

サイエンティフィック DBMS DREAM における データの導出過程について

3R-7

中田 充 宝珍輝尚 都司達夫

福井大学 工学部 情報工学科

1 はじめに

科学技術データを扱うサイエンティフィック DB では、データを格納する際にデータ定義（スキーマ）をあらかじめ定義することが困難である。これに対処するため、筆者らはボトムアップにスキーマを定義できる DREAM モデル^{1,2)}を提案し、それを採用したサイエンティフィック DBMS DREAM を設計構築している。DREAM モデルは、データ値に主眼を置いたデータモデルであり、従来のモデルと異なりモデルの中心はデータ値である。サイエンティフィック DB では、格納されているデータに対して研究者が自分の視点や仮定のもとに様々な操作を行い新たなデータを作成することが多い。DREAM では、これをデータの導出と呼ぶ。ここでデータが何を元にして、どのような操作を行って導出されたかを知ることが可能となれば、そのデータを導出した研究者の視点や仮定の理解を補助することが可能となる。

そこで本論文では、導出 DB 中のデータについて、それがどのように導出されたかを表す導出過程を提案する。以降、2 で導出 DB について述べ、3 で導出過程の定義を行う。

2 導出 DB

ここでは、データの導出過程を考えるために最低限必要な導出 DB の定義を示す。データエレメント (de) は導出 DB においてデータ値を格納する要素であり、導出 DB の基本要素である。導出 DB は基本的に de の集合として定義される。de の定義を定義 1 に示す。

[定義 1] (データエレメント de)

$de = (id, value, dp)$ ここで、id は de の識別子、value はデータ値、dp は導出過程を示す。□

Deriving Processes of data in Scientific DBMS
DREAM

Mitsuru NAKATA Teruhisa HOCHIN

Tatsuo TSUJI

Dept. of Information Science, Fukui University

3 導出過程

導出過程を定義する前に幾つかの準備を行う。導出 DB 中の de のうち、既存の de をもとに導出された de を de^A 、それ以外の de を de^B と表記する。 de^B には、ユーザが導出 DB に挿入した de や基本 DB 中のデータを参照する de などがある。 $v(de)$ は de のデータ値を返す関数とする。

3.1 導出過程のレベル

起源をどこまで遡るかを表すレベルの概念を導入する。図 1 は de の導出の様子を表している。白丸は de を表し、黒丸は関数を、矢印は導出の関係を表す。例えば、 de_3 は de_1, de_2 に関数 f_1 を適用して導出されている。 de_7 のレベル 1 の起源は de_5, de_6 であり、レベル 2 の起源は de_3, de_4, de_6 である。このように、同じ de でも、レベルに応じて導出過程が異なる。

3.2 レベル 1 の導出過程

de^A のレベル 1 の導出過程の定義を定義 2 に示す。レベル 1 の導出過程では一つ前まで遡る。

[定義 2] de^A のレベル 1 の導出過程

$$dp_{de^A}^1 = (S_{de^A}^1, F_{de^A}^1) \quad S_{de^A}^1 = (de_i, \dots, de_j)$$

$$F_{de^A}^1 = f^1(\arg_1, \dots, \arg_n) \text{ 但し } \arg_i = v(S_{de^A}^1[i])$$

$$S_{de^A}^1[i] = S_{de^A}^1 \text{ の } i \text{ 番目の要素}$$

$S_{de^A}^1$ は de^A の元になっている de の組である。

$F_{de^A}^1$ は de^A を導出するために用いた操作を表す。

関数 f^1 は導出に用いた操作を表す関数である。

$S_{de^A}^1[i]$ は、 $S_{de^A}^1$ の i 番目の de を表す。□

図 1 の de_7 は、 de_5, de_6 に関数 f_3 を適用することにより導出された de である。 de_7 のレベル 1 の導出過程 $dp_{de_7}^1$ は、 $S_{de_7}^1 = (de_5, de_6)$

$$F_{de_7}^1 = f_3(v(S_{de_7}^1[1]), v(S_{de_7}^1[2])) \text{ となる。}$$

de^B のレベル 1 の導出過程の定義を定義 3 に示す。

[定義 3] de^B のレベル 1 の導出過程

$$dp_{de^B}^1 = (S_{de^B}^1, F_{de^B}^1) \quad S_{de^B}^1 = () \quad F_{de^B}^1 = v(de^B)$$

de^B の元になる de の集合は空である。 F は de^B

自身のデータ値とする。 □

定義4に、deのレベル1の起源を求める演算の定義を示す。

[定義4] deのレベル1の起源を求める演算

$$O^1(de) = \begin{cases} S_{de}^1 & (de = de^A) \\ (de) & (de = de^B) \end{cases}$$

起源は、deが de^A の時、その導出過程のSであり、 de^B の時、自身を要素とする組である。 □

3.3 レベルmの導出過程

レベルmの導出過程の定義を示す。その前に組の連結を表す演算 Ω を以下のように定義する。

[定義5] 演算 Ω

組 $A=(a_1, a_2, \dots, a_n)$, 組 $B=(b_1, b_2, \dots, b_n)$ とする。演算 Ω は $A\Omega B=(a_1, \dots, a_n, b_1, \dots, b_n)$ $A\Omega\emptyset=(a_1, \dots, a_n)$, $\emptyset\Omega A=(a_1, \dots, a_n)$ で定義される。ここで \emptyset は空の組。また、n個の組 X_1, \dots, X_n に対して、 $\overset{n}{\Omega} X_j = X_1\Omega \dots \Omega X_n$ とする。 □

