

関係データベースデザインの数学的基礎理論について

東北大学 電気通信研究所
増 永 良 文

1. はじめに

関係データベースを論理設計する手法は大別して、1) 分解的手法 2) 合成的手法の二つがある。1) の分解手法は Codd [1] の導入した関数従属性, Fagin [2] の導入した多値従属性等の情報を使い、与えられた初期関係スキームを通常オ三正規形, Boyce-Codd 正規形, あまゝオ四正規形である基本関係スキームと呼ばれる関係スキームの集合に情報無損失分解してゆく手法である。勿論、分解の目的は良し知られてゐる様に、正規形で云わゆる storage anomalies を少なくすることが出来るからである。

さて、本稿の目的は最初段に示してゐる幾つかの分解的手法に則つたデザイン法と概指し、次いでこれらの分解法の共通の欠点である情報無損失分解が下向き（与えられに関係スキームを分解してゆく方）の過程であるということと改善す

べく、著者が既に導入している関係関数従属性存在概念を紹介し、この概念が従来の分解法より更に強力なデザイン法のデザインツールとなっていることを示す。

2. 基礎的事項.

(a). 関係スキーマは関係の構造的、意味論的枠組を与えるものであり、関係データベースの設計とは関係スキーマの組を設計することである。実際のデータはこの枠組がとり値であって、これを関係スキーマのインスタンスという。

$R = (\alpha, \mu, \pi)$, $\alpha = \{A_1, \dots, A_n\}$ は属性集合, μ はドメイン写像, $\mu(A_i) = D_i$ (D_i はドメイン), π は以下で述べる関数従属性の集合, は関係スキーマである。

R のインスタンス r は $D_1 \times \dots \times D_n$ の有限部分集合で, r の各元と両立するものを言う。

(b) π の元は $\beta, \gamma \subseteq \alpha$, とし $\beta \rightarrow \gamma$ の形の陳述である。 $\beta \rightarrow \gamma$ と r が両立するとは次の条件が成立する時を言う。 $\forall x, x' \in r, x_\beta = x'_\beta \Rightarrow x_\gamma = x'_\gamma$ 。

ここに x_β はタプル x の β -部分である。

(c) $R(\alpha)$ は関係スキーマ, $\beta \rightarrow \gamma \in R$ の関数従属性とすると, $R_1(\beta \cup \gamma) = \pi_{\beta \cup \gamma}(R)$, $R_2(\alpha \setminus \gamma) = \pi_{\alpha \setminus \gamma}(R)$, ここに π は射影演算, と定義すると, $R \subseteq R_1$ と R_2

に分解する二の分解の任一方は情報無損失である。二に情報無損失の定義は R の任意のインスタンス r に対して $r = \pi_{\{Burst\}}(r) * \pi_{\{dir\}}(r)$ が成立することである。*は結合演算を表わす。

(d) 関係スキーマ R の全ての $non\text{-}prime$ 性 (どの candidate key にも P 含まない性) が各 candidate key に全関数従属の時に α 正規形という。関係スキーマ $R(SUPPLIER, CITY, POPULATION)$, $SUPPLIER \rightarrow CITY$, $CITY \rightarrow POPULATION$, は α 正規形にある。 α 正規形であってかつ全ての $non\text{-}prime$ 性は各 candidate key に非遷移的に従属している時, β 正規形であるという。Boyce-Codd 正規形とは全ての全関数従属性の決定子が candidate key に存在する時を言う。 β 正規形であって Boyce-Codd 正規形でない有名な例は関係スキーマ $R(STUDENT, TEACHER, COURSE)$, $\{STUDENT, COURSE\} \rightarrow TEACHER$, $TEACHER \rightarrow COURSE$, である。これは β 正規形であるが $TEACHER$ が key でないため Boyce-Codd 正規形でない。 R が γ 正規形であるとは多値従属性 $\beta \twoheadrightarrow \gamma$ が存在すれば, R の全 P 性が β に関数従属の時を言う。Boyce-Codd 正規形であって γ 正規形でない例は $R(COURSE, TEACHER, TEXT)$, $COURSE \twoheadrightarrow TEACHER$, $COURSE \twoheadrightarrow TEXT$ がある[4]。

(e) 基本関係が $\alpha \equiv$ Boyce-Codd, α 正規形である関係スキーマの分解法 $\beta \equiv$ Boyce-Codd, α 正規形分解法と云う。

3. $\alpha \equiv$ 正規形分解法の能力的限界.

