Information Processing Letters 22 (1986) 281–283
North-Holland
30 May 1986

A SIMPLE CHARACTERIZATION OF DATABASE DEPENDENCY IMPLICATION

Yoshito HANATANI

Université de Paris-Sud, Centre d'Orsay, Laboratoire de Recherche en Informatique, 91405 Orsay Cedex, France

Ronald FAGIN

IBM Research Laboratory, San Jose, CA 95193, U.S.A.

Communicated by M.A. Harrison
Received 24 July 1985
Revised 20 August 1985

A simple and elegant set-theoretic characterization is given as to when a given set of functional and multivalued dependencies logically implies a given functional or multivalued dependency. A simple proof of the characterization is given which makes use of a result of Sagiv, Delobel, Parker, and Fagin (1981).

Keywords: Relational database, functional dependency, multivalued dependency, logical implication

1. Introduction

Functional dependencies [4] and multivalued dependencies [5,11] are important, widely-studied constraints on databases. A fundamental issue in database theory is the question of determining whether or not a given set of dependencies (in our case, functional and multivalued dependencies) logically implies another given dependency. We provide a simple and elegant set-theoretic characterization for dependency implication. The characterization has found several applications by the first author [6–9]. By making use of a theorem of Sagiv et al. [10], we give a simple proof of the characterization.

2. Basic definitions

We assume a fixed finite set U of *attributes* (which, intuitively, are the names of columns of a relation). A *tuple* is a mapping which associates with each attribute a value, and a *relation* is a set of tuples. If t is a tuple and $X \subseteq U$ is a set of attributes, then the *projection* of t onto X, written t[X], is the restriction of the mapping t to X.

A *functional dependency*, or FD, is a sentence of the form $X \to Y$, where both X and Y are sets of attributes. A relation r *satisfies* the FD $X \to Y$ if for every pair t_1, t_2 of tuples of r, whenever $t_1[X] = t_2[X]$, then $t_1[Y] = t_2[Y]$. In other words, the FD $X \to Y$ means that the Y values are determined by the X values. A *multivalued dependency*, or MVD, is a sentence of the form $X \twoheadrightarrow Y|Z$, where X, Y, and Z are sets of attributes, and $X \cup Y \cup Z = U$.[1] A relation r *satisfies* the MVD $X \twoheadrightarrow Y|Z$ if for every pair t_1, t_2 of tuples of r, whenever $t_1[X] = t_2[X]$, then there is a tuple t_3 of r such that (i) $t_3[X] = t_1[X]$, (ii) $t_3[Y] = t_1[Y]$, and (iii) $t_3[Z] = t_2[Z]$. In other words, the MVD $X \twoheadrightarrow Y|Z$ means that the set of Y values associated with a particular X value must be independent of the values of the rest of the attributes. If Σ is a set

[1] Normally, an MVD is simply written as $X \twoheadrightarrow Y$, where Z is implicitly assumed to be $U - (X \cup Y)$. However, for our purposes it is convenient to explicitly mention Z.

of dependencies, then we say that relation r satisfies Σ if r satisfies every member of Σ.

Let Σ be a set of dependencies, and let σ be a single dependency. When we say that Σ *logically implies* σ, we mean that whenever a relation r satisfies Σ, then also r satisfies σ. That is, there is no 'counterexample relation' or 'witness' r such that r satisfies Σ but not σ. We write $\Sigma \models \sigma$ to mean that Σ logically implies σ, and $\Sigma \not\models \sigma$ to mean that Σ does not logically imply σ.

3. Main theorem

We say that a set S of attributes is *closed with respect to the FD* $X \to Y$ if $X \subseteq S$ implies that $Y \subseteq S$ (that is, either $X \not\subseteq S$ or $Y \subseteq S$). We say that a set S of attributes is *closed with respect to the MVD* $X \twoheadrightarrow Y | Z$ if $X \subseteq S$ implies that either $Y \subseteq S$ or $Z \subseteq S$. If Σ is a set of dependencies, then we say that the set S of attributes is *closed with respect to* Σ if it is closed with respect to every member of Σ. In the case of FDs alone, Armstrong [1] called a closed set *saturated*. We get our terminology 'closed' from Beeri et al. [3], again in the case of FDs alone. Armstrong and Delobel [2] call a closed set S such that there are at least two attributes not in S an *antiroot*.

Remark. There is another way to view closed sets. For each attribute $A \in U$, let us define a distinct propositional variable $\hat{A}$. As in [3], to each FD $X \to Y$, we associate the propositional formula

$$(\wedge\{\hat{A} : A \in X\}) \Rightarrow (\wedge\{\hat{A} : A \in Y\}),$$

and to each MVD $X \twoheadrightarrow Y | Z$, we associate the propositional formula

$$(\wedge\{\hat{A} : A \in X\}) \Rightarrow ((\wedge\{\hat{A} : A \in Y\}) \vee (\wedge\{\hat{A} : A \in Z\})).$$

Let us denote by $\hat{\sigma}$ the propositional formula corresponding to the dependency σ as described above. An *atom* is a propositional formula of the form

$$(\wedge\{\hat{A} : A \in S\}) \wedge (\wedge\{\sim \hat{A} : A \in T\}),$$

where S and T are disjoint and $S \cup T = U$; we say that S is the *positive part* of the atom. The *complete disjunctive normal form* of a propositional formula φ is the (unique) disjunction of atoms such that the disjunction is logically equivalent to φ; each atom in the disjunction is called an *atom of* φ. Then the closed sets with respect to the set Σ of dependencies are precisely the positive parts of the atoms of the formula $\wedge\{\hat{\sigma} : \sigma \in \Sigma\}$.

