

On a New Class of Boolean Dependencies

Nguyen Xuan Huy
Institute of Informatics
National Center for Scientific Research of Vietnam

Le Thi Thanh
Center for Applied Mathematics
National Center for Scientific Research of Vietnam

1. Introduction

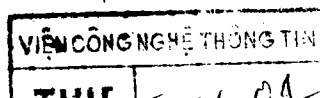
In the theory of relational databases, the connection between functional and multivalued dependencies and a certain fragment of propositional logic has been investigated in several papers. In [11], a family of Boolean dependencies is introduced. In [3,4], a large subclass of positive Boolean dependencies, that is, Boolean combinations of attributes and the logical constant TRUE in which neither negation nor FALSE occur is studied. Boolean dependencies of a special form are investigated in [6,7]. In [5], a class of equational dependencies is introduced. This class includes the class of functional dependencies as well as Boolean dependencies, positive Boolean dependencies and classes of dependencies considered in [6,7,12].

The main result of these aspects is showing equivalence theorem of consequences in the world of all relations, the world of 2-tuple relations and propositional logic. This makes available the familiar tools of truth tables, Karnaugh maps, and syntactic derivations for deciding if a given dependency is a consequence of some set of dependencies.

In the papers mentioned above, the connection between dependencies and fragment of propositional logic is built on the set of truth assignments T_R of a given relation R as follows. For each pair of distinct tuples of R , the set T_R contains the truth assignment that maps an attribute A to TRUE if the two tuples are equal on A , and to FALSE if the two tuples have different values for A .

In this paper we introduce a large class of mappings for constructing the truth assignments of relations. This class includes the equal mappings mentioned above. We call the dependencies constructing on these mappings generalized positive Boolean dependencies (GPBD).

Typeset by $\text{\AA}M\text{\S}-\text{T}\text{E}\text{X}$



The paper is structured as follows. In Section 2 we give some basic definitions. The equivalence theorem for GPBDs is proved in Section 3. Section 4 discusses the membership problem for GPBDs. The update problem and Armstrong relations for GPBDs are studied in Section 5 and Section 6, respectively. Finally, in Section 7 we list some problems of our further research.

2. Basic Definitions

We assume that the reader is familiar with the relational model of database systems and with the basic concepts of relational database theory [9,13]. In this paper we use the following notation.

Let $\mathcal{U} = \{A_1, \dots, A_n\}$ be a set of *attributes*. Corresponding to each attribute A_i is a set d_i , $1 \leq i \leq n$, called the *domain* of A_i . We assume that every d_i contains at least two elements. A *relation* R over \mathcal{U} is a subset of $d_1 \times \dots \times d_n$. Elements of R are called *tuples* and we usually denote by u, v or t . The class of all relations over \mathcal{U} is denoted by \mathcal{R} . For $k \geq 0$, \mathcal{R}_k denotes those relations in \mathcal{R} that have at most k tuples. If $R \in \mathcal{R}$, $t \in R$, $A \in \mathcal{U}$ and $X \subseteq \mathcal{U}$, then we denote by $t[A]$ the value of t for the attribute A , and by $t[X]$ the set $\{t[A] \mid A \in X\}$.

By \mathcal{F} we denote the set of all *formulas* that can be constructed from \mathcal{U} using the logical connectives $\wedge, \vee, \rightarrow, \neg$, and logical constants 1 (TRUE) and 0 (FALSE).

For $X = \{A_{i_1}, \dots, A_{i_k}\} \subseteq \mathcal{U}$, $\wedge X$ denotes the formula $A_{i_1} \wedge \dots \wedge A_{i_k}$, and $\vee X$ denotes the formula $A_{i_1} \vee \dots \vee A_{i_k}$.

Let $\mathcal{B} = \{0, 1\}$. A *valuation* is any function $x : \mathcal{U} \rightarrow \mathcal{B}$. The notation $x = (x_1, \dots, x_n) \in \mathcal{B}^n$ means that $x(A_i) = x_i$, $A_i \in \mathcal{U}$, $1 \leq i \leq n$.

