An Intelligent Approach to Improve the Query Performance using Disjoint Set Theory

Ranjith Reddy,
AP/CSE,SRIT, Anantapur.

MalavathulavaNarasimhulu,AP/IT, SRIT, Anantapur.

AshokMurugesan,
SL/IT,RIT,Chennai-124.

ABSTRACT:
An index improves the speed of data retrieval operations on a table. Index Management includes Creation, insertion, deletion & Updation. Reducing the access time of index even for more number of transactions will be our objective. Algorithms efficiency fell down when the index updation is occurring in the existing literature. Managing index is a tough task when the size of database increases. We are organizing clusters using disjoint set theory and a ranking algorithm. A back propagation based learning approach is introduced for index management. It will improve the performance of querying.

KEY WORDS
index, set, rank, query.

1. INTRODUCTION
Association rule mining discovers correlations among data items in a transactional database D. Each transaction in D is a set of data items. Association rules are usually represented in the form A! B, where A and B are item sets, i.e., sets of data items. Item sets are characterized by their frequency of occurrence in D, which is called support. Research activity usually focuses on defining efficient algorithms for item set extraction, which represents the most computationally intensive knowledge extraction task in association rule mining [2]. The data to be analyzed is usually stored into binary files, possibly extracted from a DBMS. Most algorithms exploit ad hoc main memory data structures to efficiently extract item sets from a flat file. Recently, disk-based extraction algorithms have been proposed to support the extraction from large data sets, but still dealing with data stored in flat files. To reduce the computational cost of item set extraction, different constraints may be enforced among which the most simple is the support constraint, which enforces a threshold on the minimum support of the extracted item sets. Relational DBMSs exploit indices, which are ad hoc data structures, to enhance query performance and support the execution of complex queries. In this paper, we propose a similar approach to support data mining queries. The IMine index (Item set-Mine index) is a novel data structure that provides a compact and complete representation of transactional data supporting efficient item set extraction from a relational DBMS.

2. DATA ACCESS METHODS
The existing three data Access methods[1] to load from the IMine index the following projections of the original database: 1) Frequent-item based projection, to support projection-based algorithms (e.g., FP-growth). 2) Support-based projection, to support level based (e.g., APRIORI [1]), and array-based (e.g., LCM v.2 [15]) algorithms. 3) Item-based projection, to load all transactions where a specific item occurs, enabling constraint enforcement during the extraction process.

The existing methods support minimum support of index updation. Index updation can’t be incremental in the survey.

3. PROPOSED ALGORITHM
Step1: Find out the mostly accessed data with reference to the table.
Step2: Frame separate set which are accessed frequently.
Step3: Check for intersection of subsets such that commonly accessed elements should not be present.
Step4: determine the Rank based on cardinality of the sets.
Step5: Repeat Step1 to 3 till all the data items are iterated.

4. EXPERIMENTATION & RESULTS
Table 1.1 illustrates the representation of item sets. Table 1.2 depicts the frequently accessed items from Table 1.1. Table 1.3 describes our set Ranking algorithm on the available item sets.

Table 1.1 Sample Data

Transactions ID	ITEMS ID					
001	A	B	D	F	K	P
002	B	C	H	E	G	I
003	I	J	A	C	D	F
004	D	E	F	G	I	
005	F	A	B	D	H	I
006	J	F	C	D	A	

Table 1.2 Frequent sample

ITEMS ID	No. Of Times Accessed
{A, I}	4
{B, C}	3
{D, F}	5
{E, G, H, J}	2
{K, P}	1

Table 1.3 Rank Calculations

ITEMS ID	No. Of Times Accessed	Rank
{D, F}	5	1
{A, I}	4	2
{B, C}	3	3
{E, G, H, J}	2	4
{K, P}	1	5

5. BPN SIMULATION

Fig 1.1 shows the back propagation neural network architecture. In this architecture we take the inputs as the cluster size and cluster rank based up on these we trained the BPN stimulation using Back propagation algorithm and supervised learning technique.

