Sets Sequential Emission by Transmitting Streams

Sarath Chand P.V., VenuMadhav K., Arya Bhanu M., Nagamani K., Balaram A.

Abstract— The increasing generation and collection of data have been increased rapidly in the last several decades. The contribution in the widespread of commercial products like bar coding, computerization of many business corporations, bank transactions and advancement of data collections ranges from scanned text to image platforms. The popular usage of internet as a global search engine for information system has flooded with a tremendous amount of data. The explosion of data is stored and an urgent need for new technologies and techniques should evolve day by day. The paper is the concepts and techniques for data retrieval method and promising the flourishing mechanism in database systems and new database applications. The SETS is a knowledge discovery mechanism and automated for extraction data sets which are stored in the database. The paper SETS can be viewed as a result of natural evolution of information technology. The SETS makes the user can gain the convenient and flexible data access through the queries on set of data[1]. This technology provides a great boost to the database and information industry and makes a huge amount number of databases and information repositories available for transaction management, data analyzing methods and information managements. It can be a powerful tool for the fast growing and tremendous access of data which are collected and stored in large and numerous data bases. The SETS provides the data analysis and covers the important patterns which are contributing to the business requirements.

Keywords: Transaction management, Repositories, flourishing mechanism, information industry, patterns

I. INTRODUCTION

The transmission of streams is a process to convert the operational data into informational access that includes the unwanted data is removed. The streams generally convert the heterogeneous data marts in to one common schema. This is the major problem in accessing the data from multiple heterogeneous sources. So, while transmitting the streams in a sequential manner may avoid some of the problems encounters during the accessing the huge amounts of data[2]. The transmissions of steam are similar to snap shot of the data so a multiple snapshots may be used to be emerged to create

the historical views. The summarization of data is mainly performed by the higher levels of data transmission of streams. It can also be performed by using the multiple granularities with various dimensions. The newly derived data is better to facilitate for decision support facilities and also in handling missing the erroneous data. It can entertain by using the predicated and by using the default values for the entries. Generally many applications require the summary data that are known to be needed for consolidations and pre-summarization before using the data. Different levels of summarization should be improved to increase the performance The streams technology should consider the normalization phenomenon which reduces the redundancy as well as the insertion, updating and deletion operations. However these improvements are achieved at the cost of increasing the processing time[2],[3]. Generally the performance can be improved by storing the demoralized data. Since the data ware house are not frequently updated as operational data and the negative associative with the updated operations which are not in issue.

II. LITERATURE SURVEY AND METHODS

The SETS assist the mining activities in a large amount of data in the ware houses. The fact is the ware houses are continually growing and obtaining the desires information is a challenging access. As stream queries, the search engine queries can be stated as keyword, Boolean, weighted and so on. The difference is mainly concerned with the data being searched, pages with heterogeneous data and extensive hyperlinks and the architecture involvement.

> - The SETS query will retrieve only a very small subset of data from the huge data involved.
> - It provides the result from subset of the ware house. In other words it looks to the indices that updated periodically so that the index is directly accessed.
> - It provides access based only on the keyword based searching and also in the order of data samples based on other properties such as the popularity of the data

These queries are similar to the pattern matching and recognition techniques which can be used in many diverse applications. The text matching is to find the occurrences of a string in the text being edited. The information retrieval and the search engine may use the pattern sets to find the documents containing the predefined keywords. Each data set that is a mapped to the same cluster set may be the thoughts of more similar to the other data sets in the same cluster than it is to the data sets found in the other clusters. Therefore the distance measures may be used to identify the likeness of the different items in the database [4].

The transmission of streams is a measure of the classification where the classes are predefined is somewhat than using the similarity measure for clustering where the clusters are not known in advance. The classification problem becomes one of the determining similarities among all tuple in the database but between each tuple and the query. It makes the an $O(n)$ rather than an $O(n^2)$.