If $f \in \mathcal{F}$ and $x \in \mathcal{B}^n$, then $f(x)$ denotes the truth value of f on the valuation x . For a finite subset Σ of \mathcal{F} and for a valuation x in \mathcal{B}^n , we denote $\Sigma(x) = \wedge\{f(x) \mid f \in \Sigma\}$.

Let f be a formula in \mathcal{F} . We denote $T_f = \{x \in \mathcal{B}^n \mid f(x) = 1\}$. For a subset Σ of \mathcal{F} , we denote $T_\Sigma = \bigcap\{T_f \mid f \in \Sigma\}$. Then $x \in T_\Sigma$ if and only if $(\forall f \in \Sigma)(f(x) = 1)$.

Definition 2.1. Let f and g be two formulas. f *implies* g , written $f \vdash g$, if $T_f \subseteq T_g$. f and g are *equivalent*. $f \equiv g$, if $T_f = T_g$. For $\Sigma, \Gamma \subseteq \mathcal{F}$, $\Sigma \vdash \Gamma$ if $T_\Sigma \subseteq T_\Gamma$, and $\Sigma \equiv \Gamma$ if $T_\Sigma = T_\Gamma$.

Let $e = (1, \dots, 1)$ be the valuation that consists of all 1. A formula f in \mathcal{F} is *positive* if $f(e) = 1$. Let \mathcal{F}_p denote all positive formulas on \mathcal{U} . We know that \mathcal{F}_p is equivalent to the set of all formulas that can be built using the connectives $\wedge, \vee, \rightarrow$ and constant 1 [10].

For each domain d_i , $1 \leq i \leq n$, we consider a mapping $\alpha_i : d_i^2 \rightarrow \mathcal{B}$. We assume that the mappings α_i satisfy the following properties.

- (i) $(\forall a \in d_i)(\alpha_i(a, a) = 1)$,
- (ii) $(\forall a, b \in d_i)(\alpha_i(a, b) = \alpha_i(b, a))$, and
- (iii) $(\exists a, b \in d_i)(\alpha_i(a, b) = 0)$.

Example 2.1. It is easy to see that the equal mappings on d_i ,

$$\alpha_i(a, b) = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{otherwise.} \end{cases}$$

$$a, b \in d_i, 1 \leq i \leq n$$

satisfy the properties (i) – (iii).

Example 2.2. Let $\mathcal{U} = \{A, B, C\}$, where d_A is the set of positive whole numbers, d_B is the set of real numbers and a null-value \perp , and d_C is the set of words w on a nonempty alphabet P , where the length of w is not greater than k , $k \geq 1$. We define the mappings α_A, α_B , and α_C as follows.

$$\alpha_A(a, b) = \begin{cases} 1 & \text{if both } a \text{ and } b \text{ are simultaneously odd or even numbers} \\ 0 & \text{otherwise.} \end{cases}$$

$$\alpha_B(a, b) = \begin{cases} 1 & \text{if both } a \text{ and } b \text{ are simultaneously real or } \perp \\ 0 & \text{otherwise.} \end{cases}$$

$$\alpha_C(a, b) = \begin{cases} 1 & \text{if } a \text{ and } b \text{ have the same length} \\ 0 & \text{otherwise.} \end{cases}$$

It is not hard to verify that the mappings α_A, α_B , and α_C satisfy the properties (i) – (iii).

Let $R \in \mathcal{R}$. For $u, v \in R$ we denote by $\alpha(u, v)$ the valuation

$$(\alpha_1(u[A_1], v[A_1]), \dots, \alpha_n(u[A_n], v[A_n])).$$

Now for $R \in \mathcal{R}$ we denote $T_R = \{\alpha(u, v) \mid u, v \in R\}$. Note that for every u in R , $\alpha(u, u) = e$, so e is in T_R .

Definition 2.2. Elements of \mathcal{F}_p are called *generalized positive Boolean dependencies (GPBD)*.

Definition 2.3. For $R \in \mathcal{R}$ and $f \in \mathcal{F}_p$, we say that R satisfies the GPBD f , written $R(f)$, if $T_R \subseteq T_f$.

