International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 3 Issue: 6 3858 - 3861

Query Recommender System Using Hierarchical Classification

Shruti Y. Patil
Department of Computer Science
S. S. G. B. College
Bhusawal, India
shrutiypatil@gmail.com

Prof. Dinesh D. Patil
Department of Computer Science
S. S. G. B. College
Bhusawal, India
dineshonly@gmail.com

***Abstract*—** In data warehouses, lots of data are gathered which are navigated and explored for analytical purposes. Even for expert people, to handle such a large data is a tough task. Handling such a voluminous data is more difficult task for non-expert users or for users who are not familiar with the database schema. The aim of this paper is to help this class of users by recommending them SQL queries that they might use. These SQL recommendations are selected by tracking the users past behavior and comparing them with other users. At first time, users may not know where to start their exploration. Secondly, users may overlook queries which help to retrieve important information. The queries are recorded and compared using hierarchical classification which is then re-ranked according to relevance. The relevant queries are retrieved using users querying behavior. Users use a query interface to issue a series of SQL queries that aim to analyze the data and mine it for interesting information.

Keywords- *Data Mining, Data discovery, Interactive data exploration, Query personalization.*

*****

I. INTRODUCTION

The Database systems are becoming gradually more popular in the scientific community to support the interactive exploration of large volumes of data. For example the Genome browser which provides access to the genomic database and SkyServer stores large volumes of astronomical measurements. These databases allows user to submit queries and retrieve the results.

The need for data discovery tools increases as the data are increases vastly. Over large databases, despite the availability of querying tools, the users often have difficulties in understanding the underlying complicated schema and formulating queries. For example, the study on Hive. The data warehouse platform used in Facebook, come up with the following: Because of the heavy usage, a lot of tables are generated in data warehouse which in turn vastly increased the need for data discovery tools. Even when users have the capability to issue complex queries over large data sets, the task of knowledge discovery remains a big challenge. Moreover, a complete exploration of such databases is not practically feasible due to the continuously increasing size of the data.

To support the non-expert user for retrieving interesting information, query recommender system is used. Recommender Systems are software tools and techniques which provides suggestions to users for items to be of their use. The suggestions provided help the users in various decision-making processes, such as what items to buy, which music to listen, or what news to read. For online users, recommender systems have proven to be valuable means to deal with the information overload and have become one of the most powerful and popular tools in electronic commerce.

The interest of the user is drawn out and makes recommendations accordingly by the Query Recommender systems. Those recommended queries can be used as templates and submitted as it is instead of composing new ones or they can be further edited. In an attempt to identify previous users with similar information needs, this system continuously monitors the user's querying behavior and finds matching patterns in the system's query log. Subsequently, query recommender system uses these "similar" users and their queries to recommend queries that the current user may find interesting.

In this work, the query is stored in the query log. When the active user fired a query, his query is stored in the log and matched with the past users queries in query log with cosine similarity method. If queries are matched then those queries are recommended to active user so that they may be of user interest. This motivation draws from Web Recommender System. The principle on which the system is built is simple: If user A and user B placed the same queries then the other queries of each user may be of interest of each other.

This idea is projected with the help of collaborative filtering. A collaborative Query Management System is placed on new, large scale, shared data environments. Collaborative filtering method makes automatic predictions (filtering) about the interest of the user by collecting preferences information from many users. This method requires (1) user's active participation, (2) an easiest way to represent user's interest to the system, and (3) methods that are able to match people with similar interests. The system should also mine its query log and actively recommend queries to users, thus help them further pull the previously performed analysis.

II. RELATED WORK DONE

A keyword based query interface is provided by web databases which suffers from the empty answer or too many answers problems. In such case, it is critical to provide the correct answer to each user rather than the exact one to everyone for the same query. Preferences may be expressed qualitatively or quantitatively by embedding into relational query languages a special operator or by re-ranking or filtering the results of the original query respectively. Recently, as an additional personalization parameter, context has been added in such systems. The common denominator of these works with ours is that all can be categorized under the query personalization area.

