ISSN : 0974 - 7435 Volume 10 Issue 13

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(13), 2014 [7045-7050]

An efficient materialized view selection approach for data cube utilizing evolutionary optimization

Gang Li
Huaian College Of Information Technology, Huai'an City Jiangsu Province 223003, (CHINA)

ABSTRACT

In this paper, we focus on the problem of materialized view selection for data cube, which is an important in the research field of database management. In the data warehouse, multi-dimension data can be represented as a data cube, which is a basic element in data warehouse. Particularly, each sub-cube is corresponding to an aggregation view in a specific the data cube. As the objective of materialized view selection for data cube is to minimize the sum of query cost and maintenance cost, in this paper, we converted data cube materialized view selection problem to an evolutionary multi-objective optimization problem. Afterwards, we propose a materialized view selection algorithm for data cube using evolutionary multi-objective optimization. When the stopping condition is satisfied, output of the proposed algorithm can be utilized as the data cube materialized view selection results. To testify the effectiveness of the proposed algorithm, we conduct experiments to make performance evaluation. Compared with other materialized view selection methods, the proposed algorithm performs better in the evaluation criteria “Time cost”, “Average response time”, and “Maintaining and updating time”.

KEYWORDS

Materialized view selection; Data cube; Evolutionary optimization; Elite solutions.

INTRODUCTION

As is well known that data warehouse is defined as a subject-oriented, integrated collection of data which support managers to make correct decision. Particularly, data warehouse should support complex and frequent data queries[1,2]. Hence, how to retrieve useful information from data warehouse is of great importance. Materialized view refers to a key approach to solve the information retrieving problem in data warehouse. However, high accuracy information searching need large scale of system memory and the system maintaince cost is quite high as well[3-5]. Thus, we can know that materialized view selection is an important topic in the research field of data warehouse.

In the data warehouse, the data can be viewed as a data cube. Furthermore, each cube in the data cube can be regarded as a view or query, and then these queries or views contain user interest aggregation values[6,7]. When the dimension of a data cube is n, number of data cube is 2^n. To enhance the query efficiency, the method we should study is to materialize the data cube. Unfortunately, materialized view selection need large storage space, and it is impossible to bear. In order to achieve the dual purpose of raising the query efficiency and reducing space consumption, our idea is to materialize only part of the data cubes, that is, to choose which cube to be materialized[8,9].

The materialized view can memorize the data physically, and then accelerate the speed of data searching with a lot of computing resources. Thus, how to select suitable views to be materialized refers to an important technology to tackle the problem of data warehouse query. In the former studies, existing materialized view selection methods mainly belong to static algorithms, and conflict with dynamic features of decision support system.

In this paper, we proposed a novel and efficient materialized view selection approach for data cube using evolutionary optimization under the dynamic environment. Evolutionary optimization refers to as a general term describing population-based search methods that involve some form of randomness and selection[10-13].

The rest of this paper is organized as follows. The next section gives a statement of the problem of materialized view selection for data cube. Section 3 proposed the materialized view selection approach for data cube using evolutionary optimization. In section 4, experiments are conducted to make performance evaluation. The concluding section draws the findings from the analyses of the proposed algorithm.

PROBLEM STATEMENTS

In the data warehouse, multi-dimension data can be regarded as a data cube, which is a basic data unit. In the data cube, each sub-cube is corresponding to an aggregation view. For example, a data warehouse is made up of three dimension table, which are 1) part, 2) supplier, 3) customer. Furthermore, a fact table is sales. The relationship between them is that supplier provides part to customer under the price sales. Then a three dimension data cube (Part, Supplier, Customer) is described as follows.