Definition 2.4. Let $R \in \mathcal{R}$ and $\Sigma \subseteq \mathcal{F}_p$, we say that R satisfies the set of GPBDs Σ , written $R(\Sigma)$, if $R(f)$ for all $f \in \Sigma$. This is equivalent to $T_R \subseteq T_\Sigma$.

For $\Sigma \subseteq \mathcal{F}_p$ and $f \in \mathcal{F}_p$, $\Sigma \models f$ means that, for all $R \in \mathcal{R}$, if $R(\Sigma)$ then $R(f)$. $\Sigma \models_2 f$ means that, for all $R \in \mathcal{R}_2$, if $R(\Sigma)$ then $R(f)$. In other words, $\Sigma \models f$ if and only if for all $R \in \mathcal{R}$, $T_R \subseteq T_\Sigma$ implies $T_R \subseteq T_f$.

For the equal mappings mentioned in Example 2.1 several classes of Boolean dependencies were investigated. Boolean dependencies were introduced in [11]. Positive Boolean dependencies are studied in [3,4]. Equational dependencies were introduced in [5]. Boolean dependencies of a special form are studied in [6,7]. These papers consider dependencies equivalent to the Boolean dependencies $\wedge X \rightarrow \wedge Y$ (functional dependency), $\wedge X \rightarrow \vee Y$ (weak dependency), $\vee X \rightarrow \wedge Y$ (strong dependency), and $\vee X \rightarrow \vee Y$ (dual dependency). In [4], the authors shown that the consequence relation for positive Boolean dependencies is the same as the consequence relation for propositional logic.

3. Equivalence Theorem

In [4,11] equivalence theorem was proved for the class of positive Boolean dependencies defined on the equal mappings. In this section we give a generalized form of this theorem for the mappings satisfied the properties (i) - (iii).

Theorem 3.1. (Equivalence theorem) *Let $\Sigma \subseteq \mathcal{F}_p$ and $f \in \mathcal{F}_p$. The following are equivalent:*

- (1) $\Sigma \vdash f$.
- (2) $\Sigma \models f$.
- (3) $\Sigma \models_2 f$.

Proof. (1) \Rightarrow (2). If $\Sigma \vdash f$ then $T_\Sigma \subseteq T_f$. Suppose $R \in \mathcal{R}$ and $R(\Sigma)$. Then $T_R \subseteq T_\Sigma$, so by the transitivity, $T_R \subseteq T_f$. It follows that $R(f)$, and hence $\Sigma \models f$.

(2) \Rightarrow (3) is obvious.

(3) \Rightarrow (1). Assume (3) and suppose $x = (x_1, \dots, x_n) \in T_\Sigma$. We have to show $x \in T_f$. For this, let us construct a relation R with two tuples $u = (a_1, \dots, a_n)$ and $v = (b_1, \dots, b_n)$ such that $\alpha_i(a_i, b_i) = x_i$, $1 \leq i \leq n$. The existence of a_i and b_i is guaranteed by properties (i) and (iii). Thus we have $R \in \mathcal{R}_2$ and $T_R = \{e, x\}$. Since $\Sigma \subseteq \mathcal{F}_p$ and $x \in T_\Sigma$, it follows that $T_R \subseteq T_\Sigma$. Hence $R(\Sigma)$. By (3), $R(f)$. Therefore $T_R \subseteq T_f$, so $x \in T_f$ as desired. \square

4. Membership problem for GPBDs

In general, the membership problem can be expressed as follows. Given a set of GPBDs Σ and a GPBD f , decide whether $\Sigma \models f$. In this section we give some necessary and sufficient conditions for the dependencies $\wedge X \rightarrow \wedge Y$, $\wedge X \rightarrow \vee Y$, $\vee X \rightarrow \wedge Y$, $\vee X \rightarrow \vee Y$, and $\wedge X \rightarrow (\wedge Y \vee \wedge Z)$ to be implied from a set of GPBDs Σ . Note that the similar results for equal mappings were presented in [1,2,6,7,8].

From propositional logic we know that $\Sigma \vdash g \rightarrow h$ if and only if for all x in T_Σ , either $g(x) = 0$ or $h(x) = 1$. By an application of this claim we get the following result.

