A Study on Deciding Reorganization Points for Data Bases with Quadratic Search Cost Function

Abstract

Reorganization is an essential part of data base maintenance work and the reasonable reorganization points can be determined from the trade-off between reorganization cost and performance degradation. There has been many reorganization models so far, but none of these models have assumed nonlinear search cost function. This paper presents the extensions of two existing linear reorganization models for the case where the search cost function is quadratic. The higher performance of these extended models was shown in quadratic search cost function case.

I. 序 論

대규모 데이터 베이스 시스템 (Date Base System)의 경우, 插入 (insertion), 削除 (deletion) 등의 修正 (update) 作業들이 일정기간 계속됨에 따라 別途의 再構成作業 (Reorganization) 을 수행해 주지 않으면 데이터 베이스의 探索費用이 급격히 증가하게 된다.

데이터 베이스를 再構成함으로써 探索費用의 급격한 증가를 막을 수 있으나 재구성 작업에는 적지 않은 비용이 隨伴되므로 그 時期는 任意로 결정될 수 없다.

確定的 (Deterministic) 探索費用函數을 假定한 지 급까지의 研究는 주로 探索費用의 增加가 시간變數의 一次式으로 표시되는 線形再構成模型에 대해 어

次하지 않았다.

즉, 시간이 經過함에 따라 構成 (File)의 크기가 線形의으로 증가하고 따라서 원하는 레코드 (Record)에 대한 探索時期도 線性적으로 증가한다는 가정이 前提이었다.

그러나, 實際의 데이터 베이스 시스템에서는 非線形 探索費用函數도 存在하고 있음이 발표된 바 있

으며 (4), (7), 이러한 境遇에 從来의 線性 再構成 모형들은 採용한다면 實際 탐색費用函數와의 適合

缺如에 따른 探索費用損失이 推想된다.

따라서, 本論文은  두 가지 대표적인 線形 再構成 모형들을 採用하여 探索費用 함수가 二次의 경우에

도 적응가능한 두 가지 再構成模型들을 提出하고자 한다.
Ⅱ. 연구背景

데이터 베이스의 재구성은 시기 결정에 관한 연구는 재구성을 필요로 하는 여러 요인 중에서도, 원하는 데이터에 대한 접근을 하려면 사용자(또는 사용자 구성)가 증가하여 시스템 성능이 상승에 대한 경우에도, 이러한 요인에 따라 주로 이루어져 왔다.

그 이유는 대부분 데이터의 구조의 변화가 물리적 접근방식의 변경으로 인한 재구성 요인 중은 일정한 변화 양상을 갖지 않으므로 이론적으로 하려면, 탐색비용의 증가는 어느 데이터 베이스에서나 도달할 수 있게 발생하고 있으므로 그 변화 과정은 시간에 따라 일정한 형태를 가지는 것이 대부분의 경우에 정량적 및 정량적 대비가 비교적 쉽게 가능하기 때문이다.

시간의 경과에 따른 탐색비용의 증가는 그림 - 1에서 보는 바와 같이 일정한 크기의 증가에 의한 부분과 데이터의 재구성의 여건에 따른 부분으로 구분되며, 재구성 작업에 의해 억제할 수 있는 수용 공간의 증가 또는 후자의 경우라고 하겠다. [8]

탐색비용의 증가가 임의의 (random)인 경우에 대해서는 Mendelson(10), Yechiali (10), Heyman(12) 및 Winslow(11) 등이 연구결과를 발표한 바 있으며, 탐색비용의 증가가 특정한 (deterministic)인 경우에 대해서는 1973년 Shneiderman의 "Optimum Data Base Reorganization Points"(1), 토요로 Yao & Theory (5), 투메(6) 등이 연구한 바 있다.

이들의 모형이나 알고리즘은 모두 탐색비용이 시간에 따라 선형의 증가를한다고 가정하고 있다.

본 논문은 데이터베이스의 구조 변화의 시기, T가 주어진 경우가 그렇지 않은 경우의 대표적인 두 모형들인 Shneiderman의 '최적 재구성 구조 모형 (Optimum Fixed-Interval Reorganization Model)'과 Yao의 '동적 재구성 모형 (Dynamic Reorganization Model)'을 확장하여 탐색비용이 비례의 (Quadratically)로 증가할 경우의 재구성 시기를 결정하고자 한다.

