A sound and complete axiomatization of embedded cross dependencies

François Bancilhon, Patrick Richard

CENTRE DE ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél.:(3) 954 90 20

Rapports de Recherche

N° 249

A SOUND AND COMPLETE AXIOMATIZATION OF EMBEDDED CROSS DEPENDENCIES

François BANCILHON
Philippe RICHARD

Octobre 1983

A SOUND AND COMPLETE AXIOMATIZATION OF EMBEDDED CROSS DEPENDENCIES

François. BANCILHON (1), (2)

Philippe. RICHARD (1)

(1) INRIA

(2) LRI - Bât. 490, Université de Paris Sud, 91405 - ORSAY.

A SOUND AND COMPLETE AXIOMATIZATION OF EMBEDDED CROSS DEPENDENCIES

François. BANCILHON (1), (2)

Philippe. RICHARD (1)

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 non decomposable 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) INRIA

(2) LRI - Bât. 490, Université de Paris Sud, 91405 - ORSAY.

1. - INTRODUCTION

In the context of the relational model (with which we will assume the reader familiar) many types of data dependencies have been defined : functional dependency (FD's) multivalued (MVD's), join dependencies (JD's). One of the best characterization 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 FD's, [Arms 74] for MVD's, for FD's and MVD's [Beer 77]. It was also shown that no complete and sound axiomatization could be found for embedded MVD's.

In this paper we study a special case of MVD's, 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 non decomposable sets, and the exhibition of an Armstrong Relation (a relation that satisfies exactly a set of cross dependencies).

2. - CROSS DEPENDENCIES

Definition 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 SCOPE(Y) the set

$$\mathrm{SCOPE}(Y) = \bigcup_{X \in Y} X$$

and we say that the CD is total if SCOPE(Y) = U and partial otherwise.

Definition 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

$$R(\mathrm{SCOPE}(Y)) = \prod_{X \in Y} R(X)$$

where Π denotes the cartesian product. Let F be a family of CD's, we say that R satisfies F if it satisfies each Y in F. Finally, given R we denote CDF(R) the set of CD's 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 1. The following set of axioms is sound for CD's

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

C2 : Let Y be a CD
Let $Z \subset U$
Then $Y_Z = \{X \cap Z \mid X \in Y \text{ and } X \cap Z \neq \emptyset\}$ is a CD
(Projection)

C3 : Let $Y = \{X_1, X_2, X_3,..., X_n\}$ be a CD then $Y = \{X_1X_2, X_3,..., X_n\}$ is also a CD (Branch Clustering)

C4 : Let $Y = \{X_1X_2, X_3,..., X_n\}$ and $Y' = \{X_1, X_2\}$ be two CD's then $Y'' = \{X_1, X_2, X_3,..., X_n\}$ is also a CD (Transitivity)

□

Proof : These axioms are just special cases of the FOHD's axioms given in [Delo 78] .

The main result of this paper is to show that C_1 C_2 C_3 C_4 are indeed complete. To show this we must at first study the notion of non decomposition.

3. - NON DECOMPOSITION

Definition 3. Let U be an attribute set.

A non decomposition (ND) over U is a subset $X \subseteq U$.

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

Finally we denote ND(R) the set of ND's satisfied by R.

Theorem 2. The following set of axioms is sound for the class of ND's :

N1 : $\forall A \in U$, $\{A\}$ is an ND (Trivial ND's)

N2 : If Y_1 and Y_2 are ND's and $Y_1 \cap Y_2 \neq \emptyset$ then $Y_1 \cup Y_2$ is an ND. (Transitivity).

□

Proof :

(1) N1 is trivial

(2) assume Y_1, Y_2 are ND's with $Y_1 \cap Y_2 \neq \emptyset$ and assume $Y_1 \cup Y_2$ is not an ND.

Then there is some partition of $Y_1 \cup Y_2 \neq \emptyset$ which is a cross decomposition.

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

Q.E.D.

4. - RELATIONSHIP BETWEEN CD's AND ND's

We first associate with each CDF an NDF :

Definition 4.

Let F be a family of CD's. The associated family of ND's $\bar{F}$ is defined by :

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

$\square$

Intuitively it is the greatest family of ND's that can be satisfied knowing that F is satisfied.

Definition 5. Let F be a family of ND's over U let Y U we define

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

Theorem 3. Let F satisfy N_1 and N_2 then

$\text{MAX}_F(Y)$ is a partition of Y for all $Y \subseteq U$.

