RESEARCH PAPER | Science | Volume : 4 | Issue : 8 | August 2014 | ISSN - 2249-555X

Apriori Rule Based Continuous Processing of Preference Queries

KEYWORDS: Top-k dominating queries, data streams, continuous queries, algorithms, analysis, approximation

P.Kiruthika

Assistant Professor, Dr. NGP Arts and Science College, Kalapatti, Coimbatore – 641048

ABSTRACT *During the course of this project, the frequency value is found using ranking function. This concept is used in Real Estate Property Management Application. Hence this web application is useful to the users who buying property can get the accurate rank value with highest priority. Already we have a ranking function which is based upon Top-k query and Skyline query algorithm that provides the approximate ranking value. Top-k dominating queries use an intuitive scoring function which ranks multidimensional points with respect to their dominance power, i.e., the number of points that a point dominates. The k points with the best (e.g., highest) scores are returned to the user. Both top-k and skyline queries have been studied in a streaming environment, where changes to the data set are very frequent. In comparison to continuous top-k and skyline queries, continuous top-k dominating queries pose additional challenges. Two exact algorithms (EVA [Event Based Algorithm], ADA [Advanced Algorithm]) are studied, and among them ADA, which is enhanced with additional optimization techniques, shows the best overall performance.*

I. INTRODUCTION

The Main Objective of the project is to find frequency value using ranking function. We use this concept in Real Estate Property Management Application. Hence this web application is useful to the users who buying property can get the accurate rank value with highest priority. Data mining refers to extracting or "mining" knowledge from large amounts data. The term is actually a misnomer. Remember that the mining of gold from rocks or sand is referred to as Gold mining rather than rock or sand mining. Thus ,data mining should have been more appropriately named "Knowledge mining from data" .many other terms carry a similar or slightly different meaning to data mining, such as knowledge mining from data, knowledge extraction, data/ pattern analysis, data archaeology, and dredging. Two of the most widely used preference-based queries are: 1) the top-k query and 2) the skyline query. In a top-k query, a ranking function f : IRd ! IR is required, which assigns a value to each point p. The result of a top-k query comprises the k points with the highest values with respect to f(). A nice feature of top-k queries is that the number of answers is controlled by the parameter k, although in some cases the cardinality of the result set may exceed k due to ties (i.e., two or more points may have the same value.) In such a case, either all ties will be part of the answer or a tie breaking criterion is applied to select exactly k answers. The most important limitation of top-k queries is that a ranking function is required. This function is usually user defined, where as different functions generally result in different results. Moreover, in several cases the selection of an appropriate ranking function is not intuitive. For example, in an ecommerce application, there is no straight forward way to combine the attributes CPU speed and battery autonomy to select the most interesting laptop computers. Skyline queries, on the other hand, do not require a ranking function and they have the scaling in variance property, meaning that if scaling is applied to dimension values the result remains unchanged. The result of a skyline query is composed of the points that are not dominated by any other point. The dominance relationship depends on the semantics of each attribute; in some cases, small values are preferable (e.g., price) whereas in other cases large values are desirable (e.g., quality). Without loss of generality, we focus on minimizing dimension values (the smaller the better). Therefore, a point p dominates another point q(p _ q), if and only if p is no worse than q in all dimensions and strictly better than q in at least one dimension.

Thus the fundamental property of these algorithms is that they operate in an ad hoc fashion, meaning that they initiate a query processing task only if a query is issued. This is sufficient for applications operating on static or almost static data sets, where updates are rare. These functions are usually user defined, whereas different functions generally result in different results. Moreover, in several cases the selection of an appropriate ranking function is not intuitive. In existing system they are using two algorithms like top-k query and skyline query algorithm. In existing system the most important limitation of top-k queries is that a ranking function is required. This function is usually user defined, whereas different functions generally result in different results and the selection of an appropriate ranking function is not intuitive. Skyline queries, on the other hand, do not require a ranking function and they have the scaling invariance property, meaning that if scaling is applied to dimension values the result remains unchanged. The result of a skyline query is composed of the points that are not dominated by any other point. Therefore, top-k dominating queries have the following desirable properties:

1. The number of results is controllable,
2. The result is scaling invariant,
3. No user-defined ranking function is required, and
4. Each point is assigned an intuitive score which determines its rank.

Given a database D of n objects, a scoring function F (according to which we rank the Objects in D) and the number of expected answers k, a Top-k query returns the k objects with the best score (rank) in D. In top-k the most important limitation of top-k queries is that a ranking function is required. This function is usually user defined, whereas different functions generally result indifferent results and the selection of an appropriate ranking function is not intuitive