2.1 Measures of Approach

If we have representative of each class then we can perform the classification by assigning the tuple by knowing to which it is more similar. Let us assume that each tuple in the data base is defined as a vector of numeric values. The tuple be $T_1, T_2 \ldots, T_n$ and the classes which defined by the tuple are $C_1, C_2, C_3, \ldots C_n$.
The classification problem can be restated as

Definition: The data base D= {T1, T2....., T_n} of tuples where each tuple contains a numeric values and a set of classes C={ $C_1, C_2, C_3, \ldots C_n$.} where each class contains a numeric value. The classification problem is to assign the T_i to C_j such that $(T_i, C_j) \geq$ similar $(T_i, C_l) \vee C_l \varepsilon$ C where $C_l \neq C_j$

2.2 Sample algorithm:

Inputs:
C1,…,C_m // center for every class
I // input tuples to classification
Outputs:
 C // it is the class to which t is assigned
Distances measure calculations
Dist = infinite I;
For I=1 to m do
If dist(C_I, t) < dist then
C= I;
Dist = dist (C_I, t);
End
The distance is the measure of calculating the object to find. The streams are generated when the distance was in minimum shortest path so that the required tuple is retrieved from the database.

III.   ACCESSING DATA BY SETS

The SETS system takes the data from SQL server, OLAP data cube, which itself constructed from the single and multiple relational tables of data ware house systems and also with other form of the data are spreadsheets. The Emission can be represented by many forms depending on the retrieval operations and user preferences. The data summarization and characterization generate the cross tabulation table, generalized rules, bar charts pie charts curves and other forms of graphical output displays. The system provides to view the conceptual hierarchies and data cube contents. The conceptual hierarchies are presented in the form of tree form similar to directory and subdirectory structures. The cube contents are presented in a three dimensional cube form where the size of the cuboids in the three dimensional cube represented in the summarization of the corresponding selected measures with in a set of 3-D intervals. The SETS mines the set of association rules from a n- dimensional data bases and data ware houses. The rules are meant for cross market analysis, correlation analysis and so on. The search of interesting rules can be confined so as to match the user

defined Meta data [5],[6]. The meta data are "Details(S: Student, X) ∧ P(S, Y) which implies the relation of Class(S, Y, Z). where the Details and Class are the predicates which corresponds to the attributes in the relation student.

Algorithm:

Procedure for retrieval;
Var I: integer;
Winner: integer;
Unit: float;
Find: float [ ];
Inputs: floats [ ];
Begin
Find = F2. Outs; // index to find the desired data set;

Winner = F2.W; // index to winning data
Retrieval = F2.Weights [Winner];
Inputs = F1. Outs;
For i=1 to length
Do
{ Update the units accordingly

Winner = F1.Outs * inputs[i] ;
End do
For I = 1 to length // now to retrieval
Do
Winner = F2.Outs * Find[i]. Weights [I]
Winner = Winner [I] + inputs [I];
End do
End procedure

IV.   IMPLEMENTAION OF SETS

The implementation is performed by Grouping the selected set of training data and set of objects whose class label should be known. It constructs the model for each class based on the features in the data and adjusts the model based on the test data. This model is presented in the form of decision trees and is used to classify future data and to develop a better understanding of data in the data base. The high dimensional can be performed by the multi dimensional data bases. It predicts the values for certain missing and unknown data in the selected set of objects. It involves finding the set of attributes relevant to the attribute of interest by some statistical analysis [7]. The predicting the value distribution base on the set of data similar to the selected objects, generally in any organization the employees are the potential data which can be predicted based on the salary distribution of similar employees in the company
The graph shows the predictions in the chart 1
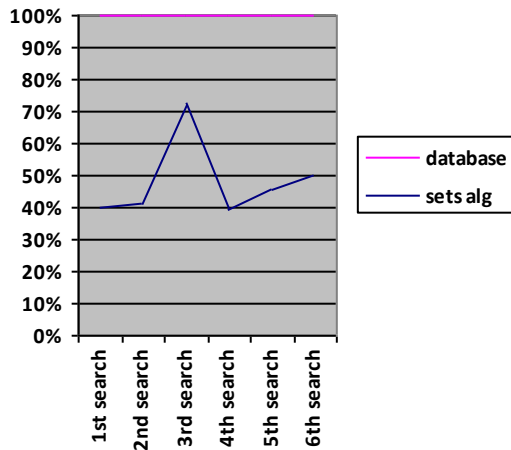
Chart 1

The chart indicates the transmission of streams in a sequential to find the sets in the data ware house. The searching mechanism is repeated in the entire database with comparisons of the data sent by the user till the required sets are available.

