

TFE 処理系における問い合わせ文の分割による最適化

2 X - 3

有澤 達也† 実政 宏幸‡ 遠山元道§

†慶應義塾大学 理工学研究科 管理工学専攻

‡サントリー株式会社 情報システム事業部 情報システム部

§慶應義塾大学 理工学部 情報工学科

1 はじめに

TFE[1]は、データの階層構造に関する情報を付加することによって関係データベースからの問い合わせに対する結果を構造化する。

しかし、TFE 処理系では、関係データベースから情報を取り出す際に対象となる属性を一度に問い合わせしているため、質問によっては重複によって中間結果が膨大になっていた。従来の研究[2]では、TFEの問い合わせ文を複数に分割して問い合わせる手法が用いられていたが、大きな問題点が残されており、分割手法のさらなる改良が必要とされていた。

本研究では、TFE 処理系における問い合わせ文の発行に着目し、従来の分割手法に改良を行い、それを実装することによって時間的・領域的コストの削減を図る。

2 従来の分割手法の問題点

TFE 処理系で従来研究されてきた分割手法では、TFEにおける階層構造の一番外側の階層によって判断することによって、分割を行ってきた。

しかし、自由な階層構造を持つ TFE ではグルーピングをネストさせることが多く、内側の階層でも独立とみなせる属性の組合せが生じていることがある。そのような属性の組合せを同時に問い合わせると、中間結果に属性間の直積を生じてしまうのであるが、従来の手法では、そのような部分に対して分割を行うことができなかった。

また、階層構造が必ずしも属性集合間の独立性を示しているとは限らず、効果のない分割をしている場合も存在した。

3 実装する分割手法

従来の問い合わせ文の分割手法を改良し、以下にあげる3つの問い合わせ文の分割方法を採用する。

3.1 表分割

TFEでは、複数の表を同時に検索しレイアウトするためにも用いられる。複数の表は、一般的に互いに独立であり、これらを一度に問い合わせた際に中間結果に属性集合間の直積を生じてしまう。

そこで、TFEが複数の表となる部分から構成されているとき、それぞれの表に対して検索を行うように問い合わせ文を再帰的に分割することによって、表の持つ属性集合間の直積を防ぐことができる。このような分割を“表分割”と呼ぶことにする。

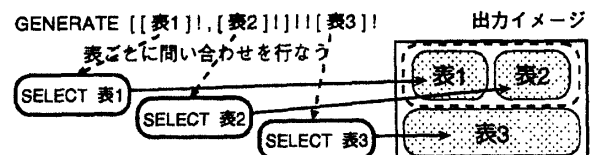


図1: 表分割

3.2 再帰的階層分割

従来の非再帰的な分割手法を拡張し、再帰的に行うことによって、問い合わせの中間結果の大きさを小さくすることが可能になる。このような分割を“再帰的階層分割”と呼ぶことにする。

この分割では、群化基準項目となる属性を問い合わせ、その結果を元に、階層構造によって独立とみなせる属性集合ごとに、さらなる問い合わせの分割を検討する。以上の操作を階層の外側から内側へ行うことによって、問い合わせ文は分割され、中間結果の冗長度を減らすことができる。

3.3 結合属性集合分割

属性集合間の独立性を考える上で、WHERE句は最も重要な要素である。問い合わせ文にあるWHERE句は、それによって関係付けられたリレーション間の

Optimization of TFE processing by query division.
ARISAWA Tatsuya†, SANEMASA Hiroyuki‡, TOYAMA Motomichi§

†Department of Administration Engineering, Faculty of Science and Technology, Keio University.

‡Suntory Ltd. Information Systems Dept.

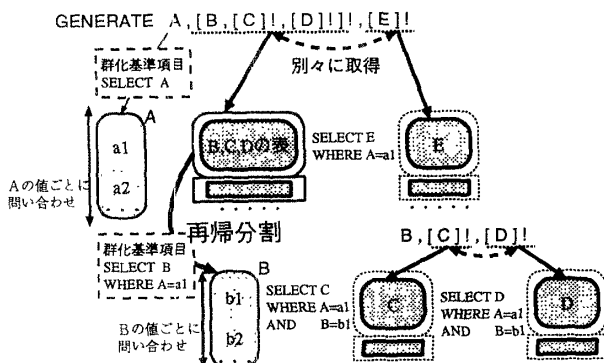


図 2: 再帰的階層分割

結合コストを大きく減少させる。逆に、関係付けられていない属性を一度に問い合わせると、その中間結果は必ず属性間の直積を生じてしまう。

そこで、TFE の中で WHERE 句によって関係付けられた属性の集合を“結合属性集合”と定義し、その集合ごとに問い合わせ文を発行することによって、属性間の直積を防ぎ、コストを減らすことができる。