Theorem 4.1. *Let Σ be a set of GPBDs on \mathcal{U} , and $X, Y, Z \subseteq \mathcal{U}$. Then*

1. $\Sigma \models \wedge X \rightarrow \wedge Y \Leftrightarrow (\forall x \in T_\Sigma)((\exists A \in X)(x(A) = 0) \vee ((\forall B \in Y)(x(B) = 1)))$.
2. $\Sigma \models \wedge X \rightarrow \vee Y \Leftrightarrow (\forall x \in T_\Sigma)((\exists A \in X)(x(A) = 0) \vee ((\exists B \in Y)(x(B) = 1)))$.
3. $\Sigma \models \vee X \rightarrow \wedge Y \Leftrightarrow (\forall x \in T_\Sigma)((\forall A \in X)(x(A) = 0) \vee ((\forall B \in Y)(x(B) = 1)))$.
4. $\Sigma \models \vee X \rightarrow \vee Y \Leftrightarrow (\forall x \in T_\Sigma)((\forall A \in X)(x(A) = 0) \vee ((\exists B \in Y)(x(B) = 1)))$.
5. $\Sigma \models \wedge X \rightarrow (\wedge Y \vee \wedge Z) \Leftrightarrow (\forall x \in T_\Sigma)((\exists A \in X)(x(A) = 0) \vee ((\forall B \in Y)(x(B) = 1) \vee ((\forall C \in Z)(x(C) = 1))))$.

For $x = (x_1, \dots, x_n) \in \mathcal{B}^n$ we define $x' = \{A_i \in \mathcal{U} \mid x_i = 1\}$, and for $T \subseteq \mathcal{B}^n$ we define $T' = \{x' \mid x \in T\}$.

Let $X \subseteq \mathcal{U}$ and $y \in \mathcal{B}^n$. From propositional logic we know that $\wedge X(y) = 1$ if and only if $X \subseteq y'$, and $\vee X(y) = 1$ if and only if $X \cap y' \neq \emptyset$. Therefore we get the following result.

Theorem 4.2. *Let Σ be a set of GPBDs. Then*

1. $\Sigma \models \wedge X \rightarrow \wedge Y \Leftrightarrow (\forall E \in T'_\Sigma)(X \subseteq E \Rightarrow Y \subseteq E)$.
2. $\Sigma \models \wedge X \rightarrow \forall Y \Leftrightarrow (\forall E \in T'_\Sigma)(X \subseteq E \Rightarrow Y \cap E \neq \emptyset)$.
3. $\Sigma \models \forall X \rightarrow \wedge Y \Leftrightarrow (\forall E \in T'_\Sigma)(X \cap E \neq \emptyset \Rightarrow Y \subseteq E)$.
4. $\Sigma \models \forall X \rightarrow \forall Y \Leftrightarrow (\forall E \in T'_\Sigma)(X \cap E \neq \emptyset \Rightarrow Y \cap E \neq \emptyset)$.
5. $\Sigma \models \wedge X \rightarrow (\wedge Y \vee \wedge Z) \Leftrightarrow (\forall E \in T'_\Sigma)(X \subseteq E \Rightarrow ((Y \subseteq E) \vee (Z \subseteq E)))$.

Let f be a GPBD and Σ be a set of GPBDs on \mathcal{U} . Following a result of [10] we can assume that f and Σ do not contain the connective \neg . For $X \subseteq \mathcal{U}$, denote by $f \setminus X$ the formula constructed from f by replacing all the occurrences of symbols in X by 1. For Σ we denote $\Sigma \setminus X = \{f \setminus X \mid f \in \Sigma\}$. Note that $\Sigma \models \wedge X \rightarrow \wedge Y$ if and only if $\Sigma \models \wedge X \rightarrow A$ for every $A \in Y - X$. Thus $f \setminus X$ and $\Sigma \setminus X$ can be considered as formulas on $\mathcal{U} - X$.

