ON ARMSTRONG DATABASES FOR CONDITIONAL GENERALIZED EMBEDDED DEPENDENCIES

Felea, Victor
Faculty of Computer Science ”Al. I. Cuza” University of IASI

https://doi.org/10.5109/13471

出版情報：Bulletin of informatics and cybernetics. 30 (1), pp.79-91, 1998-03. Research Association of Statistical Sciences
バージョン：
権利関係：

Bulletin of Informatics and Cybernetics, Vol. 30, No. 1, 1998

ON ARMSTRONG DATABASES FOR CONDITIONAL GENERALIZED EMBEDDED DEPENDENCIES

By

Victor FELEA *

Abstract

The family of conditional generalized embedded dependencies about relations in relational databases is introduced. These constraints are expressible as formulas in second-order logic and they are constructed with generalized embedded dependencies as the components. For a class of this family a characterisation of a given constraint logically implied by a certain subclass of the class is given.

The existence of the Armstrong-like databases for special classes of conditional generalized embedded dependencies is proved. A method for the construction of an Armstrong model, when it exists and certain conditions are fulfilled, is given.

CR categories: 4.33, 5.21, 5.27, 5.32

Key Words and Phrases: Armstrong database, Armstrong relation, conditional dependency, conditional generalized embedded dependency, functional dependency, generalized embedded dependency, logical consequence, mathematical logic, relational database.

1. Introduction

The purpose of this paper is to investigate several aspects of a large family of dependencies. We define the family of conditional generalized embedded dependencies using the second-order logic. These contraints are Horn formulas of second-order having as the components, generalized embedded dependencies.

This family of dependencies contains as special cases the class of conditional-functional dependencies, studied by P. de Bra and J. Paredaens (1988), the family of generalized embedded dependencies, studied by J. Grant and B. Jacobs (1982) and the class of conditional embedded implicational dependencies studied by the author (1995).

In the paper, for a class of this family we give a characterisation of a given constraint logically implied by a certain subclass of the class.

For special classes of conditional generalized embedded dependencies we prove the existence of the Armstrong-like databases, using the direct products of databases.

Finally, we give a method for the construction of an Armstrong model, when it exists and under certain conditions.

* Faculty of Computer Science "Al. I. Cuza" University of IAŞI Berthelot 16, 6600 - Iaşi, Romania

2. Preliminaires

In this section, we specify several principal notations used throughout the paper. We assume that the usual notations in first-order and second-order logic (Chang and Lee 1973, Robbin 1969, Yasuhara 1971) are known.
Let U be a finite nonempty set of distinct attributes. Let R_i be a relational symbol, which corresponds to a relation r_i over U_i, where U_i is a subset of U. Let S_i be a predicate symbol, universally quantified, which represents a subset s_i of r_i. Let us denote by u_i a tuple ordered of variables over U_i. The atomic formulas are typed; they are either of the form $R_i(u_i)$, $S_i(u_i)$, or of the form $x = y$, where x and y are individual variables with the same type. A generalized embedded dependency statement (ged) is a sentence of the form (Grant and Jacobs 1982):

$$H(R_{j_1}, \ldots, R_{j_{k+1}}) \equiv \forall x_1 \ldots \forall x_n \exists z_1 \ldots \exists z_m \left[R_{j_1}(t_1) \wedge \ldots \wedge R_{j_k}(t_k) \Rightarrow R_{j_{k+1}}(t_{k+1})\right]$$

where each R_{j_i} is an n_i-ary predicate symbol, $1 \leq i \leq k+1$, t_j is a tuple ordered of terms, $1 \leq j \leq k+1$, every x-variable in t_{k+1} is contained in some t_i, $1 \leq i \leq k$, no t_i, for $1 \leq i \leq k$, contains any z-variable.

Let X_i be a subset of U_i and r_i a relation over U_i. A subset s_i of tuples in r_i is called X_i-complete in r_i if for all $t_1 \in s_i$ and $t_2 \in r_i - s_i$, we have $t_1[X_i] \neq t_2[X_i]$. A conditional functional dependency (De Bra and Paredaens 1988) is denoted by $X \to Y \supset\!\!- X \to Z$. We say that r obeys dependency $X \to Y \supset\!\!- X \to Z$ if in every X-complete set of tuples of r in which the functional dependency $X \to Y$ holds, the functional dependency $X \to Z$ must hold too.

The family of conditional functional dependencies has been introduced and studied by P. de Bra and J. Paredaens (1988). These authors used conditional functional dependencies for horizontal decompositions of relation schemes.

If we take embedded multivalued dependency instead of functional dependencies, we obtain a conditional embedded multivalued dependency of the form: $X \to\!\!\!\to Y/T_1 \supset -X \to\!\!\!\to Z/T_2$. We say that r obeys this dependency, if in every X-complete set of tuples of r in which the embedded multivalued dependency $X \to\!\!\!\to Y/T_1$ holds, the embedded multivalued dependency $X \to\!\!\!\to Z/T_2$ must hold too. It is known (Grant and Jacobs 1982) that for an embedded multivalued dependency $X \to\!\!\!\to Y/T$ a generalized embedded dependency statement is associated.

3. The Family of Conditional Generalized Embedded Dependencies

In this section we define the family of conditional generalized embedded dependencies.