 ISSN: 2321-8169
3858 - 3861

A multidimensional query recommendation system is already present. In this work, for generating OLAP query recommendations, the authors propose a framework for the users of a data warehouse. Although this work has some similarities to our, the techniques and algorithms used in the multidimensional development are very different to the ones we propose.

III. PROPOSED SYSTEM

Database systems provide the critical infrastructure to access and analyze large volumes of data in a variety of applications. For accessing the data from database, user uses querying tools. However new or not expert users face difficulties in understanding the underlying schema and formulating queries. This system provide help to database user, such as recommending the past user queries and also giving authority to execute or to edit recommended queries. When the structure of the queries is wrong then also our system provides similar recommend queries and gives authority to execute these queries.

The abstract framework is fundamentally a workflow, as shown in Fig. 1.

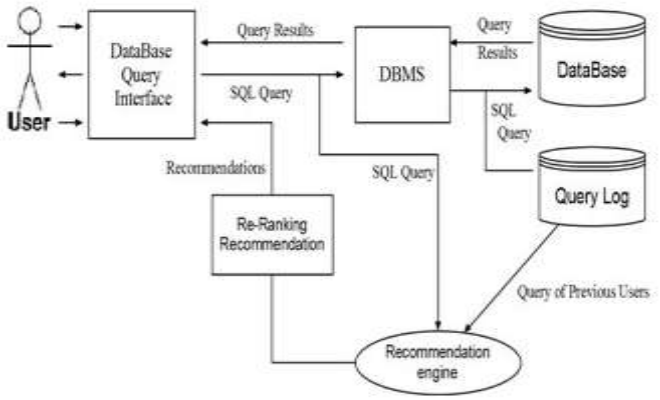

Figure 1. Architecture of Query Recommender System

The active user's queries are forwarded to both the database management system and the Recommendation Engine. The Database Management System processes each query and returns a set of results. At a time, the query is stored in the Query Log. The Recommendation Engine combines the current user's input with information gathered from the database interactions of past users, as stored in the Query Log, and generates a set of query recommendations that are returned to the user. If the query is nested then the recommended queries are re-ranked before returned to the user.

IV. FLOW OF SYSTEM

The Query recommender system recommends the query as per user interest. The flow of proposed system is as shown in fig.2.

