資　料

リレーショナル・データベースにおける第3正規形について*

穂鷹　良介**

Abstract

Several definitions of normality conditions appeared in the relational data base are investigated. Especially, the difference between the Boyce-Codd condition and the Codd's original definition of Third Normal Form is presented.

はじめに

リレーショナル・データベース・モデルにおける第3正規形 (Third Normal Form) はよく使用される概念であるが，定義方法がやや複雑で直観的な意味が仲々把握しにくい．

Codd は 1) において従来の第3正規形に代るものとしていわゆる Boyce-Codd の条件を新しい第3正規形として示したが，この定義が従来の定義とどのような関係にあるかを示さなかった．

本稿では従来の Codd の定義[2]，Kent が与えた定義[3]，Boyce-Codd の定義[1]を比較検討し，それらの間の差を明確にする．

1. 予備考察

原則として記号法，使用概念等は 2) によるが，本稿で便宜のため設ける記号法，2)で明確には述べられていない概念で正確な定義を必要とするものについてはここで改めて述べておく．

今対象としているリレーションが R でその属性が $A_1, A_2, \cdots, A_n$ であるとき $\Omega(R)=\{A_1, A_2, \cdots, A_n\}$ とする．前後関係でリレーションとして何が指定されているかが明確である場合には $\Omega(R)$の代りに単に Ω とも書く．

x_1, x_2 を $\Omega(R)$ の空でない部分集合とする．属性集合 x_2 が x_1 に関数従属 (functionally dependent)[2] なるとき $x_1 \to x_2$ と書く．x_2 が x_1 に関数従属ではないときには $x_1 \nrightarrow x_2$ と書く．特に x_1 などが $A_1, A_2, \cdots, A_k$ などの要素から成ることを示す場合には $\{A_1, A_2, \cdots, A_k\}$ とも書く．

［定義1］ リレーション R の一つの属性 A_i が R のいかなる候補キー $K(\subset\Omega)$ に対しても $A_i \notin K$ となるときキー無縁 (non-prime) であるという．

以下は種々の第3正規形を定義するための条件である．これらの意味については **2.** 以下で検討するがここで一括列挙しておく．

［定義2］ (2NP): 2nd Normality Property

$A\in\Omega$ かつ A は，キー無縁とするとき，ある候補キー $K\subset\Omega$，属性集合 $x\subset\Omega$ が存在して

$$K\supsetneq x,\quad x\to\{A\}$$

となることはない．

［定義3］ (2NPK): 2nd Normality Property by Kent

ある候補キー $K\subset\Omega$ を考える．このキー K に属さないすべての属性 A に対して，（ある属性集合 $x\subset\Omega$ が存在して

$$K\supsetneq x,\quad x\to\{A\}$$

となる）***ことはない．

［定義4］ (3NP): 3rd Normality Property

$A\in\Omega$ を勝手に選ばれたキー無縁な属性とする．このときある候補キー $K\subset\Omega$，属性集合 x が存在して

$$K\cap x=\phi,\ x\cap\{A\}=\phi,\ K\to x,\ x\to\{A\},\ x\nrightarrow K$$

となることはない．

［定義5］ (3NPK): 3rd Normality Property by Kent

ある候補キー $K\subset\Omega$ を考える．

このキー K に属さないすべての属性A に対して，（ある属性集合 $x\subset\Omega$ が存在して

* On the Third Normal Form in Relational Data Base by Ryosuke HOTAKA (Social Science Laboratory Ltd.).