DEFINITION 3.1. A conditional generalized embedded dependency (shortly CGED) is a sentence in second-order logic of the form:

$$\begin{gathered}\alpha(R_1, \ldots, R_k, X_1, \ldots, X_k) \equiv (\forall S_1 \ldots \forall S_k)\, [\varphi_1(S_1, R_1) \wedge \ldots \wedge \varphi_k(S_k, R_k) \wedge \\ \wedge \psi_1(S_1, R_1, X_1) \wedge \ldots \wedge \psi_k(S_k, R_k, X_k) \wedge \sigma_1(S_1, \ldots, S_k) \wedge \ldots \wedge \sigma_h(S_1, \ldots S_k) \Rightarrow \\ \Rightarrow \sigma(S_1, \ldots, S_k)]\end{gathered} \quad (1)$$

where:

a) R_i is a predicate symbol, which represents a relation r_i over U_i, $1 \leq i \leq k$;

b) X_i is a fixed subset of U_i, $1 \leq i \leq k$;

c) S_i is a predicate symbol, universally quantified, which represents a subset s_i of r_i, $1 \leq i \leq k$;

d) $\varphi_i(S_i, R_i)$ is a first-oder formula of the form:

$$\varphi_i(S_i, R_i) \equiv (\forall u_i)[S_i(u_i) \Rightarrow R_i(u_i)],\ 1 \leq i \leq k;$$

The pair (s_i, r_i) of relations over U_i satisfies $\varphi_i(S_i, R_i)$ whenever s_i is a subset of r_i, $1 \leq i \leq k$;

e) $\psi_i(S_i, R_i, X_i)$ is a formula of the form:

$$\psi_i(S_i, R_i, X_i) \equiv (\forall x_i)(\forall t_i)\left[S_i(x_i, t_1^i) \wedge R_i(x_i, t_2^i) \Rightarrow S_i(x_i, t_2^i)\right];$$

x_i is a tuple of variables associated to the attributes of X_i, t_1^i, t_2^i are tuples of variables associated to the attributes in $U_i - X_i$ and $t_i = (t_1^i, t_2^i)$. The pair (s_i, r_i) of relations over U_i satisfies the formulas $\psi_i(S_i, R_i, X_i)$ and $\varphi_i(S_i, R_i)$ iff the relation s_i is X_i-complete in r_i;

f) $\sigma_j(S_1, \ldots, S_k)$ is a generalized embedded dependency (ged) with the relation symbols $S_1, \ldots, S_k$, $1 \leq j \leq h$, $\sigma(S_1, \ldots, S_k)$ is also a generalized embedded dependency;

g) $h \geq 0$; if $h = 0$ then the formula $\alpha(R_1, \ldots, R_k, X_1, \ldots, X_k)$ is equivalent with

$$(\forall S_1 \ldots \forall S_k)\sigma(S_1 \ldots S_k)$$

EXAMPLE 3.1. If in the Definition 3.1, we take $k = 1$, $h = 1$ and $\sigma_1(S_1), \sigma(S_1)$ as the formulas associated to the embedded multivalued dependencies $X \twoheadrightarrow Y/T_1$, $X \twoheadrightarrow Z/T_2$, respectively, then it obtains a formula that represents the conditional embedded multivalued dependency $X \twoheadrightarrow Y/T_1 \supset\!\!- X \twoheadrightarrow Z/T_2$.

EXAMPLE 3.2. Let f be a conditional functional dependency of the form: $f = X \rightarrow Y \supset\!\!- X \rightarrow Z$. For $X \rightarrow Y$ we can construct a generalized embedded dependency statement (Grant and Jacobs 1982), denoted by $\sigma_1(S)$, where $\sigma_1(S)$ does not contain z-variables. Similarly, for $X \rightarrow Z$ there is a generalized embedded dependency statement, denoted by $\sigma(S)$. Thus, for the dependency f, we can associate a CGED of the form (1).

REMARK 3.1. *Let us denote by $\alpha_1(R_1, \ldots, R_k)$ the formula $\alpha(R_1, \ldots, R_k, \emptyset, \ldots, \emptyset)$ of the form (1), where $\emptyset$ is the empty set. Moreover, let α' be the following formula:*

$$\alpha' \equiv [\sigma_1(R_1, \ldots, R_k) \wedge \ldots \wedge \sigma_h(R_1, \ldots, R_k) \Rightarrow \sigma(R_1, \ldots, R_k)]$$

We have:

$(r_1, \ldots, r_k)$ satisfies α' iff $(r_1, D_1), \ldots, (r_k, D_k)$ satisfies $\alpha_1(R_1, \ldots, R_k)$ for every $(D_1, \ldots, D_k)$ such that $D_j \neq \emptyset$, $1 \leq j \leq k$.

The relation r_i corresponds to the relation symbol R_i, $1 \leq i \leq k$; D_i is a set of relations over U_i (the universe of R_i), and the domain of values for S_i, $1 \leq i \leq k$.

EXAMPLE 3.3. If in the formula α' from the Remark 3.1 we consider $h = 1$ and $\sigma(R_1, \ldots, R_k)$ as a tautology, then the new α' is equivalent with $\sigma(R_1, \ldots, R_k)$, which is a generalized embedded dependency (these dependencies have been defined and studied by J. Grant and B.E. Jacobs in 1982).

EXAMPLE 3.4. We consider in the formula (1) as $\sigma_i(S_1 \ldots S_k)$ the embedded implicational dependencies over an unique relational symbol ($k = 1$). The new formula (1) has the form of a conditional embedded implicational dependency (these dependencies have been defined and studied by the author in 1995).

EXAMPLE 3.5. Let us consider an X_i-complete set S_i of tuples of R_i with one X_i-value. Such a set is called X_i-unique and complete. Let us denote by $\sigma_1(S_i)$ the following formula:

$$\sigma_1(S_i) \equiv (\forall t_1 t_2 v_1 v_2)[S_i(t_1 v_1) \wedge S_i(t_2 v_2) \Rightarrow (t_1 = t_2)],$$

where t_1 and t_2 are tuples of variables corresponding to X_i, and v_1, v_2 are tuples of variables corresponding to $U_i - X_i$. We have:

s_i is X_i-unique and complete in r_i iff

(s_i, r_i) satisfies $\varphi_i(S_i, R_i)$, $\psi_i(S_i, R_i, X_i)$ and $\sigma_1(S_i)$.

