

利用维护代价约束的物化视图选择方法

冯少荣

(厦门大学 信息科学与技术学院, 福建 厦门 361005)

摘要: 为了在一定维护代价约束条件下, 使查询过程中花费的总查询成本最优化, 提出了最小 V 最大候选集变换算法. 该算法构造最大候选视图集和最小候选视图集, 基于最小有效的极大基数配比技术, 并通过单位维护代价内的查询收益而设计的代价计算模型来获得最佳物化视图集. 理论分析和实验结果表明该算法是高效、动态、近似最优的. 与以往算法相比, 在数据维度大、维层次复杂的情况下, 该算法有着更优的执行效率.

关键词: 物化视图; 视图选择; 维护代价; 候选集; 查询收益

中图分类号: TP311 文献标识码: A 文章编号: 1674-4512(2009)05-0057-04

Materialized view selection under maintenance cost constraint

Feng Shaorong

(College of Information Science and Technology, Xiamen University, Xiamen 361005, Fujian China)

Abstract: In order to minimize the total query processing cost for a given set of queries under maintenance cost constraint, min/max candidate set transforming algorithm was proposed, in which MACVS (maximum candidate views set) and MICVS (minimum candidate views set) were constructed. The best materialized views set was gained by the cost calculation model which was designed using maximum matching method of minimizing effective and in order of query benefit per unit maintenance cost. Both the theoretical analysis and the experimental results showed that the algorithm is efficient, flexible and approximately optimal. The algorithm is more efficient than the previous algorithm under high dimension situations.

Key words: materialized view; view selection; maintenance cost; candidate set; query benefit

数据库中的视图选择^[1-6]关系到数据库的效率和维护代价, 是数据库研究中的重要问题. 物化视图选择问题的关键是如何在满足存储空间限制的条件下, 使得查询处理代价和视图维护代价的总和最小. 该问题为组合优化问题, 是一个 NP 完全问题^[7]. 目前存在许多算法^[1-10], 基于各自不同的代价计算模型, 通过各种途径求解该问题的近似最优解.

在实际应用中, 视图选择主要有存储空间的约束和维护代价的约束两个方面的约束条件. 存储空间的约束是指在给定磁盘空间约束下选择相关的视图集合进行物化, 使总的查询代价最小; 维护代价的约束是指在给定维护代价条件下如何选

择物化视图. 随着数据存储技术的发展, 视图的维护时间逐渐成为限制数据库物化所有视图的主要因素. 所以, 这里只讨论维护代价约束条件下物化视图的选择问题.

为解决维护代价约束下物化视图的选择问题, 文献[1, 2]在 OR 图和 AND/OR 图上提出了反向树贪婪算法和 A* 启发算法. 文献[11]提出了两相贪婪和集成贪婪算法. 然而在实际应用中, 特别是当数据库的数据维数超过 10 时, 这些算法效率往往得不到有效保证. 本文通过设计基于单位维护代价内的查询收益的代价模型和采用基于最小有效的极大基数配比技术, 提出了最小 V 最大候选集变换算法. 该算法高效、动态, 还可以得

收稿日期: 2008-09-27.

作者简介: 冯少荣(1964-), 男, 副教授, E-mail: shaorong@xmu.edu.cn.

基金项目: 国家自然科学基金资助项目(50474033).

到近似最优的可行方案.

1 相关知识

1.1 有向无环图

根据文献[1, 2, 12, 13], 假设在数据库中的视图用有向无环图(DAG) $G=(V, E)$ 来表示, 其中: V 是视图的集合; E 是有向边的集合. 有向边 (v, u) 指视图 u 可以从视图 v 中求得. 数据库的所有数据模型均可以由有向无环图来表示. 有向无环图的具体构造过程见文献[13].

对于有向无环图 $G=(V, E)$ 中的顶点和边, 分别有不同的权. $v(v \in V)$ 有 3 个权: r_v 为 v 的初始扫描代价, 即所占的空间大小; f_v 为 v 的查询频率; g_v 为 v 的更新频率.

边 $(v, u) \in E$ 有 2 个权: $w_q(u, v)$ 为利用 v 查询 u 的查询处理代价; $w_m(u, v)$ 为利用 v 更新 u 的更新代价. 给定物化视图 v 和 u , 函数 $q(u, v)$ 为从 v 到 u 相关边的最短路径查询处理代价的总和减去 v 的初始扫描代价 r_v . 当且仅当 v 不能回答 u 时, 在 $q(u, v)$ 中用原始表代替 v . 同理, $m(u, v)$ 指从 v 到 u 相关边的最短路径维护代价之和.

