

超グラフ的手法を用いた第三正規化データグループ
作成手法に関する一検討

3G-7

山田光博

NTT情報通信網研究所

1. まえがき

通信網管理等, 同一の対象を複数の観点から把握し得る分野においては, データベース化の際, 関数従属性の分析が行われる過程で, 互いに関数従属する(同値な)ものが多出すという特徴がある.

従来, 様々な第三正規化手法が提案されてきたが^(1,2,3,4), 上記の分野の概念設計に適用した場合は, 次に示す問題が生じる. 即ち, 冗長な関数従属性の排除を行い, 同値なデータ項目をまとめた後に, それに伴い新たに発生した冗長な関数従属性を再度排除するため手順の無駄が多い.

この問題に対しては, 冗長な従属性の排除に先立ち, 同値な項目(候補キー)を一つにまとめる方法が有効であると考えられる. この方法を確立するには, 同値な項目の条件及び, それらをまとめる際に満たされているべき条件を明確にする必要がある.

本稿では, これらの条件, 関数従属性, および同値な関係などを超グラフを用いて容易に表現する手法を提案する. 更に, 変形規則, 及び変形規則を用いて冗長な関数従属性の排除を一度で済ます正規化手法を提案する.

2. 関数従属性超グラフによる関数従属性の表現

本提案では, 正規化の各過程とそれが適用される場合の条件を明確化する一手段として, 与えられる関数従属性の集合Fを超グラフで表現し, 第三正規化に必要な各過程をその上の変形として捉える. 但し, 左辺が複数項目からなる関数従属性の表現等を容易にするため, 通常超グラフを変形したものを用いる. これを以下では関数従属性疑似超グラフ(FDHG)と呼ぶ. FDHG上では, 単一項目(A, B等と記す)を頂点で, 複数項目からなる関数従属性の左辺を超枝で, 関数従属性($f \in F$, 但しFは関数従属性の集合)を有向枝で表現する. 有向枝が存在する点, 超枝を始点とする有向枝が存在する点が通常の超グラフとは異なる.

FDHG上での表現の便宜上以下の仮定をおく. FDHGでの表現を容易にするため仮定1.を, 同じ内容の関数従属性の表現を同じにするため仮定2.を置く. これらを導入しても一般性は損なわれない. 更に, 同一の左辺と右辺を持つ異なる関数従属性が存在する場合に, 右辺の名前を変える等で区別することを要請するため仮定3.を置く.

[関数従属性に関する仮定]

1. $f \in F$ に対し, その右辺は単一項目からなる.

2. $\forall f : X \rightarrow Y \in F$ に対し, $X \cap Y = \phi$
3. $f : X \rightarrow B, g : X' \rightarrow B' \in F$ に対し,
 $X = X', B = B' \Rightarrow f = g$

また, 正規化の最終段階では, 同一の左辺に直接関数従属する右辺をまとめて第三正規化データグループ作成を行う. この時, 互いに関連のあるデータ項目が異なるデータグループにあることに起因する, データグループ間の関連を捉え易くするため, データグループ間にそれらのキー間の従属関係に基づく連関を設ける.

[データグループ間連関]

データグループGiはデータグループGjに従属する
 $\Leftrightarrow \text{key}(Gi) \rightarrow \text{key}(Gj)$

以下, $Gi \rightarrow Gj$ と略記する.

以上に基づきFDHGの要素を, 以下にまとめる.
[FDHGによる関数従属性等の表現]

- HG = (V, HE1, HE2, E1, E2)
- V = [単一データ項目]
- HE1 = [複数項目からなる関数従属性の左辺]
- HE2 = [第三正規化データグループ]
- E1 = [関数従属性] (=E1' \cup E1'')
- E2 = [データグループ間連関]

複数項目からなる候補キーから, その成分であるデータ項目への自明な関数従属性を通常関数従属性と区別するため, E1を以下の様に分類する.

[関数従属性を表す有向枝の分類]

- E1' = [自明な関数従属性] : ラベル1'
- E1'' (=E1 - E1') = [通常関数従属性] : ラベル1

FDHGを定義した理由の一つに, 同値な候補キーを特徴づけることがあった. 複数項目を超枝で表現したことにより, これらの候補キーも複数項目の場合には超枝で捉えられる. そこで同値な候補キーは, 下記のように有向グラフ上の単純閉路に似た概念で捉えることが可能になる.