Thus the formula (1) contains certain restrictions defined over the X_i-unique and complete sets.

EXAMPLE 3.6. Let be the following relation schemes:

$ORDERS(NAME, ITEM, QUANTITY)$,

$SUPPLIERS(SNAME, SITEM, SQUANTITY)$

The first schema represents the orders for certain items and the second represents the posibilities of the suppliers (the supplier with name $SNAME$ can supply the item $SITEM$ with the maximum quantity $SQUANTITY$). We assume that two different suppliers can supply the sets of items disjoined.

Let us consider the following restriction:

For every person who makes an order, if there exists a supplier who contains all items necessary to this person, then the quantities for these items are sufficient.

Let X be the set containing the attribute $NAME$ and Y containing $SNAME$. Let R_1 be ORDERS and R_2 be $SUPPLIERS$. Let S_1 be a relational variable corresponding to R_1 and X. Let S_2 be a relational variable that corresponds to R_2 and Y.

The formula $\alpha(R_1, R_2, X, Y)$ expresses the restriction above:

$$\alpha(R_1, R_2, X, Y) \equiv (\forall S_1 S_2)[\varphi_1(S_1, R_1) \wedge \varphi_2(S_2, R_2) \wedge \psi_1(S_1, R_1, X) \wedge$$

$$\wedge \psi_2(S_2, R_2, Y) \wedge \gamma_1(S_1) \wedge \gamma_2(S_2) \wedge \xi_1(S_1, S_2) \Rightarrow \xi_2(S_1, S_2)],$$

where

$\varphi_i(S_i, R_i) \equiv (\forall u)[S_i(u) \Rightarrow R_i(u)],\ i = 1, 2$

$\psi_1(S_1, R_1, X) \equiv (\forall x y_1 z_1 y_2 z_2)[S_1(x, y_1, z_1) \wedge R_1(x, y_2, z_2) \Rightarrow S_1(x, y_2, z_2)]$

$\psi_2(S_2, R_2, Y) \equiv (\forall t v_1 w_1 v_2 w_2)[S_2(t, v_1, w_1) \wedge R_2(t, v_2, w_2) \Rightarrow S_2(t, v_2, w_2)]$

$\gamma_1(S_1) \equiv (\forall x_1 y_1 z_1 x_2 y_2 z_2)[S_1(x_1, y_1, z_1) \wedge S_1(x_2, y_2, z_2) \Rightarrow (x_1 = x_2)]$

$\gamma_2(S_2) \equiv (\forall t_1 v_1 w_1 t_2 v_2 w_2)[S_2(t_1, v_1, w_1) \wedge S_2(t_2, v_2, w_2) \Rightarrow (t_1 = t_2)]$
$\xi_1(S_1, S_2) \equiv (\forall x_1 p_1 c_1)[S_1(x_1, p_1, c_1) \Rightarrow (\exists t c_2) S_2(t, p_1, c_2)]$
$\xi_2(S_1, S_2) \equiv (\forall x_1 p_1 c_1 t c_2)[S_1(x_1, p_1, c_1) \wedge S_2(t, p_1, c_2) \Rightarrow R_3(c_1, c_2)]$
$R_3(c_1, c_2)$ is the relation $c_1 \leq c_2$.

Let T be a class of generalized embedded dependency statements. Let us define the class T^* such that

$$T^* = \{\sigma | T \models \sigma, \sigma \text{ is a generalized embedded dependency statement}\}$$

Let α be a CGED formula of the form (1). We denote this formula by $\alpha(\overline{R}_k, \overline{X}_k)$. For a class $\mathcal{C}$ of CGEDs and α a CGED formula of the form (1), we define recursively a set of generalized embedded dependency constraints, denoted by $T(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k))$, where $\alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)$ is the left-side part of $\alpha(\overline{R}_k, \overline{X}_k)$:

$$\begin{aligned}\alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k) \equiv \varphi_1(S_1, R_1) \wedge \ldots \wedge \varphi_k(S_k, R_k) \wedge \psi_1(S_1, R_1, X_1) \wedge \ldots \\ \ldots \wedge \psi_k(S_k, R_k, X_k) \wedge \sigma_1(S_1, \ldots, S_k) \wedge \ldots \wedge \sigma_h(S_1, \ldots, S_k)\end{aligned} \tag{2}$$

DEFINITION 3.2. Let $\mathcal{C}$ be a class of CGEDs and α be a CGED of the form (1). Let $\alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)$ be the left-side part of $\alpha(\overline{R}_k, \overline{X}_k)$. We define $T_i(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k))$ in the following manner:

$$\begin{aligned}T_0(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) = \{\varphi_1(S_1, R_1), \ldots, \varphi_k(S_k, R_k), \psi_1(S_1, R_1, X_1), \ldots \\ \ldots, \psi_k(S_k, R_k, X_k), \sigma_1(\overline{S}_k), \ldots, \sigma_h(\overline{S}_k)\}\end{aligned}$$

