Reengineering Probabilistic Relational Databases with Fuzzy Probability Measures into XML Model

Zongmin Ma, College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing, China
Chengwei Li, College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China
Li Yan, College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China

ABSTRACT

This paper concentrates on modeling probabilistic events with fuzzy probability measures in relational databases and XML (Extensible Markup Language). Instead of crisp probability degrees or interval probability degrees, fuzzy sets are applied to represent imprecise probability degrees in relational databases and XML. A probabilistic XML model with fuzzy probability measures is introduced, which incorporates fuzzy probability measures to handle imprecision and uncertainty. In particular, the formal approach to reengineering the relational database model with fuzzy probability measures into the DTD (document type definition) model with fuzzy probability measures is developed in the paper.

KEYWORDS

Fuzzy Probability Measures, Fuzzy Sets, Mapping, Probabilistic XML Model, Probabilistic Relational Databases

1. INTRODUCTION

The prompt development and wide applications of the Web technologies have resulted in the availability of huge amounts of Web data. The data representation and exchange on the Web are increasingly important. Being a standard recommended by W3C (World Wide Web Consortium), XML (eXtensible Markup Language) has been the de-facto standard of data representation and exchange on the Web (XML, 2015), which allows the easy development of applications that exchange data over the Web. In the context of deep Web, data in relational databases are extracted to XML documents, which are then published on the Web. Mapping relational databases into XML is essential for XML data management (Fong & Shiu, 2012; Fong, Wong & Cheng, 2003; Lee, Mani, Chiu & Chu, 2001; Liu, Vincent & Liu, 2006).

Data in the real-world applications are diverse and some data are inherently uncertain. Human knowledge and natural language, for example, have a big deal of uncertainty. To deal with uncertain data, two major foundations have been developed (Parsons, 1996), which are probability theory for objective uncertainty and fuzzy sets for subjective uncertainty, respectively. Also, Dempster-Shafer theory (also known as belief theory or evidence theory (Yager & Alajlan, 2015)) is applied. Subjective uncertainty comes from subjective estimation and judgment made by human, for example, “young” for age and “hot” for temperature. Objective uncertainty comes from objective statistics instead of

DOI: 10.4018/JDM.2017070102

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
subjective estimation and judgment, for example, possible orientation of landing on for a coin being tossed.

Fuzzy sets and probability theory have widely been applied to extend various database models (Meged & Gelbard, 2012), including the uncertain relational database model (e.g., the fuzzy one (Buckles & Petry, 1982; Ma, 2005; Prade & Testemale, 1984; Tayal & Saxena, 2015; Yang, Zhang, Liu, Wu, Yu, Nakajima & Rishe, 2001) and the probabilistic one (Barbara, Garcia-Molina & Porter, 1992; Dalvi, Re & Suciu, 2011; Dey & Sarkar, 1996; Lakshmanan, Leone, Ross & Subrahmanian, 1997; Pittarelli, 1994; Murthy, Ikeda & Widom, 2011; Zhu, Zhang, Cao & Tang, 2016) and the uncertain object-oriented database model (e.g., the fuzzy one (Delgado, Marín, Pérez & Vila, 2016; Ma, Zhang & Ma, 2004; Yan & Ma, 2014a) and the probabilistic one (Eiter, Lu, Lukasiewicz & Subrahmanian, 2001; Kornatzy & Shimony, 1994; Nuray-Turan, Kalashnikov, Mehrotra & Yu, 2012)). In the context of XML, some recent efforts have devoted on the uncertain XML (e.g., the probabilistic one (Kimelfeld & Senellart, 2013) and the fuzzy one (Ma & Yan, 2007; Ma, Liu & Yan, 2010)) to handle uncertain data on the Web. More important, several efforts have been carried out to map uncertain relational databases (Amarilli & Senellart, 2013; Ma & Yan, 2007) from or to uncertain XML model. Being similar with the classical XML data management (Fong & Shiu, 2012), mapping uncertain relational databases to XML serves as uncertain data representation and exchange on the Web.

The uncertain databases and uncertain XML have been developed separately to deal with subjective uncertainty and objective uncertainty, respectively. So, each of fuzzy data models and probabilistic data models actually suffer from the inability to simultaneously deal with two kinds of uncertain data together. But some data in the real life typically contain both probabilistic and fuzzy information (Baldwin, Lawry & Martin, 1996), for example, the description of “the probability that it will be mainly cloudy tomorrow is very high”. In such an example, the probability measure of a probabilistic event is described by a fuzzy value instead of a crisp value. Then we have a kind of hybrid uncertain data named fuzzy probability (Buckley, 2005; Ralescu, 1994; Rebiasz, 2016; Zadeh, 1984).

Currently, few efforts have been carried out to introduce fuzzy probability into databases. A deductive probabilistic and fuzzy object-oriented database language is proposed in (Cao & Rossiter, 2003) and an object base model is proposed in (Cao & Nguyen, 2011). In these models, fuzzy sets are applied to represent uncertain attribute values and uncertain probability degrees are represented by interval values rather than crisp values and fuzzy sets. Interval probabilities are originally applied in the probabilistic relational databases in (Lakshmanan, Leone, Ross & Subrahmanian, 1997) and then the probabilistic object-oriented databases in (Eiter, Lu, Lukasiewicz & Subrahmanian, 2001). Actually, a fuzzy set is more informative than an interval value and it is a natural way to use fuzzy sets to represent uncertain probability measures. So, a probabilistic relational database model and a probabilistic object-oriented database model with fuzzy probability measures are recently proposed in (Yan & Ma, 2013) and (Yan & Ma, 2015), respectively, where fuzzy sets are used to represent uncertain probability measures rather than uncertain attribute values in (Cao & Nguyen, 2011; Cao & Rossiter, 2003).

In this paper, we concentrate on a kind of hybrid uncertain data, in which probabilistic events are associated with fuzzy probability measures. We concern the hybrid uncertain data modeling in relational databases and XML. We investigate how to map the relational databases with the hybrid uncertain data into XML model. For this purpose, we propose an extended XML model with the hybrid uncertain data. On the basis, we develop the formal approach to mapping the relational database model with the hybrid uncertain data proposed in (Yan & Ma, 2013) into the XML model with the hybrid uncertain data. We finally implement a prototype system to verify the proposed mapping. Although
Theoretical vs. Practical Complexity: The Case of UML
[www.igi-global.com/article/theoretical-practical-complexity/3336?camid=4v1a](www.igi-global.com/article/theoretical-practical-complexity/3336?camid=4v1a)

Data Integration: Introducing Semantics
[www.igi-global.com/chapter/data-integration-introducing-semantics/20731?camid=4v1a](www.igi-global.com/chapter/data-integration-introducing-semantics/20731?camid=4v1a)