** ソーシアル・サイエンス・ラボラトリ(株)

*** 日本語の表現があいまいなので文中に（　）を用いた．(3NPK)の定義の場合も同じ．

$K\cap x=\phi$, $x\cap\{A\}=\phi$, $K\rightarrow x$, $x\rightarrow\{A\}$, $x\nrightarrow K$ となる）ことはない．

定義2と定義3（定義4と定義5）の定義方法の差は，前者においてはキー無縁である属性，したがって，どの候補キーにも属さない属性 A について論じているのに対し，後者では他の候補キーには属しているかもしれないがすくなくとも一つの候補キーに関して見ればそれには属していない属性 A について論じているところにある．

［定義6］ (BC)：　Boyce-Codd Condition

$x\subset\Omega$ をある属性の集合，$A\in\Omega$ を x には属さないある属性（つまり $x\cap\{A\}=\phi$ となる）とするとき

$$x\rightarrow\{A\}$$

ならば，常に

$$x\rightarrow\Omega$$

となる．

［定義7］ (3NF)：　Third Normal Form

リレーション R が第1正規形[2]であり，さらに(2NP), (3NP) を満足するとき (3NF) であるという．

［定義8］ (3NFK) Third Normal Form by Kent

リレーション R が第1正規形であり，更に(2NP), (3NPK) を満足するとき (3NFK) であるという．

［定義9］ (3NFKM) Modified (3NFK)

リレーション R が第1正規形であり，更に(2NPK), (3NPK) を満足するとき (3NFKM) であるという．

2. 各種正規条件の比較

Codd は 2) において第3正規形という概念をはじめて述べたが，これは上記の (3NF) に当る概念であった．彼は (2NP) を用いて第2正規形を，(3NP) を用いて第3正規形を説明したわけであるが，Codd がこれらの正規形を導いた原理はどんなものであったかを反省してみよう．

これは **Fig. 1** によって説明できよう．この場合 $\{A\}\rightarrow\{B\}$，$\{A\}\nrightarrow\{C\}$ となっている．属性 A と B の間に存在する関数従属関係を表現するには R のように A, B のほかに属性 C も含まれているリレーションを用いることは適当ではない．たとえば a_1 に対して b_1 が対応するということを表現するには Fig. 1 の R では a_1 に対応する C の値が3個ある以上 (a_1, b_1, c_1)，(a_1, b_1, c_2)，(a_1, b_1, c_3) の3個の組 (tuple) を必要としてしまうが，属性が A と B から成るリレーション S を考えれば (a_1, b_1) 1個の組で十分である．従って情報の表現としては初めから2つのリレーション S，T に分解しておくことが考えられる．

R(A, B, C)

A	B	C
a_1	b_1	c_1
a_1	b_1	c_2
a_1	b_1	c_3
a_2	b_2	c_1
a_3	b_3	c_2

$\Rightarrow$

S(A, B)

A	B
a_1	b_1
a_2	b_2
a_3	b_3

T(A, C)

A	C
a_1	c_1
a_1	c_2
a_1	c_3
a_2	c_1
a_3	c_2

Fig. 1 Functional dependence and the decomposition of the relation

つまり属性間に $\{A\}\rightarrow\{B\}$ のような関係があるときには，A はそれが属するリレーションの候補キーになっていることが求められる．この気持を素直に表現したものは，実は (BC) であった．

Codd はしたがって後の文献 1) では (BC) のことを古い第3正規形に代えて“第3正規形と呼ぶ”といっている．

一方正規形に関しては，時間的に 2) と 1) との間に Kent が 3) において別の考察を行っており，“第3正規形”として (3NFK) を提案している．Codd は，(BC) を述べるに当って，これは Kent の与えた (3NFK) に論理的には同等であると述べた．しかし，これは我々がこれから示すように論理的に同等の概念ではない．実は，(BC) ならば (3NFK) が成立するが逆は成立しないのである．もう一つややこしいことに，Kent は“第2正規形”の別案として (2NPK) を持ち出している．Kent は彼の第3正規形の定義には (2NPK) ではなく (2NP) を用いると明言しているが，(2NPK) は (2NP) よりも狭い概念なので，(2NPK) と (3NPK) とを組合せると (3NFK) より狭い(3NFKM) になる．もしかすると(BC) が (3NFKM) と等しいのではないかという疑問が残るが，実は (BC) であるなら (3NFKM) であるがその逆は成立しない．

この辺のややこしい関係は次の定理につくされる．

［定理］

（1） (BC) ならば (3NFKM)

（2） (3NFKM) ならば (3NFK)

