

УДК 519.68

И.А.Пшеничный (5 курс, каф. ПМ), С.Ю.Жуков, к.ф.-м.н., доц.

ПАКЕТНАЯ ОПТИМИЗАЦИЯ ЗАПРОСОВ К СУБД

Целью данной работы являлся анализ проблемы построения оптимального способа вычисления пакета запросов к СУБД (при этом на вид запросов, входящих в пакет, могут накладываться различные ограничения) и построения алгоритмов ее для решения.

Результат произвольного запроса Q можно получить, выполнив последовательность базисных реляционных операций (объединение, пересечение, вычитание, произведение, выборка, проекция, естественное соединение, деление). Заметим, что в общем существует множество различных способов (последовательностей базисных операций) вычислить один запрос. Будем называть каждый способ планом выполнения, а базисную операцию задачей.

Формализация задачи.

Заданы:

1. Множество запросов $\mathcal{Q} = \{Q_1, Q_2, \dots, Q_n\}$.
2. Каждому запросу Q_i соответствует множество планов выполнения $P_i = \{p_{ij}\}$. Это множество задается в виде графа, вершины которого соответствуют отношениям, а ребра – операциям над этими отношениями.
3. Каждый из планов p_{ij} состоит из задач $(t_{ij}^1, \dots, t_{ij}^m)$.
4. Каждой задаче t_{ij}^l сопоставлена ее стоимость c_{ij}^l и выгода от хранения ее в кэше b_{ij}^l .
5. Размер кэша C .
6. Константа K .

Вопрос: существует ли общий план выполнения стоимостью меньше K ?

Исследована сложность задачи в общем случае и доказана ее NP-трудность. Построен точный (экспоненциальный) алгоритм решения задачи.

Рассмотрен частный случай проблемы с кэшем нулевого объема, показано, что этот частный случай также является NP-трудным. Для данного частного случая доказана невозможность существования приближенного полиномиального алгоритма A : $A/Opt < Const$ (в данном случае под A понимается стоимость плана, построенного приближенным алгоритмом, а под Opt – стоимость оптимально плана) - т.е. алгоритма, находящего решение с точностью до мультипликативной константы - в предположении, что $P \neq NP$, что доказывает и невозможность существования приближенного алгоритма для общей задачи.

Рассмотрен частный случай проблемы с некоторыми упрощениям: процесс оптимизации разбивается на две независимые части (построение общего плана выполнения и определение оптимального порядка выполнения подзадач), и оптимизируется только вторая часть, дан порядок, в котором запросы из пакета будут выполняться и выгода от сохранения в кэше результата некоторой задачи не зависит от содержимого кэша. Доказана, что и эта проблема NP-трудна, и приведен приближенный полиномиальный алгоритм.

Приведены полиномиальные алгоритмы для частных случаев проблемы:

– для случая, когда граф выполнения состоит из m компонент связности, $m = \theta(n)$, n – общее число уникальных задач;

– для случая, когда кэш имеет нулевой объем и граф выполнения каждого из запросов в пакете является деревом.

Выводы. В работе была рассмотрена в достаточно общем виде задача пакетной оптимизации запросов на СУБД. Было доказано, что эта задача является NP-трудной. Также так же было показано, что некоторые упрощения этой проблемы также NP-трудны. Была доказана невозможность получить для данной задачи приближенный алгоритм, получающий ответ, отличающийся от оптимального не более, чем в фиксированное число раз. Для частного случая задачи (так же NP-трудного) был получен приближенный полиномиальный

алгоритм. Были построены полиномиальные алгоритмы для двух частных случаев исходной задачи.

ЛИТЕРАТУРА:

1. Timos Sellis, Subrata Ghosh "On the Multiple-Query Optimization Problem", TKDE June 1990 (Vol. 2, No. 2). pp. 262-266.
2. P. Roy, S. Seshadri, S. Sudarshan, S. Bhoje "Efficient and Extensible Algorithms for Multi Query Optimization", in *Procs. of the ACM SIGMOD Conf. on Management of Data*, May 2000, pages 249-260.
3. N.Dalvi, S.Sanghai, R. Roy, S. Sudarshan "Pipelining in Multi-Query Optimization", in *Procs. of the ACM Symposium on Principles of Database Systems (PODS)* May 2001.
4. T.Ibaraki and T.Kameda. "Optimal Nesting for Computing N -relational joins." *ACM Trans. on Database Systems*, 9(3):482-502, 1984.
5. Abdel-Wahab, H.M., Kamdeda, T. "On strictly optimal schedules for the cumulative cost-optimal scheduling problem." *Computing* 24 (1980), 61-86.
6. Lawler, E.L. "Sequencing jobs to minimize total weighted completion time subject to precedence constraints." *Ann. Discrete Math.* 2 (1978), 75-90.
7. Monma, Cl., Sidney, J.B. "Sequencing with series-parallel precedence constraints." *Math Oper. Res.* 4 (1979), 215-224.
8. A. Gupta, S. Sudarshan, S. Vishwanathan "Query Scheduling in Multi Query Optimization", IDEAS 2001: 11-19.
9. S. Chatterji, S.S.K. Evani, S. Ganguly, M.D. Yemmanuru „On the complexity of approximate query optimization“ *Principles of Database Systems*, 2002.
10. S. Arora "Probabilistic Checking of Proofs and Hardness of Approximation Algorithms", Technical Report TR-476-94, Department of Computer Science, Princeton University, 1994.
11. Garey, Johnson "Вычислительные машины и труднорешаемые задачи", М.Мир, 1982.
12. Т. Cormen, Ch. Leiserson, R. Rivest „Алгоритмы: построение и анализ“ МЦНМО, Москва, 1999.
13. P. Selinger, M. Astrahan, D. Chamberlin, R. Lorie, T. Price "Access path selection in a relation database management system", in *Procs. of the ACM SIGMOD Conf. on Management of Data*, 1979, pp. 23-34.