![그림 - 1 탐색비용 및 재구성](image)
III. 연구 방법

여기서는 탐색비용이 1의 무로 정의되는 데이터 베이스의 노선화 결정문제로써, 정량적 날짜가 독립적이고 매번 고정할 경우에 탐색비용

III-1 FIXED_QUAD 모형

본 모형은 II장에서 소개된 바 있는 '최적노선화 데이터 모형' (1)의 몇 가지 개념들을 기초로 하여, 이 중 특이한 경우, \( S_t = S + \theta t \) \( S = S + \theta t^2 + \theta t \) 수 변동이 존재하지 않음의 경우에는 DYNAMIC_QUAD 모형을 사용한다.

따라서, 다음 식들이 가로된다.

\[
\begin{align*}
& i) \gamma_{11}(t) = \gamma(t) + \delta(t) + \theta(t) + \theta(t) \\
& ii) \delta_{11}(t) = \delta(t) + \theta(t) + \theta(t) \\
& iii) R_{11}(t) = R(t) + \mu t \\
& iv) \theta \geq \theta \geq 0, \quad \theta \geq \theta \geq 0
\end{align*}
\]

여기서, iv) 식은 다음 조건에서 유래된다.

\[
d(t) = \gamma(t) - \delta(t) = (\theta(t) - \theta(t)) + (\theta(t) - \theta(t)) t > 0 \quad \forall t > 0
\]

이러한 조건된 가정을 하여 화일에 대한 탐색

\[
\begin{align*}
& C(T) = \sum_{i=1}^{N} \left[ \int_{0}^{T} (\gamma(t) - \delta(t)) \, dt + R(t) \right] \\
& = N \left[ \frac{(\theta_1 - \theta_2)}{3} \, t^3 + \frac{(\theta_2 - \theta_2)}{2} \, t^2 + \frac{(\theta_2 - \theta_2)}{2} \, t + NR + \frac{N(N+1)}{2} \mu t \right] \\
& = \left( N = \frac{T}{t} \text{ 대입} \right) \\
& = T \left[ \frac{(\theta_1 - \theta_2)}{3} \, t^3 + \frac{(\theta_2 - \theta_2)}{2} \, t + \frac{1}{t} \, (TR + \mu t) \right] + \frac{1}{2} \mu T
\end{align*}
\]

따라서, \( C(T) \)는 \( t \)가 양수일 때 \( t \)에 대한 근면 (c
onvex) 함수이다. 그러므로, 구하고자 하는 적절한 구조
성 관계, \( t_{eq} \) 는 다음과 같다.

i) \( \theta_i = \theta_s, \theta_i > \theta_s \) 일 때
\[
t_{eq} = \left[ \frac{2R_s + \mu T}{\theta_i - \theta_s} \right]^{\frac{1}{2}}
\]

ii) \( \theta_i > \theta_s, \theta_i \geq \theta_s \) 일 때
\( t_{eq} \)는 \( 0 \)과 끝 일 때, \( T \) 사이의 값으로 다음과
동계시켜서 \( t \) 가 된다.
\[
F(t) = \frac{2(\theta_i - \theta_s)}{3} t^3 + \frac{(\theta_i - \theta_s)}{2} t^2 - \left( R_s + \frac{\mu T}{2} \right) t
\]
\( -0, \quad 0 < t < T \)

여기서, \( t \)는
\[
t = 0, \quad \frac{(\theta_i - \theta_s)}{2(\theta_i - \theta_s)}(0 < 0)
\]
의 두 점에서 극값을 갖게 되며, \( \theta_i = 3 \)과 같은
형태를 갖는다.

따라서, \( F(t) = 0 \)은 \( t > 0 \)인 범위에서 하나
의 균, \( t^* \)을 가정을 알 수 있다.

이 \( t^* \)은 \( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
또는 \( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
\( t \)의 \( \pm \)의 값으로, \( t \)의 \( \pm \)의 값으로,
이 구해진다.
그래므로, 구하고자 하는 $t_{0q}$는 다음과 같다.

i) $\theta_i = \theta_1, \theta_i > \theta_1$ 일 때

$$t_{0q} = \mu + \left[ \frac{\mu^2 + 4R_n(\theta_1 - \theta_i)}{2(\theta_i - \theta_1)} \right]^{\frac{1}{2}}$$

ii) $\theta_i > \theta_1, \theta_i > \theta_1$ 일 때

$t_{0q}$는 다음 식을 만족시키는 $t_n$이다.

$$D(t_n) = (\theta_i - \theta_1) t_n^2 + (\theta_1 - \theta_{\theta_2}) t_n^2 - \mu t_n - R_n = 0$$

여기서, 다음 재구성 시기를 결정하기 위해서는 (2) 식의 $R_n$에 당시의 (Current) 재구성 비용, $R_n$을 대입하여 (3), (4) 식을 새로운 적용해야 한다.