给定有向无环图 $G=(V, E)$ 和查询的集合 $Q=\{q_1, q_2, \dots, q_n\}$, 维护代价视图选择问题是选择一个视图集合 M , 使查询处理代价 $\tau(G, M)$ 最小化^[7], 即

$$\tau(G, M) = \sum_{q_i \in Q} f_{q_i} q_i(q_i, M),$$

在总维护代价^[7] $U(M)$ 小于约束 S 的条件下,

$$U(M) = \sum_{v \in M} g_v m(v, M),$$

式中: f_{q_i} 为查询 q_i 的查询频率; $q(q_i, M)$ 为当物化视图集合 M 存在时, 回答查询 $q_i (q_i \in Q)$ 处理代价的最小值; $m(v, M)$ 为物化视图集合 M 存在时, 维护物化视图 v 代价的最小值.

1.2 单位维护代价内的查询收益

设 M 为物化视图集合, ΔT_v 为维护物化视图 v 在总维护代价中的代价增量, 则有^[7]

$$\Delta T_v = U(M \cup \{v\}) - U(M).$$

假设 v_1, v_2, \dots, v_e 是在视图 v 之后物化的视图, 并将成为 v 的后代, T'_{v_i} 和 T_{v_i} 分别为 v_i 视图物化之前和之后的维护代价. 当 $T_{v_i} < T'_{v_i}$ 时, 在维护代价没有提高的条件下, v_i 不能改变 v , 表示为 $\nabla_{v_i} = T'_{v_i} - T_{v_i}$, 则有^[8]

$$\Delta T_v = T_v - \sum_{i=1}^e \nabla_{v_i}.$$

单位维护代价内的查询收益^[11, 12] 可以表示

增加的物化视图. 在有向无环图中, 先估计没有物化视图的代价总和, 使用单位维护代价内的查询收益估计每个节点的收益代价, 再用更高的单位维护代价内的查询收益作为物化的候选视图.

在可获得的维护代价 S 的限制下, 如 $U(M \cup \{v\}) \leq S$, 根据增加物化视图的数量导致查询响应时间的总和减少, 可以计算单位维护代价内的查询收益, 精确定义为^[8] $[\tau(Q, M) - \tau(Q, M \cup \{v\})] / \Delta T_v$. ΔT_v 在大多数情况下是非负的, 若 $\Delta T_v < 0$, 则意味着物化视图 v 的维护代价总和没有增加反而减少了.

2 算法描述

最小/最大候选集变换算法 (minimum/maximum candidate set transforming algorithm, MMCSTA) 主要使用物化视图的最小候选集 (MICVS) 和最大候选集 (MACVS), MICVS 只包含可以回答所有查询的基本表; MACVS 包含用最小查询处理时间能回答所有查询的视图. 算法的策略是使用单位维护代价内的查询收益从 MACVS 中将最有用的视图不断插入到 MICVS 中, 并将 MICVS 中没用的视图删除, 直到维护代价不能满足约束条件为止. MMCSTA 算法流程图如图 1 所示.

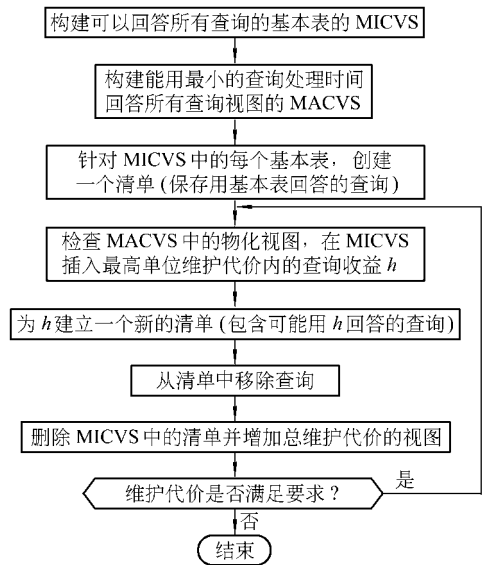


图 1 MMCSTA 算法流程图

3 实验与性能分析

3.1 实验设计

将 MMCSTA 算法与反向树贪婪算法 ITGA (inverted tree greedy algorithm)、A* 启发式算法

$A^* HA$ (A^* heuristic algorithm)、双相贪婪算法 TPGA (two phase greedy algorithm)、集成贪婪算法 IGA (integrated greedy algorithm) 进行比较. 实验的数据源为维度为 10 的数据仓库, 实验的硬件平台为在奔腾 IV 1.6 GHz CPU, 512 Mbyte 内存, 以 Oracle 8i for NT 为数据平台, 实现语言为 Visual C++ 6.0.

实验方案: **a.** 视图数量一定, 维护代价不同的情况下, 比较查询处理代价和维护代价; **b.** 维护代价一定, 视图数量改变的情况下, 比较查询处理代价和维护代价.