（3） (3NFK) ならば (3NF)

であっておのおの逆は成立しない．

証明は次の一連の補題及び反例から導かれる．

補題1

(BC) ならば (2NPK) である．

［証明］

(2NPK) が仮に成立しないと仮定してみる．このとき勝手に選ばれた候補キー K に対して，ある $A \notin K$, ある属性集合 $x \subset \Omega$ が存在して

$$K \supsetneq x,\quad x \to \{A\}$$

となる．$x \cap \{A\} = \phi$ であるから (BC) の前提条件が満され $x \to \Omega$ となるが，これは候補キー K の極小性と矛盾する．

補題 2

(BC) ならば (3NPK) である．

［証明］

(3NPK) が仮に成立しないと仮定してみる．このとき勝手に選ばれた候補キー K に対してある $A \notin K$, ある属性集合 $x \subset \Omega$ が存在して

$K \cap x = \phi,\ x \cap \{A\} = \phi,\ K \to x,\ x \to \{A\},\ x \nrightarrow K$

となる．$x \to \{A\}$, $x \cap \{A\} = \phi$ であるから (BC)の前提条件が満され $x \to \Omega$ となるが，これは上記の $x \nrightarrow K$ と矛盾する．

補題 3

(2NPK) ならば (2NP) である．

［証明］ 略

補題 4

(3NPK) ならば (3NP)

［証明］ 略

［反例 1］

(3NFKM) であって (BC) でない例がある．

R(A,	B,	C,	D,	E)
0	0	0	0	0
1	0	1	0	0
0	1	1	0	0
1	1	0	1	1
2	0	2	0	0
3	0	2	0	1
2	1	2	2	0

$\{A, B\}$ は唯一の候補キーである．

$\{B, C\} \to \{D\}$ であり，それ以外の関数従属関係はない．$\{B, C\} \to \Omega(R)$ ではないから (BC) は満足されていない．また候補キーに属さない属性(この場合 C, D, E) について (2NPK), (3NPK) が満足されている．

［反例 2］

(3NFK) であって (3NFKM) でない例がある．

R(A,	B,	C)
0	0	0
1	1	0
2	0	1
2	1	1

$\{A, B\}$, $\{B, C\}$ が候補キーであり，かつ $A \to C$ である．(2NP) だが (2NPK) ではない．

［反例 3］

(3NF) であって (3NFK) でない例がある．

R(A,	B,	C,	D)
1	1	1	0
2	2	0	1
3	3	0	0
4	1	1	1

$\{A\}$, $\{B, D\}$, $\{C, D\}$ が候補キーであり，

$A \to B$, $B \to C$, $B \nrightarrow A$ となるケースである．(3NP) だが (3NPK) ではない．

3. おわりに

「第3正規形」と称するものには以上見たような数種類のものが存在するが，(BC) は過去の第2正規形，第3正規形を導いたセマンティカルな考慮を素直に表現した分り易い正規条件であるということがいえよう．過去の第3正規形には (2NP) で導入した条件 (キー無縁，あるいは full dependence) や，(3NP) で導入した条件 (transitive dependence) など複雑な条件を考慮する必要があったのに対し，(BC)はこれらの概念を全く必要としない．このため，正規形かどうかの決定アルゴリズムも簡単になるなどの利点を有している．

本稿の正規性の議論に関してはソーシアル・サイエンス・ラボラトリの菅田光次長に大変有益な助言を頂いたことを感謝する．

参考文献

1) E. F. Codd: Recent Investigations in Relational Data Base Systems, Information Processing 74, pp. 1017～1021.
2) E. F. Codd: Further Normalization of the Data Base Relational Model, in Data Base Systems, Courant Computer Science Symposia Series, Vol. 16, Prentice-Hall, Englewood Cliffs, N. J., (1972)
3) W. Kent: A Primer of Normal Forms, IBM System Development Division, San Jose, California: Technical Report TR 02.600(December 17, 1973)

(昭和51年5月21日受付)
(昭和52年2月10日再受付)