$D(t_n)$은

$$t_n = \frac{-\theta_i - \theta_1 \pm \sqrt{(\theta_i - \theta_1)^2 + 3\mu(\theta_i - \theta_1)}}{3(\theta_i - \theta_1)}$$

의 두정에서 근값을 갖게 되어 그림 - 4 의 곡선 A

의 같은 형태를 갖는다. 또한, 다음의 재구성 시기, $t_{n^{**}}$ 도 $R_n$만 $R_n$으로 바뀌면 되 같은 방식으로 찾 아낼 수 있다.

이때, $D(t_n)$은 그림 - 4의 곡선 B의 형태를 갖는다.

### IV.適用例

선형探索費用 함수를 가정한 Yao 등의 재구성 모형에서는 다음과 같은費用파라미터들을 사용해 왔다. [5, 6]

- $\theta_1 = 5,460$ (access seconds/day)
- $\theta_1 = 3,000$ (access seconds/day)
- $R_n = 75,100$ (access seconds)
- $S_i = 681,000$ (access seconds/day)
- $\mu = 75,100$ (access seconds/day)

여기서 비용은 각接近 (access)시의 待機時間 (단위: 초) 으로 가정된다.

본 사례는 二手 탐색비용 함수를 가정하고 있으므로 두 가지 탐색비용 곡선,

$$S_i = S_0 + \theta_i t' + \theta_i t$$와 $$S_i = S_0 + \theta_i t + \theta_i t$$가 회귀곡선 $S_i = S_0 + 5460 t$와 $S_i = S_0 + 3300 t$로 적합되는 $\theta_i (i=1,2,3,4)$ 값들을 취한다.

![그림 - 4 D(t_n) 그래프](image-url)
표 - 1은 본 사례의 탐색비용계수들로서 기간 T 기, β이다. 별로 1회분은 바쳐 선택된 파라미터들이고, 관호 표 - 2는 본 사례에 대한 6 가지 정책들 인의 값들은 적합된 회귀식 (y - S + β T)의 기울의 결과를 수명주기별로 비교해 본 결과이다.

<table>
<thead>
<tr>
<th>T</th>
<th>C_out</th>
<th>정책</th>
<th>재구성시기</th>
<th>C_saving</th>
<th>% Save</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>199.0</td>
<td>NO</td>
<td>LF 83, 80, 167, 61</td>
<td>37.3</td>
<td>19.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LD 35.74, 92.87, 169, 97</td>
<td>30.7</td>
<td>17.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>QF 91.54, 155.62</td>
<td>29.2</td>
<td>21.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>QD 52.36, 131.29</td>
<td>23.8</td>
<td>36.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MO 30, 60, 90, 120, 150, 180</td>
<td>51.5</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>520.7</td>
<td>NO</td>
<td>LF 118.22, 236.45, 354.67</td>
<td>144.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LD 35.74, 92.87, 169, 97, 266.39, 381.75</td>
<td>91.3</td>
<td>36.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>QF 139.25, 278.49</td>
<td>65.5</td>
<td>54.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>QD 63.99, 161.08, 286.32</td>
<td>64.3</td>
<td>55.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MO 30, 60, 90, ......, 390</td>
<td>212.3</td>
<td></td>
</tr>
<tr>
<td>800</td>
<td>1452.1</td>
<td>NO</td>
<td>LF 166.99, 333.97, 500.96, 667.94</td>
<td>608.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LD 35.74, 92.87, 169, 97, 266.39, 381.75, 668.35</td>
<td>166.0</td>
<td>72.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>QF 224.30, 449.60, 628.03</td>
<td>185.7</td>
<td>69.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>QD 116.58, 279.99, 480.29, 712.22</td>
<td>155.7</td>
<td>74.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MO 30, 60, 90, ......, 780</td>
<td>168.1</td>
<td>72.4</td>
</tr>
<tr>
<td>1,600</td>
<td>4751.7</td>
<td>NO</td>
<td>LF 236.01, 472.01, 708.02, 944.03, 1180.03, 1416.04</td>
<td>2304.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LD 35.74, 668.35, 839.27, 1028.46, 1235.83, 1461.31</td>
<td>452.0</td>
<td>80.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>QF 361.17, 722.33, 1083.50, 1372.43</td>
<td>776.3</td>
<td>66.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>QD 167.45, 401.65, 688.30, 1019.88, 1391.46</td>
<td>407.4</td>
<td>82.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MO 30, 60, 90, ......, 1590</td>
<td>390.5</td>
<td>83.1</td>
</tr>
</tbody>
</table>

