ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

Memory Optimization Using Genetic Algorithm of Relational Keyword Search Techniques

A.Ravichandiran, A.Vijayan, K.Ravikumar

Post Graduate Student, Dept of CSE, Rrase College of Engineering, Chennai, India.

Assistant Professor, Dept of C.S.E Rrase College of Engineering Chennai, India.

Professor, Dept of CSE, Rrase College of Engineering, Chennai, India.

ABSTRACT: Currently the relational keyword based searches techniques consider the large number of data's to provide efficient result while the user searching. There is an issue of limited memory hence there is a need of the implementation of the novel techniques/ algorithm. To improve the search technique process by optimizing the query from that has to attain the memory optimization with the help of the genetic algorithm. Here the genetic algorithm plays the major role to select the optimized query to execute in the final result execution process based on the user given query. Here the process is executed in the dynamic manner which is considered as the real time scenario in that have to execute the whole process as the dynamic based on the user given query. The proposed system called MOGA. It means Memory Optimization with Genetic Algorithm. Here the memory is optimized based on the user given query to search the particular term. Then the User search goal is attained by applying the efficient clustering technique called Fuzzy C Means Clustering on the Clicked URL Data. The performance will be executed in terms of the space consumption and the time complexity of the searching process. In the past decade, extending the keyword search paradigm to relational data has been an active area of research within the database and information retrieval (IR) community. A large number of approaches have been proposed and implemented, but despite numerous publications, there remains a severe lack of standardization for system evaluations.

KEYWORDS: Keyword search, relational database, information retrieval, empirical evaluation

I. INTRODUCTION

Data Mining is defined as extracting the data from large datasets. The main goal of data mining is extracting the information from data sets and transforms it to an understandable structure for future works. The actual data mining task is the automatic analysis of large data sets. It has been classified into cluster analysis (group of records), unusual records (anomaly detection), and dependencies (association rules mining). Data mining uses information from past data to analyze the outcome of a particular problem or situation that may arise. Now the main aim of the process is to improve the searching performance in terms of the memory by the query optimization. Because the query optimization is the main need of the relational keyword search technique because there is the process depends of the above two data set based on the query or the single information. Hence there is a need of the novel technique to attain the memory optimization by the query optimization that will be takes place by the help of the genetic algorithm. In that they provide the optimized query based on the user given query to search the data from the database. And also have to attain the user assumption on the searching process. By applying the fuzzy c means clustering technique on the data after the user attained result by re ranking the search result. It has to phases they are training and testing process. In that have to train the data by applying the clustering technique In that they test the process by giving query again the same query as given in the trained process.

II. RELATED WORK

In user query keyword Search with ranking in IR techniques to get the data from database schema the information retrieval the queries execution time consumption is less. In this process they implement file length and execution time can be seen. At last the ranking can be seen by using chart of processed event. The experimental results were more scalability and improvements were most efficient. In particular, memory consumption precludes many search techniques from scaling beyond small datasets with tens of thousands of vertices. Also explore the relationship

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

between execution time and factors varied in previous evaluations; our analysis indicates that these factors have relatively little impact on performance. In summary, our work confirms previous claims regarding the unacceptable performance of these systems and underscores the need for standardization—as exemplified by the IR community—when evaluating these retrieval systems.

III. PROPOSED ALGORITHM

we propose the novel approach based on the genetic algorithm in the query optimization process called MOGA. Mainly here we consider the join ordering problems in the complex database queries which have considers the two joins. In our process which is used to provide the near-optimal execution pan for the user given query based on the considered database. The important advantage is the application of the genetic algorithm to find the optimal query in the given possible user given query. Here we experimentally evaluate using the MySQL database and the user given query with the GALib tool for the genetic algorithm. The proposed system will be implemented based on the dataset which we considered and then the query have to be given by the user from that dynamically the query will be formed and also get the efficient query to attain the optimized result. Then to enhance the process have to attain the user assumption process by applying the efficient clustering process using the click url data set.

