Theoretical Computer Science 34 (1984) 343-350
North-Holland

NOTE

A SOUND AND COMPLETE AXIOMATIZATION OF EMBEDDED CROSS DEPENDENCIES

François BANCILHON* and Philippe RICHARD

Institut National de Recherche d'Informatique et d'Automatique, Domaine de Voluceau, Rocquencourt, 78150 Le Chesnay, France

Communicated by M. Nivat
Received January 1984

Abstract. A sound and complete axiomatization of embedded cross dependencies is given. It is proven to be complete through the study of the dual structure of nondecomposable sets and the exhibition of an Armstrong Relation for a family of cross dependencies.

Résumé. Un système d'axiomes valide et complet pour les dépendances produit des relations n-aires est donné. La preuve de la complétude est obtenue par l'étude de la structure duale des ensembles d'attributs non décomposables. On exhibe une 'Relation d'Armstrong' pour toute famille de dépendances produit.

1. Introduction

In the context of the relational model (with which we will assume the reader is familiar) many types of data dependencies have been defined: functional dependency (FDs) multivalued (MVDs), join dependencies (JDs). One of the best characterizations one can give to such dependencies is a sound and complete set of axioms: A set A of axioms is sound if, given a set F of dependencies, every new dependency f that can be derived from F using A is implied by F; it is complete if every f that is implied by F is derivable from F using A.

Complete and sound sets of axioms have been given for FDs [1] and for FDs and MVDs [2]. It was also shown that no complete and sound axiomatization could be found for embedded MVDs.

In this paper we study a special case of MVDs, cross dependencies, and we give a sound and complete axiomatization for embedded cross dependencies.

This is done through the study of the dual structure of nondecomposable sets, and the exhibition of an Armstrong Relation (a relation that satisfies exactly a set of cross dependencies).

* Also at LRI, Bâtiment 490, Université de Paris Sud, 91405 Orsay, France.

2. Cross dependencies

Definition 2.1. Let U be a set of attributes. A *cross dependency* Y over U is a set of sets such that

(i) $X \in Y \Rightarrow X \subseteq U$ and $X \neq \emptyset$,

(ii) $X \in Y, X' \in Y$ and $X \neq X' \Rightarrow X \cap X' = \emptyset$.

Let Y be a cross dependency (CD) over U; we denote by $\textsc{Scope}(Y)$ the set

$$\textsc{Scope}(Y) = \bigcup_{X \in Y} X$$

and we say that the CD is *total* if $\textsc{Scope}(Y) = U$ and partial otherwise.

Definition 2.2. Let Y be a CD over U and R be a relation over U; we say that R *satisfies* Y ($R \models Y$) iff

$$R(\textsc{Scope}(Y)) = \prod_{X \in Y} R(X),$$

where $\prod$ denotes the cartesian product. Let F be a family of CDs; we say that R satisfies F if it satisfies each Y in F. Finally, given R we denote by $\mathrm{CDF}(R)$ the set of CDs satisfied by R. We are concerned here with the characterization of the class of cross dependency families (CDF) that are satisfied by relations, i.e.,

$$\mathscr{C} = \{F \mid \exists R \text{ such that } \mathrm{CDF}(R) = F\}$$

Such a characterization is usually done through a set of axioms that must be both *sound* and *complete*.

Axioms are sound if, $\forall F \in \mathscr{C}$, F verifies the axioms and they are complete if

$\forall F$ satisfying the axioms, $F \in \mathscr{C}$.

Theorem 2.3. *The following set of axioms is sound for CDs:*

(C1) $\forall X \subseteq U$, $\{X\}$ *is a CD* (***trivial CD***).

(C2) *Let* Y *be a CD. Let* $Z \subseteq U$. *Then* $Y_Z = \{X \cap Z \mid X \in Y \text{ and } X \cap Z \neq \emptyset\}$ *is a CD* (***projection***).

(C3) *Let* $Y = \{X_1, X_2, X_3, \ldots, X_n\}$ *be a CD; then* $Y' = \{X_1X_2, X_3, \ldots, X_n\}$ *is also a CD* (***branch clustering***).