$T_{i+1}(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) = T_i^*(\mathcal{C}; \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) \cup$
$\cup \{\sigma'(\overline{S}'_t)\theta |$ there is $\beta(\overline{Q}_t, \overline{Y}_t) \in \mathcal{C}$, such that:
$\beta(\overline{Q}_t, \overline{Y}_t) \equiv (\forall \overline{S}'_t)[\varphi'_1(S'_1, Q_1) \wedge \ldots \wedge \varphi'_t(S'_t, Q_t) \wedge \psi'_1(S'_1, Q_1, Y_1) \wedge \ldots$
$\ldots \wedge \psi'_t(S'_t, Q_t, Y_t) \wedge \sigma'_1(\overline{S}'_t) \wedge \ldots \wedge \sigma'_q(\overline{S}'_t) \Rightarrow \sigma'(\overline{S}'_t)], \{Q_1, \ldots, Q_t\} \subseteq \{R_1, \ldots, R_k\}$
and there is a substitution θ: $\theta = (S'_1/S_{j_1}, \ldots, S'_t/S_{j_t})$ such that:
$T_i(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) \models \varphi'_j(S'_j, Q_j)\theta, \psi'_j(S'_j, Q_j, Y_j)\theta$, $j = \overline{1,t}$,
$T_i(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) \models \sigma'_j(\overline{S}'_t)\theta$, $j = \overline{1,q}\}$

If $Q_i = R_{j_i}$, then S'_i and S_{j_i} correspond to the same relational symbol R_{j_i}, $i = \overline{1,t}$.

$$T(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) = \bigcup_{i=0}^{\infty} T_i(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k))$$

With the notations used in Definition 3.2, we say that $\beta(\overline{Q}_t, \overline{Y}_t)$ from $\mathcal{C}$ "appears" in the generation of $T(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k))$.

THEOREM 3.3. *Let $\mathcal{C}$, $\alpha(\overline{R}_k, \overline{X}_k)$, $\alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)$ be defined as in the Definition 3.2. Let $\mathcal{C}_1$ be the set of all elements from $\mathcal{C}$ which "appear" in the generation of $T(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k))$. We have:*

$$\mathcal{C}_1 \models \alpha(\overline{R}_k, \overline{X}_k) \text{ iff } \sigma(\overline{S}_k) \in T(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) \tag{3}$$

PROOF.

($\Leftarrow$). Let $\sigma(\overline{S}_k) \in T(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k))$. We must show, by induction on i the following implication:

$$\mathcal{C} \cup T_0(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) \models T_i(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) \tag{4}$$

If $i = 0$, the assertion (4) is true.
Now, assume that (4) is true for a natural number i. Let γ be from $T_{i+1}(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k))$.
We have two situations:

i) $\gamma \in T_i^*(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k))$ or

ii) $\gamma \in T_{i+1}(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) - T_i^*(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k))$

In the case i), it results:

$$T_i(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) \models \gamma \tag{5}$$

From (4) and (5) it obtains:

$$\mathcal{C} \cup T_0(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) \models \gamma$$

In the case ii) there are $\beta(\overline{Q}_t, \overline{Y}_t) \in \mathcal{C}$ and $\theta = (S_1'/S_{j_1}, \ldots, S_t'/S_{j_t})$ such that:

$$\begin{array}{c} T_i(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) \models \varphi_j'(S_j', Q_j)\theta,\ \psi_j'(S_j', Q_j, Y_j)\theta,\ j = \overline{1,t} \\ T_i(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) \models \sigma_j'(\overline{S}_t')\theta,\ j = \overline{1,t} \end{array} \tag{6}$$

Using (4) and (6), we obtain:

$$\begin{array}{c} \mathcal{C} \cup T_0(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) \models \varphi_j'(S_j', Q_j)\theta,\ \psi_j'(S_j', Q_j, Y_j)\theta,\ j = \overline{1,t} \\ \mathcal{C} \cup T_0(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) \models \sigma_j'(\overline{S}_t')\theta,\ j = \overline{1,t} \end{array} \tag{7}$$

Since $\beta(\overline{Q}_t, \overline{Y}_t) \in \mathcal{C}$, it results:

$$\mathcal{C} \cup T_0(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) \models \sigma'(\overline{S}_t')\theta \tag{8}$$

We have $\gamma = \sigma'(\overline{S}_t')\theta$, hence

$$\mathcal{C} \cup T_0(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) \models \gamma \tag{9}$$

Thus, the relation (4) is true for every natural number i. If $\sigma(\overline{S}_k) \in T(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k))$, then there is i such that $\sigma(\overline{S}_k) \in T_i(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k))$.
We have: $\mathcal{C} \cup T_0(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) \models \sigma(\overline{S}_k)$ which implies $\mathcal{C} \models \alpha(\overline{R}_k, \overline{X}_k)$.

($\Rightarrow$). We have: $\mathcal{C}_1 \models \alpha(\overline{R}_k, \overline{X}_k)$.

We must show that $\sigma(\overline{S}_k) \in T(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k))$. Assume the contrary, that means

$$\sigma(\overline{S}_k) \notin T(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) \tag{10}$$

$\sigma(\overline{S}_k)$ has the form:

$$\sigma(\overline{S}_k) \equiv \forall x_1 \ldots \forall x_n \exists z_1 \ldots \exists z_m [S_{j_1}(t_1) \wedge \ldots \wedge S_{j_h}(t_h) \Rightarrow S_{j_{h+1}}(t_{h+1})]$$

The elements of $T(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k))$ and $\sigma(\overline{S}_k)$ are generalized embedded dependency statements. For a class $\mathcal{E}$ of generalized embedded dependencies and an S a generalized embedded dependency, J. Grant and B.E. Jacobs (1982) have associated a set of atomic formulas, denoted by $Y(\mathcal{E}; S)$ and they have shown that: $\mathcal{E} \models S$ iff there is a substitution θ on the z-variables of S such that $R_{j_{h+1}}(t_{h+1})\theta \in Y(\mathcal{E}; S)$, where $R_{j_{h+1}}(t_{h+1})$ is the right-hand side of S.

