analysis, McGraw Hill International Edition
- N. Venkatesan Differential Link Analysis in Health Care using Data Mining MPhil. thesis
- N. Venkatesan Shared Differential Link Association Analysis for Mining Intricate Medical Database National Conf. on Data Mining 04
- J. Han J. Pei and Y.Yin Mining Frequent Patterns with candidate Generation Proc. 2000 ACM SIGMOD Int;l conf. on Mgt. of Data (SIGMOD 2000) Dallas TX, May 2000.
- Dr. E. Ramaraj and N. Venkatesan A High Performance Efficient Pattern Mining in Health Care Database Computing NCICCA2005 at Kalasalingam college of Engineering, April 2005
- Margaret H. Dunham Data Mining and Advanced Topics Pearson Education 2004
- Coenen, F., Leng, P., Goulbourne, G. Tree Structures for Mining Association Rules. Journal of Data Mining and Knowledge Discovery, Vol 15 (7), pp391-398.
- Survey on Frequent Pattern Mining, Bart Goethals, HIIT Basic Research Unit, University of Helsinki, Finland.
- R. Agrawal and R. Srikant. Fast algorithms for mining association rules. Proceedings 20th International Conference on Very Large Data Bases, pages 487–499. Morgan Kaufmann, 1994.
- Mingju Song and Sanguthevar Rajasekaran A transaction mapping for frequent itemsets mining IEEE transactions on Knowledge and Data Engineering 18(4):472-480, April 2006.