V. CONCLUSION AND FUTURE WORK

The SETS algorithm is mainly proposed to the for the competitive market, many of the companies needs the innovate and requires the efficient data retrieval methods. The Sequential Emission by Transmission of Streams performs the comparisons in the form of sets. It makes the retrieval process very easy because at a time a group of data is shown by the algorithm in the form of sets which can build the companies to extend their Ware houses and data bases[8],[9]. The main theme of the algorithm is for banking and on line sectors because the data base is updated rapidly and increases day by day with the new customers. The paper is mainly proposed for online analytical transaction The algorithm constitutes the sets mechanism but to a limited extent. The future work has to elaborate the clustering techniques. This algorithm penetrates for 2D versions and should be extended to the 3D versions and also for multi dimensional. So, it is also recommended to bring the algorithm for minimal iteration for searching the sets and for 3D applications. So that it can to improve to increase the performance of the retrieval of data or information with more speed and can viewed in all sorts of dimensional methods.

REFERENCE:

1. http//www.google.com for study materials.
2. Data Mining by Margaret H.Dunham.Introductory and Advanced Topics Margaret H. Dunham.2002 Publisher: Prentice Hall pp 85-88
3. Data Mining Concepts and Techniques second edition by Jiawei Han and Micheline Kamber pp 494-499
4. J. Han, M. Kamber, and A. K. H. Tung. Geographic Data Mining and Knowledge Discovery, chapter Spatial Clustering Methods in Data Mining: A Survey, pages 15-29. Taylor and Francis, 2001
5. Data Mining: Concepts and Techniques, 2 Edition ... August 26, 2010.
6. "iSAX 2.0: Indexing and Mining One Billion Time Series," icdm, pp.58-67, 2010 IEEE International Conference on Data Mining, 2010
7. In Proceedings of the 12th ACM SIGKDD in the International conference on Knowledge discovery and data mining (KDD '06). ACM, New York, NY, USA, 748-753.
8. M.J.Zaki Efficient Enumeration of frequent sequences. In proc 7th int. conf information and knowledge Management 98 pages 68-75 washing ton DC, nov 1998.
9. Fabio Aioli, Ricardo Cardin, Fabrizio Sebastiani, Alessandro Sperduti,"Preferential Text Classification: Learning Algorithms and evaluation measures", Springer – Inf Retrieval 2009.

AUTHORS PROFILE

Prof. P.V. SarathChand, B.E (C.S.E), M. Tech(CS), and pursuing PhD(CSE).He is having 14 years of Teaching Experience. He is working as a Professor and Head of I.T Dept for Indur Institute of Engineering and Technology,Siddipet, Medak, Dist.

K. Venu Madha. is having 13 years of teaching experience and presently working as Senior Lecturer in the Department of Applied Information Systems in the University of Johannesburg, South Africa.

Mahendra Arya Bhanu, M.Tech(cse)(PH.D). He is having 20 years of teaching experience and Working as Professor and HOD in Vidya Jyothi Institute of Technology, Hyderabad in the Dept. of Information   Technology.

K. Nagamani M.Sc (cs), M.Tech(CSE).She is having 7 years of teaching experience. Presently working as Assistant Professor in **Indur Institute of Eng. & Technology**, Siddipet, Medak Dist.A.P. India.

Mr. Balaram A., M.Sc Mathematics from Andhra University Visakhapatnam and M.Tech (cse) from Andhra University Visakhapatnam. Presently working as a Asst. Prof & M.Tech Coordinator in CSE dept at Indur Institute of Eng & Tech. He is having 5 years of teaching.