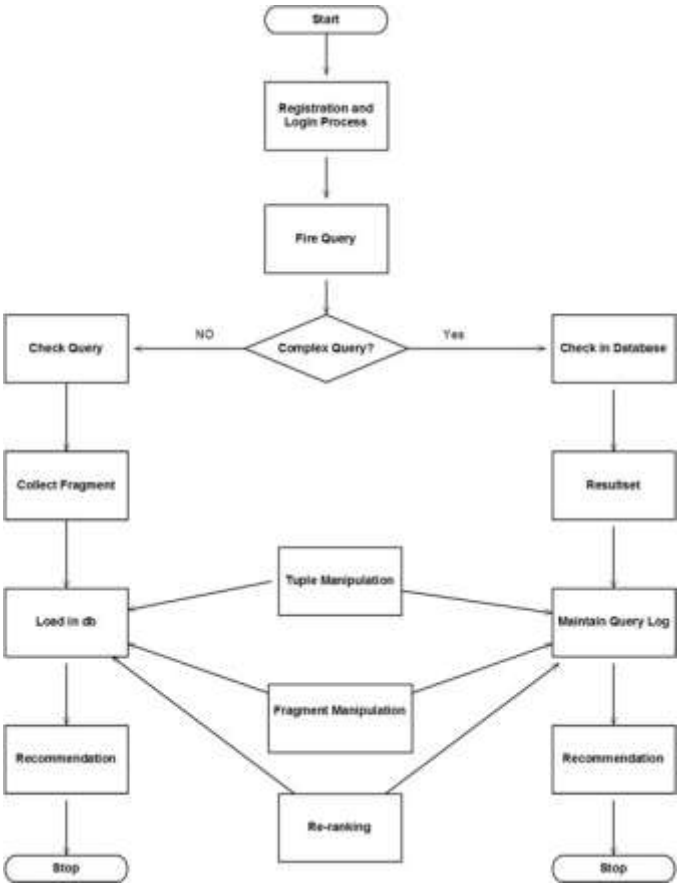

Figure 2. Flow of Query Recommender System

User needs to first register and login to a system. After that, he enters his query. If query is not valid then system gives message that not valid query. If fired query is nested query then it will check similarity in database and recommend relevant queries. User can edit query which he enters and corrects his query or choose option of recommendation. The recommended queries can also be edited. When query is fired then DBMS check the attribute and syntax of the query. If attribute and syntax are wrong then DBMS give the error message. Each user query stored in database independently. Query log avoids redundancy and duplication. When user query is simple, it is fragmented into different attribute using fragment method. Using fragment attribute recommendation engine find similar queries and recommend queries to user.

V. IMPLEMENTATION

A. *Fragment Based Recommendation*

This approach is based on the pair-wise similarity of query fragments i.e. on attributes, tables, joins and predicates. We need to identify fragments that co-appear in several queries posed by different users.

- Session Summary
  For a user i, the Session Summary vector S_i consists of all the query fragments ϕ of the users past queries. Let the set of queries posed by user i during a session is represented by Q_i .The set of all distinct query fragments recorded in the query logs represented by F. Let assume that a single query Q_i is represented by the vector SQ. For a given fragment ϕ F, we define SQ[ϕ] as a binary variable that represents the presence or

 ISSN: 2321-8169
3858 - 3861

absence of in a query Q. Then $S_i[\phi]$ represents the importance of in session S_i.

Weighted Scheme

$$S_i = \sum_{Q \in Q_i} S_Q$$

- Recommendation Seed Computation
  For generating recommendations, the framework computes a "predicted" summary S captures the predicted degree of interest of the active user S serves as the "seed" for the generation of recommendations. The predicted summary is defined as follows

$$S_0^{pred} = f(\alpha * S_0, (1-\alpha) * \{S_1, \ldots\ldots S_h\}).$$

  "Mixing factor" α [0, 1] that determines the importance of the active user's queries

  With the Use of the session summaries of the past users and a vector similarity metric, we construct the (|F| x |F|) fragment-fragment matrix that contains all similarities sim(ρ,ϕ) ρ,ϕ F.
- Generation of Query Recommendation
  Once the predicted summary S_{pred} has been computed, the top-n fragments Fn (i.e. the fragments that have received the higher weight) are selected. Then all past queries Q, Q U Q_i receive a rank QR with respect to the top-n fragments:

$$QR(Q) = \frac{|F_Q \cap F_n|}{|F_Q|} * \frac{|F_Q \cap F_n|}{n},$$

B. Recommendation for Nested Query Using Hierarchical Classification

One of the most powerful features of a query language is the nesting of queries means the possibility of writing in a single expression a query which uses the output of other queries. The recommendation for the nested queries is done by matching the query of active users with the recorded queries in query log of past users with the help of cosine similarity method. Cosine similarity is a similarity measure which can be computed amongst arbitrary vectors:

$$\cos \ell = \frac{a \cdot b}{\|a\| \|b\|}$$

Where a and b are vectors we want to compare.

The cluster of the most matching queries is formed for recommendation. For the clustering, we are using hierarchical clustering algorithm:

Agglomerative clustering method is one of the Hierarchical Clustering algorithm's methods. The algorithm forms clusters in bottom to up manner as follows:

1. Initially, put each article in its own cluster.
2. Among all current clusters, pick the two clusters with the smallest distance.
3. Replace these two clusters with a new cluster, formed by combining the two original ones.
4. Repeat the above two steps until there is only one remaining cluster in the pool.

Thus, the agglomerative clustering algorithm will result in a binary cluster tree with single article clusters as its leaf nodes and a root node containing all the articles.

VI. PROPOSED FRAMEWORK

The query recommender system recommends queries as per user interest. This system helps to non-experts users for analyzing and extracting the information of their interest in data warehouse.

User needs to first register himself for the use of this system. As shown in figure 3, after clicking on the button 'Execute a Query', user can fired a query. At the time of firing of query, user gets recommendation according to his interest. He can select any of the recommended queries to execute or he can be able to edit that query as per need.

