

Chapter 83

Comparison of Linguistic Summaries and Fuzzy Functional Dependencies Related to Data Mining

Miroslav Hudec

University of Economics in Bratislava, Slovakia

Miljan Vučetić

University of Belgrade, Serbia

Mirko Vujošević

University of Belgrade, Serbia

ABSTRACT

Data mining methods based on fuzzy logic have been developed recently and have become an increasingly important research area. In this chapter, the authors examine possibilities for discovering potentially useful knowledge from relational database by integrating fuzzy functional dependencies and linguistic summaries. Both methods use fuzzy logic tools for data analysis, acquiring, and representation of expert knowledge. Fuzzy functional dependencies could detect whether dependency between two examined attributes in the whole database exists. If dependency exists only between parts of examined attributes' domains, fuzzy functional dependencies cannot detect its characters. Linguistic summaries are a convenient method for revealing this kind of dependency. Using fuzzy functional dependencies and linguistic summaries in a complementary way could mine valuable information from relational databases. Mining intensities of dependencies between database attributes could support decision making, reduce the number of attributes in databases, and estimate missing values. The proposed approach is evaluated with case studies using real data from the official statistics. Strengths and weaknesses of the described methods are discussed. At the end of the chapter, topics for further research activities are outlined.

DOI: 10.4018/978-1-4666-7456-1.ch083

INTRODUCTION

The increasing use of information systems by business and governmental agencies has created large amounts of data that contain potentially valuable knowledge (Rasmussen & Yager, 1997). This amount of data should be processed and interpreted to be useful. Modern corporations are becoming more and more dependent on information generated from databases (Vucetic, Hudec, & Vujošević, 2013). Accordingly, decision makers are often not interested in large sheets of figures, but in relational knowledge that is usually overshadowed by large amount of data in relational databases. Hence, data with meaning are more important than pure data. Traditional knowledge discovery in a database provides precise information from the data rather than providing a global review of the whole database. Therefore, it is important to develop methods able to handle imprecision, uncertainty and partial truth, and present revealed information in an understandable way to users.

In this chapter we use fuzzy concept to induce associational rules hidden in the data. Discovering these useful rules from databases is seen as a data mining technique. The fuzzy sets and fuzzy logic are used to find association among attributes of relational database and the character of the discovered dependencies. Approaches presented in this chapter are bio-inspired techniques because of similarity of the fuzzy concept to human reasoning (computing with words instead of crisp numbers and precise measurement). More recently, these techniques are increasingly present as suitable methods for mining knowledge from different kinds of databases.

Dependencies and relations between attributes convey relevant information for users. It is obvious that they may also exist between particular parts of attributes domains which could not have clear boundaries. Feil and Abonyi (2008) pointed out that there is an urgent need for a new generation of computational techniques to assist humans in extracting useful information (knowledge) from

the constantly growing volumes of collected data. Mining this information also requires approaches which enable ambiguity and imprecision in data to be easily handled (Ansari, Biswas, & Aggarwal, 2012).

Initially, Linguistic Summaries (LSs) have been developed to express a relational, concise and easily understandable knowledge about the data (Rasmussen & Yager, 1997). LSs mimic human reasoning in looking for the information by natural language questions and processing data without precise measurements. The concept of LSs has been initially introduced in Yager (1982) and further developed in Hudec (2013b), Kacprzyk and Yager (2001), Kacprzyk and Zadrozny (2009), Rasmussen and Yager (1997), Yager (1989).

Our research uses the LS of the following structure:

Q R entities in database are (have) S

where S is a summarizer defined as a linguistic term on the domain of the examined attribute, R is a linguistic term adding some constraints and Q is a fuzzy quantifier as in Zadeh (1983). The truth value of a summary is usually called *validity* and gets value from the $[0, 1]$ interval. For example, the rule: *most of well-paid employees are middle aged*, may have a higher truth value than other rules describing a particular company and its employees.

Data summarization is one of the basic capabilities of any “intelligent” system (Kacprzyk & Zadrozny, 2009). We could say that the same holds for the Fuzzy Functional Dependencies (FFDs). The main aim of FFDs is detecting attributes that have a high value of dependency.

A fuzzy functional dependency, denoted by $X \overset{\theta}{\rightarrow} Y$ expresses that a relation exists between the two sets of attributes X and Y . It can be stated as follows: if t and t' have similar values on X , they also have the similar value on Y with linguistic strengths θ (Sözat & Yazici, 2001).

26 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/comparison-of-linguistic-summaries-and-fuzzy-functional-dependencies-related-to-data-mining/124580

Related Content

Comparing the Behaviour of Two Topic-Modelling Algorithms in COVID-19 Vaccination Tweets: LDA vs. LSA

Jordan Thomas Bignell, Georgios Chantziplakis and Alireza Daneshkhah (2022). *International Journal of Strategic Engineering* (pp. 1-20).

www.irma-international.org/article/comparing-the-behaviour-of-two-topic-modelling-algorithms-in-covid-19-vaccination-tweets/292445

Citation Essentials in Scholarly Communication

S. Sudarshan Rao (2017). *Scholarly Communication and the Publish or Perish Pressures of Academia* (pp. 11-19).

www.irma-international.org/chapter/citation-essentials-in-scholarly-communication/169455

Hole Drilling Route Optimization in Printed Circuit Boards Using Far-to-Near Metaheuristics: Optimizing the Hole Drilling Route via Far-to-Near Metaheuristic

Souhail Dhoub (2022). *International Journal of Strategic Engineering* (pp. 1-12).

www.irma-international.org/article/hole-drilling-route-optimization-in-printed-circuit-boards-using-far-to-near-metaheuristics/301568

Factor Analysis of the Sydney Metro's ITS: System Integration Iteration

Koorosh Gharehbaghi, Kenneth Farnes, Kathryn M. Robson and Neville Hurst (2021). *International Journal of Strategic Engineering* (pp. 1-13).

www.irma-international.org/article/factor-analysis-of-the-sydney-metros-its/279642

An Integrated Heuristic for Machine Sequencing With Specific Reference to the Permutation Flow-Shop Scheduling Problem

Kaveh Sheibani (2019). *International Journal of Strategic Engineering* (pp. 1-8).

www.irma-international.org/article/an-integrated-heuristic-for-machine-sequencing-with-specific-reference-to-the-permutation-flow-shop-scheduling-problem/230933