(C4) *Let* $Y = \{X_1X_2, X_3, \ldots, X_n\}$ *and* $Y' = \{X_1, X_2\}$ *be two CDs; then* $Y'' = \{X_1, X_2, X_3, \ldots, X_n\}$ *is also a CD* (***transitivity***).

Proof. These axioms are just special cases of the FOHDs axioms given in [5]. □

The main result of this paper is to show that (C1), (C2), (C3), (C4) are indeed complete. To show this we must at first study the notion of nondecomposition.

3. Nondecomposition

Definition 3.1. Let U be an attribute set. A *nondecomposition* (ND) over U is a subset $X \subseteq U$.

We say that $R(U)$ satisfies the ND X ($R \models X$) iff it does not satisfy any nontrivial CD with scope X.

Finally we denote by $\mathrm{ND}(R)$ the set of NDs satisfied by R.

Theorem 3.2. *The following set of axioms is sound for the class of NDs:*

(N1) $\forall A \in U$, $\{A\}$ *is an ND* (***trivial NDs***).

(N2) *If* Y_1 *and* Y_2 *are NDs and* $Y_1 \cap Y_2 \neq 0$, *then* $Y_1 \cup Y_2$ *is an ND* (***transitivity***).

Proof. (1) (N1) is trivial.

(2) Assume Y_1, Y_2 are NDs with $Y_1 \cap Y_2 \neq 0$ and assume $Y_1 \cup Y_2$ is not an ND. Then there is some partition of $Y_1 \cup Y_2 \neq 0$ which is a cross decomposition.

Projecting this partition on Y_1 and Y_2 we will generate a nontrivial CD with scope Y_1 or Y_2, hence a contradiction. □

4. Relationship between CDs and NDs

We first associate with each CDF and NDF as follows:

Definition 4.1. Let F be a family of CDs. The *associated family* of NDs $\bar{F}$ is defined by

$$\bar{F} = \{Y \mid Y \subseteq U \text{ and } (\forall Z \in F,\ \mathrm{Scope}(Z) = Y \Rightarrow Z \text{ is trivial})\}.$$

Intuitively it is the greatest family of NDs that can be satisfied knowing that F is satisfied.

Definition 4.2. Let F be a family of NDs over U let $Y \subseteq U$; we define

$$\mathrm{Max}_F(Y) = \{Z \mid Z \in F, Z \subseteq Y \text{ and } (\forall Z' \in F, Z \subseteq Z' \subseteq Y \Rightarrow Z' = Z)\}.$$

Theorem 4.3. *Let F satisfy* (N1) *and* (N2); *then*

$$\mathrm{Max}_F(Y) \textit{ is a partition of } Y \textit{ for all } Y \subseteq U.$$

Proof. (1) Let Z_1 and $Z_2 \in \mathrm{Max}_F(Y)$ and assume $Z_1 \cap Z_2 \neq 0$. Then $Z_1 \in F$ and $Z_2 \in F$ and F satisfies (N2) $\Rightarrow Z_1 \cup Z_2 \in F$ and we have

$$Z_1 \subseteq Z_1 \cup Z_2 \subseteq Y \Rightarrow Z_1 \cup Z_2 = Z_1.$$

Symmetrically, $Z_1 \cup Z_2 = Z_2 = Z_1$.

(2) To see that $\mathrm{Max}_F(Y)$ is a covering of Y we just have to remember that each $A \in Y$ is in F, i.e., each element A will be covered by some element of F. □

We can now associate with each NDF a CDF as follows.

Definition 4.4. Let F be a family of NDs satisfying (N1) and (N2). We define the *associated family* of CDs $\bar{F}$ by

$$\bar{F} = \{\mathrm{Max}_F(Y) \mid Y \subseteq U\}.$$

Let us now prove a set of results concerning the relationship between F and $\bar{F}$.

Theorem 4.5. *Let F be a family of CDs satisfying* (C1), (C2), (C3), (C4). *Then $\bar{F}$ satisfies* (N1) *and* (N2).