在方案 **a** 中, 视图数量设定均为 200, 它们属于最大候选集 MACVS, 维护代价的约束条件 S 的范围为 $0.7C_{min} \sim 1.0C_{min}$, 条件 C_{min} 是指允许 200 个视图均被选择物化的情况下的维护代价. 方案 **b** 中维护代价的约束条件 S 分别为 $0.8C_{min}$ 和 $0.9C_{min}$, 视图数量在 50~350 之间.

3.2 性能分析

方案 **a** 和 **b** 都采用 MMCSTA 算法, 按照实验规定的条件进行了维护代价和查询代价实验各 7 次, 将 MMCSTA 算法作为比较的基准, 根据每次对 5 种算法获取的数据, 得到的结果分别如图 2~图 4 所示, 图中: t_w 为维护代价; t_c 为查询代价

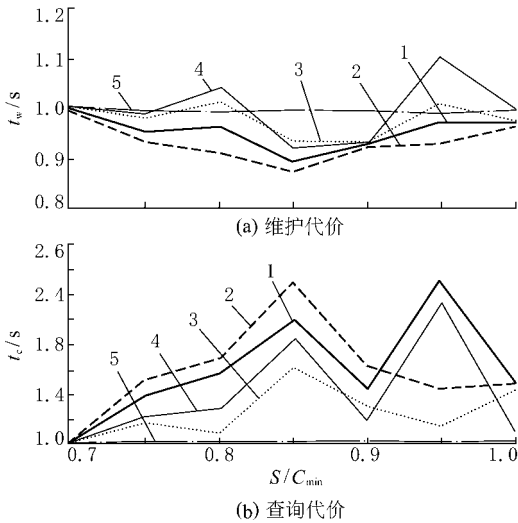


图 2 维护代价约束条件下各算法维护代价和查询代价

1—ITGA; 2—TPGA; 3—IGA; 4— $A^* HA$; 5—MMCSTA

(查询处理时间); N 为视图数量. 可以看出:

a. 在视图数量 N 一定, 除了 $S=0.7C_{min}$ 的情况下, MMCSTA 算法的性能均优于其他算法.

b. 在 $S=0.8C_{min}$ 情况下, MMCSTA 算法优于其他算法. 视图数量增加越多, 算法 MMCSTA 优势表现得越明显. 随着视图数量的增多, 5 个算

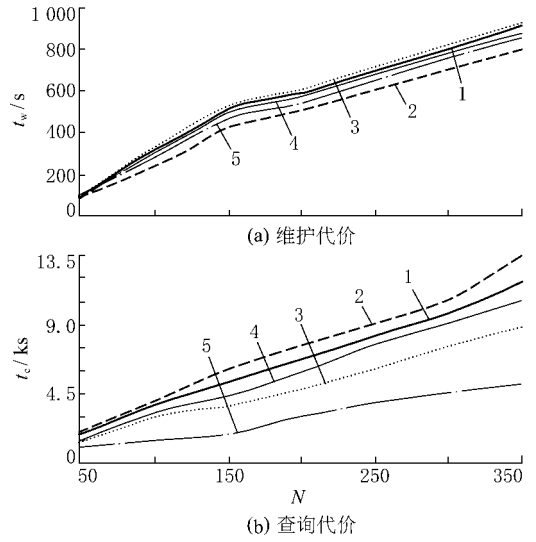


图 3 $S=0.8C_{min}$ 时各算法产生的维护代价和查询代价

1—ITGA; 2—TPGA; 3—IGA; 4— $A^* HA$; 5—MMCSTA

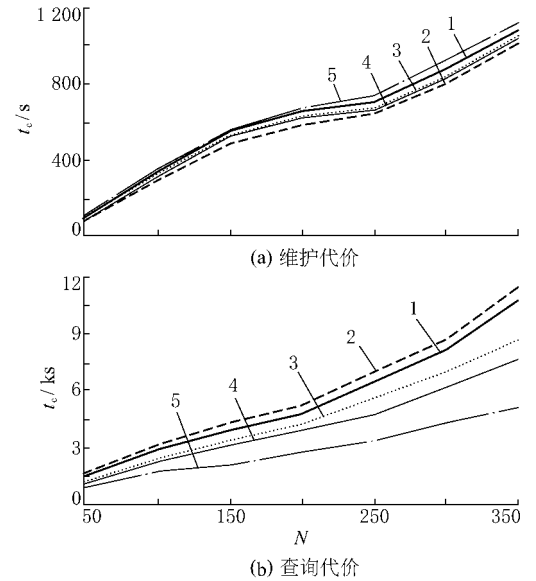


图 4 $S=0.9C_{min}$ 时各算法产生的维护代价和查询代价

1—ITGA; 2—TPGA; 3—IGA; 4— $A^* HA$; 5—MMCSTA

法产生的查询处理代价和维护代价均为单调增加.

c. 当维护代价约束条件 $S=0.9C_{min}$ 时, 每种算法产生的查询代价比 $S=0.8C_{min}$ 时低, 而维护代价比 $S=0.8C_{min}$ 时高. 同样 MMCSTA 算法在所有情况下优于其他的算法.