Proof :

(1) Let Z_1 and $Z_2 \in MAX_F(Y)$ and assume $Z_1 \cap Z_2 \neq 0$. Then $Z_1 \in F$ and $Z_2 \in F$ and F satisfies $N_2 \Rightarrow$ $Z_1 \cup Z_2 \in F$ and we have

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

symmetrically $Z_1 \cup Z_2 = Z_2 = Z_1$

(2) To see that $MAX_F(Y)$ 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.

Q.E.D.

We can now associate with each NDF a CDF :

Definition 6. Let F be a family of ND's satisfying N_1 and N_2. We define the associated family of CD's $\bar{F}$ by

$$\bar{F} = \{MAX_F(Y) \mid Y \subseteq U\}$$

□

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

Theorem 4. Let F be a family of CD's satisfying C_1 C_2 C_3 C_4. Then $\bar{F}$ satisfies N_1 and N_2.

Proof :

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

N1 : $\{A\} \in \bar{F}$ for all $A \in A$
$\forall A \in U$, $\forall Z \in F$ if SCOPE(Z) = $\{A\}$. Then Z is trivial.

Q.E.D.

N2 : Assume $X_1 \in \bar{F}$ and $X_2 \in \bar{F}$ and $X_1 \cap X_2 = \emptyset$ then

$$\forall \; Z_1 \in F \text{ s.t. SCOPE}(Z_1) = X_1 \quad Z_1 = \{X_1\}$$

$$\forall \; Z_2 \in F \text{ s.t. SCOPE}(Z_2) = X_2 \quad Z_2 = \{X_2\}$$

Let Z be such that 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 non trivial at least one of these projections will be non trivial, hence a contradiction, therefore Z is trivial and $X_1 \cup X_2 \in F$.

Q.E.D.

Lemma 1. If R satisfies a family of ND's F and $\bar{F}$ then NDF(R)=F.

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

Q.E.D.

Theorem 5. If R satisfies a family of ND's F and $\bar{F}$ then

$$\text{CDF}(R) = (\bar{F})^{*}.$$

Proof : Let $Y = (X_1,X_2,\ldots,X_n) \in$ CDF(R) and consider

$$Y' = \bigcup_i MAX_F(X_i)$$

(1) from transitivity axiom : $Y' \Rightarrow Y$

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

$NDF(R) = F \Rightarrow \forall X \in F, X \subseteq SCOPE(Y)$
$X \in MAX_F(X_i)$ otherwise X would be decomposable

therefore $Y' = MAX_F(\bigcup_i X_I)$
i.e. $\forall Y \; \bar{F} \Rightarrow Y$

Q.E.D.

Theorem 6. Let F be a family of CD's satisfying C_1 C_2 C_3 C_4 then $(\bar{\bar{F}})^{\times} = F$.

Proof : $F \subseteq (\bar{\bar{F}})^{\times}$

Let $Y = \{X_1, X_2, \ldots, X_n\} \in F$ and let $Y' = MAX_{\bar{\bar{F}}}(SCOPE(Y))$

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

proof : assume it is not 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 generalises easily)
$(X_1, X_2) \in F$ 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 $MAX_{\bar{\bar{F}}}(SCOPE(Y))$.

Q.E.D.

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

Q.E.D.

Proof : $(\bar{\bar{F}})^{\times} \subseteq F$

F beeing closed under C_1 C_2 C_3 C_4 it is sufficient to show $\bar{\bar{F}} \subseteq F$

Let $Y \in \bar{\bar{F}}$ then there is some Z s.t.

$$Y = MAX_{\overline{F}}(Z) = (X_1 X_2 \ldots X_n)$$

assume $Y \in F$ then

$$\exists\, X_i \text{ s.t. } (X_i, \bigcup_{j \neq i} X_j) \notin F$$

(otherwise by projection and transitivity $Y \in F$) therefore there exists some $T \in \overline{F}$ that intersects both X_i and $\bigcup_{j \neq i} X_j$ and X_i is not a maximal undecomposable hence a contradiction.

Q.E.D.

5. - COMPLETENESS OF ND's

Theorem 7. Given a family of ND's F there exists a relation R_F such that

$$R_F \models F \quad \text{and} \quad R_F \models \overline{F}$$

Proof : We build the relation as follows :

(1) define some coding $c : F \rightarrow \mathbf{N}$

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

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

(3) Define

$$R_o(A_1 A_2 \ldots A_n) = D(A_1) \times D(A_2) \times \ldots \times D(A_n)$$

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

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

Thus obtaining relation R_F.