Theorem 4.3. *Let F be a GPBD set of the form $\wedge Y \rightarrow \wedge Z$, $Y, Z \subseteq \mathcal{U}$, and let $X \subseteq \mathcal{U}$, $A \in \mathcal{U} - X$. Then $F \models \wedge X \rightarrow A$ on \mathcal{U} if and only if $F \setminus X \models 1 \rightarrow A$ on $\mathcal{U} - X$.*

Proof. Let $\mathcal{V} = \mathcal{U} - X$ and $G = F \setminus X$.

If: Assume $G \models 1 \rightarrow A$ on \mathcal{V} and suppose that x is a truth assignment for F on \mathcal{U} such that $F(x) = 1$ and $\wedge X(x) = 1$. Form y from x by removing the values $x(B)$ for all $B \in X$. Clearly, y is an assignment on \mathcal{V} and $G(y) = 1$. Therefore $y(A) = x(A) = 1$, since $G \models 1 \rightarrow A$. It follows that $F \models \wedge X \rightarrow A$.

Only if: Assume $F \models \wedge X \rightarrow A$ and suppose that y is a truth assignment for G on \mathcal{V} such that $G(y) = 1$. Form x from y by adding the values $x(B) = 1$ for all $B \in X$. Clearly, x is the assignment on \mathcal{U} . Moreover, $F(x) = 1$ and $\wedge X(x) = 1$. Since $F \models \wedge X \rightarrow A$, it follows that $y(A) = x(A) = 1$. Therefore $G \models 1 \rightarrow A$. \square

Theorem 4.3 gives a basis for the concept of translations of relation schemes [8]. The main purpose of this concept is to transform a given set of functional dependencies by removing some attributes that seem to be unimportant for computing several objects in the relational model of databases. Note that Theorem 4.3 was proved for a more general case, where F may contain multivalued and join dependencies besides functional dependencies (for the definitions of multivalued and join dependencies see [9,13].)

5. Update Problem

In general, the update problem can be expressed as follows. Given a relation R that satisfies a set of GPBDs Σ . After performing one of the following update operations on R : add a tuple to R , delete a tuple from R , and change the contents of a tuple in R we get a new relation, say R' . Does R' satisfy Σ ?

If R satisfies Σ , then $T_R \subseteq T_\Sigma$. Since the delete operation reduces the number of tuples in R , we have $T_{R'} \subseteq T_R$, so $T_{R'} \subseteq T_\Sigma$, by transitivity. Hence R' satisfies Σ . Thus the delete operation does

not change the semantics of R . The result of the change operation can be obtained with a delete followed by an add. Therefore, our aim will be focussed on the add operation. Let R be a relation that satisfies a set of GPBDs Σ and t be a tuple in $d_1 \times \cdots \times d_n$. We say that t can be added to R if $R \cup \{t\}$ satisfies Σ . We know that R satisfies Σ if and only if $T_R \subseteq T_\Sigma$. Therefore we get the following result.

Theorem 5.1 *Let R be a relation satisfying a set of GPBDs Σ , and let t be a tuple in $d_1 \times \cdots \times d_n$. Then t can be added to R if and only if $(\forall u \in R)(\alpha(t, u) \in T_\Sigma)$.*

6. Armstrong relations

Let Σ and Γ be the sets of GPBDs. Denote by Σ^+ the set $\{f \mid \Sigma \models f\}$. Clearly $\Sigma^+ \supseteq \Sigma$, and $\Sigma \equiv \Gamma$ if and only if $\Sigma^+ = \Gamma^+$. Let R be a relation on \mathcal{U} , denote by $LD(R)$ the set of all GPBDs on \mathcal{U} that hold in R . Clearly $(LD(R))^+ = LD(R)$.

Definition 6.1. Let Σ be a set of GPBDs on \mathcal{U} and let R be a relation on \mathcal{U} . R represents Σ if $LD(R) \supseteq \Sigma^+$. R exactly represents Σ if $LD(R) = \Sigma^+$. If R exactly represents Σ then we also say that R is an *Armstrong relation* for Σ . e also say

In [1,2,8] the structure and size of Armstrong relations are investigated for functional dependencies. The next theorem gives a characterization of Armstrong relations for generalized positive Boolean dependencies.