Our main result is as follows.

Theorem 3.1. *Let Σ be a set of dependencies (FDs and/or MVDs), and let σ be a single dependency. Then $\Sigma \models \sigma$ if and only if every set S of attributes that is closed with respect to Σ is also closed with respect to σ.*

Our proof makes crucial use of the following result of Sagiv et al. [10], where a *two-tuple relation* is a relation with exactly two distinct tuples.

Theorem 3.2 ([10]). *Let Σ be a set of dependencies (FDs and/or MVDs), and let σ be a single dependency. Then $\Sigma \models \sigma$ if and only if every two-tuple relation that satisfies Σ also satisfies σ.*

We note that Theorem 3.2 is obvious if we restrict our attention to FDs, but is not at all clear when MVDs are also allowed.

Before proving our main theorem, we need another definition and a lemma. Let r be a two-tuple relation with tuples t_1 and t_2. The *agreement set* of r is the set of attributes where the tuples agree, that is, $\{A : t_1[A] = t_2[A]\}$.

Lemma 3.3. *Let r be a two-tuple relation with agreement set S. Then r satisfies an FD or MVD τ if and only if S is closed with respect to τ.*

Proof. *Case* 1: *τ is an FD* $X \to Y$. If $X \not\subseteq S$, then r satisfies τ, and S is closed with respect to τ. So assume that $X \subseteq S$. It is clear that r satisfies τ if and only if S is closed with respect to τ.

Case 2: *τ is an MVD* $X \twoheadrightarrow Y$. It is easy to verify that since r is a two-tuple relation, r satisfies the MVD $X \twoheadrightarrow Y$ if and only if either r satisfies the FD $X \to Y$, or r satisfies the FD $X \to Z$. Case 2 then easily follows from Case 1. □

Proof of Theorem 3.1. ($\Rightarrow$): Assume that there is a set S of attributes which is closed with respect to Σ, but which is not closed with respect to σ. Let r be a two-tuple relation with agreement set S (for example, let the first tuple of r consist of all 0's, and let the second tuple have 0's in the attributes S, and 1's elsewhere). By Lemma 3.3 we know that r satisfies Σ but not σ. Therefore, $\Sigma \not\models \sigma$.

($\Leftarrow$): Assume that every set S of attributes that is closed with respect to Σ is also closed with respect to σ. Let r be an arbitrary two-tuple relation that satisfies Σ. Let S be the agreement set of r. By Lemma 3.3, it follows that S is closed with respect to Σ. So, by assumption, S is closed with respect to σ. By Lemma 3.3 again, it follows that r satisfies σ. We have just shown that every two-tuple relation that satisfies Σ also satisfies σ. By Theorem 3.2, this implies that $\Sigma \models \sigma$. □

4. Historical remarks

Armstrong [1] first proved our main theorem in the special case where only FDs are allowed. Armstrong and Delobel [2] obtained results that are related to our main theorem, but their results do not seem to directly imply ours. The main theorem of this paper was first given by the first author in [7], but the proof was quite elaborate, and much longer than the current proof.

Acknowledgment

We would like to thank Nicolas Spyratos and Moshe Vardi for helpful suggestions.

References

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

[2] W.W. Armstrong and C. Delobel, Decompositions and functional dependencies in relations, ACM Trans. Database Systems 5 (4) (1980) 404–430.

[3] C. Beeri, M. Dowd, R. Fagin and R. Statman, On the structure of Armstrong relations for functional dependencies, J. ACM 31 (1) (1984) 30–46.

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

[5] R. Fagin, Multivalued dependencies and a new normal form for relational databases, ACM Trans. Database Systems 2 (3) (1977) 262–278.

[6] Y. Hanatani, Graph-theoretical classification and generation of JD's by means of bases for MVD-structures, Rapport de Recherche INRIA No. 170, 1982.

[7] Y. Hanatani, Bases for FD–MVD-structures, Rapport de Recherche INRIA No. 236, 1983.

[8] Y. Hanatani, Eliminating cycles in data base schemas, in: H. Gallaire, J. Minker and J.-M. Nicolas, eds., Advances: Data Base Theory, Vol. 2 (Plenum Press, New York, 1984).

[9] Y. Hanatani, Join-decomposition of MVD-constraints using the characterization by "bases", Proc. Internat. Conf. on Foundations on Data Organization, Kyoto Univ., 1985.

[10] Y. Sagiv, C. Delobel, D.S. Parker, Jr., and R. Fagin, An equivalence between relational database dependencies and a fragment of propositional logic, J. ACM 28 (3) (1981) 435–453.

[11] C. Zaniolo, Analysis and design of relational schemata for database systems, Ph.D. Dissertation, Tech. Rept. UCLA-ENG-7669, UCLA, 1976.