以下例題が示す様に $\alpha \equiv$ 正規形分解法による設計法は初期デザインをデザイナーがどの様に与えるかに強く依存してしまふ。これは Boyce-Codd, α 正規形 いづれの分解法にも共通した欠点であり、次に示す様に分解法が top-down の性格を有するといふことから来る能力的限界に於る。

[例題] 科学研究所のデータベースの設計例: 研究所には研究者、タスク、プログラム、予算の概念があるとする。

次の二つの初期デザインを与えよ。

初期デザイン A: $R_{\text{研-タ}}(SNAME, TNAME, TLEADER, TBUDGET)$

, $SNAME \rightarrow TNAME$, $TNAME \rightarrow TLEADER$, $TNAME \rightarrow TBUDGET$,

$R_{\text{研-プ}}(SNAME, PNAME, PLEADER, PBUDGET)$, $SNAME \rightarrow PNAME$,

$PNAME \rightarrow PLEADER$, $PNAME \rightarrow PBUDGET$, からなる。

初期デザイン B: $R_{\text{研-タ}}$, $R_{\text{タ-プ}}(TNAME, PNAME, PLEADER,$

$PBUDGET)$, $TNAME \rightarrow PNAME$, $PNAME \rightarrow PLEADER$, $PNAME$

$\rightarrow PBUDGET$, からなる。

これら初期デザインに定義されている関数従属性を用いて α

三正規形分解を行なうと次の最終デザインをうる。最終デザイン A, B は右の初期デザイン A, B から得られたもの。

最終デザイン A: $R_{st}(S.NAME, T.NAME), R_t(T.NAME, T.LEADER, T.BUDGET), R_{sp}(S.NAME, P.NAME), R_p(P.NAME, P.LEADER, P.BUDGET)$.

最終デザイン B: $R_{st}, R_t, R_{tp}(T.NAME, P.NAME), R_p$.

問題点は最終デザイン A は初期デザイン B から, 最終デザイン B は初期デザイン A からいほどの様な分解をしても決して得られないという点である。このことからデザイナーは初期デザインの選定に苦悩するがもしない。これが従来の分解法の欠点であり, よりフレキシブルなデザイン法が要求される原因でもある。この問題を次に関係関数従平性の概念を入れて打破する。

4. 関係関数従平性の導入と応用.

$R_1 = (d_1, \mu_1, \pi_1), R_2 = (d_2, \mu_2, \pi_2)$ が結合可能とは次の二つの条件が成立するときを云う。

$$(1) d_1 \cap d_2 \neq \emptyset \quad (2) \mu_1 / d_1 \cap d_2 = \mu_2 / d_1 \cap d_2.$$

更に (3) $\beta \rightarrow \gamma, \beta \cup \gamma \subseteq d_1 \cap d_2$, が共に π_1 と π_2 の元であるならこれらの二つの関数従平性は時変関数の意味で一貫する。この条件を関数従平性の一貫性の保証の為課する。

そして改めて R_1 と R_2 は結合可能と云う。注意するところは二つの結合の概念は極めて一般的であることである。その意味は本章終りに記す。 $R_1 * R_2$ で R_1 と R_2 の結合を表わす。 $R_1 * R_2 = (d_1 \cup d_2, \mu_{12}, \pi_1 \cup \pi_2)$, $\mu_{12}|_{d_1} = \mu_1$, $\mu_{12}|_{d_2} = \mu_2$ である。

次に R_1 と R_2 を結合可能な関係スキーマとし, $\beta, \gamma \subseteq d_1 \cup d_2$, しかも $\beta, \gamma \subseteq d_1$ でも $\beta, \gamma \subseteq d_2$ でもなり得る。この時 R_1 と R_2 の間の関係関数従属性 (以下 IFD と略記) は $\beta \rightarrow \gamma$ なる形の陳述である。二の意味は組 (R_1, R_2) の任意の時刻 t におけるインスタンスの組 (r_1, r_2) に対して, もし $r_1 * r_2 \neq \emptyset$ ならば, $\forall u, u' \in r_1 * r_2$ に対して $u_\beta = u'_\beta \Rightarrow u_\gamma = u'_\gamma$ が成立することである。IFD は R_1 と R_2 の間に任意に定義されるものではなく, 次の命題群で規定される制約を有する。