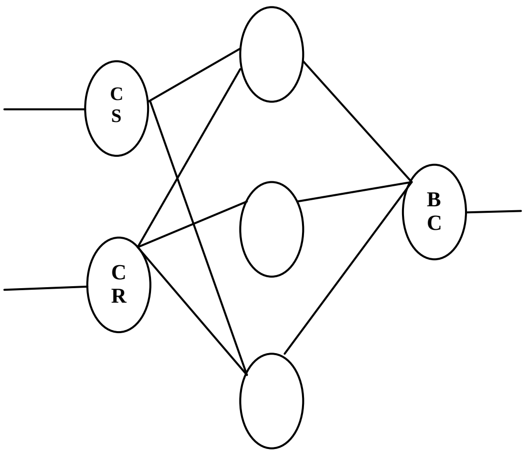

Fig. 1.1 Back Propagation Network

5.1. EXPLANATION

Fig. 1.1 shows the multi layer architecture of neural networks. In this multi layer architecture one layer treated as input layer also called inner layer where CS and CR inputs are given. CS means Cluster size of the cluster which will be taken from Table 1.3 and CR means cluster Rank which is evaluated from Table 1.1 and Table 1.2 and the evaluated resulted are collected in Table 1.3. The inputs from the inner layer to hidden layer are transformed and from hidden layer to output layer are transformed during transformation the BPN simulation detect the best cluster which is collected at the output layer. In this section, we are providing an intelligent approach to find the best cluster for index storage. Cluster rank and cluster size are the parameters termed as input units. A back propagation network based learning approach is incorporated with IUDE. On completion of learning and training, best clusters are identified on the long run.

6. EXPERIMENTATION

Table 1.4 BPN Simulation Analysis

Number of Iterations	Total Inputs	Number of best clusters
1000	10000	8
2000	10000	5
3000	10000	3
4000	10000	1

6.1. EXPLANATION

Table 1.4 shows the experimentation Testing and Training where we collect some set of input samples, these input samples consists of cluster size and cluster rank and theses samples are trained using a specified number of iterations and we evaluate the best clusters as output. As we go on increasing number of iteration detection of a particular cluster is best will be notices through graph which is shown in Fig. 1.2.

7. PERFORMANCE PROJECTION IN LEARNING

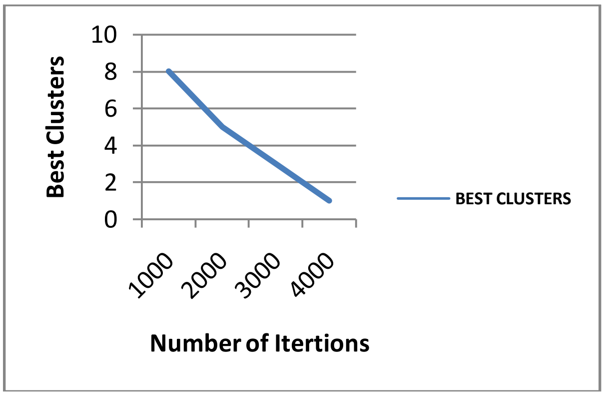

Fig. 1.2: Performance Evaluation of Learning

7.1. EXPLANATION

In the above section we have shown the graph in Fig. 1.2 for performance projection in learning concluded that as number of iteration with sample inputs going on increasing the BSN stimulation System detects best clusters.

8. CONCLUSION & FUTURE WORK

We have designed a Set Ranking algorithm based on Disjoint Sets. We revealed better index updation rate. Also we have integrated a learning approach for improving the query performance. On our future work, we are planning to introduce an agent for intelligence.

9. REFERENCES

[1] M. Ranjit Reddy, and A. Suresh Babu "A New Dimension to improve query processor using disjoint set theory." proc. International Conference on Emerging Technology Trends in Cloud Computing (ICETT-2011), MAR 2011.

[2] Elena Barallis, Tania Carquitelli and silviachiusane, "IMime: Index support for Item set Mining, " proc. IEEE trans. knowledge and Data Eng., vol, 21 no. 4, April 2009.

ABOUT THE AUTHORS

M.Ranjith Reddy received his M.Tech from JNTU, Anantapur. He completed his B.Tech from SKD engineering college, Gooty. He is currently working as Assistant Professor in Computer Science and Engineering Department at SrinivasaRamanujan Institute of Technology, Anantapur. His area of research includes data mining and computer networking.

M. Narasimhulu received his M.Tech from JNTUACEA, Anantapur. He completed his B.Tech from Annamacharya institute of technology and sciences, Rajampet. He is currently working as Assistant professor in the Department of information Technology at SrinivasaRamanujan Institute of Technology, Anantapur. His area of research includes Software Testing, Neural Networks, Computer Networks and Data Base Management Systems. he presented several papers in National and International Conference.

M.Ashok received his M.Tech from JNTU, Anantapur. He completed his BE from Madurai Kamaraj University, Madurai. He is working as Senior Lecturer in the Department of Information Technology at Rajalakshmi Institute of Technology,Chennai. He is the academic consultant for Fabmax Semiconductors Research Labs Private Limited, Chennai. His area of research includes Grid Computing, Cloud Computing, Wireless Networks, Neural Networks and Bioinformatics. He is the member of ISTE. He presented several papers in National and International Conferences.