MMCSTA 算法不仅考虑在每一个阶段选择最好的视图进行物化不同, 同时还考虑从已存在的物化视图集合中删除没用的视图, 这将显著减少视图的维护代价.

参 考 文 献

- [1] Gupta H, Mumi k I S. Sele tion of views to materialize in a data warehouse[J]. IEEE Transa tions on Knowledge and Data Engineering, 2005, 17(1): 24-43.
- [2] 顾军华,赵秀丽,谭 庆. 蚁群算法在物化视图选择问题中的应用[J]. 计算机应用, 2007, 27(11): 2 763-2 765.
- [3] 林子雨,杨冬青,宋国杰,等. 实时主动数据仓库中多维数据实视图的选择[J]. 软件学报, 2008, 19(2): 301-313.
- [4] Gang G, Jeffrey X Y, Lu H J. A* sear h: an efficient an flexible appra h to materialized view sele tion[J]. IEEE Transa tions on Systems, Man, and Cyberneti s (Part C): Appli ations and Reviews, 2006, 36(3): 411-425.
- [5] Shukla A, Deshpande P M, Naughton J F. Materialized view sele tion for multidimensional datasets [C] // Pro eeding of the 24th VLDB Conferen e. New York: IEEE Press, 1998: 488-499.
- [6] Gupta H, Mumi k I S. Sele tion of views to materialize under a maintenanc e ost onstraint[C] // Pro eedings of the International Conferen e on Extended Database Theory (ICDT'99). [s. n.], 1999: 453-470.
- [7] Liang Weifa, Wang Hui, Orlowska M E. Materialized view sele tion under the maintenanc e time onstraint[J]. Data & Knowledge Engineering, 2001, 37: 203-216.
- [8] 崔晓军,薛永生,张东,等. 基于多维护策略的物化视图选择方法[J]. 计算机科学, 2006, 33(2): 114-117.
- [9] Lee M, Hammer J. Speeding up materialized view sle tion in data warehouses using a randomized algorithm[J]. International Journal of Cooperative Information Systems, 2001, 10(3): 327-353.
- [10] Harinarayan V, Rajaraman A, Ullman J D. Implementing data ubes efficiently[C] // Pro eedings of the 1996 ACM SIGMOD International Conferen e on Management of Data. [s, n.], 1996: 205-216.
- [11] Gupta H, Harinarayan V, Rajaraman A, et al. Index sele tion for OLAP[C] // Pro of the 13th Int'l Conf on Data Engineering. Birmingham: IEEE Computer So iety Press, 1997: 208-219.
- [12] 张柏礼,孙志挥,胡文瑜,等. 一种物化 Web 视图的在线选择算法[J]. 东南大学学报: 自然科学版, 2007, 37(1): 1-4.
- [13] Zhang C, Yao X, Yang J. An evolutionary appra h to materialized views sele tion in a data warehouse environment[J]. IEEE Transa tions on Systems, Man and Cyberneti s (Part C): Appli ations and Reviews, 2001, 31(3): 282-294.

“紧凑型 4 kW 轴快流 CO₂ 激光器”通过专家鉴定

日前,湖北省科技厅在武汉光谷科威晶激光技术有限公司召开了由华中科技大学、武汉光谷科威晶激光技术有限公司、武汉华工恒信激光有限公司、武汉华工团结激光技术有限公司共同完成的“紧凑型 4 kW 轴快流 CO₂ 激光器”项目的技术成果鉴定会。

专家们认为该激光器是一个结构紧凑的模块化系统,与 CNC 加工机床有良好的接口,其主要创新点为:基于激光能量反馈的功率监控与稳定系统,使其功率波动 $< \pm 0.5\%$;对放电结构的优化设计,提高了大体积均匀辉光放电的稳定性和注入电功率密度;基于高功率高频开关电源轴快流气体放电激励技术及调制技术,提高了电源效率和系统的响应速度;设计了一种稳定的激光谐振腔结构,保证了激光的高度稳定性;采用光纤传输技术在高频开关电源中的应用,提高了系统的稳定性和抗干扰能力.该激光器设计合理、功能齐全、操作简单、维护方便,在国内首次实现了稳定的 4 kW 低阶模激光输出,其关键技术已申请了国家专利,形成了自主知识产权。

专家们认为该项目填补了国内空白,达到了国际先进水平,建议进一步加快该激光器的生产规模以满足市场需求。