Expert Systems with Applications 41 (2014) 3615–3627

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Formal translation from fuzzy EER model to fuzzy XML model

Li Yan [a,*], Z.M. Ma [b]

[a] School of Software, Northeastern University, Shenyang 110819, China
[b] College of Information Science and Engineering, Northeastern University, Shenyang 110819, China

ARTICLE INFO

Keywords:
Data modeling
Fuzzy information
Conceptual data model
XML model
Mapping

ABSTRACT

XML has been the de facto standard of data representation and exchange over the Web. In addition, fuzzy data are inherent in the real-world applications. Although fuzzy information has been extensively investigated in the context of relational database model, the classical relational database model and its fuzzy extension to date do not satisfy the need of modeling and processing complex objects with imprecision and uncertainty on the Web. Based on fuzzy sets, this paper concentrates on fuzzy information modeling in the EER (enhanced entity-relationship or extended entity-relationship) model and the fuzzy XML model. In particular, the formal approach to mapping the fuzzy EER model to the fuzzy DTD (document type definition) model is developed.

1. Introduction

With the prompt development and wide applications of the Internet, the requirement of managing Web data has attracted much attention both from academia and industry. XML is widely regarded as the next step in the evolution of the World Wide Web, and has been the de facto standard. Aiming at enhancing content on the World Wide Web, XML and related standards are flexible and allow the easy development of applications which exchange data over the Web such as e-commerce (EC) and supply chain management (SCM). But this flexibility makes it challenging to develop an XML management system. To manage XML data, it is necessary to integrate XML and databases (Bertino and Catania, 2001). Various databases, including relational, object-oriented, and object-relational databases, have been applied for mapping to and from the XML document. In addition, XML lacks sufficient power in modeling real-world data and their complex interrelationships in semantics. So it is necessary to use other methods to describe data paradigms and develop a true conceptual data model, and then transform this model into an XML encoded format. Conceptual data modeling of XML document schema (Conrad et al., 2000; Elmasri et al., 2005; Mani et al., 2001; Psaila, 2000; Xiao et al., 2001) and XML Schema (Bernauer et al., 2004) have been studied in the recent past. In (Conrad et al., 2000), for example, UML is used for designing XML DTD (document type definition). At the same time, some data are inherently imprecise and uncertain since their values are subjective in the real-world applications. For example, considering the values which represent the satisfaction degree for a film, different individuals may have very different satisfaction degree. Information fuzziness on the Web has been investigated in the context of EC and SCM (Petrovic et al., 1999; Yager, 2000; Yager and Pasi, 2001). It is shown that fuzzy sets are very useful in Web-based business intelligence.

Viewed from data modeling, fuzzy information has been extensively investigated in the context of relational database model (Buckles and Petry, 1982; Prade and Testemale, 1984; Raju and Majumdar, 1988; Umano and Fukami, 1994). However, the classical relational database model and its fuzzy extension cannot satisfy the need of modeling complex objects with imprecision and uncertainty. The requirements of modeling complex objects and information imprecision and uncertainty can be found in many application domains (e.g., multimedia applications) and have challenged the current database technology (Aygun and Yazici, 2004; Chamorro-Martínez et al., 2007). To model uncertain data and complex-valued attributes as well as complex relationships among objects, current efforts have concentrated on the fuzzy object-oriented databases (Bordogna et al., 1999; Dubois et al., 1991; George et al., 1996; Gyseghem and Caluwe, 1998; Ma et al., 2004) and the fuzzy conceptual data models (Chen and Kerre, 1998; Galindo et al., 2004; Zvieli and Chen, 1986). In (Zvieli and Chen, 1986), fuzzy sets are applied to some of the basic ER (entity-relationship) concepts, introducing fuzzy entity sets, fuzzy relationship sets and fuzzy attribute sets (the first level of fuzziness) as well as fuzziness in entity and relationship occurrences (the second level of fuzziness) and in attribute values (the third level of fuzziness). Without including graphical representations, the fuzzy extensions of several major EER (enhanced entity-relationship or extended entity-relationship) concepts (e.g., superclass/subclass, generalization/specialization, category and

* Corresponding author.
E-mail address: yanl@swc.neu.edu.cn (L. Yan).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.12.014

the subclass with multiple superclasses) are introduced in Chen et al. (1998). In Galindo et al., 2004, the fuzzy EER model is extended by relaxing some constraints with fuzzy quantifiers. Also there are efforts to conceptually design the fuzzy databases using the fuzzy conceptual data models (Ma, 2005; Ma and Shen, 2006; Ma et al., 2001; Yazici et al., 1999; Ma et al., 2012, 2011). More recently, the fuzzy object-relational databases are proposed (Cuevasa et al., 2008) which combine both characters of fuzzy relational databases and fuzzy object-oriented databases. Ones can refer to (Ma and Yan, 2008, 2010) for recent surveys of these fuzzy data models.

Despite fuzzy sets have been applied to model and handle imprecise and uncertain information in databases since Zadeh introduced the theory of fuzzy sets (Zadeh, 1965), relative little work has been carried out in extending XML towards the representation of imprecise and uncertain concepts. Abiteboul et al. Abiteboul et al. (2001) provide a model for XML documents and DTDs and a representation system for XML with incomplete information. The representations of probabilistic data in XML are proposed in other previous research papers such as (Hung et al., 2003; Nierrman and Jagadish, 2002; Senellart and Abiteboul, 2007; Van Keulen et al., 2005). Without presenting XML representation model, data fuzziness in XML document is discussed directly according to the fuzzy relational databases in Gaurav and Alhajj (2006), and the simple mappings from the fuzzy relational databases to fuzzy XML document are provided also. In (Oliboni and Pozzani, 2008), an XML Schema definition is given for representing fuzzy data, which adopts data type classification for the XML data context. A fuzzy XML data model based on XML DTD (Document Type Definition) is proposed in Ma and Yan (2007), in which the mappings of the fuzzy XML DTD from the fuzzy UML data model and to the fuzzy relational database model are discussed, respectively. In (Yan, 2009), a fuzzy XML data model based on XML Schema is developed. Ones can refer to (Ma and Yan, 2010) for recent research work of soft computing in XML data management.