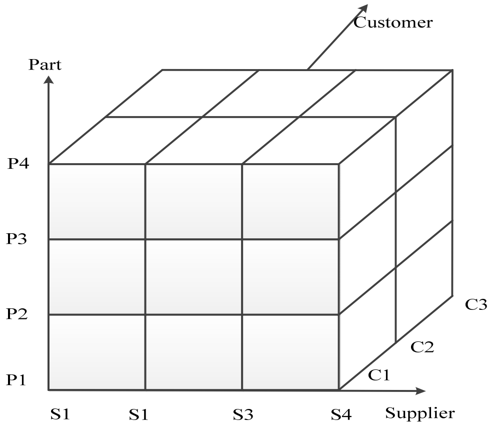

Figure 1 : An example of the data cube

According to data dimension characteristics, eight combinations can be obtained ($8=2^3$), which are (P,S,C), (P,S), (S,C), (P,C), (P), (S), (C), and (Null). P, S, and C denote the part, supplier and customer respectively. Particularly, the dependency relationships between the views part, supplier and customer can be described in Figure 2.

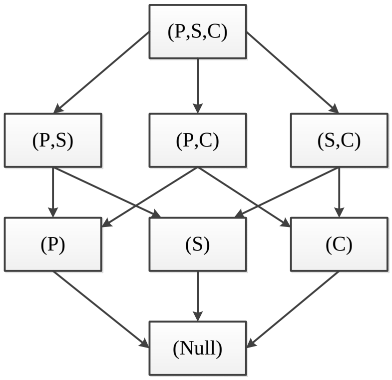

Figure 2 : Dependency relationship between different views for the example in Figure 1.

The proposed algorithm

The objective of the problem of materialized view selection for data cube can be regarded as minimizing the sum of query cost and maintenance cost and then the objective or fitness of evolutionary optimization is represented as maximization. Hence, we should transform the cost function to the fitness in evolutionary optimization as follows.

$$f(x)=\begin{cases} C_{\max}-c(x), if\ c(x)<C_{\max} \\ \quad 0, \quad otherwise \end{cases} \tag{1}$$

where $c(x)$ represents the cost function and $C_{\max}$ is an input coefficient.

As the problem of materialized view selection for data cube can be regarded as a multi-objective optimization problem, in this paper we use the evolutionary multi-objective optimization to solve it. Next, we will give the formal description of multi-objective optimization.

Definition 1 (Multi-objective optimization problem) Supposing that there is a feasible solution Ω and objective functions $f_1, f_2, \cdots, f_m$, the maximum multi-objective optimization is to seek the solution Ω^* satisfying the following condition:

$$\Omega^* = \arg\max_{x\in\Omega} f(x) = \arg\max_{x\in\Omega}\left(f_1(x), f_2(x), \cdots f_m(x)\right) \tag{2}$$

Where $f(x)=f_1(x), f_2(x), \cdots f_m(x)$ refers to the objective vector of the solution x. Afterwards, the method to select materialized view for data cube is described as following algorithm.

Algorithm 1 : Materialized view selection algorithm for data cube using evolutionary multi-objective optimization.

(1) Initializing the population P_t via randomly sampling N solutions in the solution space Ω, and then evaluating the F function. Moreover, parameter t is set to 1.

(2) When the stopping condition is satisfied, stopping this algorithm and then the solutions P_t can be regarded as the data cube materialized view selection results.

(3) Dividing P_t to K disjoint subpopulations via the $m-1$ dimensional local Principal Component Analysis algorithm, and then construct a probability model for each subpopulation.

(4) Building a population of offspring solutions Q_t through sampling from the probability model from (3), and then give the Gaussian noise for each population.

(5) Evaluating the F function in the solutions space Q_t.

(6) Choosing N elite solutions from the set $P_t \cup Q_t$ to construct the new set P_{t+1}. Let $t = t+1$, and then jump to (2).

EXPERIMENTS

In this section, experiments are conducted to make performance evaluation. To compare the performance of our algorithm, other two typical materialized view selection methods are chosen (that is YKL[14] and IMDVS[15]). Particularly, the hardware platform used in this experiment is based on Pentium 4 3.0 GHz and 2G RAM. The OS and DBMS we choose are Windows 2003 server and Microsoft SQL Server 2007 respectively.