命題1. r_1, r_2 は結合可能な二つの関係スキーマ R_1 と R_2 のインスタンスとし $r_1 * r_2 \neq \emptyset$ とする。もし IFD $\beta \rightarrow \gamma$ (以下一般性を失うことなく $\gamma \subseteq d_2$ と仮定) が R_1 と R_2 の間に存在するならば, $(d_1 \cup \beta) \cap d_2 \rightarrow \gamma$ は $\pi_{d_2}(r_1 * r_2)$ に対して成立しなければならない関係関数従属性である。

命題2. 上述と同じ条件のもとに次が成立する。

(i) もし $d_2 \cap \beta \neq \emptyset$ でかつ $(d_2 \cap \beta) \rightarrow \gamma$ が $\pi_{d_2}(r_1 * r_2)$

に対して成立するならば、 $\pi_{(\alpha_1 \cap \beta)}(r_1 * x_{(\alpha_1 \cap \alpha_2)}) \cap \pi_{(\alpha_1 \cap \beta)}(r_1 * x'_{(\alpha_1 \cap \alpha_2)}) = \emptyset$ が $x_{(\alpha_2 \cap \beta)} = x'_{(\alpha_2 \cap \beta)}$ で $x \neq x'$ なる条件を満足する様な $\pi_{\alpha_2}(r_1 * r_2)$ の全々7-ツプル x と x' に対して成立する。

(ii) もし $d_2 \cap \beta = \emptyset$ ならば, $\pi_\beta(r_1 * x_{(d_1 \cap d_2)}) \cap \pi_\beta(r_1 * x'_{(d_1 \cap d_2)}) = \emptyset$ が $x_f \neq x'_f$ なる条件をみたす可 $\pi_{d_2}(r_1 * r_2)$ の全 $f \in \Gamma$ 上 x と x' に対して成立する。

命題3. r_1 と r_2 は結合可能な二つの関係スキーマ R_1 と R_2 のインスタンスとし $r_1 * r_2 \neq \emptyset$ とする。 $\beta \in r \subseteq d_1 \cup d_2$ の部分集合で $\gamma \subseteq d_2$ しか $\beta \subseteq d_2$ ではないとする。この時、もし命題1及び2の条件が成立するならば $\beta \rightarrow \gamma$ は (r_1, r_2) に対して成立するIFDである。

定理 1. R_1 と R_2 は組合可能な二つの関係スキーマとする。
 $\beta, \gamma \in \mathcal{R} \leq \mathcal{A}_2$ だけあるが $\beta \leq \mathcal{A}_2$ だけない $\mathcal{A}_1 \cup \mathcal{A}_2$ の部
 分集合とある。この時、 $\beta \rightarrow \gamma$ なる形の陳述が $\text{rel}(r_1, r_2)$ 、
 ことに r_1 と r_2 は R_1 と R_2 のインスタンスで $r_1 \times r_2 \neq \emptyset$ 、
 に対して成り立つ為の EFD である為の必要かつ十分条件は
 命題 1 及び 2 の条件が成り立つことである。

定理 2. 定理 1 と同様の条件のもとで, $\beta \rightarrow \gamma$ なる形の陳述
が R_1 と R_2 の間の ZFD である為の必要かつ十分条件は $\beta \rightarrow$
 γ が $r_1 * r_2 \neq \emptyset$ なる R_1 と R_2 の全てのインスタンスの組

(r_1, r_2) に対して成立する IFD であることである。

さて、上述の結果のデータベースデザインへの応用を以下考察する。まず R_1 と R_2 を結合可能な関係スキーマとして、全ての時刻 t に対して、そのインスタンスの組 (r_1, r_2) が $\pi_{(d_1 \cap d_2)}(r_1) = \pi_{(d_1 \cap d_2)}(r_2)$ を満たす時、 R_1 と R_2 は情報無損失結合可能と云うことにする。

