

再帰有向超グラフデータベースに対する問合せについて

4K-1

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

ネットワーク管理, CAD/CAM, VLSI 設計, ハイパーメディアシステムやマルチメディアデータの管理といった高度なデータベース応用分野では, 操作対象を直接グラフを用いて記述したいという要求があり, グラフに基づくいくつかのデータモデルが提案されてきている. 筆者らも, ラベル付き有向グラフに超グラフと再帰グラフの概念を取り込んだ再帰有向超グラフデータモデルを提案してきている [1, 2]. 本モデルでは, 枝の始末端に, 点やグラフの集合を含むことができる.

本論文では, 再帰有向超グラフデータモデルに基づくデータベースの問合せ処理の明確化を目的として, 問合せがデータの連結関係と構成関係を利用することで複合値を考慮した Datalog プログラムとして処理できることを示す.

以下, 2. で, 再帰有向超グラフデータモデルについて概説する. 3. で, データの構造的な性質について考察し, 4. で, まとめを行う.

2 再帰有向超グラフデータモデル

データ表現グラフは以下のように定義される.

- データ表現グラフ (drg) は, 原子 drg, または, 標準 drg である. drg はラベル $(d_{ID}, name, d)$ を持つ. ここで, d_{ID} は識別子, $name$ は名前, d はデータで組 (データ型, 値) である.
- 原子 drg は点である.
- 標準 drg g は $(V, E, L_v, L_e, \phi_v, \phi_e, \phi_{connect}, \phi_{comp})$ である. ここで, V, E は, 各々, drg の集合, 枝の集合であり, L_v, L_e は, 各々, drg のラベルの集合, 枝のラベルの集合であり, $\phi_v : V \rightarrow L_v, \phi_e : E \rightarrow L_e, \phi_{connect} : 2^V \times 2^V \rightarrow E, \phi_{comp} : 2^V \cup 2^E \rightarrow V$.

[例 1] 図 1 にデータ表現グラフの例を示す. 図では, 原子 drg (点) を小円で表している. 枝を楕円で表し, 楕円中の矢印で枝の向きを表している. また, 標準 drg を

点線で表している. この図ではマニュアルの文が順序付けられて表され, データとしては, レベル (上級, 初級, 共通) と文字列が格納されている.

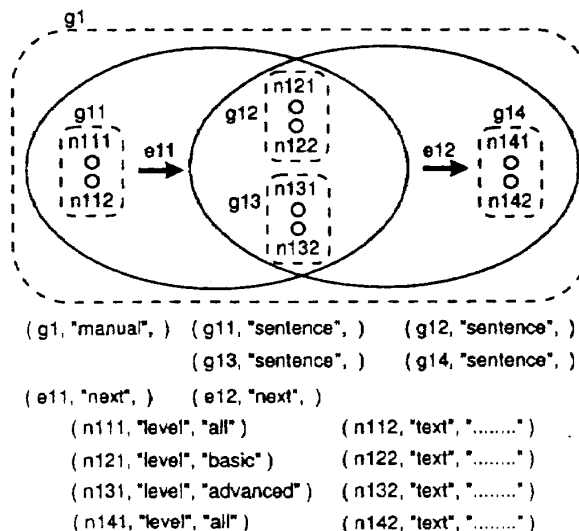


図 1: データ表現グラフの例

また, データ表現グラフを成分とするグラフを考え, これをデータグラフという. さらに, あるデータグラフの構造を表す, 要素名のみをラベルとする 1 以上の有向グラフをスキーマグラフと呼ぶ.

次に, 再帰有向超グラフデータモデルにおいて問合せに使用する演算である *rewrite* について概説する.

- $rewrite(target, sl)$ は, sl で指定された問合せグラフを $target$ で指定された問合せグラフに書き換える. ここで, $target$ は割当て仕様, sl は問合せ仕様の列である.
- 割当て仕様は組 (qg, al) であり, 問合せ仕様は (qg, vs) である. ここで, qg は問合せグラフ, al は変数に対する値の割当ての列, vs は変数に対する検索条件である.
- 問合せグラフには, 原子問合せグラフと標準問合せグラフがある. 問合せグラフはラベルを持つ. ラベルは 3 つ組 $(v_{id}, v_{name}, v_{val})$ で表される. ここで, v_{id} は識別子用変数, v_{name} は名前用変数, v_{val} は値用変数である.