The standalone MOGA and the hybridized with the other local search techniques. Despite their well-recognized benefits, the extra functionality of the simulated annealing (i.e. escaping local optima) and the tab search (keeping record of previously visited solutions) were unnecessary adds-ons in the context of most of these problems and consequently redundant computational effort was allocated to these local search functionality at the expense of the evolutionary global search. It was further noted that finetuning either all the individuals of the genetic population or just the dominated individuals was similarly effective. Either approach was significantly better than applying the local search process to just the non-dominated solutions in terms of convergence towards and spread along the Pareto front. It was notable that the process of fine-tuning the non-dominated individuals resulted in unwanted genetic drift and premature convergence. Note that due to the stochastic nature of the evolutionary strategies, a well-based judgment concerning the performance of a specific algorithm cannot be stated unless the whole optimization process is repeated number of times.

The hybrid algorithm has ensured to the decision maker enhanced objectives values previously undetected by the MOGA, reflecting better fitness values for some of the ZDT functions. On the other hand it was very obvious that although the hybrid algorithm were balanced interm of objective function evaluation all along the experimentations, the hybrid optimizer was much faster to achieve its results compared to the computational time needed by the MOGA, a characteristic well appreciated, especially when dealing with real world applications. MOGA-II is an improved version of MOGA (Multi-Objective Genetic Algorithm) by Poloni and is not to be confused with MOGA by Fonseca and Fleming with which it shares only the same acronym. MOGA uses smart multi search elitism for robustness and directional crossover for fast convergence. Its efficiency is ruled by its operators (classical crossover, directional crossover, mutation and selection) and by the use of elitism. Encoding in MOGA is done as in classical genetic algorithms and it uses four different operators for reproduction: classical crossover, directional crossover, mutation and selection. At each step of the reproduction process, one of the four operators is chosen (with regard to the predefined operator probabilities) and applied to the current individual.

IV. PSEUDO CODE

Fuzzy c-means clustering algorithm

Let X = {x_1, x_2, x_3 ..., x_n} be the set of data points and V = {v_1, v_2, v_3 ..., v_c} be the set of centers.

1) Randomly select *'c'* cluster centers.

2) Calculate the fuzzy membership $'\mu_{ij}'$ using:

$$\mu_{ij} = 1 / \sum_{k=1}^{c} (d_{ij} / d_{ik})^{(2/m-1)}$$

3) Compute the fuzzy centers $'v_j'$ using:

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

$$v_j = (\sum_{i=1}^{n} (\mu_{ij})^m x_i) / (\sum_{i=1}^{n} (\mu_{ij})^m), \forall j = 1, 2, ...c$$

4) Repeat step 2) and 3) until the minimum *'J'* value is achieved or $||U^{(k+1)} - U^{(k)}|| < \beta$.
where,

'k' is the iteration step.
'β' is the termination criterion between [0, 1].
'$U = (\mu_{ij})_{n*c}$' is the fuzzy membership matrix.
'*J*' is the objective function.

V. SIMULATION RESULTS

The simulation studies involve user has to give the query for the further propose and to obtain the optimized query. Here we consider the static tables and data's. The table names and attributes are emp, dept, acnt, bank and the attributes of emp is name, age , sal, dno then dept table attributes are dno, dname, floor, budget, mgr, ano then acnt tables contains the following attributes are ano, type, balance, bno then bank table contains the following attributes are bno, bname, address. In this module, have to rewrite the user given query into the representation format based on the selection, project and joint. Based on this rewrites query only have to prepare the execution plans.

There are two ways to generate the execution plans. They are user given plans or can generate the execution plans based on the parameter representation such as selection, projection and joint. In this module, have to apply the GALib tool to obtain the best chromosome with the help of fitness calculation value. Which is based on the genetic algorithm in this algorithm can get more number of chromes based on the execution plans and have to calculate the fitness nothing but cost estimation of chrome. select the optimized query by selecting the best fitness value, hence we can choose that query as the optimized query and based on the query evaluation and result obtain have to show the performance evaluation for the proposed algorithm for getting the optimized query.