We use the simulator to generate 2000 queries, and the principle of query distribution meets the 2/8 rule, which means that 80% query times generated from 20% queries. For the design of base table updating frequency, there are 10% basic tables exist in each update, and the ratio of base table updating is from the range of 1%-10%. Firstly, we will test the time cost of different methods with the number of dimension table in the data warehouse, and the related experiment results are shown in Figure 3-Figure 5 as follows.

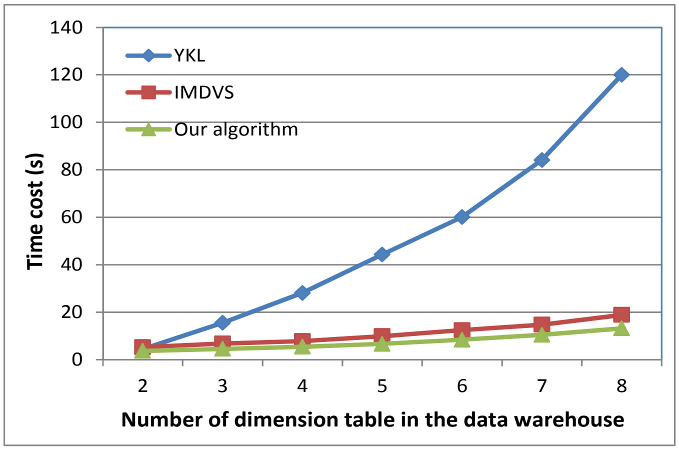

Figure 3 : Time cost for different number of dimension table in the data warehouse

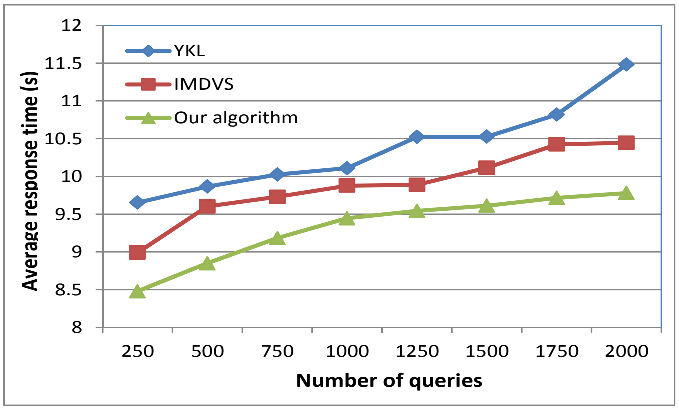

Figure 4 : Average response time for different number of queries

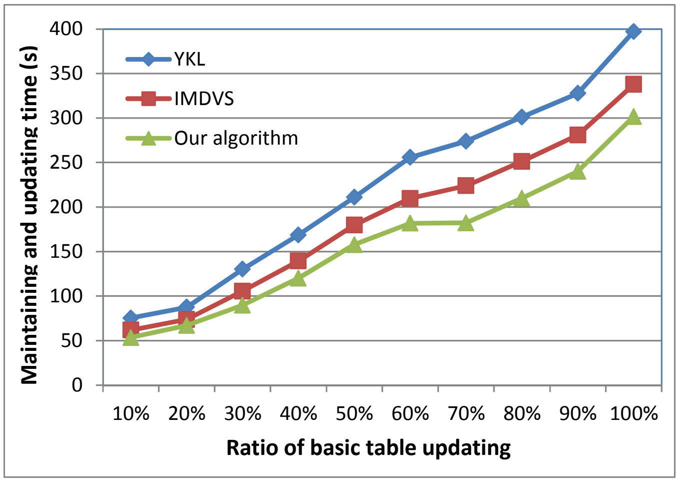

Figure 5 : Maintaining and updating time for different ratio of basic table updating

As is shown in Figure 3-Figure 5, it can be seen that our proposed method better than YKL IMDVS in the domain of "Time cost", "Average response time", and "Maintaining and updating time". Thus the conclusions can be drawn that our proposed can provide efficient materialized view selection scheme, and then performance of data cube management can be obviously enhanced.