Taking $T(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k))$ instead of $\mathcal{E}$, and $\sigma(\overline{S}_k)$ instead of S, we obtain:

$T(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) \models \sigma(\overline{S}_k)$ iff there is a substitution θ on z-variables of $\sigma(\overline{S}_k)$, such that

$$S_{j_{h+1}}(t_{h+1})\theta \in Y(T(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)),\ \sigma(\overline{S}_k)) \tag{11}$$

The relation (10) implies that:

$$T(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) \not\models \sigma(\overline{S}_k) \tag{12}$$

From (11) and (12) it results that:

for every substitution θ on the z-variables of $\sigma(\overline{S}_k)$, we have

$$S_{j_{h+1}}(t_{h+1})\theta \notin Y(T(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)), \sigma(\overline{S}_k)) \tag{13}$$

Let us define a model M:

$$M \models V_i(t) \text{ iff } V_i(t) \in Y(T(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)), \sigma(\overline{S}_k))$$

where V_i is either R_j or S_j.

By the Theorem 4.2 (Grant and Jacobs 1982), it results that M obeys every element in $T(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k))$ and M does not satisfy $S_{j_{h+1}}(t_{h+1})\theta$, for every substitution θ on the z-variables of $\sigma(\overline{S}_k)$.

Since M satisfies $T(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k))$ it obtains that M satisfies $\mathcal{C}_1$. On the other hand, the model M does not satisfy $\sigma(\overline{S}_k)$, hence a contradiction. □

REMARK 3.2. *We have* $T(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)) = T(\mathcal{C}_1, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k))$.

REMARK 3.3. *If* $\sigma(\overline{S}_k) \in T(\mathcal{C}, \alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k))$ *then* $\mathcal{C} \models \alpha(\overline{R}_k, \overline{X}_k)$.

4. Armstrong-like databases

In the following we prove the existence of the Armstrong-like databases for certain classes of conditional generalized embedded dependencies.

In the sequel we use the direct product defined by R. Fagin (1982).

The direct product for a nonempty family of relations $\langle r_i, i \in I\rangle$ is denoted by $\otimes\langle r_i, i \in I\rangle$.

DEFINITION 4.1.

We say that a sentence σ in first-order logic is upward faithful

(with respect to direct products) if whenever $\langle r_i, i \in I\rangle$ is a family of nonempty relations such that σ holds for every r_i, then σ holds for $\otimes\langle r_i, i \in I\rangle$.

We say that the sentence σ in first-order logic is downward faithful (with respect to direct products) if whenever $\langle r_i, i \in I\rangle$ is a family of nonempty relations such that σ holds for $\otimes\langle r_i, i \in I\rangle$, then σ holds for every r_i, $i \in I$.

We say that σ is faithful (with respect to direct products) iff it is both upward and downward faithful.

Let $\mathcal{C}$ be a class of CGEDs. For the sake of simplicity, assume that the R-relational symbols are $R_1, \dots R_l$. A model for the class $\mathcal{C}$ has the form

$$M = (r_1, \dots, r_l, \mathcal{D}_1, \dots, \mathcal{D}_l)$$

where r_i are relations on R_i and $\mathcal{D}_i$ is a set of relations on R_i, $i = \overline{1,l}$.
If $\mathcal{D}_i$ and $\mathcal{D}'_i$ are sets of relations on R_i, then their direct product, denoted by $\mathcal{D}_i \otimes \mathcal{D}'_i$ is defined thus:

$$\mathcal{D}_i \otimes \mathcal{D}'_i = \{s_i \otimes s'_i | s_i \in \mathcal{D}_i,\ s'_i \in \mathcal{D}'_i\}.$$

If $M^j = (r^j_1, \dots, r^j_l, \mathcal{D}^j_1, \dots, \mathcal{D}^j_l)$ are models, then the direct product of M^j, $j \in J$, $J \neq \emptyset$, denoted by $\otimes\langle M^j, j \in J\rangle$ is defined by:

$$\otimes\langle M^j, j \in J\rangle = (\otimes\langle r^j_1, j \in J\rangle, \dots, \otimes\langle r^j_l, j \in J\rangle,$$

$$\otimes\langle \mathcal{D}^j_1, j \in J\rangle, \dots, \otimes\langle \mathcal{D}^j_l, j \in J\rangle)$$

A model M is called relationwise nonempty if every r_i is nonempty, $i = \overline{1,l}$, every $\mathcal{D}_i$ is nonempty, $i = \overline{1,l}$ and $\emptyset$ is not an element in $\mathcal{D}_i$, $i = \overline{1,l}$.

DEFINITION 4.2. Let σ be a sentence in second-order logic. We say that σ is upward faithful (with respect to direct products) if whenever $\langle M^j, j \in J\rangle$ is a family of relationwise nonempty models, such that σ holds for every M^j, $j \in J$, then σ holds for $\otimes\langle M^j, j \in J\rangle$.

We say that σ is dowward faithful (with respect to direct products) if whenever $\langle M^j, j \in J\rangle$ is a family of relationwise nonempty models, such that σ holds for $\otimes\langle M^j, j \in J\rangle$, then σ holds for every M^j, $j \in J$.
The sentence σ is faithful iff it is both upward and downward faithful.

PROPOSITION 4.3. *Let $\alpha(\overline{R}_k, \overline{X}_k)$ be a CGED statement as in Definition 3.1. Let $\alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)$ his left-hand side and $\sigma(\overline{S}_k)$ his right-hand side. We have:*

i) If $\alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)$ is downward faithful and $\sigma(\overline{S}_k)$ is upward faithful, then $\alpha(\overline{R}_k, \overline{X}_k)$ is upward faithful.

ii) If $\alpha_1(\overline{R}_k, \overline{X}_k, \overline{S}_k)$ is upward faithful and $\sigma(\overline{S}_k)$ is downward faithful, then $\alpha(\overline{R}_k, \overline{X}_k)$ is downward faithful.

iii) The formulas $\varphi_i(S_i, R_i)$, $\psi_i(S_i, R_i, X_i)$ are faithful, $i = \overline{1,k}$.

iv) The formulas $\sigma_i(\overline{S}_k)$ and $\sigma(\overline{S}_k)$ are upward faithful, $i = \overline{1,h}$.