Proof

$$\bar{F} = \{Y \mid \forall Z \in F, \mathrm{Scope}(Z) = Y \Rightarrow Z \text{ trivial}\}.$$

We first prove that $\bar{F}$ satisfies (N1): $\{A\} \in \bar{F}$ for all $A \in A$, $\forall A \in U, \forall Z \in F$ if $\mathrm{Scope}(Z) = \{A\}$. Then Z is trivial.

Now we prove that $\bar{F}$ satisfies (N2): Assume $X_1 \in \bar{F}$ and $X_2 \in \bar{F}$ and $X_1 \cap X_2 = \emptyset$; then

$$\forall Z_1 \in F \text{ s.t. } \mathrm{Scope}(Z_1) = X_1: \quad Z_1 = \{X_1\},$$

$$\forall Z_2 \in F \text{ s.t. } \mathrm{Scope}(Z_2) = X_2: \quad Z_2 = \{X_2\}.$$

Let Z be such that $\mathrm{Scope}(Z) = X_1 \cup X_2$; by the projection axioms, Z can be projected on X_1 and X_2. It is clear that if Z is nontrivial, at least one of these projections will be nontrivial, hence a contradiction; therefore, Z is trivial and $X_1 \cup X_2 \in \bar{F}$. □

Lemma 4.6. *If R satisfies a family of NDs F and $\bar{F}$, then* $\mathrm{NDF}(R) = F$.

Proof. $X \in \mathrm{NDF}(R)$. Assume $X \notin F$: then $\mathrm{Max}_F(X)$ has at least two elements, i.e., there is a (X_1, X_2) in $\bar{F}$ nontrivial and with scope X, i.e., X is decomposable, hence a contradiction.

Theorem 4.7. *If R satisfies a family of NDs F and $\bar{F}$, then*

$$\mathrm{CDF}(R) = (\bar{F})'.$$

Proof. Let $Y = (X_1, X_2, \ldots, X_n) \in \mathrm{CDF}(R)$ and consider

$$Y' = \bigcup \mathrm{Max}_F(X_i).$$

(1) From the transitivity axiom we have $Y' \leq Y$.

(2) $Y' \in \bar{F}$ because

$$\text{NDF}(R) = F \Rightarrow \forall X \in F,\ X \subseteq \text{Scope}(Y),\ X \in \text{Max}_F(X_i) \text{ otherwise } X \text{ would be decomposable;}$$

therefore $Y' = \text{Max}_F(\bigcup_i X_i)$, i.e., $\forall Y, \bar{F} \Rightarrow Y$. □

Theorem 4.8. *Let F be a family of CDs satisfying* (C1), (C2), (C3), (C4); *then* $(\bar{\bar{F}})^* = F$.

Proof. We first prove $F \subseteq (\bar{\bar{F}})^*$.

Let $Y = \{X_1, X_2, \ldots, X_n\} \in F$ and let $Y' = \text{Max}_F(\text{Scope}(Y))$.

Claim. $\forall X \in Y', \exists X_i \in Y$ s.t. $X \subseteq X_i$.

Proof of the Claim. Assume this is not to be true. Then there exists some $X \in Y'$ not included in any X_i. It is therefore covered by more than one X_i, assume by X_1 and X_2 (the proof easily generalizes). $(X_1, X_2) \in F$ and, by the projection axiom, $(X_1 \cap X, X_2 \cap X) \in F$; therefore $X \notin \bar{F}$ and X cannot be in $\text{Max}_F(\text{Scope}(Y))$. The claim is proved.

Therefore Y' is a refinement of Y. By the clustering axiom, $Y' \Rightarrow Y$ and $Y' \in \bar{\bar{F}} \Rightarrow Y \in (\bar{\bar{F}})^*$.

We now prove the reverse inclusion, $(\bar{\bar{F}})^* \subseteq F$.

