Chapter 8
Prediction of Uncertain Spatiotemporal Data Based on XML Integrated With Markov Chain

Luyi Bai
Northeastern University, China

Nan Li
Northeastern University, China

Chengjia Sun
Northeastern University, China

Yuan Zhao
Northeastern University, China

ABSTRACT

Since XML could benefit data management greatly and Markov chains have an advantage in data prediction, the authors study the methodology of predicting uncertain spatiotemporal data based on XML integrated with Markov chain. To accomplish this, first, the researchers devise an uncertain spatiotemporal data model based on XML. Then, the researchers put forward the method based on Markov chains to predict spatiotemporal data, which has taken the uncertainty into consideration. Next, the researchers apply the prediction method to meteorological field. Finally, the experimental results demonstrate the advantages the authors approach. Such a method of prediction could broaden the research field of spatiotemporal data, and provide a significant reference in the study of forecasting uncertain spatiotemporal data.

DOI: 10.4018/978-1-5225-8446-9.ch008

INTRODUCTION

Spatiotemporal data is a kind of special data which can contain time information and spatial information simultaneously (Pfoser, Tryfona, & Jensen, 2005). Spatiotemporal data has the characteristics of multi-source, large scale and fast update. It can describe the information of the object more accurately because it contains the attributes of time and space. It also can look up the characteristic of the stage behavior of the object in real-time, and can observe and predict the probability of the occurrence of a specific stage behavior with reference to the spatiotemporal association constraint model. Spatiotemporal database is a complex system to store and process spatiotemporal data, and it is an important branch in the field of database querying. However, spatial database and temporal database are independent of each other and there is no intersection of both research fields before 1990s. With the development of temporal database and spatial database, researchers found the relationship between the two kinds of databases and combined the two to study gradually, resulting in a spatiotemporal database. The content of spatiotemporal database is very complex and huge, which can be used to manage spatiotemporal data. It can express the time and change, and can solve the problem of storage and management of spatiotemporal data in general database. Therefore, the spatiotemporal database has more research value than other forms of databases, and can be applied to a wider range of fields, such as the meteorological information management systems (Kurte, Durbha, King, Younan, & Potnis, 2017), environmental changes monitoring (Li et al., 2017), vehicle detection and tracking (Huang, Lee, & Lin., 2017; Ramanathan, & Chen, 2017), video surveillance (Hampapur et al., 2005), online estimation of temperature (Xu, Li, & Liu, 2018), even gesture recognition (Zhang, Zhu, Shen, & Song, 2017), spatiotemporal distribution of birds (Ferreira et al., 2011), and so on.

Unfortunately, the uncertainty of spatiotemporal data is widespread and cannot be avoided due to the existence of data distortion, loss or network delay in the transmission process (Tossebro, & Nygård, 2002). Therefore, a series of works on uncertain spatial-temporal data emerge. Such as modeling, querying and predicting uncertain spatiotemporal data and so on. As for modeling, Yazici et al. (2001) introduce a semantic data modeling approach for spatiotemporal database applications, which utilizes unified modeling language (UML) for handling spatiotemporal information, uncertainty and fuzziness, especially at the conceptual level of database design; Bai et al. (2018) propose an uncertain spatiotemporal data model based on XML, and present a series of algebraic operations based on the model they proposed to capture uncertain spatiotemporal information; Emrich et al. (2012) model the uncertain object by a homogeneous Markov chain with an initial probabilistic density function. As for querying, Emrich et al. (2012) reduce all queries on the uncertain spatiotemporal object model to simple matrix multiplications, and present two approaches towards efficient query processing as well. They are called object-based approach and query-based approach, and the latter is more efficient. Along with window queries (Emrich, Kriegel, Mamoulis, Renz, & Zufle, 2012; Tao, Papadias, Sun, 2003), there are also range queries (Trajcevski, Choudhary, Wolfson, Ye, & Li, 2010), similarity queries (Niedermayer et al., 2013a) and kNN queries (Niedermayer et al., 2013b; Trajcevski, Tamassia, Ding, Scheuermann, & Cruz, 2009) can be solved using both proposed solution approaches (i.e. equivalent worlds and monte-carlo sampling). There are also event queries on probabilistic data streams (Ré, Letchner, Balazinksa, & Suci, 2008). However, the research on the prediction of uncertain spatiotemporal data has not received too much attention. There is a system that models spatiotemporal events through the combination of kernel density estimation for event distribution and seasonal trend decomposition by loess smoothing for temporal predictions (Maciejewski et al., 2011). There is also a congestion prediction for urban areas (Wang,