On the Query Processing on Directed Recursive Hypergraph Databases
 Teruhisa HOCHIN, Tatsuo TSUJI
 Fukui University, 3-9-1 Bunkyo, Fukui-Shi 910-8507, Japan

- 原子問合せグラフは点である。標準問合せグラフは、 $(V_q, E_q, P_q, L_{v_q}, L_{e_q}, L_{p_q}, \phi_{v_q}, \phi_{e_q}, \phi_{p_q}, \phi_{connect_q}, \phi_{comp_q}, \phi_{path_q})$ で表される。ここで、 V_q は問合せグラフに含まれる問合せグラフの集合、 E_q は問合せグラフに含まれる枝の集合、 P_q は問合せグラフ中のパスの集合、 L_{v_q} は問合せグラフのラベル集合、 L_{e_q} は枝のラベル集合、 L_{p_q} はパスのラベル集合、 $\phi_{v_q} : V_q \rightarrow L_{v_q}$ 、 $\phi_{e_q} : E_q \rightarrow L_{e_q}$ 、 $\phi_{p_q} : P_q \rightarrow L_{p_q}$ 、 $\phi_{connect_q} : 2^{V_q} \times 2^{V_q} \rightarrow E_q$ 、 $\phi_{comp_q} : 2^{V_q} \cup 2^{E_q} \rightarrow V_q$ 、 $\phi_{path_q} : 2^{E_q \cup P_q} \rightarrow P_q$ 。パスはいわゆる正規表現が許される [1]。

[例2] 図2に rewrite の例を示す。この例では、"database" が含まれる文字列が "graph" が含まれる文字列の次にあるようなデータ表現グラフ manual を求めている。ただし、図では、名前用変数は変数の下に name を付けて表し、値用変数は変数の下に val を付けて表している。

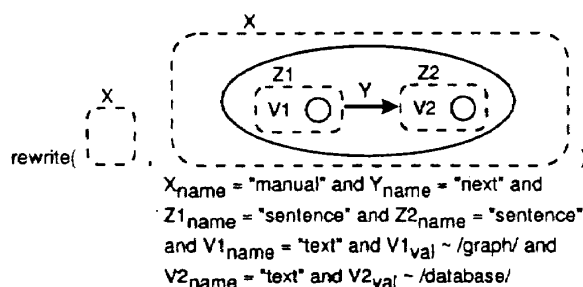


図 2: rewrite の例

3 問合せについて

2 で示したように、標準 drg の構造を決定するのは $\phi_{connect}$ と ϕ_{comp} である。 $\phi_{connect}$ は接続関係を表し、 ϕ_{comp} は構成関係を表す。ここでは、これらの写像を述語として記述してみる。 $\phi_{connect}$ に対しては、枝 e_i 、始要素集合 S_i 、終要素集合 S_j に対して $connect(e_i, S_i, S_j)$ と表現する。例えば、 $connect(e_{11}, \{g_{11}\}, \{g_{12}, g_{13}\})$ といった具合である。同様に、 ϕ_{comp} に対しては、標準 drg g_i と、 g_i に直接包含される要素の集合 S に対して $contain(g_i, S)$ と表現する。例えば、 $contain(g_{11}, \{n_{111}, n_{112}\})$ といった具合である。以上、標準 drg について述べたが、同様の表現を問合せグラフに対しても施すことができる。

このようにすると、rewrite 演算は、複合値を扱うように拡張した datalog プログラム [3] で記述することができる。例えば、図2に示した rewrite は以下のように表現できる。

$$X \leftarrow contain(X, Q), Q \ni Y, connect(Y, S, T), S \ni Z1, contain(Z1, R1), R1 \ni V1,$$

$$T \ni Z2, contain(Z2, R2), R2 \ni V2, X_{name} = "manual", Y_{name} = "next", Z1_{name} = "sentence", Z2_{name} = "sentence", V1_{name} = "text", V1_{val} \sim "graph", V2_{name} = "text", V2_{val} \sim "database".$$

以下に変換の概要を示す。

- 標準 drg (変数 X) は、その直接構成要素 [2] の集合を表す変数 (Q とする) を設け、 $contain(X, Q)$ とする。問合せグラフに指定された X の直接構成要素を表す変数 (Y_i とする) の各々に対して $Q \ni Y_i$ なる述語を加える。
- 枝 (変数 Y) は、その始点集合と終点集合を表す変数 (各々、 S と T とする) を設け、 $connect(Y, S, T)$ とする。問合せグラフに指定された枝 Y の始点集合の要素を表す変数 (Z_i とする) の各々に対して $S \ni Z_i$ なる述語を加える。終点集合についても同様とする。
- 検索条件は datalog プログラムの形式に直して加える。ただし、論理和で結合されている場合はルール節を分ける。値割り当ても datalog プログラムの形式に直して加える。

4 おわりに

本論文では、再帰有向超グラフデータモデルに基づくデータベースに対する問合せについて考察した。本データベースに対する問合せが、データの連結関係と構成関係を利用することで複合値を考慮した Datalog プログラムとして処理できることを示した。

今後は、再帰有向超グラフデータモデルに基づくデータベース管理システムの実現が課題である。

謝辞

本研究は、一部、文部省科学研究費奨励研究(A) (課題番号 09780258) による。

参考文献

- (1) 宝珍, 都司: 超グラフに基づくデータモデルについて, 情処研報 DBS-113-1, pp. 1-6 (1997).
- (2) 宝珍, 都司: 再帰有向超グラフデータモデルにおけるデータの構造的な性質について, 情処第 56 回全大, 1Aa-02, pp. 3-231 - 3-232 (1998).
- (3) S. Abiteboul et al: Foundation of Databases, Addison-Wesley Publishing Company, 1995.