Volume : 4 | Issue : 8 | August 2014 | ISSN - 2249-555X

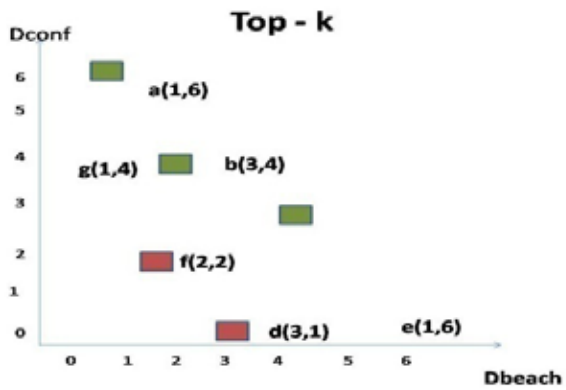
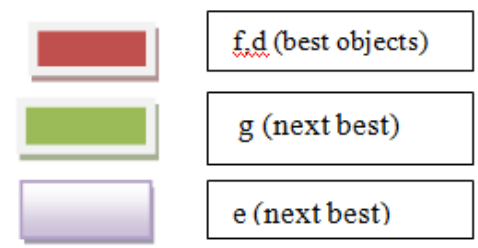


Fig.1.Top-K Query Algorithm



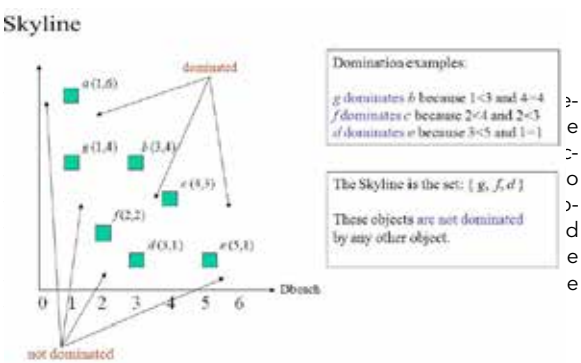


Fig.2.Skyline Query Algorithm

In existing system they are using two algorithms like top-k query and skyline query algorithm. In existing system the most important limitation of top-k queries is that a ranking function is required. This function is usually user defined, whereas different functions generally result in different results and the selection of an appropriate ranking function is not intuitive. Skyline queries, on the other hand, do not require a ranking function and they have the scaling invariance property, meaning that if scaling is applied to dimension values the result remains unchanged. The result of a skyline query is composed of the points that are not dominated by any other point.

- The number of objects in the answer (k) is user-defined.
- The "best" score is either the lowest or the highest depending on user preferences.
- The ranking function F, may involve more than two attributes.
- Sorting is an expensive operation requiring a complexity.
- Requires scanning the whole database for each object. This is not convenient in systems with large volumes of data.

II. RELATED WORK

In top-k the most important limitation of top-k queries is that a ranking function is required. This function is usually user defined , whereas different functions generally result in different results and the selection of an appropriate ranking function is not intuitive. The result of a skyline query is composed of the points that are not dominated by any other point. In 2003 Babcock, et al in "Models and Issues in Data Stream Systems" In this paper we motivate the need for and research issues arising from a new model of data processing. In this model, data does not take the form of persistent relations, but rather arrives in multiple, continuous, rapid, time-varying *data streams*. In addition to reviewing past work relevant to data stream systems and current projects in the area, the paper explores topics in stream query languages, new requirements and challenges in query processing, and algorithmic issues. In 2009 Kontaki, et al in "Continuous Top-*k* Dominating Queries in Subspaces" describing the Dominating queries are significant tools for preference based query processing in databases and decision support applications. An important preference-based query is the top-k dominating query, which reports the k most important objects according to their domination capabilities (score).In this paper, we address the following issues to tackle two limitations of previously proposed approaches: (i) we allow dominating queries to be expressed in a subset of the available dimensions and (ii) we provide the necessary techniques to enable continuous processing of multiple queries. We use a grid-based indexing scheme to facilitate efficient search and update operations, avoiding expensive reorganization costs. In addition, several optimizations are proposed to enhance efficiency. Performance evaluation results, based on real-life and synthetic data sets, show the efficiency and scalability of the proposed scheme.

PROPOSED SYSTEM

Proposed system which combines the concept of dominance with the notion of Ranking functions. This new query is termed top-k dominating query and in a sense it is a combination of top-k and skyline queries: it uses a ranking function to rank points (as in top-k queries) and it uses the dominance relationship (as in skyline queries). The score associated with a point pi, denoted as score, equals the number of points that pi dominates. The motivation behind this idea is to define a preference query that maintains the advantages and eliminates the limitations of both top-k and skyline queries.

Advantage

- Proper ranking function
- dominance relationship
- eliminates the limitations

Continuous top-k dominating queries

This new query is termed top-k dominating query and in a sense it is a combination of top-k and skyline queries: it uses a ranking function to rank points (as in top-k queries) and it uses the dominance relationship (as in skyline queries). The score associated with a point pi, denoted as score (pi), equals the number of points that pi dominates. The motivation behind this idea is to define a preference query that maintains the advantages and eliminates the limitations of both top-k and skyline queries.

EVA (Event based Algorithm)

In computer programming, event-driven programming or event-based programming is a Programming paradigm in which the flow of the program is determined by events—

Volume : 4 | Issue : 8 | August 2014 | ISSN - 2249-555X

e.g., sensor Outputs or user actions (mouse clicks, key presses) or messages from other programs or threads.