PROOF. The assertions i)–iii) follow by the Definitions 4.1 and 4.2. The part iv) follows from the fact that these formulas are generalized embedded dependencies and by the Theorem 2.4 pg. 961 (Fagin 1982). □

DEFINITION 4.4. Let $\mathcal{C}$ be a class of CGEDs and α a single CEID. We say that α is nonempty logical consequence of $\mathcal{C}$, denoted by $\mathcal{C} \models_{\text{nlc}} \alpha$, if every relationwise nonempty model M that obeys $\mathcal{C}$, it also obeys α.

THEOREM 4.5. (Fagin 1982) *Let $\mathcal{S}$ be a set of sentences. The part a) implies the part b):*

a) Existence of a faithful operator. There is an operator $\oplus$ that maps nonempty families of models into models, such that if σ is a sentence in $\mathcal{S}$ and $\langle R_i : i \in I\rangle$ is a nonempty family of models, then σ holds for $\oplus\langle R_i : i \in I\rangle$ iff σ holds for each R_i.

b) Existence of Armstrong models. Whenever Σ is a consistent subset of $\mathcal{S}$ and Σ^* is the set of sentences in $\mathcal{S}$ that are logical consequences of Σ, then there is a model (an "Armstrong model") that obeys Σ^* and no other sentences in $\mathcal{S}$

THEOREM 4.6. *Let $\mathcal{C}$ and $\mathcal{D}$ be two classes of CGEDs, such that $\mathcal{C} \subseteq \mathcal{D}$. Let $\mathcal{C}^*_{nlc}$ be the set of all CGED that are nonempty logical consequence of $\mathcal{C}$. Assume that for every $\gamma \in \mathcal{D}$, both left-hand side and right-hand side of γ are downward faithful. Then there exists a model M that obeys $\mathcal{C}^*_{nlc}$ and no other CGEDs in $\mathcal{D}$.*

PROOF. By the assumption of the theorem and by the Proposition 4.3, it results that every sentence γ from $\mathcal{D}$ is faithful. Now, we consider the Theorem 4.5. Let $\mathcal{S}$ be the class $\mathcal{D}$, let a "model" be a relationwise nonempty model, and let $\oplus$ be the product $\otimes$ of models. Thus, the part a) of Theorem 4.5 holds. It obtains that the part b) holds too. Every class $\mathcal{C}$ of CGEDs is consistent. Hence there is a model M_0 that satisfies $\mathcal{C}^*_{nlc}$ and no other CGEDs in $\mathcal{D}$. □

REMARK 4.1. *The model M_0 from the theorem 4.6 is called "Armstrong-like" rather than "Armstrong" because here we are using "nonempty logical consequence" instead of "logical consequence".*

It was known that even for extended embedded implicational dependencies (Fagin 1982) there is an Armstrong-like database for a class of extended embedded implicational dependencies and there is not necessarily an Armstrong database.

REMARK 4.2. *The result of the Theorem 4.6 remains true if we consider "the left-part side of σ is downward faithful for every $\sigma \in \mathcal{C}$ and the right-hand side of γ is downward faithful for every $\gamma \in \mathcal{D} - \mathcal{C}^*_{nlc}$" instead of "for every $\gamma \in \mathcal{D}$, both left-hand side and right-hand side of γ are downward faithful".*

5. Armstrong Databases

Let $\mathcal{C}$ and $\mathcal{D}$ be two classes of CGEDs, such that $\mathcal{C} \subseteq \mathcal{D}$.

$\mathcal{C}^+ = \{\alpha | \alpha \in \mathcal{D},\ \mathcal{C} \models \alpha\}$.
A model M is an Armstrong model for $\mathcal{C}$ with respect to $\mathcal{D}$ if $M \models \alpha$, for every α in $\mathcal{C}$ and $M \not\models \beta$, for any $\beta \in \mathcal{D} - \mathcal{C}^+$.

We assume that every component of α contains no constant symbols for every α in $\mathcal{D}$. Moreover, we assume that no variable of type S occurs in more than one sentence in $\mathcal{D}$ (on the contrary, we use appropiate substitutions).

For $\mathcal{C}$ and $\mathcal{D}$ we define recursively a set of generalized embedded dependency constraints, denoted by $T(\mathcal{C};\mathcal{D})$. This definition extends Definition 3.2.

DEFINITION 5.1.
$T_0(\mathcal{C};\mathcal{D}) = \{\varphi | \varphi$ is a left-hand side component in some α from $\mathcal{D} - \mathcal{C}^+\}$.
$T_{i+1}(\mathcal{C};\mathcal{D}) = T_i(\mathcal{C};\mathcal{D}) \cup \{\sigma'(\overline{S}'_t)\theta |$ there is $\beta(\overline{Q}_t, \overline{Y}_t) \in \mathcal{C}$,
such that:

$$\beta(\overline{Q}_t, \overline{Y}_t) \equiv (\forall \overline{S}'_t)[\varphi'_1(S'_1, Q_1) \wedge \ldots \wedge \varphi'_t(S'_t, Q_t) \wedge$$
$$\wedge \psi'_1(S'_1, Q_1, Y_1) \wedge \ldots \wedge \psi'_t(S'_t, Q_t, Y_t) \wedge$$
$$\wedge \sigma'_1(\overline{S}'_t) \wedge \ldots \wedge \sigma'_q(\overline{S}'_t) \Rightarrow \sigma'(\overline{S}'_t)],$$