定理 3. 情報無損失結合可能な二つの関係スキーマ R_1 と R_2 の間に IFD $\beta \rightarrow \gamma$ ($\gamma \subseteq d_2$) が存在したとすると、この時、 R_1 と R_2 は情報を失うことなく次の三つの関係スキーマに置き換えられる。 $R'(\beta \cup \gamma)$, $R''(d_1 \cup (\beta \cap d_2))$ と $R'''(d_2 \setminus \gamma)$ 。

[略証] IFD $\beta \rightarrow \gamma$ により存在を保証した R_2 の関数従属性 $((d_1 \cup \beta) \cap d_2) \rightarrow \gamma$ を使って R_2 を $R_{21} = \pi_{((d_1 \cup \beta) \cap d_2)}(R_2)$ と $R_{22} = \pi_{(d_2 \setminus \gamma)}(R_2)$ に分解する。この時 R_1 と R_{21} は情報無損失結合可能で、かつ $\beta \rightarrow \gamma$ は R_1 と R_{21} 間の IFD となるので R_1 と R_{21} は情報を失うことなく結合して $R_1 * R_{21}$ を構成し、 $\beta \rightarrow \gamma$ はこの関数従属性となる。そこで $R_1 * R_{21}$ を $\beta \rightarrow \gamma$ で情報無損失分解すると $\pi_{(\beta \cup \gamma)}(R_1 * R_{21})$, $\pi_{(d_1 \cup (\beta \cap d_2))}(R_1 * R_{21})$ となる。□

本結果は IFD を使ったデータベースのデザイン法を論ずる場合本質的である。つまり、定理 3 の条件のもとに、最

初に $R_2 \leq ((\alpha_1 \cup \beta) \cap \alpha_2) \rightarrow \gamma$ で分解せよに、 $R_1 * R_2$ を
 作ってしまふと、その後 $R_1 * R_2$ の関数従属性として定義し
 れた $\beta \rightarrow \gamma$, あまり $((\alpha_1 \cup \beta) \cap \alpha_2) \rightarrow \gamma$ をどの様に使
 っても定理中の R', R'', R''' の類は得られな。これは
 初期デザインとして $R_1(\alpha_1), R_2(\alpha_2), \text{IFD } \beta \rightarrow \gamma$ を
 構成するのと $R(\alpha_1 \cup \alpha_2), \beta \rightarrow \gamma, ((\alpha_1 \cup \beta) \cap \alpha_2)$
 $\rightarrow \gamma$ なる構成を有するとの根本的差異である。又後者の
 構成から導出される最終デザインは全て前者の構成からも導
 出出来かつ上述のことから、IFD_s を導入した関係デー
 タベースのデザイン法は、それを導入しな。デザイン法に比べ
 真にデザイン能力に於いて強力であることが判る。オミ正
 規形分解法であつて、初期デザインとして被教門の関係スキ
 ームと IFD_s を有するデザイン法を拡張したオミ正規形
 分解法と呼ぶことにする。この分解法では容易に確かめら
 れる様に、オミ草の例題を使へば、最終デザイン A と初期デ
 ザイン B から、又最終デザイン B と初期デザイン A から得る
 ことが出来、拡張したオミ正規形分解法は真にデザイン能
 力に於いて、従来のオミ正規形分解法より強力な方法である
 ことが検証される。

5. 結論.

関係データベースデザインの分解的技法が提唱され、それらに共通な欠点である下向き過程がデザイン能力を限定していることを指摘した。この欠点を改善めく、関係関数従平性を新しい概念が定義導入された。この結果得られた拡張されたオミ正規形分解デザイン法は従来のオミ正規形分解法よりもデザイン能力で強くなる技法であることが示された。

[文献]

- [1] Codd, E. F. Further normalization of the data base relational model. Courant Comp. Sci. Symp. (1971)
- [2] Fagin, R. Multivalued dependencies and a new normal form for relational databases, TODS 2.3 (1977)
- [3] Fagin, R. The decomposition versus the synthetic approach to relational database design. 3rd VLDB Conf. Proc. (1977)
- [4] Date, C. J. An Introduction to database Systems. 2nd edition. Addison-Wesley. (1977)

以上。