Chapter XXIV
Incremental Discovery of Fuzzy Functional Dependencies

Shyue-Liang Wang
New York Institute of Technology, USA

Ju-Wen Shen
Chunghwa Telecom Lab, Taiwan

Tuzng-Pei Hong
National University of Kaohsiung, Taiwan

ABSTRACT

Mining functional dependencies (FDs) from databases has been identified as an important database analysis technique. It has received considerable research interest in recent years. However, most current data mining techniques for determining functional dependencies deal only with crisp databases. Although various forms of fuzzy functional dependencies (FFDs) have been proposed for fuzzy databases, they emphasized conceptual viewpoints and only a few mining algorithms are given. In this research, we propose methods to validate and incrementally search for FFDs from similarity-based fuzzy relational databases. For a given pair of attributes, the validation of FFDs is based on fuzzy projection and fuzzy selection operations. In addition, the property that FFDs are monotonic in the sense that \( r_1 \subseteq r_2 \) implies \( FD_a(r_1) \supseteq FD_a(r_2) \) is shown. An incremental search algorithm for FFDs based on this property is then presented. Experimental results showing the behavior of the search algorithm are discussed.

INTRODUCTION

Functional dependencies (FDs) are relationships among attributes of a database relation. A functional dependency states that the value of an attribute is uniquely determined by the values of some other attributes. It is often used as a guideline for the design of relational schemas that are conceptu-
ally meaningful and free of update anomalies. It therefore plays a crucial role in enforcing integrity constraints for database design. Due to its importance in database technology, extensions and variations of functional dependency have attracted much attention.

To extend functional dependency, there are many approximate dependencies proposed and defined on the crisp relational data model. For example, Haux and Eckert (1985) have extended functional dependency to probabilistic dependency. Saharia and Barron (1995) have extended functional dependency to cluster dependency. To extend functional dependency to fuzzy databases, there are different forms of fuzzy functional dependency on various types of fuzzy relational data models (Galindo, Urrutia, & Piattini, 2006). For example, Raju and Majumdar (1998), Bosc, Lietard, and Pivert (1997), and Chen (1998) have defined various forms of fuzzy functional dependencies on the possibility-based fuzzy relational data model. For the similarity-based fuzzy relational data model, Buckles and Petry (1982), Sachar (1986), Yazici and George (1999), and Yazici, Gocmen, Buckles, George, and Petry (1999) have defined a fuzzy functional dependency based on conformance. Wang, Shen, and Hong (2000) and Wang, Tsai, and Hong (2001) proposed a form of fuzzy functional dependency based on equivalence classes induced by level values on given similarity relations. For equivalence-class-based fuzzy relational data models, Shenoi, Melton, and Fan (1990) defined a fuzzy functional dependency based on redundancy. These fuzzy functional dependencies are basically intended to preserve the functionality of classical functional dependency on the new fuzzy relational data models so that semantics between fuzzy attributes can be captured.

Mining functional dependencies from relational databases, also called dependency discovery, has been identified as an important database analysis technique. It has received considerable research interest in recent years. For example, from a database of chemical compounds, it is valuable to discover compounds that are functionally dependent on a certain structure attribute (Huhtala, Karkkainen, Porkka, & Toivonen, 1999). In addition, as a kind of data dependency, a large data set can be losslessly decomposed into a set of smaller data sets using the discovered functional dependencies.

To find all functional dependencies from a given database relation \( r \), we need to search for all possible functional dependencies and test their validity. There are essentially two approaches for searching for all possible functional dependencies: top-down searching and bottom-up searching. The top-down approach starts with the set of trivial functional dependencies and adds those functional dependencies that are satisfied by \( r \). It can be further classified into depth-first search and level-wise (breadth-first) search. The dependency discovery techniques proposed in Bell and Brockhausen (1995), Huhtala et al. (1999), Lopes, Petit, and Lakhal (2000), Mannila and Rääihä (1994), Mannila and Toivonen (1997), Schlimmer (1993), Wyss, Gianella, and Robertson (2001), and Yao, Hamilton, and Butz (2002) are top-down approaches. The bottom-up approach starts with the set of all possible functional dependencies and removes those functional dependencies that are contradicted by \( r \). The discovery techniques proposed in Flach (1990), Flach and Savnik (1999), and Mannila and Rääihä (1994) are bottom-up approaches. To test the validity of a given possible functional dependency, the pairwise comparison method between any two tuples, decision-tree method, and tuple-partition method have been proposed. However, current mining techniques of determining functional dependencies deal only with crisp databases. Although various forms of fuzzy functional dependencies have been proposed for fuzzy databases, they emphasized conceptual viewpoints, and few mining algorithms are given.

Mining extended functional dependencies has been studied in recent years. Currently, most of the discovered dependencies are used for answering imprecise or fuzzy queries. For example, Bosc, Lietard, and Pivert (1998), Bosc, Lietard, and Ughetto (1999), and Cubero, Medina, Pons, and Miranda (1999) proposed mining extended
Related Content

Business Data Warehouse: The Case of Wal-Mart
Indranil Bose (2009). Selected Readings on Database Technologies and Applications (pp. 112-133).
www.igi-global.com/chapter/business-data-warehouse/28549?camid=4v1a

Conceptual Modeling for XML: A Myth or a Reality
www.igi-global.com/chapter/conceptual-modeling-xml/28551?camid=4v1a

INDUSTRY AND PRACTICE: Why Not SISP Too?
www.igi-global.com/article/indstry-practice-not-sisp-too/51167?camid=4v1a

UB2SQL: A Tool for Building Database Applications Using UML and B Formal Method
www.igi-global.com/article/ub2sql-tool-building-database-applications/3363?camid=4v1a