[FDHG上の擬閉路]

有向枝による以下の有向道を擬閉路Cと呼ぶ

$$C : A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_n \rightarrow A_1 \quad (A_i \in V \cup HE1)$$

ここで擬閉路上にある A_i, A_j に対し, $A_i \rightarrow A_j \in F^*$ が成り立つ. 但し, F^* はFの元に関数従属性の導出則を適用して得られる関数従属性の集合とする.

3. 第三正規化に必要なFDHG上の変形

次に, 第三正規化に必要な変形を用意する. これらは全て第三正規化の各過程と関連する. 下記の規則中2.から6.が正規化の各段階をカバーする. 規則1.は各段階をFDHG上の単純な変形に置き換えるのに先だって行う変形である.

[変形規則]

規則1: 超枝HE1の生成および変形

以下の順序で, 超枝および有向枝を追加する.

- ① 超枝 $(A_1, \dots, A_n) = \{A \in V \mid A \rightarrow A_i \in F\}$ を生成し, ラベル1'を持つ有向枝 $A \rightarrow A_i$ をE1に加える.

A Study on a Method of Synthesizing
Third Normal Form Datagroups
Using Hypergraph Representations

Network Information Systems Labs. NTT

② $\exists he1 \in HE1, \forall Ai \in he1$ に対し、ラベル 1' を持つ有向枝 $he1 \rightarrow Ai$ を $E1$ に加える。
 規則 1 の①および②はそれぞれ、関数従属性の反射律及び合併律を、疑似超グラフ上で表現するための変形規則である。この変形を行った後の FDHG に対し、超枝を頂点と見なした有向グラフ G^{10} 上で、通常の有向グラフの推移閉包 G^{10} を求め、その結果導出される有向枝を加えることにより、関数従属性集合 F から F' を求める操作が実現される。

但し、 $G^{10} = (V', E')$, $V' = V \cup VHE1$, $E' = E$

規則 2: 擬閉路縮退

擬閉路 $C: A1 \rightarrow A2 \rightarrow \dots \rightarrow An \rightarrow A1$ を一頂点 $A1A2 \dots An$ に縮退する。ここに、 Ai は頂点または超枝。これは同値な候補キーのマージに対応する。

規則 3: 推移的なリンクの排除

FDHG に対し、 G^{10} 上で推移的簡約をとりその場合に削除される $e \in E'$ を求める。それぞれに対応する $e \in E$ を FDHG から削除する。冗長な関数従属性の排除等に必要の変形である。

規則 4: 超枝内の FD 有向枝の排除

超枝 $he1(A1, A2, \dots, An) ((A1, A2, \dots, An) \rightarrow Am)$ に対し、有向枝 $(A\alpha, \dots, A\zeta) \rightarrow A\omega \in T(G^{10} - h)$ ($Ai \in he1$ かつ $h = \{e \in E1 \mid e \text{ の始点は } he1\}$) の場合には有向枝の終点を $he1$ から排除する。複数項目からなる候補キーで、当該従属関係に余分な項目を削除する部分に対応する。

規則 5: データグループを示す超枝 HE2 の作成

各 $A (\in V \cup HE1)$ に対し、HG 上で以下を満たす V の部分集合を $HE1$ に加える。

$he2 = \{ A1 \in V \mid A1 (A1 \in V \cup HE1) \text{ を終点とし、} A \text{ を始点とするラベル 1 の有向枝を持つ。} \}$

A を $he2$ のキーとする。更に超枝 $he2'$ のキーが以下を満たすとき、有向枝 $he2 \rightarrow he2'$ を $E2$ に加える

・有向枝: $key(he2') \rightarrow A (A \in V \cap HE1, A \in he2)$

規則 6: 孤立した超枝の除去

超枝 he が、ラベル 1 の始点にも終点にもなっていないとき、 he を HE から排除する。左辺が複数項目からなり、しかも冗長な関数従属性の排除に必要である。

4. 正規化手法

従来様々な正規化手法が提案されてきたが、大略以下の手順を行う^(1, 2, 4)ものが多い。

[従来の代表的な手法の概略]