- Objective: reduce domination checks
- interval of a tuple
- ✓ Ignore tuple for this interval
- ✓ It depends on its score and the k-th score
- End of safe interval -> event
- Event
- ✓ Try to compute new safe interval, else
- ✓ Compute score from scratch
- New tuple
- ✓ Find another tuple that dominates the new one
- ✓ Estimate a lower bound of the safe interval

THE ADVANCED ALGORITHM (ADA)

- Advanced computation of safe interval
- ✓ Depends on the number of tuples that dominate this tuple and expire later
- Candidate tuples
- ✓ Tuples with scores close to k-th score are updated in each time instance

Advantages

- As Top-k dominating queries intuitive ranking function is used, the result is unaffected by dimension scaling and the number of results is bounded by k.
- Less complexity when compared to top-k and skyline.
- The best overall performance can be achieved.
- Faster computation
- Better accuracy compared to top-k and skyline queries.

PERFORMANCE MEASURES

To measure the performance Continuous Top-K Dominating query algorithm we compare it with existing algorithm. The comparison of both the algorithm is given in Table 5.1.Which shows that Accuracy consumed by Continuous Top-K Dominating query algorithm is less compared to existing algorithm. In top-k query ranking function is not intuitive, Skyline query is difficult to return dominating value. Proposed Algorithm have number of attractive properties so the three algorithms, ADA, AHBA, and AMSA can work with combination.EVA Algorithm used to calculated the spatial's data information such as hospital, market, college, school, and transport.

Algorithms	Ranking Function	Dominating value
Top-K Query	Medium	Medium
Skyline Query	Medium	Low
EVA	High	Medium
AMSA+AHBA	Medium	High
ADA	High	High

Table 1.1 Measuring Accuracy

PERFORMANCE ANALYSIS

To analyze the performance based according to the performance measures comparison table 5.1. In this performance analysis the existing algorithm provides the approximate results. And the ranking function is not intuitive but in the proposed algorithm gives the accurate value for the ranting function. In the proposed algorithm we are using the dominating function to find the highest priority value to manipulate rank to display on the top most of the list.

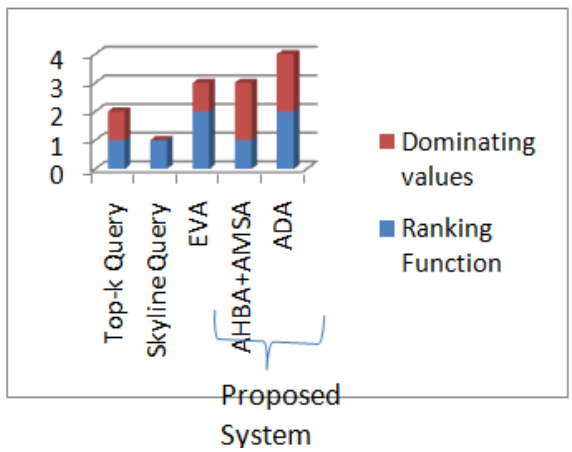


Fig 3. Performance Analysis

CONCLUSION

Thus the proposed system Top-k dominating queries have number of attractive properties so the three algorithms, ADA, AHBA, and AMSA, can work in combination. If we have to be strict regarding accuracy then ADA must be used. In case a heavy system load is detected and accuracy may be compromised then AHBA may be activated. Finally, if an even more fast solution is required AMSA may be used. Since all algorithms are based on the event based framework, they can operate in a synergetic manner, offering significant flexibility

REFERENCE

1 I.F. Ilyas, G. Beskales, and M.A. Soliman, "A Survey of Top-k Query ProcessingTechniques in Relational Databases," ACM Computing Surveys, vol. 40, no. 4, pp. 1-58, 2008. | 2W. Kiessling, "Foundations of Preferences in Database Systems," Proc. 28th Int'l Conf. Very Large Data Bases (VLDB), pp. 311-322, 2002. | 3M.Kontaki, A.N. Papadopoulos, and Y. Manolopoulos, "Continuous Top-k Dominating Queries in Subspaces," Proc. Panhellenic Conf. Informatics (PCI), 2008. | 4 X. Lian and L. Chen, "Top-k Dominating Queries in Uncertain Databases" Proc. 12th Int'l Conf. Extending Database Technology: Advances in Database Technology (EDBT), 2009. | 5 X. Lin, Y. Yuan, Q. Zhang, and Y. Zhang, "Selecting Stars: The k Most Representative Skyline Operator," Proc. IEEE 23rd Int'l Conf. Data Eng. (ICDE), pp. 86-95, 2007. | Books references | ➢ DATA MINING Concepts and Techniques book written by Jiawwi Han|Micheline Kamber. | | Web reference | ➢ www.indaiproperty.com | ➢ www.sulekha.com | | BIOGRAPHY | | Ms.P.Kiruthika, received her professional degree MCA from Hindustan university, Chennai. Her area research includes data mining. She has presented 2 papers in national conferences. She has 1.5 years of teaching experience in arts and Science College. |