Before proceeding to the proof let us look at an example

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

define c as follows :

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

Then

$D(A) = \{1,4,5\}$ $(A \in \{A\}, \{AB\}, \{ABC\})$
$D(B) = \{2,4,5\}$ $(B \in \{B\}, \{AB\}, \{ABC\})$
$D(C) = \{3,4\}$ $(C \in \{C\}, \{ABC\})$

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

The final relation is

$$R_o(ABC) - \begin{bmatrix} 4 & 4 & 4 \\ 5 & 5 & 3 \\ 5 & 5 & 4 \end{bmatrix}$$

$[4\ 4\ 4]$ 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 proving that R_F is indeed the good candidate.

Claim 1.

$Y \in F$ and $X \subsetneq Y$

Then $(c(Y),\ c(Y).....c(Y)) \in R_F(X)$

$\square$

Proof : $\forall A \in X \quad c(Y) \in R_o(A) \quad$ (because $A \subseteq X \subseteq Y \in F$)

$\forall A \notin X \quad c(A) \in R_o(A) \quad$ (because $\{A\} \in F$)

Denote $U\text{-}X = B_1 B_2 ... B_p$ (single attributes).

Let x_o be the following tuple :

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

$x_o \in R_o(U)$

more over $\forall\ Y' \in F$, $|Y'| > 1$

$$x_o(Y') \neq (c(Y')...c(Y'))$$

therefore $x_o \in R_F$ and $x_o(X) \in R_F(X)$.

Q.E.D.

Claim 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 $\exists\ Z \subseteq XY$, $Z \in F$ such that

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

Proof :

$x \in R_F(X) \Rightarrow x \in R_o(X)$
$y \in R_F(Y) \Rightarrow y \in R_o(Y)$
therefore $xy \in R_o(XY)$
Let $T = U - XY = B_1B_2 \dots B_p$
Define tuple v as follows

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

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

$$v(z) = (c(z) \dots c(z))$$

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

Claim 3.

Let $Y \subseteq U$ and $MAX_F(Y) = \{X_1, X_2, \dots, X_n\}$
Let $x_i \in R_F(X_i) \quad \forall\, i = 1,2,\dots,n$
then $x_1\ x_2 \dots x_n \in R_F(X_1\ X_2 \dots X_n)$.

Proof :

Assume $x = x_1\ x_2 \dots x_n \notin R_F(X_1 \dots X_n)$
then by Claim 2, $\exists\, Z \subseteq X_1\ X_2 \dots X_n$ such that

$$x(Z) = c(z)\ c(z) \dots c(z)$$

(1) assume $Z \subseteq X_i$ for some i

then $x_i(z) = (c(z) \dots c(z))$
and $x_i \notin R_f(X_i)$ which is a contradiction.

(2) Then 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 N_2 : $X_1 \cup X_2 \in F$
i.e. X_1 and X_2 are not maximal in Y which is a contradiction

Q.E.D.

This claim is clearly equivalent to R satisfies $\overline{F}$.

Claim 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 $(c(x) \ldots c(x)) \notin R_F(X_1,A)$

Q.E.D.

6. - COMPLETENESS OF CD

We can now state the main theorem of this paper :

Theorem 8

Axioms C_1, C_2, C_3, C_4 are complete for CD's.

Proof :

Let F satisfy C_1, C_2, C_3, C_4
By theorem 4, $\overline{F}$ satisfies N_1 and N_2
By theorem 7, there exists $R_{\overline{F}}$ such that

$$R_{\overline{F}} \models \overline{F} \text{ and } R_{\overline{F}} \models \overline{\overline{F}}^{*}$$

By theorem 6, $\bar{\bar{F}}^{*} = F$ i.e.

$R_{\bar{\bar{F}}}$ satisfies F and $\bar{\bar{F}}$

Finally by theorem 5

$CDF(R_{\bar{\bar{F}}}) = F$

Q.E.D.

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

References

Arms 74 : Armstrong W.W., "Dependency Structures of Database Relationships", Proc. IFIP 74 (1974), pp. 580-583.

Beer 77 : Beeri C., Fagin R., Howard J.D., "A Complete Axiomatization for Functional and Multivalued Dependencies", Proc. ACM SIGMOD Conf. (1977), pp. 47-61.

Codd 70 : Codd E.F., "A Relational Model for Large Shared Data Bases", Communications of the ACM, Vol. 13, N°6, June 1970, pp. 377-387.

Codd 71 : Codd E.F., "Further Normalizations of the Data Base Relational Model", Courant Comp. Sc. Symp. 6 : Data Base Systems, Prentice Hall, N.I. May 1971, pp. 65-98.

Delo 78 : Delobel C., "Normalization and Hierarchical Dependencies in the Relational Data Model", ACM Transactions on Database Systems, Vol. 13, N°3, September 1978, pp. 201-222.

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique