データベースにおける最適再編成方策

3Q−3

安井一民　中川俊夫　本告光男
愛知工業大学

1. はじめに

データベースは，運用後におけるデータの追加や削除等の更新使用によって次第に格納状態の整然性が乱され，システムの性能低下の原因となる．従って，適切な時期をとらえてデータベースを最適な格納状態に整理することが望ましい．これは，データベースの再編成またはガーベジコレクション[1]として知られている．ここでは，データベースの再編成時期に関する問題を考察する．すなわち，ある確率分布に従って使用されるデータベースが更新処理を伴うとき，ある確率で領域にガーベジが発生して格納効率や処理効率を低下させるようなモデルを設定する．そのとき，一定間隔でデータベースのガーベジコレクションを実施し，期待費用を最小にする最適再編成方策を考察する．

2. モデルと期待費用

データベースは平均値関数 $R(t)$ をもつ非定常ポアソン過程[2]で使用されるものとする．各々の使用において，確率 α（$0<\alpha\leqq 1$）で更新処理が行われ，確率 $1-\alpha$ で非更新処理が行われる．j 回目の更新使用では確率 p_j（$j=0,1,2,..$）でガーベジが発生し，データベースにおける格納状態の整然性が乱されるとする．ここで，$0=p_0<p_1<p_2\cdot\cdot<p_j<\cdot\cdot\cdot$ と仮定する．このようなシステムに対して，一定間隔Tでデータベースの再編成処理を実施し，データベースを最適な格納状態に復元するものとする．

さて，データベースは非定常ポアソン過程で使用され，確率 α で更新されるから，時刻Tまでに j 回の更新使用が発生する確率は，

$$H_j[\alpha R(T)]=\{[\alpha R(T)]^j/j!\}e^{-\alpha R(T)}, \qquad (1)$$

である．従って，$(0,T]$ 間におけるガーベジの発生する平均回数は，

$$\sum_{j=0}^{\infty} H_j[\alpha R(T)]\cdot\sum_{i=0}^{j} p_i=\sum_{j=0}^{\infty} P(j)\,H_j[\alpha R(T)], \qquad (2)$$

で与えられる．ここで，$P(j)\equiv\sum_{i=0}^{j} p_i$，とおく．

次に，$(0,T]$ 間における更新使用回数の上限Nをしきい値として設定し，N回未満の使用の場合とN回以上の使用の場合とを区別して期待費用を導入する．データベースの１回の使用に要する費用を c_0 とし，更新使用においてガーベジが発生した場合のリソースの損失費用を c_1 とおく．また，時刻Tで再編成を実施する場合の費用を，N回未満の更新使用のとき c_2，N回以上の更新使用のとき c_3 とし，$c_2\leqq c_3$ と仮定する．

Optimum Reformation Policy for a Database System
Kazumi YASUI, Toshio NAKAGAWA and Mitsuo MOTOORI
Aichi Institute of Technology

以上のような仮定における単位時間あたりの期待費用は，（０，Ｔ］間におけるデータベースの平均使用回数は $R(T)$ であるから，次式で与えられる．

$$C(T) = \left\{ c_0 R(T) + c_3 - (c_3 - c_2) \sum_{j=0}^{N-1} H_j [\alpha R(T)] + c_1 \sum_{j=0}^{\infty} P(j) H_j [\alpha R(T)] \right\} / T, \quad (3)$$

３．最適再編成方策

(3) 式の期待費用Ｃ(Ｔ) を最小にする問題を考えよう．Ｃ(Ｔ) をＴで微分して零とおくことによって，

$$Tr(T) \left\{ c_0 + \alpha (c_3 - c_2) H_{N-1} [\alpha R(T)] + \alpha c_1 \sum_{j=0}^{\infty} p_{j+1} H_j [\alpha R(T)] \right\}$$
$$+ (c_3 - c_2) \sum_{j=0}^{N-1} H_j [\alpha R(T)] - \left\{ c_0 R(T) + c_1 \sum_{j=0}^{\infty} P(j) H_j [\alpha R(T)] \right\} = c_3, \quad (4)$$

を得る．ここで，$R(t) \equiv \int_0^t r(u)du$，とおく．

(4) 式を満たす最適再編成時間間隔T^* を理論的に議論することは難しい．とくに$c_0 = 0$，$c_3 = c_2$ のとき，(4) 式は次式のように簡単化される．

$$T \alpha r(T) \sum_{j=0}^{\infty} p_{j+1} H_j [\alpha R(T)] - \sum_{j=0}^{\infty} P(j) H_j [\alpha R(T)] = c_3 / c_1. \quad (5)$$

(5) 式の左辺を$Q_1(T)$ とおく．もし，$r(t)$ が微分可能で，$r(t)$ およびp_j がともに増加関数かつどちらか一方が単調増加関数ならば，$Q_1(T)$ はＴの単調増加関数となることが示される．よって，もし$Q_1(\infty) > c_3 / c_1$ ならば，(5) 式を満たすT_1^* が唯一に存在する．以上の考察から，(4) 式の左辺を$Q(T)$ とおくと，$c_3 \geqq c_2$，$Tr(T) \geqq R(T)$ であるから，$Q(T) \geqq Q_1(T)$ を得る．従って(4) 式を満たす解をT^* とおくと，T_1^* が存在するならば$T^* \leqq T_1^*$ であり，最適再編成時間間隔T^* も存在することが示される．

４．数値例と考察

仮定として，データベースの更新使用率を$\alpha r(t) = \alpha \lambda$とし，ｊ回目の更新使用によるガーベジの発生確率を$p_j = 1 - p^j$ ($j = 0, 1, ...; 0 < p < 1$) とおく．さらに，ガーベジが発生したときのリソースの損失費用をシステムの単位費用として$c_1 = 1$とおく．以上の仮定のもとで，(4)式を満たす最適再編成時間間隔T^* の数値例を表１に示す．例えば，N＝700，$c_2 / c_1 = 50$，$c_3 / c_2 = 2$，とすると，T^* は約128時間となり，いわば100回／日の更新使用に対して，およそ週１回のガーベジコレクションが最適であることがわかる．

表１　$\alpha\lambda = 5$，$p = 1 - 10^{-4}$ のとき，期待費用Ｃ(Ｔ) を最小にする再編成時間間隔T^* の数値例

N	c_2/c_1	c_3/c_2 = 1	2	5	10
500	10	91.0	87.0	85.6	84.8
	20	129.4	89.7	87.6	86.6
	50	207.0	90.5	88.2	87.1
700	10	91.0	90.9	90.9	90.9
	20	129.4	124.1	122.5	121.5
	50	207.0	127.8	125.2	123.9
1000	10	91.0	91.0		
	20	129.4	129.4		
	50	207.0	207.0		

参考文献

(1) 情報処理学会編："新版情報処理ハンドブック"，p.1166，オーム社(1982).
(2) S.M.Ross："Stochastic Processes"，p.309，John Wiley & Son, New York(1980).