- 80 -
표-1 사례의 \( \theta \) 추정치

<table>
<thead>
<tr>
<th>( T )</th>
<th>( \theta_1 )</th>
<th>( \theta_2 )</th>
<th>( \theta_3 )</th>
<th>( \theta_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>15</td>
<td>3,000</td>
<td>10</td>
<td>1,800</td>
</tr>
<tr>
<td></td>
<td>(5,357)</td>
<td>(3,371)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>9</td>
<td>2,500</td>
<td>6</td>
<td>1,500</td>
</tr>
<tr>
<td></td>
<td>(5,328)</td>
<td>(3,385)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>800</td>
<td>7</td>
<td>1,000</td>
<td>4</td>
<td>700</td>
</tr>
<tr>
<td></td>
<td>(5,399)</td>
<td>(3,214)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,600</td>
<td>3.9</td>
<td>500</td>
<td>2.4</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>(5,402)</td>
<td>(3,317)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

여기에 사용된 용어들은 다음과 같다.

\( T \) : 데이터베이스의 수명 (단위: 일)

\( C_{int} \) : 화일크기 증가에 따른 입출력 시간 증가 (단위: 10^4 access seconds)

\( C_{dis, reorg} \) : 재구성 비용 + 저장소 구조 손상에 인한 탐색비용의 증가 (단위: 10^4 access seconds)

\% Save : 재구성에 의한 비용 감소 비율 (%)

NO : 재구성 없음 수행

LF : 최적 재구성 점검 모형 (선형 탐색비용 가정)

LD : 판별 재구성 모형 (선형 탐색비용 가정)

QF : FIXED QUAD 모형

QD : DYNAMIC QUAD 모형

MO : 매월 재구성 수행

총 비용, \( C(T) \) 은 \( C_{int} \) 와 \( C_{dis, reorg} \) 의 합이며, 이들은 다음과 같이 구해진다.

\[
C_{int} = S_t + \frac{1}{2} \theta_1 T^2 + \frac{1}{3} \theta_3 T^3
\]

\[
C_{dis, reorg} = \frac{1}{3} (\theta_1 - \theta_3) \sum_{i=1}^{k+1} (t_i - t_{i-1})^2 + \frac{1}{2} (\theta_2 - \theta_3) \sum_{i=1}^{k+1} (t_i - t_{i-1})^2 + \sum_{i=1}^{k} (R_0 + \mu t_i)
\]

여기서, \( t_0 = 0, t_{k+1} = T \)이고 \( K \) 는 재구성 횟수이다.

또한, \% Save는 다음과 같다.

\[
\% \text{Save} = 100 \times \frac{C_{dis, reorg}(\text{NO}) - C_{dis, reorg}(\text{재구성})}{C_{dis, reorg}(\text{NO})}
\]

V. 결론

본 연구에서는, 데이터베이스의 구성 모형들이 총합을 주석 후에 길고의 한 결과를 구정해 온 집의 확장 하여, 탐색비용가격이 낮을 경우에도 적용 가능한 2 가지 재구성 모형 (FIXED QUAD, DYNAMIC QUAD) 등을 제시하였다.

또한, 작업시간을 통해, 탐색비용가격이 낮은 데이터 베이스 시스템의 측면, FIXED QUAD 모형은 (선형) 최적 재구성 점검 모형보다 동 모양의 모형보다 좋은 성능을 가진 모형을 보였다. 본 연구의 요구 연구 사항으로는, 최적 재구성, 탐색비용가격이 낮은 비선형 함수의 재구성, 모형, 로 가고, 재구성 점검의 본질적 집합 값이 흥미로운 결과를 얻을 수 있었다.

참고 문헌


7. Söderlund, L.: "Concurrent Database Reorganization-Assessment of a Powerful Technique

8. Sockut, G. H., Golaberg, R. P.: "Database Re-
organization: Principles and Practices." Com-

9. Sockut, G. H.: "A performance model for com-
puter database reorganization performed con-

10. Mendelson, H., Yechiali, U.: "Optimal policies
for Data Base Rerogization." O. R., 29 Jan.

11. Winslow, L. E., Lee, J. c.: "Optimal Choice of
353 – 363.

12. Heyman, D. P.: "Mathematical Models of Data-
base Degradation," ACM Trans. Database Syst

ll, 1981.

14. Atrc, S.: Data Base: structured Techniques
for Design, Performance, and Management, John

15. Date, C. J.: An Introduction to Database Sys-

박영사, 1983