F being closed under (C1), (C2), (C3), (C4), it is sufficient to show $\bar{\bar{F}} \subseteq F$. Let $Y \in \bar{\bar{F}}$; then there is some Z s.t. $Y = \text{Max}_F(Z) = (X_1 X_2 \ldots X_n)$. Assume $Y \notin F$; then $\exists X_i$ s.t. $(X_i, \bigcup_{j \neq i} X_j) \notin F$ (otherwise, by projection and transitivity, $Y \in F$). Therefore, these exists some $T \in \bar{F}$ that intersects both X_i and $\bigcup_{j \neq i} X_j$ and X_i is not maximal undecomposable, hence a contradiction. □

5. Completeness of NDs

Theorem 5.1. *Given a family of NDs F, there exists a relation R_f such that*

$$R_f \models F \quad \textit{and} \quad R_f \models \bar{F}.$$

Proof. We build this relation as follows:

(1) Define some coding $c: F \to \mathbb{N}$.

(2) For each $A \in U$ define

$$D(A) = \{c(Y) \mid Y \in F, A \in Y\}.$$

(3) Define

$$R_0(A_1 A_2 \ldots A_n) = D(A_1) \times D(A_2) \times \cdots \times D(A_n).$$

(4) For each $Y \in F$, $|Y| > 1$, delete from R_0 all tuples x such that

$$x(Y) = (c(Y), c(Y), \ldots, c(Y)).$$

Thus we have obtained relation R_f.

Before proceeding with the proof of this theorem, let us look at an example:

$$U = ABC, \ldots \quad F = \{\{A\}, \{B\}, \{C\}, \{AB\}, \{ABC\}\}.$$

Define c as follows:

$$c(\{A\}) = 1; \quad c(\{B\}) = 2; \quad c(\{C\}) = 3; \quad c(\{ABC\}) = 4; \quad c(\{AB\}) = 5.$$

Then

$$D(A) = \{1, 4, 5\} \quad (A \in \{A\}, \{AB\}, \{ABC\}),$$

$$D(B) = \{2, 4, 5\} \quad (B \in \{B\}, \{AB\}, \{ABC\}),$$

$$D(C) = \{3, 4\} \qquad (C \in \{C\}, \{ABC\}),$$

$$R_0(ABC) = \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix} \times \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix} \times \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$

The final relation is

$$R_0(ABC) = \begin{bmatrix} 4 & 4 & 4 \\ 5 & 5 & 3 \\ 5 & 5 & 4 \end{bmatrix}.$$

[4 4 4] is deleted because $\{ABC\} \in F$; [5 5 3] and [5 5 4] are deleted because $\{AB\} \in F$.

It is easy to check that by removing these 3 tuples we have forbidden any decomposition of ABC on AB, while all others are satisfied. (A, C) and (B, C) are satisfied cross decompositions.

We now proceed to prove that R_f is indeed the good candidate. This will be proved in a number of claims.

Claim 5.1.1. *Let $Y \in F$ and $X \subsetneq Y$. Then $(c(Y), c(Y), \ldots, c(Y)) \in R_f(X)$*

Proof

$$\forall A \in X: \quad c(Y) \in R_0(A) \quad (\text{because } A \subseteq X \subsetneq Y \in F),$$

$$\forall A \notin X: \quad c(A) \in R_0(A) \quad (\text{because } \{A\} \in F).$$

Denote $U - X = B_1 B_2 \ldots B_p$ (single attributes). Let x_0 be the following tuple:

R	X	$B_1\ B_2 \ldots B_p$
	$c(Y) \ldots c(Y)$	$c(B_1) \ldots c(B_p)$

$x_0 \in R_0(U)$. Moreover, $\forall Y' \in F, |Y'| > 1$,

$$x_0(Y') \neq (c(Y') \ldots c(Y')).$$

Therefore, $x_0 \in R_f$ and $x_0(X) \in R_f(X)$.

Claim 5.1.2. *Let $X \subseteq U$, $Y \subseteq U$, $X \cap Y = \emptyset$. If $x \in R_F(X)$, $y \in R_F(Y)$ and $xy \notin R_F(XY)$, then $\exists Z \subsetneq XY$, $Z \in F$ such that*

$$xy(z) = (c(z) \ldots c(z)).$$

Proof

$$x \in R_F(X) \Rightarrow x \in R_0(X), \qquad y \in R_F(Y) \Rightarrow y \in R_0(Y).$$

Therefore, $xy \in R_0(XY)$.

