

リレーショナルDBMSにおける
時系列データ管理方式

3W-4

井出祐二* 山下廣太郎* 松谷 淳**

(* (株) 日立製作所 システム開発研究所 **同 大森ソフトウェア工場)

1. はじめに

金融機関では、実績情報の活用を目的とした情報システムの構築が進められている。実績情報を格納するデータベースとして、アクセス・ボタンが多様で、表形式で取扱うことが可能なリレーショナルデータベース (RDB) が有効である。また、実績情報として、主に、人事情報や顧客情報などの時系列形式の統計データ (時系列データ) が、扱われている。

このように、金融機関で構築が進められている情報システムでは、膨大な時系列データを、柔軟に検索・運用することが求められている。本報告では、RDB上での時系列データ管理と、時系列データの間合せ方式について述べる。

2. 時系列データのモデル化

2.1 統計データのモデル

ここでは、統計データを、「 C_1 別、 C_2 別、 \dots 、 C_n 別 S 」とモデル化する。上記モデルで、 C_1, C_2, \dots, C_n は、統計データの対象となった観察値を、どういう基準に従って、グループ分けしたかを示す指標であり、「分類属性」(category attribute)と呼ぶことにする。また、 S は、分類属性に従って、類別された観察値の統計的要約値を、代表するものであり、「集約属性」(summary attribute)と呼ぶ²⁾。

統計データの例として、預金高を、銀行、預金者、金額階層、年度で表現した統計データを、表1に示す。

表1. 統計データ

銀行	預金者	金額階層	年度	預金高
A行	個人	10万未満	57	83
A行	個人	10万未満	58	90
~	~	~	~	~
B行	法人	500万以上	57	352
B行	法人	500万以上	58	363
~	~	~	~	~

2.2 RDB上での時系列データのモデル化

統計データのうち、最もよく使用されるのは、時系列形式のデータ (時系列データ) である。ここで、時系列データは、例えば、昭和63年度の預金高といった時点のキーとなる値を持つ。この時点のキーを、どのようにRDBの構造上に反映させるか

によって時系列データのモデルを検討する。

時点のキーを分類属性、集約属性のいずれかに反映することにより、3つの時系列データモデルに類別する (表2)。

表2. 時系列データモデル

	(a)	(b)	(c)
モデル (R)	$R_1(C, Ct, S)$	$R_2(C, S_1, \dots, S_m)$	$R_3(C, V)$
RのRの属性	$Ct \in C$ $C = \{C_1, \dots, C_n\}$	$r_1[S_1] = s_1$ ~ $r_1[S_m] = s_m$	$r_3[V] = \{(t_1, s_1), \dots, (t_m, s_m)\}$
説明	$r_1 \in R_1$ 時点のキーを分類属性 (Ct) とするモデル	$r_2 \in R_2$ 集約属性 (S) の各時点 ($t_1 \sim t_m$) を、個々の属性 ($S_1 \sim S_m$) とするモデル	$r_3 \in R_3$ 集約属性 (S) の各時点 ($t_1 \sim t_m$) の値を、1系列 (V) とするモデル

3. 時系列データモデル

3.1 時系列データモデルの構成

本報告では、上記のモデルのうち、以下の点から、(c)の時系列データモデルを選択する。

(1) 容量上の問題

時点のキーを分類属性 (Ct) とした場合には、他の分類属性値 (C) を、重複して格納するため、データ量が増加する。

(2) 構造上の問題

集約属性 (S) の各時点 ($t_1 \sim t_m$) を個々の属性とすると、時点数 m の時、集約属性数が m 倍化され、属性数が増加する。

時系列データモデルの構成例を図1に示す。

銀行	預金者	金額階層	年度	預金高	貸出金
A行	個人	10万未満	50	52	33
			51	58	42
			~	~	~
			58	67	55
			59	86	56