- SI $f: X \rightarrow B$ の左辺から余分な項目 A を除く。
(但、 A が余分 $\Leftrightarrow X - A \rightarrow B \in F^+$)
- SII F から冗長な従属性を排除する。
(但、 f が冗長 $\Leftrightarrow f \in [F - f]^+ \wedge f: X \rightarrow B$)
- SIII 同一の左辺を持つ項目をデータグループ化し、 f の左辺をそのキーとする。
- SIV 同値な項目キーを持つグループをマージする。
- SV IV により生じる冗長な従属性を排除し、III を行う。

ところが、SII を行った後、SIV を行うため SV で再度冗長な従属性の排除が必要な場合がある。一方、以下に示す本手法では、SV を SII にをまとめた。また、SI を詳細に見ると、以下の様に分析される。即ち余分な項目 $A (\subset X)$ の排除は次の(1), (2)に分類される。

- (1) $X - A \rightarrow A \in [F - \{X \rightarrow B\}]^+$
 - (2) $X - A \rightarrow B \in [F - \{X \rightarrow B\}]^+$
- (1) を排除する部分を FDHG で明確化したものが規則 4 である。一方、(2) の場合は f 自身が冗長な場合なので SII に含めて行う。この部分は FDHG 上では、規則 3 および規則 6 に対応する。更に、SIII は規則 5 に、SIV は規則 2 の適用に対応する。

以上により、下記の正規化手法を得る。
 [第三正規化データグループ群作成手法]

- S0 関数従属性の集合から FDHG を作成
(この状態で、 $HE2 = E2 = \phi$)
- S1 関数従属性の合併律・反射律を実現可能な形に変形
FDHG を規則 1 により、変形し FDHG1 を得る。
- S2 関数従属性の左辺から余分な項目を排除
FDHG1 に規則 4 を適用し、FDHG14 を得る
- S3 同値な候補キーをまとめる
FDHG14 に規則 2 を適用し、FDHG142 を得る
- S4 冗長な関数従属性を排除する
FDHG142 に規則 3 を適用し、FDHG1423 を得る
- S5 左辺が複数項目からなる関数従属性で冗長でありしかも、左辺が他のいかなる関数従属性の右辺にも含まれないものの削除
FDHG1423 に規則 6 を適用し、FDHG14236 を得る
- S6 データグループ化
FDHG14236 に規則 5 を適用し、FDHG142365 を得る
一例として $F = \{ AB \rightarrow C, AB \rightarrow F, EF \rightarrow A, EF \rightarrow D, AC \rightarrow D, BD \rightarrow E, E \rightarrow C \}$ に対し、従来手法を適用すると、S4 で、同値なキーを持つデータグループ $R1(AB, C, E)$, $R2(EF, A, B)$ を統合するために、冗長な従属性 $EF \rightarrow C$ が生成され、再度これを排除するため無駄が生じるのに対し、本方式の場合 S3 で AB, EF が同値になることを検出した後、S4 で正規化を一度限り行えばよい。

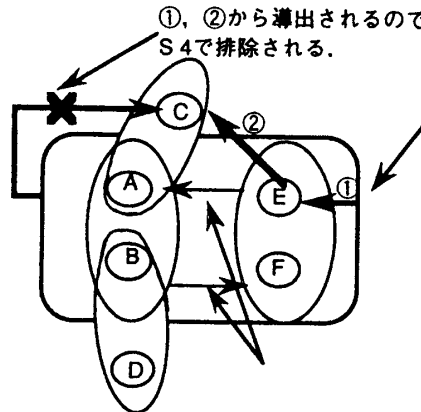


図 1 同値な関係が存在する例

5. むすび

同値な候補キーをまとめた後に、冗長な関数従属性の排除を一度で行う正規化手法を提案した。同値なキーが、データグループ化以前にまとめられているため、作成されたデータグループ群はグループ間連関に関して閉路を持たないという特徴を持つ。

参考文献

- (1) Berri, C and Bernstein, P.A: Computational Problem Related to the Design of Normal Form Relational Schemes, *ACM TODS*, 4, 1979.
- (2) Maier, D: Minimum Covers in the Relational Database model, *J.ACM*, 27, 1980.
- (3) Ausiello, G, Da'tri, A and Sacca, D: Graph Algorithms for Functional Dependency Manipulations, *J.ACM*, 30, 1983.
- (4) Diederich, J and Milton, J: New Methods and Fast Algorithms for Database Normalization, *ACM TODS*, 13, 1988.
- (5) Atkins, J: A note Minimum Covers, *ACM SIGMOD RECORD*, 17, 4, 1988.