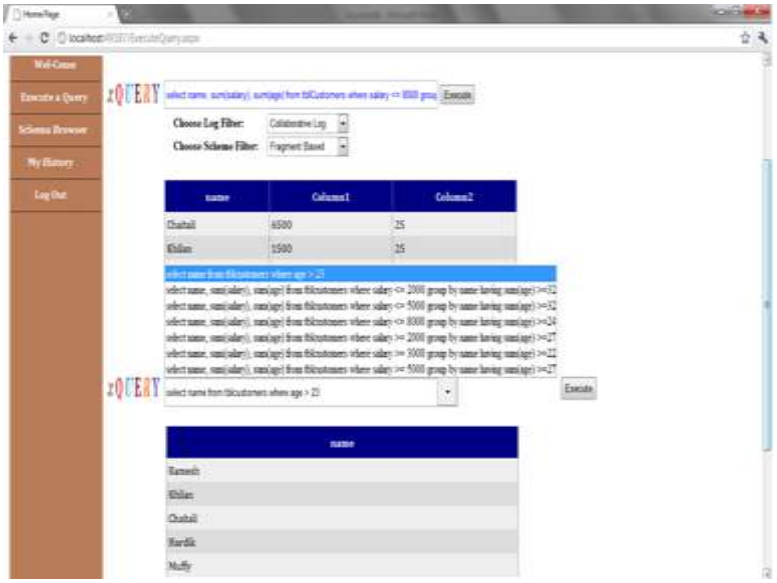

Figure 3. Execution of Query with Collaborative Log and Fragment Based Filter

When user fired nested query then it will match with database query with the help of cosine similarity method. And all the matching queries are recommended to the user Here, we are not using any filter. It is as shown in fig. 4.

The users who is first time going to fired a query may not know the definition schema of the selected table. The user first needs to select the table name of which user wants to see the definition. After clicking on the button 'Schema Browser',it will displayed the column names and data types of that column of the selected table.

If user wants to see the history of his queries i.e. if he want to see which queries he fired previously then it can shown by 'My History'. By clicking on button 'My History' , it will display all the queries fired by that user previously.

 ISSN: 2321-8169
3858 - 3861

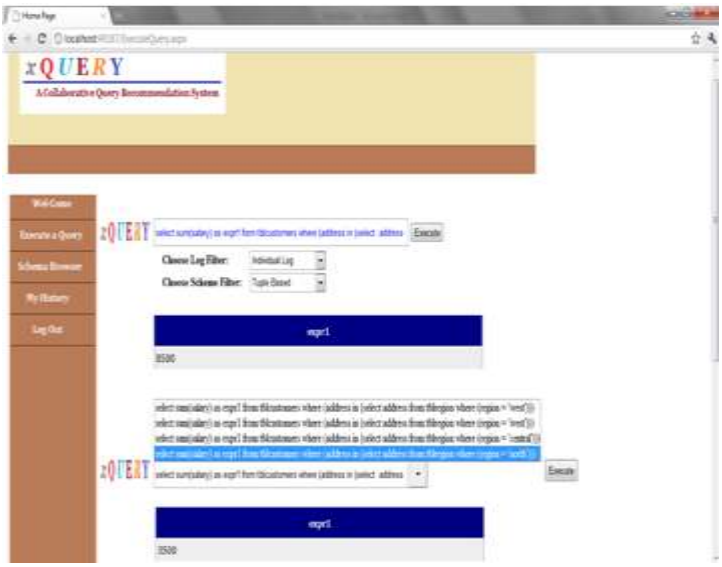

Figure 4. Recommendation for Nested Query Using Hierarchical Classificatioin

VII. Performance Evaluation

In order to evaluate the performance of the proposed technique, we have used two measures which are Precision and Recall. Precision and recall are the most popular metrics which are widely used in evaluating search strategies.

Precision is defined as the ratio of the number of queries that correctly retrieved to the total number of queries retrieved.

$$\text{Precision} = \frac{\text{Number of relevant queries retrieved}}{\text{Total Number of queries retrieved}}$$

While recall is the ratio of the number of queries that retrieved correctly to the total number of queries in log.

$$\text{Recall} = \frac{\text{Number of relevant queries retrieved}}{\text{Total relevant queries in log}}$$

High precision means less irrelevant queries are returned or more relevant queries are retrieved. The performance of the system, by plotting precision and recall graph is as shown in figure 5, in which precision values are plotted against recall values.

