Characterization and Armstrong Relations for Degenerate Multivalued Dependencies Using Formal Concept Analysis*

Jaume Baixeries and José Luis Balcázar

Dept. Llenguatges i Sistemes Informàtics,
Universitat Politècnica de Catalunya,
c/ Jordi Girona, 1-3,
08034 Barcelona
{jbaixer, balqui}@lsi.upc.es

Abstract. Functional dependencies, a notion originated in Relational Database Theory, are known to admit interesting characterizations in terms of Formal Concept Analysis. In database terms, two successive, natural extensions of the notion of functional dependency are the so-called degenerate multivalued dependencies, and multivalued dependencies proper. We propose here a new Galois connection, based on any given relation, which gives rise to a formal concept lattice corresponding precisely to the degenerate multivalued dependencies that hold in the relation given. The general form of the construction departs significantly from the most usual way of handling functional dependencies. Then, we extend our approach so as to extract Armstrong relations for the degenerate multivalued dependencies from the concept lattice obtained; the proof of the correctness of this construction is nontrivial.

1 Introduction

It is well-known [19] that, from the Concept Lattice associated to a given binary relation, one can extract a number of implications that hold in the relation, for instance via the Duquenne-Guigues basis or, alternatively, by using minimal hypergraph transversals of the predecessors of each closed set ([26], [27]). Actually, the implications obtained in that way are known to characterize a Horn axiomatization of the given relation if one exists (otherwise, they provide a well-defined Horn approximation to the data; see [3], where a complete proof may be found).

Moreover, in [19] we also find an interesting application to database theory, since the syntactical similarity between implications and functional dependencies is more than a mere syntactical similarity; and there is a precise method (that we will call here "comparison-based binarization") to change a given database

* This work is supported in part by MCYT TIC 2002-04019-C03-01 (MOISES) and by the PASCAL-NETWORK Project.

B. Ganter and R. Godin (Eds.): ICFCA 2005, LNAI 3403, pp. 162–175, 2005.

relation r into a binary relation, or scaling, whose implications (its Horn axiomatization) provide exactly the functional dependencies that hold in r. Specifically, for each pair of tuples in r the values for each attribute are compared so as to yield a binary result, and therefore a binary relation (of quadratic size) is obtained.

There are other forms of dependencies in database theory, and we consider it interesting to find methods to obtain or characterize them on the basis of Formal Concept Analysis (FCA). Indeed, we have seen that this is possible for some of them, but the task turns out to be far from trivial. In [4] we have developed a careful semantic study of the relations or propositional theories where formulas of these sorts do hold: namely, multivalued dependencies, degenerate multivalued dependencies, and a family of propositional formulas introduced by Sagiv [28] that parallel them in the same sense as Horn clauses parallel functional dependencies. There we have identified precise meet operators, that is, various forms of combining objects to replace the standard intersections in the closure property, that characterize semantically these dependencies; but these do not readily give as yet a formal concept lattice. Here we consider instead an alternative approach based on defining Galois connections on classes, or partitions, of tuples and of attributes.

Along this line, in [1] we actually demonstrate how to define a Galois connection between attributes and partitions of tuples, inspired by the algorithms in [21] for computing functional dependencies, and thus propose a closure operator giving another precise characterization of functional dependencies in terms of FCA; and our recent, still unpublished work [2] proves that this approach, generalized to handle partitions of attributes instead of single attributes, actually does handle adequately multivalued dependencies.

Associated to each sort of dependency in databases is a notion of Armstrong relation: for a given set of rules, according to whatever syntax the sort of dependency considered allows for, an Armstrong relation is a relation that obeys that set of rules, and of course their logical consequences, but absolutely no other rule. They are useful in database design since they exemplify a set of dependencies, for instance those designed for a database schema prior to actually populating the database with tuples, and the analysis of such an Armstrong relation (seen as a potential database that the rules do allow, but where no other rules hold) is valuous for the prior understanding of the schema designed. Indeed, if the schema should have included a dependency that was forgotten, or expected to follow from the others, but actually does not, the Armstrong relation will point it out by including tuples that violate it, and this is usually a good method for the database engineer to realize the omission.

Our goal in this paper is twofold: first, we complete the existing characterizations by exhibiting an FCA-based characterization of degenerate multivalued dependencies, that did not follow from the previous study; and then, taking advantage of the crisper semantics of these dependencies in comparison with multivalued dependencies proper, we set out to explore the possibility of using FCA methods to construct appropriate Armstrong relations. For the case