Theorem 6.1 *Let Σ be a set of GPBDs on \mathcal{U} , and let R be a nonempty relation on \mathcal{U} . Then R is an Armstrong relation for Σ if and only if $T_R = T_\Sigma$.*

Proof. *Only if:* If R is an Armstrong relation for Σ , then by definition 6.1 $LD(R) = \Sigma^+$. It follows that $LD(R)$ and Σ are equivalent. Hence $T_{LD(R)} = T_\Sigma$. But R satisfies $LD(R)$, then $T_R \subseteq T_{LD(R)}$. Therefore, by the transitivity, $T_R \subseteq T_\Sigma$.

Suppose that $x \in T_\Sigma$ but $x \notin T_R$. From the theory of Boolean functions we know that there is a formula f on \mathcal{U} with $T_f = T_R$. Note that $e \in T_R$, since $R \neq \emptyset$. It follows that f is a positive Boolean formula. Since $T_R = T_f$, R satisfies f , so $f \in LD(R)$. Therefore $f \in \Sigma^+$, since $LD(R) = \Sigma^+$. Since $x \in T_\Sigma$ and $\Sigma \vdash f$, it follows that $f(x) = 1$. Hence $x \in T_f$, and so $x \in T_R$; a contradiction.

If: Assume $T_R = T_\Sigma$. Then R satisfies Σ . Since $\Sigma \vdash \Sigma^+$, R satisfies Σ^+ , so $LD(R) \supseteq \Sigma^+$. If $f \in LD(R)$, then R satisfies f . It follows that $T_R \subseteq T_f$. Since $T_R = T_\Sigma$, $T_\Sigma \subseteq T_f$. Hence $\Sigma \vdash f$, and so $f \in \Sigma^+$. \square

7. Conclusion

Our further research will be dedicated to the following problems.

Let Σ be a set of GPBDs on a set of attributes \mathcal{U} .

1. Construct an Armstrong relation for Σ .
2. What is about the size (number of tuples) of minimal Armstrong relations for Σ .

References

1. Armstrong W.W., *Dependency Structures of Database Relationships*. Proc. IFIP 74, North Holland, Amsterdam, 1974, 580-583.
2. Beeri C., Dowd M., Fagin R. & Statman R., *On the Structure of Armstrong Relations for Functional Dependencies*, J. ACM, **31** (1984), 1, 30-46.
3. Berman J. & Blok W.J., *Generalized Boolean dependencies*, Abstracts of AMS, **6** (1985), 163.
4. Berman J. & Blok W.J., *Positive Boolean dependencies*, Inf. Processing Letters, **27** (1988), 147-150.
5. Berman J. & Blok W.J., *Equational dependencies*, Manuscript (1990).
6. Czedli G., *On dependencies in the relational model of data*, J. EIK **17** (1981), 103-112.
7. Demetrovics J. & Gyepesi Gy., *Some generalized type functional dependencies formalized as equality set on matrices*, Discrete Applied Math. **6** (1983), 35-47.
8. Demetrovics J. & Huy N.X., *Closed sets and translations of relation schemes*, Computers Math. Applic. **21** (1991), 13-23.
9. Maier D., *The Theory of Relational Databases*. Computer Science Press, Rockville, Md., 1983.
10. Post E.L., *The two-valued Interactive Systems of Mathematical Logic*, Annals of Math. Studies, **5** (1941) Princeton University.
11. Sagiv Y., Delobel C., Parker D.S., & Fagin R., *An Equivalence Between Relational Database Dependencies and a Fragment of Propositional Logic*, J. ACM **28** (1981), 435-453. Also a correction to this paper in J. ACM **34** (1987), 1016-1018.
12. Thalhei B., *Funktionale Abhängigkeiten in relationalen Datenstrukturen*, J. EIK **21** (1985), 23-33.
13. Ullman J.D., *Principles of Database Systems*, (Second Edition.) Computer Science Press, Potomac, Md., 1982.

Abstract

A class of generalized positive Boolean dependencies (GPBD) is introduced. The membership problem, update problem and Armstrong relations for a given set of GPBDs are investigated.