Let $T = U - XY = B_1 B_2 \ldots B_p$. Define tuple v as follows.

R	X	Y	$B_1 \cdots B_p$
v	x	y	$c(B_1) \ldots c(B_p)$

Then $v \in R_0$ and $v(XY) = xy$ and, since $xy \notin R_F$, $v \notin R_F$, i.e., it was erased because for some $Z \in F$, $|Z| > 1$,

$$v(z) = (c(z) \ldots c(z)).$$

Since all B_i's have size 1, Z is necessarily included in XY. □

Claim 5.1.3. *Let $Y \subseteq U$ and $\text{MAX}_F(Y) = \{X_1, X_2, \ldots, X_n\}$. Let $x_i \in R_F(X_i)$ $\forall i = 1, 2, \ldots, n$. Then $x_1 x_2 \ldots x_n \in R_F(X_1 X_2 \ldots X_n)$.*

Proof. Assume $x = x_1 x_2 \ldots x_n \notin R_F(X_1 \ldots X_n)$. Then, by Claim 5.1.2, $\exists Z \subsetneq X_1 X_2 \ldots X_n$ such that

$$x(Z) = c(z)c(z) \ldots c(z).$$

(1) Assume $Z \subseteq X_i$ for some i. Then, $x_i(z) = (c(z) \ldots c(z))$ and $x_i \notin R_F(X_i)$ which is a contradiction.

(2) In that case it is necessarily the case that Z intersects two X_i's say X_1 and X_2. Then, since $X_1 \in F$, $X_2 \in F$ and $Z \in F$ by rule (N2) we have $X_1 \cup X_2 \in F$, i.e., X_1 and X_2 are not maximal in Y which is a contradiction. □

This claim is clearly equivalent to R satisfies F.

Claim 5.1.4. *Let $X \in F$. Then $R_F(X)$ is not decomposable (which is equivalent to say $R \models F$).*

Proof. Assume $|X| > 1$ (otherwise the result is trivial). Then $X = AX_1$ with $|X_1| \geq 1$, $(c(x)) \in R_F(A)$, $(c(x) \ldots c(x)) \in R_F(X_1)$ and $(c(x) \ldots c(x)) \notin R_F(X_1, A)$. □

This completes the proof of Theorem 5.1. □

6. Completeness of CD

We can now state the main theorem of this paper.

Theorem 6.1. *Axioms* (C1), (C2), (C3), (C4) *are complete for CDs.*

Proof. Let F satisfy (C1), (C2), (C3), (C4). By Theorem 4.5, $\bar{F}$ satisfies (N1) and (N2). By Theorem 5.1 there exists an $R_{\bar{F}}$ such that

$$R_{\bar{F}} \models \bar{F} \quad \text{and} \quad R_{\bar{F}} \models \bar{\bar{F}}^*$$

By Theorem 4.8, $\bar{\bar{F}}^* = F$, i.e., $R_{\bar{F}}$ satisfies F and $\bar{F}$.

Finally, by Theorem 4.7,

$$\mathrm{CDF}(R_{\bar{F}}) = F. \qquad \square$$

The reader should note that $R_{\bar{F}}$ is indeed an Armstrong Relation for F.

References

[1] W.W. Armstrong, Dependency structures of database relationships, in: J.L. Rosenfeld, ed., *Proc. IFIP 74* (North-Holland, Amsterdam, 1974) pp. 580–583.

[2] C. Beeri, R. Fagin and J.D. Howard, A complete axiomatization for functional and multivalued dependencies, *Proc. ACM SIGMOD Conf.* (1977) pp. 47–61.

[3] E.F. Codd, A relational model for large shared data bases, *Comm. ACM* **13** (6) (1970) 377–387.

[4] E.F. Codd, Further normalizations of the data base relational model, *6th Courant Computer Science Symp.: Data Base Systems* (Prentice-Hall, Englewood Cliffs, NJ, 1971) pp. 65–98.

[5] C. Delobel, Normalization and hierarchical dependencies in the relational data model, *ACM Trans. Database Systems* **13** (3) (1978) 201–222.