$\{Q_1, \ldots, Q_t\} \subseteq \{R_1, \ldots, R_k\}$ and there is a substitution θ,
$\theta = (S'_1/S_{j_1}, \ldots, S'_t/S_{j_t})$ such that:
$T_i(\mathcal{C};\mathcal{D}) \models (\varphi'_j(S'_j, Q_j)\theta, \psi'_j(S'_j, Q_j, Y_j)\theta,\ j = \overline{1,t}$
$T_i(\mathcal{C};\mathcal{D}) \models \sigma'_j(\overline{S}'_t)\theta,\ j = \overline{1,q}\}$

$$T(\mathcal{C};\mathcal{D}) = \bigcup_{i=0}^{\infty} T_i(\mathcal{C};\mathcal{D}).$$

In the following we need some notations. For a formula $\varphi(S_1, \ldots, S_k)$ from $T(\mathcal{C};\mathcal{D})$ we define a set of CGED formulas, denoted by $GEN(\varphi)$, that contains all the elements in $\mathcal{D} - \mathcal{C}^+$ which are implied in the generation of φ.
Formally, $GEN(\varphi)$ is defined recursively:

a) If $\varphi \in T_0(\mathcal{C};\mathcal{D})$, then $GEN(\varphi) = \{\alpha\}$, where α is the formula from $\mathcal{D} - \mathcal{C}^+$ that contains the left-hand side the formula φ.

b) Let us consider $T_i(\mathcal{C};\mathcal{D}) \models \varphi(S_1, \ldots, S_k)$ and assume that $GEN(\psi)$ has been defined for every ψ in $T_i(\mathcal{C};\mathcal{D})$.
There is a derivation of $\varphi(S_1, \ldots, S_k)$ from $T_i(\mathcal{C};\mathcal{D})$.
Let $\varphi_1, \varphi_2, \ldots, \varphi_h$ be the elements from this derivation which belong to $T_i(\mathcal{C};\mathcal{D})$ and they are different from $\varphi(S_1, \ldots, S_k)$.
Then we define:

$$GEN(\varphi(S_1, \ldots, S_k)) = GEN(\varphi_1) \cup \ldots \cup GEN(\varphi_h)$$

c) If $\varphi(S_1, \ldots, S_k)$ is obtained by generation from $\beta(\overline{Q}_t, \overline{Y}_t)$ in $\mathcal{C}$ as in Definition 5.1, then we define:

$$GEN(\varphi(S_1, \ldots, S_k)) = \bigcup_{j=1}^{t} GEN(\varphi'_j) \cup \bigcup_{j=1}^{t} GEN(\psi'_j) \cup \bigcup_{j=1}^{q} GEN(\sigma'_j),$$

where:

$\varphi_j' = \varphi_j'(S_j', Q_j)\theta$, $\psi_j' = \psi_j'(S_j', Q_j, Y_j)\theta$, $j = \overline{1,t}$
$\sigma_j' = \sigma_j'(\overline{S_t'})\theta$, $j = \overline{1,q}$.

LEMMA 5.2. *Let $\varphi \in T(\mathcal{C};\mathcal{D})$ and $LEFT(GEN(\varphi))$ be all elements in left-hand size of all formulas in $GEN(\varphi)$. We have*

$$\mathcal{C} \cup LEFT(GEN(\varphi)) \models \varphi(S_1, \ldots, S_k)$$

PROOF. By the induction on i, such that $\varphi \in T_i(\mathcal{C};\mathcal{D})$. □

THEOREM 5.3. *Let $\varphi(S_1, \ldots, S_k)$ be from $T(\mathcal{C};\mathcal{D})$. Let $\xi_1, \ldots, \xi_h$ be all elements from $\mathcal{D} - \mathcal{C}^+$ that contain at least an S_j ($1 \leq j \leq k$). Let $\varphi_1, \ldots, \varphi_p$ be the elements from left-hand side of the elements ξ_l, $1 \leq l \leq h$. If $GEN(\varphi(S_1, \ldots, S_k)) = \{\psi_1, \ldots, \psi_q\}$, then we have:*

$$\mathcal{C} \models \psi_1 \vee \ldots \vee \psi_q \vee (\forall S_1 \ldots S_k S_{k+1} \ldots S_{k+s})[\varphi_1 \wedge \ldots \wedge \varphi_p \Rightarrow \varphi(S_1, \ldots, S_k)],$$

where S_i, $1 \leq i \leq k+s$ are all S-variables that appear in $\varphi_1, \ldots, \varphi_p$.

PROOF. We have $\{\xi_1, \ldots, \xi_h\} \subseteq \{\psi_1, \ldots, \psi_q\}$. Let I be an index set for relations R_i. Let $M = (r_i, \mathcal{D}_i, i \in I)$ a model that obeys $\mathcal{C}$ and does not satisfy $\psi_1, \ldots \psi_q$. We must show that M satisfies

$$\psi \equiv (\forall S_1 \ldots S_k S_{k+1} \ldots S_{k+s})[\varphi_1 \wedge \ldots \wedge \varphi_p \Rightarrow \varphi(S_1, \ldots, S_k)]$$