For the classical XML, conceptual data models are generally applied to construct XML model through mapping conceptual data models into XML model because XML lacks sufficient power in modeling real-world data and their complex inter-relationships in semantics. Similarly in order to conceptually design the fuzzy XML model, the fuzzy UML data model is applied to formally map to the fuzzy XML DTD model in Ma and Yan (2007). It should be pointed out that the conceptual data models for the conceptual design of XML model are diverse and different designers may use different conceptual data model. In the context of database modeling, for example, in addition to the UML data model, the EER model is able to capture and represent rich and complex semantics at a high abstract level and can be used for conceptual design of databases as well as XML. So in order to conceptually design the fuzzy XML model using different fuzzy conceptual data model, in this paper, we present a full-fledged fuzzy extension to the EER model and the corresponding graphical representations. In particular, we develop the formal approach to mapping the fuzzy EER model to the fuzzy XML DTD mode. With the proposed approach, the construction of the complex fuzzy XML model can start from the design of the fuzzy EER model, which is then mapped into the fuzzy XML model automatically. Note that the fuzzy XML model discussed in this paper only focuses on the fuzzy XML DTD. Limited by the expressive power of XML DTD, the fuzzy XML model mapped from the fuzzy EER model cannot support a much richer set of structures, types and constraints for describing fuzzy data.

The remainder of this paper is organized as follows. Section 2 presents fuzzy information modeling in the fuzzy XML model and the fuzzy EER model. In Section 3, the formal approach to mapping the fuzzy EER model to the fuzzy XML DTD model is developed. A small case study is given in Section 4 to show the proposed formal translation. Section 5 concludes this paper.

2. Fuzzy information modeling in XML and EER models

The concept of fuzzy sets was originally introduced by Zadeh, (1965). Let U be a universe of discourse and F be a fuzzy set in U. A membership function μ_F: $U \rightarrow [0,1]$ is defined for F, where $\mu_F(u)$, for each $u \in U$, denotes the membership degree of u in the fuzzy set F. Thus, the fuzzy set F is described as follows:

$$F = \{(u_1, \mu_F(u_1)), (u_2, \mu_F(u_2)), \ldots, (u_n, \mu_F(u_n))\}$$

The fuzzy set F is consisted of some elements just like a conventional set. But, not being the same as the conventional set, each element in F may or may not belong to F, having a membership degree to F which needs to be explicitly indicated. So in F, an element (say u_i) is associated with its membership degree (say $\mu_F(u_i)$), and they occur together in form of (u_i, $\mu_F(u_i)$). When the membership degrees that all elements in F belong to F are exactly 1, the fuzzy set F reduces to a conventional set.

When the membership degree $\mu_F(u)$ above is explained to be a measure of the possibility that a variable X has the value u, where X takes values in U, a fuzzy value is described by a possibility distribution π_X (Zadeh, 1978).

$$\pi_X = \{(u_1, \pi_X(u_1)), (u_2, \pi_X(u_2)), \ldots, (u_n, \pi_X(u_n))\}$$

Here, $\pi_X(u_i)$, $u_i \in U$, denotes the possibility that u_i is true. Let π_X be the possibility distribution representation for the fuzzy value of a variable X. It means that the value of X is fuzzy, and X may take one from some possible values u_1, u_2, …, and u_n and each one (say u_i) taken possibly is associated with its possibility degree (say $\pi_X(u_i)$).

2.1. Fuzzy XML model

Two kinds of fuzziness can be identified in XML documents (Ma and Yan, 2007). The first one is the fuzziness in elements (we use membership degrees associated with such elements) and the second one is the fuzziness in attribute values of elements (we use fuzzy sets to represent such values). For the latter, there exist two interpretations on it (i.e., disjunctive semantics and conjunctive semantics) and they may occur in child elements with or without further child elements in the ancestor–descendant chain. The basic data structure of fuzzy XML data model is the data tree (Ma et al., 2010).

Definition. Let V be a finite set (of vertices), $E \in V \times V$ be a set (of edges) and ℓ: $E \rightarrow \Gamma$ be a mapping from edges to a set Γ of strings called labels. The triple $G = (V, E, \ell)$ is an edge labeled directed graph.

Based on the data tree, we introduce the definition of fuzzy XML data tree.

Definition. Fuzzy XML data tree F is a 6-tuple, $F = (V, \psi, \ell, \tau, \kappa, \delta)$ where

- $V = \{V_1, \ldots, V_n\}$ is a finite set of vertices.
- $\psi \subset \{(V_i, V_j) \mid V_i, V_j \in V\}$, (V, ψ) is a directed tree.
- ℓ: $V \rightarrow (L \cup \{\text{null}\})$, here L is a set of labels. For each object $v \in V$ and each label $\nabla \in L$, $\ell(v, \nabla)$ specifies the set of objects that may be children of v with label ∇.
- $\tau \rightarrow T$, T is a set of types.
- κ is mapping which constrains the number of children with a given label. κ associates with each object $v \in V$ and each label $\nabla \in L$, an integer-valued interval function. $\kappa(v, \nabla) = [\min, \max]$, where $\min \geqslant 0$, $\max \geqslant \min$. We use κ to represent the lower and upper bounds.