CONCLUSIONS

This paper proposes a novel materialized view selection method for data cube using evolutionary multi-objective optimization. Considering the objective of materialized view selection for data cube is to minimize the sum of query cost and maintenance cost, we solve the proposed problem through evolutionary multi-objective optimization. Firstly, we initialize the population P_t through randomly sampling N solutions in the solution space Ω, and then evaluating the F function. Secondly, P_t is divided to several disjoint subpopulations by the local Principal Component Analysis algorithm. Finally, if the stopping condition of the evolutionary optimization is satisfied, the materialized selection results can be obtained.

REFERENCE

[1] Hung Ming-Chuan, Huang Man-Lin, Yang Don-Lin; Efficient approaches for materialized views selection in a data warehouse, Information Sciences, **177(6)**, 1333-1348 **(2007)**.

[2] Gou Gang, Yu Jeffrey Xu, Lu Hongjun; A* search: An efficient and flexible approach to materialized view selection, IEEE Transactions On Systems Man and Cybernetics Part C-applications and Reviews, **36(3)**, 411-425 **(2006)**.

[3] Aouiche Kamel, Jouve Pierre-Emmanuel, Darmont Jerome; Clustering-based materialized view selection in data warehouses, Lecture Notes in Computer Science, **4152**, 81-95 **(2006)**.

[4] J.X.Yu, X.Ya, Choi Ch; Materialized view selection as constrained evolutionary optimization, IEEE Transactions On Systems Man and Cybernetics Part C-applications and Reviews, **33(4)**, 458-467 **(2003)**.

[5] J.T.Horng, Y.J.Chang, B.J.Liu; Applying evolutionary algorithms to materialized view selection in a data warehouse, Soft Computing, **7(8)**, 574-581 **(2003)**.

[6] Perkins Simon, Questiaux Jacques, Finniss Stephen; Scalable desktop visualisation of very large radio astronomy data cubes, New Astronomy, **30**, 1-7 **(2014)**.

[7] Kaya Mehmet, Alhajj Reda; Development of multidimensional academic information networks with a novel data cube based modeling method, Information Sciences, **265**, 211-224 **(2014)**.

[8] Bianco Gianfranco, Bruno Fabio; Muzzupappa Maurizio Multispectral data cube acquisition of aligned images for document analysis by means of a filter-wheel camera provided with focus control, Journal of Cultural Heritage, **14(3)**, 190-200 **(2013)**.

[9] Park Daihee, Yu Jaehak, Park Jun-Sang; NetCube: a comprehensive network traffic analysis model based on multidimensional OLAP data cube, International Journal of Network Management, **23(2)**, 101-118 **(2013)**.

[10] D.Vijayakumar, V.Malathi; A real-time management and evolutionary optimization scheme for a secure and flexible smart grid towards sustainable energy, International Journal of Electrical Power & Energy Systems, **62**, 540-548 **(2014)**.

[11] Jiao Yi; Evolutionary genetic optimization of the injector beam dynamics for the ERL test facility at IHEP, Chinese Physics C, **38(8)**, Article No. 088102, **(2014)**.

[12] Basafa Ehsan, Armand Mehran; Subject-specific planning of femoroplasty: A combined evolutionary optimization and particle diffusion model approach, Journal of Biomechanics, **47(10)**, 2237-2243 **(2014)**.

[13] Hussain Sajid, A.Gabbar Hossam, A multi-objective evolutionary optimization of fuzzy controller for energy conservation in air conditioning systems, International Journal of Energy Research, **38(7)**, 847-859 **(2014)**.

[14] Yang Jian, Karlapalem Kamalakar, Li Qing; Algorithms for materialized view design in data warehousing environment, Processings of 23rd International Conference on Very Large Data Based, 25-29 **(1997)**.

[15] Yousri Noha Ar, Ahmed Khalil M, El-Makky Nagwa M, Algorithms for selecting materialized views in a data warehouse, In proceedings of The 3rd ACS/IEEE International Conference on Computer Systems and Applications, 27-35 **(2005)**.