図1. 時系列データモデルの構成

3.2 時系列データモデルの属性

ここで、時系列データモデルの各属性を、統計データモデルの概念より、以下の通り分類する。

(1) 分類属性...分類分けの指標となる属性

(2) 時点属性...時点のキーとなる属性

(3) 集約属性...統計的要約値となる属性

また、データベースの属性構造より、時系列データモデルの属性は、以下の2種類に分類できる。

- (1) シングルオカレンス…単一の値を持つ属性
- (2) マルチオカレンス…複数の値を持つ属性

図1の時系列データの各属性を、表3に示す。

表3. 時系列データモデルの属性

属性	銀行	預金者	金額階層	年度	預金高	貸出金
統計データ 概念	分類属性			時点属性	集約属性	
属性構造	シングルオカレンス			マルチオカレンス		

4. 時系列データの検索

4.1 検索処理の課題

時系列データの検索では、統計データモデルの概念を用いて、分類属性、時点属性、及び、集約属性を用いた検索条件を指定することにより行う。

統計データモデルの概念を用いた時系列データの検索では、次のような処理が必要である。

- (1) 分類属性(C)の条件により、分類属性の値を検索する。
- (2) 指定時点(t)の集約属性の値を検索する。
- (3) 指定時点(t)の集約属性(S)の条件により、集約属性の値を検索する。

一方、データベース問合せ言語を用いて、時系列データを検索する場合には、時系列データの各属性に関する条件を、シングルオカレンス、及び、マルチオカレンスの条件として指定することにより、検索を実行することになる。ここで、データベース問合せ言語でのマルチオカレンスの検索においては、予めオカレンスの順位を指定しなければならず、上記(2),(3)のような任意の順位での検索がなされていない。

4.2 検索方式

本検索方式では、検索条件で指定された各属性について、分類属性、時点属性、集約属性を判別し、各属性の条件に従って、条件検索、及び、時点検索を行う。

時系列データの検索プロセスは、以下の通りである。

- (1) 利用者より、検索要求が入力される。

- <Step1> 入力された検索要求により、条件検索を行う。条件検索では、指定された条件に従い、各属性ごとに、分類属性、あるいは、指定時点の集約属性の検索を行う。
- <Step2> <Step1>の検索結果に対し、時点検索を行う。時点検索では、指定された時点の検索を行う。
- <Step3> <Step2>の結果をテーブル表示する。

- (2) 利用者の判断により、再度検索要求が入力される。(以下、<Step4>~<Step5>を実行) 利用者の検索要求が、終了した時点で、処理が終了する。

<Step4> 再度入力された検索要求より、時点検索を行う。時点検索では、<Step2>の検索結果について、指定時点の検索を行う。

<Step5> <Step4>の結果を、テーブルとして再表示する。

上記検索処理のうち、分類属性、集約属性、時点の処理は、次の通りである。

(a) 分類属性の検索

分類属性の検索条件により、条件に合致しているレコードの集合を得る。

(b) 指定時点の集約属性の検索

時点属性のマルチオカレンスの値と、条件で指定された時点の値との一致・不一致を判定する。一致していれば、そのマルチオカレンスの順位により、集約属性の比較条件に合致しているかどうかを判定する。これにより、条件に合致しているレコードの集合を得る。

(c) 時点の検索

時点属性のマルチオカレンスの値と、条件で指定された時点の値との一致・不一致を判定する。一致していれば、そのマルチオカレンスを識別することにより、集約属性の値を検索する。

本提案方式の検索プロセスを、図2に示す。

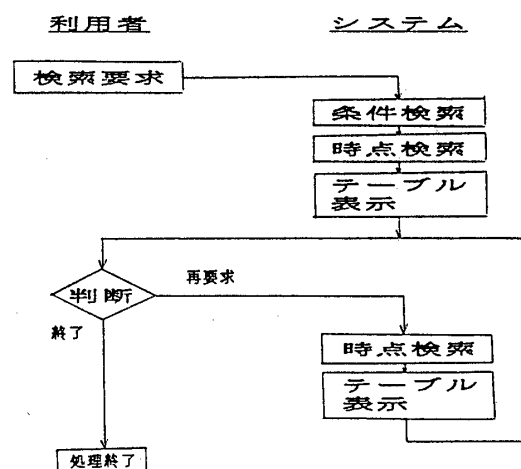


図2. 検索プロセス

5. おわりに

RDB上での時系列データの管理方式を検討した。その検索方式については、リレーショナルDBMSでの動作確認を行い、提案方式の実用性を確認した。

<参考文献>

- 1) A. Shoshani: "Statistical Database: Characteristics, Problems and Some Solutions", VLDB, 1982
- 2) E. F. Codd: "A Relational Model of Large Shared Data Banks", Communications of ACM, vol. 13, no. 6, 1970
- 3) 樋: 「リレーショナルデータベースと時系列データベース」, Computer Today, No. 7, pp. 18-21, 1985