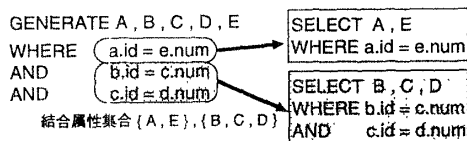


図 3: 結合属性集合分割

4 評価・検討

本研究で実装した分割手法の有効性を確認するために、従来の分割手法では不十分だと考えられる問い合わせ文について、実際に MSQL を用いて構造化データの取得をする実験を行った。各分割に対して、以下の問い合わせ文 (括弧内はタプル数) に対してそれぞれ属性 X のタプル数を変更することにより、処理時間と中間結果の大きさ (領域) を調べたところ、結果として次の表 1, 2, 3 を得た。

1. 表分割
GENERATE [A(10)]!![[B(10)]!], [X]!;
2. 再帰的階層分割
GENERATE A(5), [B(25), [C(100)]!], [X]!!, [D(25)]!
3. 結合属性集合分割
GENERATE t1.A(100), [t2.X]!
WHERE T1 t1, T2 t2

この実験では、分割をしないで問い合わせた場合、中間結果が非常に大きくなる。例えば 2 では、x が 200 タプルの時で中間結果は 25000 タプルにもなる。

表 1 からは、表分割が結合するタプル数が大きくなる程、従来の分割手法より有利になることがわかる。これは、従来の分割手法では分割が再帰的に

表 1: 表分割における比較

X のタプル数	10	20	30	40	50
時間	従来 5.05	13.31	19.63	33.33	35.14
領域	従来 35.4	70.6	105.7	140.9	176.0
	表分割 1.49	1.55	1.48	1.51	1.47
	表分割 0.9	1.1	1.2	1.3	1.5

(単位: 時間...秒, 領域...kbyte)

表 2: 再帰的階層分割における比較

X のタプル数	100	200	300	400	500
時間	従来 34.49	76.98	137.53	191.88	251.55
領域	従来 85.33	178.9	271.9	341.9	423.5
	再帰的 54.34	80.94	108.21	130.41	154.29
	再帰的 16.7	20.4	24.1	26.9	30.1

(単位: 時間...秒, 領域...kbyte)

表 3: 結合属性集合分割における比較

X のタプル数	10	20	30	40	50
時間	従来 3.74	8.13	12.49	20.29	31.42
領域	従来 24.6	49.0	73.5	97.9	122.3
	結合属性 3.65	6.64	10.05	13.83	17.79
	結合属性 1.8	1.9	2.1	2.2	2.3

(単位: 時間...秒, 領域...kbyte)

なわれないため、属性 B と属性 X 間に直積が生じてしまうからである。

表 2 から、結合するタプル数が小さいときは、再帰的な分割が時間的に不利であるが、大きくなるにつれて格段に有利になることがわかる。再帰的な分割は、内側の階層の分割によって多くの問い合わせ文を発行しなければならず、問い合わせの結果が小さいとデータベースの起動回数が多くなり不利になる。

表 3 から、結合属性分割は従来の分割手法では分割できなかった部分に対して分割を行なうため、表分割と同様に結合するタプル数が大きくなる程、従来の分割手法に対して有利になることがわかる。

また全体を通じて、分割を行うほど中間結果の領域は格段と小さくなるということがいえる。

5 おわりに

本研究では、従来の TFE 問い合わせ文の分割手法の改良を行った。そして、その分割手法を実装し実験を行うことによって、多くのケースにおいて従来より時間的・領域的成本を減少させることが確認された。

今後の課題として、多くの問い合わせ手法から適切なものを選ぶ評価関数の設定や、新たな問い合わせ文の処理手法の実現等、さらなる負荷の減少を目指す必要がある。

参考文献

- [1] 遠山元道他: レイアウト式 TFE の拡張, 情報処理学会データベースシステム研究会資料, 95-DBS-104 pp.217-224
- [2] 実政宏幸他: TFE 処理系におけるマルチセッション高速化, 情報処理学会第 51 回全国大会論文集, 2D-6