

Chapter 3

Dynamic Discovery of Fuzzy Functional Dependencies Using Partitions

Shyue-Liang Wang

National University of Kaohsiung, Taiwan

Ju-Wen Shen

Chunghwa Telecom Co., Ltd., Taiwan

Tuzng-Pei Hong

National University of Kaohsiung, Taiwan

ABSTRACT

Discovery of functional dependencies (FDs) from relational databases has been identified as an important database analysis technique. Various mining techniques have been proposed in recent years to deal with crisp and static data. However, few have emphasized on fuzzy data and also considered the dynamic nature that data may change all the time. In this work, the authors propose a partition-based incremental data mining algorithm to discover fuzzy functional dependencies from similarity-based fuzzy relational databases when new sets of tuples are added. Based on the concept of tuple partitions and the monotonicity of fuzzy functional dependencies, we avoid re-scanning of the database and thereby reduce computation time. An example demonstrating the proposed algorithm is given. Computational complexity of the proposed algorithm is analyzed. Comparison with pair-wise comparison-based incremental mining algorithm (Wang, Shen & Hong, 2000) is presented. It is shown that with certain space requirement, partition-based approach is more time efficient than pair-wise approach in the discovery of fuzzy functional dependencies dynamically.

INTRODUCTION

A functional dependency describes the relationship between attributes in a database relation. It states that the value of an attribute is uniquely determined

by the values of some other attributes. It serves as a constraint between the attributes and is being used in the normalization process of relational database design. Therefore the discovery of functional dependencies from databases has been identified as an important technique for analyzing attribute

DOI: 10.4018/978-1-61520-757-2.ch003

semantics and redesign of relational schemas. It has received considerable research interest in recent years. For examples, 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, 1998). In addition, as a kind of data dependency, a large dataset can be losslessly decomposed into a set of smaller datasets using the discovered functional dependencies.

To find all functional dependencies from a given database relation r , we need to search for all possible 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 dependencies discovery techniques proposed in Bell et al. (1995), Huhtala et al. (1998), Mannila et al. (1997), Yao et al. (2002) are top-down approach. 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 & Savnik (1999), Mannila et al. (1997) are bottom-up approach. To test the validity of a given dependency, pair wise comparison method between any two tuples, decision-tree method and a tuple-partition method have been proposed. However, current data 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.

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. For example, Raju and Majumdar (1988), Bosc et al. (1997), Chen (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 et al. (1982), Sachar (1985), and Yazici et al. (1999) have defined a fuzzy functional dependency based on conformance. Wang et al. (2001) proposed a form of fuzzy functional dependency based on equivalence classes induced by level values on given similarity relations. Belohlavek et al. (2006) proposed a fuzzy functional dependency using fuzzy predicate logic on ranked database relations over domains with similarities. For equivalence-class-based fuzzy relational data model, Shenoit et al. (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 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 et al. (1997) proposed mining extended functional dependencies for linguistic summaries of data content and fuzzy-rule-based functional dependency as knowledge from databases. Rasmussen et al. (1999) proposed mining techniques of fuzzy functional dependency using SummarySQL for linguistic summaries of database. Nambiar et al. (2004) proposed mining approximate functional dependencies to answer imprecise queries. Carrasco et al. (2000) and Galindo (2006) proposed FSQL for obtaining fuzzy dependencies and data summarization. However, similar to the application

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/dynamic-discovery-fuzzy-functional-dependencies/42355

Related Content

Generational Differences Relative to Data-Based Wisdom

Gino Maxiand Deanna Klein (2016). *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 1613-1627).

www.irma-international.org/chapter/generational-differences-relative-to-data-based-wisdom/150232

Application of Machine Training Methods to Design of New Inorganic Compounds

Nadezhda Kiselyova, Andrey Stolyarenko, Vladimir Ryazanov, Oleg Sen'koand Alexandr Dokukin (2013). *Diagnostic Test Approaches to Machine Learning and Commonsense Reasoning Systems* (pp. 197-220).

www.irma-international.org/chapter/application-machine-training-methods-design/69410

Mining Top-k Regular High-Utility Itemsets in Transactional Databases

P. Lalitha Kumari, S. G. Sanjeeviand T.V. Madhusudhana Rao (2019). *International Journal of Data Warehousing and Mining* (pp. 58-79).

www.irma-international.org/article/mining-top-k-regular-high-utility-itemsets-in-transactional-databases/223137

SLOD-BI: An Open Data Infrastructure for Enabling Social Business Intelligence

Rafael Berlanga, Lisette García-Moya, Victoria Nebot, María José Aramburu, Ismael Sanzand Dolores María Llidó (2015). *International Journal of Data Warehousing and Mining* (pp. 1-28).

www.irma-international.org/article/slod-bi/130664

Knowledge Exchange in Organizations is a Potential, Not a Given: Methodologies for Assessment and Management of a Knowledge-Sharing Culture

Richard E. Potterand Pierre A. Balthazard (2004). *Organizational Data Mining: Leveraging Enterprise Data Resources for Optimal Performance* (pp. 79-91).

www.irma-international.org/chapter/knowledge-exchange-organizations-potential-not/27909