Create a dataset related to our requirement. By considering a general search engine for particular duration from that they have to collect information like Anonymous ID, query, query time, URL. From this dataset they have to extract required information related to our next step process. The feedback session is considered as the combination of both clicked and unclicked URLs and ends with the last URL that was clicked in a session from user click-through logs. If it is not considered there will be noisy search results that are not clicked by any users may be analyzed as well. So feedback session is better that direct URL clicking search. It is actually based on single session. These data may vary a lot for different click-throughs and queries, yet it is not suitable for direct usage. Hence they go for representation method. There are different ways of representation of feedback session data. The basic way of representation is binary vector. But this method is not more efficient to get more information from this representation. Because it has information in o's and 1's from this they can't get information to cluster the keywords. So they go for pseudo-documents.

The data taken from the user search goals are used to produce the reorganized search result. To know its performance here they are using a new criteria named as Classified Average Precision (CAP). From this evaluation they are not only know about the performance of reorganized result and also it can used to optimize the parameter in the clustering method when assuming user rifle goals. And also the performance is evaluated based on the time consumption and the space consumption for the process.

VI. CONCLUSION AND FUTURE WORK

The simulation results showed that the proposed technique called the memory optimization with the genetic algorithm technique as the frame work for providing the optimization from the data with the memory optimization based on the query. Presented the approach of the query optimization it was very useful for the relation keyword searching systems. Here the proposed schemes provide the optimal and less time consumption for the detection of the best chromosome and the best fitness after the number generations of the individuals by the given user query. Then the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

process of attain the user search goal is also attained in the process by reducing the time consumption and the space consumption of the process. They have been executed the selected best query. It is mainly used to get the optimized query with the memory optimization..

REFERENCES

1. D. Fallows, "Search Engine Use," technical report, Pew Internet and Am. Life Project, http://www.pewinternet.org/Reports/2008/Search-Engine-Use.aspx. Aug. 2008.
2. comScore, "Global Search Market Grows 46 Percent in 2009, " http://www.comscore.com/ Press Events/Press_Releases/2010/1/Global_Search_Market_Grows_46_%_in_2009, Jan. 2010.
3. J. Coffman and A.C. aver, "A Framework for Evaluating Database Keyword Search Strategies,"Proc. 19th ACM Int'l Conf.Information and Knowledge Management (CIKM '10), pp. 729-738,Oct. 2010.
4. Y. Chen, W. Wang, Z. Liu, and X. Lin, "Keyword Search on Structured and Semi-Structured Data,"Proc. ACM SIGMOD Int'lConf. Management of Data (SIGMOD '09),pp. 1005-1010, June 2009.
5. W. bber, "Evaluating the Effectiveness of Keyword Search,"IEEE Data Eng. Bull.,vol. 33, no. 1, pp. 54-59, Mar. 2010.
6. A. Baid, I. Rae, J. Li, A. Doan, and J. Naughton, "Toward Scalable Keyword Search over Relational Data,"Proc. VLDB Endowment,vol. 3, no. 1, pp. 140-149, 2010.
7. Q. Su and J. Widom, "Indexing Relational Database Content Offline for Efficient Keyword-Based Search,"Proc. Ninth Int'lDatabase Eng. and Application Symp. (IDEAS '05),pp. 297-306, July2005.
8. V. Kacholia, S. Pandit, S. Chakrabarti, S. Sudarshan, R. Desai, andH. Karambelkar, "Bidirectional Expansion for Keyword Search on Graph Databases,"Proc. 31st Int'l Conf. Very Large Data Bases (VLDB '05), pp. 505-516, Aug. 2005.
9. H. He, H. Wang, J. Yang, and P.S. Yu, "BLINKS: ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '07),pp. 305-316, June 2007.
10. G. Kasneci, M. Ramanath, M. Sozio, F.M. Suchanek, and G.ikum, "STAR: Steiner-Tree Approximation in Relationship Graphs,"Proc. Int'l Conf. Data Eng. (ICDE '09),pp. 868-879, Mar.2009.

BIOGRAPHY

Ravichandiran is a post graduate student in the department of computer science engineering, rrase college of engineering, Anna University. He received Bachelor of Engineering degree (B.E) in 2006 from Anna University, Chennai, India. His research interests are Data Mining, Algorithms, Database Management System, etc.