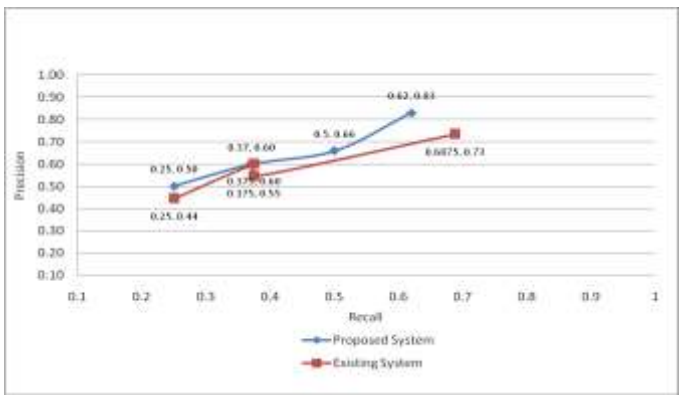

Figure 5. Comparision Graph of Existing and Proposed System

Conclusion

Recommendation system is a well-known field of the data mining used to pull out the essential knowledge from a huge amount of user query logs. Query recommender system supports users for analyzing and mining interesting information. A query recommendation system using fragment based approach helps the users to execute simple SQL query and proposed recommender system using hierarchical classification helps users to execute nested queries. The proposed system is shown to outperform previous fragment based approach.

Acknowledgment

The completion of this paper would not have been possible without the support and guidance of Prof. Dinesh D. Patil. He has been a constant of motivation and has encouraged me to pursue this project. With my deep sense of gratitude, I thank my respected teachers for supporting this of my project. This project report provides me with an opportunity to put in the knowledge of advanced technology. I thereby take the privilege opportunity to thank my guide and my friends whose helps and guidance made this study possibility.

References

[1] Magdalini Eirinaki, Suju Abraham, NeoklisPolyzotis, and NaushinShaikh, "QueRIE: Collaborative Database Exploration" IEEE transactions on knowledge and data engineering, vol. 26, no. 7, July 2014.

[2] A. Thusooet al., "Hive - A petabyte scale data warehouse using hadoop," in Proc. IEEE 26th ICDE, Long Beach, CA, USA, Mar.2010, pp. 996–1005.

[3] G. Chatzopoulou, M. Eirinaki, and N. Polyzotis, "Collaborative filtering for interactive database exploration," in Proc. 21st Int.Conf. SSDBM, New Orleans, LA, USA, 2009, pp. 3–18.

[4] S. Mittal, J. S. V. Varman, G. Chatzopoulou, M. Eirinaki, and N. Polyzotis, "QueRIE: A recommender system supporting interactive database exploration," in Proc. IEEE ICDM, Sydney, NSW, Australia, 2010.

[5] J. Akbarnejadet al., "SQL QueRIE recommendations," PVLDB, vol. 3, no. 2, pp. 1597–1600, 2010.

[6] N. Alon, Y. Matias, and M. Szegedy, "The space complexity of approximating the frequency moments," in Proc. 28th STOC, New York, NY, USA, 1996.

[7] N. Koudas, C. Li, A. K. H. Tung, and R. Vernica, "Relaxing join and selection queries," in Proc. 33nd Int. Conf. VLDB, Seoul, Korea, 2006, pp. 199–210.

[8] A. Giacometti, P. Marcel, and E. Negre, "Recommending Multidimensional Queries," in Proc. Of the 11th Intl. Conf. on Data Warehousing and Knowledge Discovery(DaWaK'09), 2009.

[9] B. Sarwar, G. Karvpis, J. Konstan, and J. Riedl."Item-based collaborative filtering recommendation algorithms". In 10th Intl WWW Conf., 2001

[10] G. Koutrika, B. Bercobitz, and H. Garcia-Molina, "Flexrecs: expressing and combining flexible recommendations" in SIGMOD conference, June 2009.

+