[定義6] レベルmの起源を求める演算

$$O^m(de) = \begin{cases} \overset{n}{\Omega} O^{m-1}(S_{de}^1[j]) & (de = de^A) \\ (de) & (de = de^B) \end{cases}$$

[定義7] de^B のレベルmの導出過程 $dp_{de^B}^m = dp_{de^B}^1$ 従って $S_{de^B}^m = S_{de^B}^1$, $F_{de^B}^m = F_{de^B}^1$. □

[定義8] de^A のレベルmの導出過程

$$dp_{de^A}^m = (S_{de^A}^m, F_{de^A}^m) \quad S_{de^A}^m = O^m(de^A)$$

$$F_{de^A}^m = f_{de^A}^m(\arg_1, \dots, \arg_n) \text{ 但し, } \arg_i = v(S_{de^A}^m[i])$$

$$f_{de^A}^m(v(S_{de^A}^m[1]), \dots, v(S_{de^A}^m[n])) = f_{de^A}^1(F_{S_{de^A}^m[1]}^{m-1}, \dots, F_{S_{de^A}^m[n]}^{m-1})$$

$$S_{de^A}^m[i] = S_{de^A}^m \text{ の } i \text{ 番目の要素} \quad \square$$

定義7,8より、 de_7 のレベル2の導出過程は、 $S_{de_7}^2 = (de_3, de_4, de_6)$, $F_{de_7}^2 = f_{de_7}^2(de_3, de_4, de_6)$ で表される。ここで、関数 $f_{de_7}^2$ は de_3, de_4, de_6 を引数として de_7 を導出する関数である。

3.4 レベルmaxの導出過程

de^A はその起源をさかのぼると最終的に de^B にいきつく。そこで、起源を最後まで遡るレベルmaxの導出過程を定義する。

[定義9] レベルmaxの起源を求める演算

$$O^{\max}(de) = \begin{cases} \overset{n}{\Omega} O^{\max}(S_{de}^1[j]) & (de = de^A) \\ (de) & (de = de^B) \end{cases} \quad \square$$

[定義10] de^B のレベルmaxの導出過程

$$dp_{de^B}^{\max} = dp_{de^B}^1 \text{ 従って } S_{de^B}^{\max} = S_{de^B}^1, \quad F_{de^B}^{\max} = F_{de^B}^1. \quad \square$$

[定義11] de^A のレベルmaxの導出過程

$$dp_{de^A}^{\max} = (S_{de^A}^{\max}, F_{de^A}^{\max}) \quad S_{de^A}^{\max} = O^{\max}(de^A)$$

$$F_{de^A}^{\max} = f_{de^A}^{\max}(\arg_1, \dots, \arg_n) \text{ 但し } \arg_i = v(S_{de^A}^{\max}[i])$$

$$f_{de^A}^m(v(S_{de^A}^{\max}[1]), \dots, v(S_{de^A}^{\max}[n])) = f_{de^A}^1(F_{S_{de^A}^{\max}[1]}^{\max}, \dots, F_{S_{de^A}^{\max}[n]}^{\max})$$

$$S_{de^A}^{\max}[i] = S_{de^A}^{\max} \text{ の } i \text{ 番目の要素} \quad \square$$

3.5 抽象化度

図1の de_7 は、 de_3, de_4, de_6 に関数 f_4 で表される操作を適用して導出したと考えることもできる。

$$dp_{de_7}^1 = ((de_5, de_6), f_3(de_5, de_6)) \quad (1)$$

$$dp_{de_7}^1 = ((de_3, de_4, de_6), f_4(de_3, de_4, de_6)) \quad (2)$$

このように、導出過程はそれを見るユーザの視点によって粒度が変化する。そこで一つの手続きとして捕らえる関数の粒度の違いにより抽象化度を考える。式(2)で表される導出過程は式(1)で表されるそれよりも抽象化度が大きい。このような、抽象化度の異なる複数の導出過程を扱うために、データエレメントを複数の導出過程を格納できるように再定義する。

[定義12] (データエレメント de (再定義))

$de = (id, value, DP)$ ここで、idはdeの識別子、valueはデータ値、DPは抽象化度の異なる複数の導出過程の集合を示す。 □

4 おわりに

導出DB中のデータが、何をもとにして、どのように作成されたかを管理するためにデータの導出過程を定義した。これにより、データ導出時のユーザの視点や仮定の理解を補助できるシステムの構築が可能となった。

参考文献

- 1) 中田他:サイエンティフィックデータベースのためのデータモデルの一提案, 情処学 DBS 研究会 101-9(1995).
- 2) 中田他:サイエンティフィックデータモデルの一評価, 信学 DE 95-64(1995).

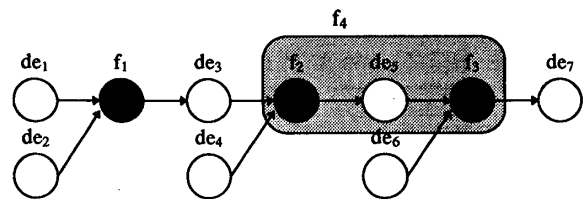


図1 データエレメントの導出