Chapter 5
Modeling and Querying Fuzzy Data:
Current Approaches and Future Trends

Li Yan
Nanjing University of Aeronautics and Astronautics, China

Z. M. Ma
Northeastern University, China

ABSTRACT

Imperfect information extensively exists in data and knowledge intensive applications, where fuzzy data play an import role in nature. Fuzzy set theory has been extensively applied to extend various database models and resulted in numerous contributions. The chapter concentrates on two main issues in fuzzy data management: fuzzy data models and fuzzy data querying based on the fuzzy data models. A full up-to-date overview of the current state of the art in fuzzy data modeling and querying is provided in the chapter. In addition, the relationships among various fuzzy data models are discussed in the chapter. The chapter serves as identifying possible research opportunities in the area of fuzzy data management in addition to providing a generic overview of the approaches to modeling and querying fuzzy data.

INTRODUCTION

One of the major areas of database research has been the continuous effort to enrich existing database models with a more extensive collection of semantic concepts. Database models have developed from hierarchical and network database models to the relational database model. As computer technology moves into non-traditional applications such as CAD/CAM, knowledge-based systems, multimedia, GIS and Internet systems and modeling and manipulation of complex objects and semantic relationships are required by such applications, many software engineers feel the limitations of relational databases in these data- and knowledge-intensive applications. Therefore, some non-traditional data models have been proposed for databases, such as the entity-relationship (ER) model (Chen, 1976), the enhanced (or extended) entity-relationship (EER) model (e.g., Embley and Ling, 1989), the object-oriented (OO) da-

DOI: 10.4018/978-1-4666-8767-7.ch005

tabase model, the object-relational database model and the logic database model. Among these database models, the object relational database model combines the robustness of the relational database model with the powerful modeling capabilities of the object-oriented paradigm (Stonebraker and Moore, 1996).

With the prompt development of the Internet, the requirement of managing information based on the Web has attracted much attention both from academia and industry. XML (eXtensible Markup Language) is widely regarded as the next step in the evolution of the World Wide Web, and has been the de-facto standard. It aims at enhancing content on the World Wide Web. XML and related standards are flexible that allow the easy development of applications which exchange data over the web such as e-commerce and supply chain management.

While traditional data models can provide efficient data management capabilities, they often suffer from some inadequacy of necessary semantics. One of these inadequacies can be generalized as the inability to handle imprecise and uncertain information (Sicilia and Mastorakis, 2004). In real-world applications, information is often imperfect. One of the semantic needs not adequately addressed by the traditional data models is that of uncertainty. Traditional data models assume that the models are a correct reflection of the world and further assume that the stored data is known, accurate and complete. It is rarely the case in real life that all or most of these assumptions are met. So fuzzy sets theory (Zadeh, 1965) and possibility theory (Zadeh, 1978) have been used to extend various data models in order to enhance the traditional data models such that fuzzy data can be represented and manipulated.

Management of fuzzy data typically involves two primary technical issues: storage and queries. In addition, to serve a given query more effectively, it is necessary to index fuzzy data. These three issues are actually closely related. Indexing of fuzzy data is enabled based on fuzzy data storage, and efficient querying of fuzzy data is supported by the storage and indexing structure. Among these three issues, fuzzy data modeling provides the infrastructure for fuzzy data management and fuzzy data querying is one major goal of fuzzy data management. Since fuzzy database approaches were first created in the late 1970s, much work has been done in the area of fuzzy databases and a number of research efforts have been undertaken to address these issues. This has resulted in numerous contributions, mainly with respect to the popular relational model or to some related form of it (Petry, 1996; Chen, 1999; Galindo, Urrutia and Piattini, 2006). Since classical relational database model and its extension of fuzziness do not satisfy the need of modeling complex objects with imprecision and uncertainty, currently many studies have been concentrated on fuzzy conceptual data models and fuzzy object-oriented database models in order to deal with complex objects and fuzzy data together (Yazici and George, 1999; Ma, 2005a). More recently, some work has been carried out in extending XML towards the representation of fuzzy concepts on the Web (Yan, Ma and Zhang, 2014b).

Although there have been a lot of fuzzy database papers published, ones only find few comprehensive review papers of fuzzy databases. Two early overview papers in this area are finished by Yazici, Buckles and Petry (1992) and Kerre and Chen (1995). In (Yazici, George, Buckles and Petry, 1992), conceptual and logical data models for uncertainty management are reviewed, in which the conceptual data model is the fuzzy IFO data model (IFO is a graph-based conceptual data model proposed in (Abiteboul and Hull, 1987)) and the logical data model is the fuzzy relational database model. In (Kerre and Chen, 1995), only the fuzzy ER (entity-relationship) model and the fuzzy relational databases (exactly data representation, queries, and design) are discussed. In (Urrutia and Galindo, 2010), an overview of different fuzzy database modeling definitions by different authors is presented, but they only highlight the fuzzy EER model, which is an extension of an EER model incorporating fuzzy semantics and notation.