Let $M_1 = (r_i, s_i, i = \overline{1, k+s})$ be a model that satisfies φ_j, $1 \leq j \leq p$ and $s_i \in \mathcal{D}_i$, $i = \overline{1, k+s}$.
For two distinct elements from $\mathcal{D} - \mathcal{C}^+$, the set of S-variables are disjoined.
Let $S_{j_1}, \ldots, S_{j_{m_j}}$ the S-variables that appear in ψ_j, $j = \overline{1,q}$. Since M does not satisfy $\psi_1, \ldots, \psi_q$, it results that there is $s_{j_t} \in \mathcal{D}_{j_t}$, for every $t = \overline{1, m_j}$ and $j = \overline{1,q}$, such that $(r_{j_l}, s_{j_l}, l = \overline{1, m_j})$ satisfies the left-hand side from ψ_j, $j = \overline{1,q}$. Let $M' = (r_{j_l}, s_{j_l}, l = \overline{1,q})$.
In M' we consider for a relation symbol R_i a single relation r_i.
If $\xi_i = \psi_{j_i}$ for $i = \overline{1,h}$, then we replace in M', r_{j_l}, s_{j_l}, $l = \overline{1, m_v}$, $v = j_1, j_2, \ldots, j_h$ with $(r_i, s_i, i = \overline{1, k+s})$.
The resulting model will be denoted by M''. We have:
M'' satisfies $LEFT(GEN(\varphi(S_1, \ldots, S_k)))$ and M'' satisfies $\mathcal{C}$, hence by Lemma 5.2 it follows that M'' satisfies $\varphi(S_1, \ldots, S_k)$.
It obtains that M_1 satisfies $\varphi(S_1, \ldots, S_k)$ and this means that M satisfies ψ.
The relation r_i may be empty and $\mathcal{D}_i$ may contain the empty set, for every $i \in I$. □

THEOREM 5.4. *Let $\mathcal{C}$ and $\mathcal{D}$ be two classes of CGEDs, such that $\mathcal{C} \subseteq \mathcal{D}$. Let I an index set for all elements in $\mathcal{D} - \mathcal{C}^+$, that means that $\mathcal{D} - \mathcal{C}^+ = \{\alpha_i | i \in I\}$. Let $\mathcal{C}_i$ be the elements from $\mathcal{C}$ which "appear" in the generation of α_i. Let REL_i be the set of*

R-symbols that belong to $\mathcal{C}_i$ *or* α_i. *We assume that* $\mathcal{C} = \bigcup_{i \in I} \mathcal{C}_i$ *and* $REL_i \cap REL_j = \emptyset$, $i \neq j$, $i, j \in I$. *Let* T *be defined thus:* $T = \bigcup_{i \in I} T(\mathcal{C}_i, \alpha_i)$. *Let* M *be a model defined by:*

$$M \models V(t) \text{ iff } V(t) \in \bigcup_{i \in I} Y(T, \sigma_i)$$

where σ_i *is the right-hand side of* α_i, $i \in I$. *Then if the pair* $(\mathcal{C}, \mathcal{D})$ *has a model Armstrong, then* M *is an Armstrong database.*

PROOF. We have that M satisfies $\mathcal{C}$. Assume that M is not an Armstrong model for $\mathcal{C}$ and $\mathcal{D}$. Then there exists $i \in I$, such that M satisfies α_i. Because the elements from left-hand side of α_i belong to T, it follows that the right-hand side of α_i, namely σ_i belongs to T. By Theorem 5.4, there are $\psi_1, \ldots, \psi_q$ from $\mathcal{D} - \mathcal{C}^+$, such that $\mathcal{C} \models \psi_1 \vee \ldots \vee \psi_q \vee \alpha_i$.

This relation implies that there is no Armstrong database for $\mathcal{C}$ and $\mathcal{D}$. □

6. Conclusions

We defined the family of conditional generalized embedded dependencies. These constraints are expressible as formulas in second-order logic. For a class we gave a characterisation of a given constraint logically implied by a certain subclass of the class. For special classes of conditional generalized embedded dependencies we proved the existence of the Armstrong-like databases.

Finally, we gave a method for the construction of an Armstrong model, when it exists and under certain conditions.

References

Atzeni, P., De Antonellis, V. (1993): *Relational Database Theory*, The Benjamin/ Cummings Publishing Company Inc.

Chang, C.L., Lee, R.C.T. (1973): *Symbolic Logic and Mechanical Theorem Proving*, Academic Press, New York.

De Bra, P., Paredaens, J. (1988): *Conditional dependencies for horizontal decompositions*, LNCS, 326, ICDT '88, pg. 67–82.

Fagin, R. (1982): *Horn Clauses and Databases Dependencies*, JACM, vol. 29, No 4, October 82, pg. 952–985.

Fagin, R. (1982): *Armstrong Databases*, R. Report, R.J.3440, 4/5/82.

Felea, V. (1993): *On the Family of Conditional Implicational Dependencies*, Proceedings of the 9th ROSYCS '93, "Al.I. Cuza" University of Iaşi, pg. 178–190 and Fundamenta Informaticae, vol. 24/3, 1995, pg. 903–912.

Felea, V. (1994): *On the Family of Conditional Generalized Dependencies*, in An. Univ. "Al. I. Cuza" Iaşi.

Felea, V. (1995): *On the Family of Conditional Embedded Implicational Dependencies*, in Computer Science Journal of Moldova, vol. 3, nr. 3.

Grant, J., Jacobs, E.B. (1982): *On the Family of Generalized Dependency Constraints*, JACM, vol. 29, No 4, October 1982, pg. 986–997.

Robbin, W.J. (1969): *Mathematical logic*, W.A Benjamin, Inc, New York, Amsterdam.

Thalheim, B. (1991): *Dependencies in Relational Databases*, B.G. Teubner Verlagsgesellschaft Stuttgart-Leipzig.

Ullman, J.D. (1980): *Principles of Database Systems*, Computer Science Press, Woodland Hills, Calif.

Yasuhara, Ann (1971): *Recursive Function Theory & Logic*, Academic Press, New York and London.

Received February 26, 1996