Towards The Use of Probabilistic Spatial Relation Databases in Business Process Modeling

Haizhou Li, LIMOS, Université Blaise Pascal, Aubière, France

François Pinet, Irstea Centre de Clermont-Ferrand, Aubière, France

Farouk Toumani, LIMOS, Université Blaise Pascal, Aubière, France

ABSTRACT

In this paper, the authors combine the methods of probabilistic databases, GIS (Geographical Information Systems) and business process modelling to evaluate the agricultural activities. The authors provide a technique to evaluate the risk of agricultural activities among the hydrological objects (lakes, rivers, etc.) and agricultural plots. The spatial relation is an important information needed in the evaluation processes. This type of information is usually uncertain and the available data are often not precise. Consequently, probabilistic database is used to capture the uncertainty of the spatial objects in order to estimate the level of possible water and soil contamination (by agricultural inputs). Probabilistic spatial relations provide information on the layout of spatial objects. Probabilities are stored in a probabilistic database. Probabilistic database is a finite number of complete databases that are assigned with a set of probabilities. Probabilistic data-aware business processes integrate the theory of probabilistic database with business processes modeling methods. This new formalism of business processes helps the experts to model the environmental risks in terms of probabilistic spatial relations.

Keywords: Agricultural Risk Evaluation, Geographical Information Systems (GIS), Probabilistic Data-Aware Business Process, Probabilistic Database, Probabilistic Spatial Relation

1. INTRODUCTION

In this paper, we introduce the new concept of probabilistic spatial relation database. The goal of this paper is to show (1) how probabilistic databases can store spatial relations and (2) how it is possible to exploit this new type of databases in business process modeling applied to agriculture and environment.

Academic and industrial worlds have shown great interest in probabilistic databases (Dalvi & Suciu, 2007; Green, 2009; Pittarelli, 1990; Sen, Deshpande, & Getoor, 2008) but only few attentions have been paid on the application of this technology in geomatics and business process modeling; for example, the authors of (Li, Pinet, & Toumani, 2014) propose a method to use business process with probabilistic databases. Our aim is to show how the information technologies related to

DOI: 10.4018/IJAEIS.2015070104

probabilistic databases, geomatics (Gomarasca, 2009) and business process modeling (Havey, 2005; van der Aalst, 2013; Weske, 2012) can be combined. In this paper, we propose an agri-environmental case study to illustrate the joint use of these technologies to evaluate the environmental impact of agricultural activities. Our goal is to provide a realistic example to show how a probabilistic spatial relation database and a business process model can be used in an environmental application.

A large range of methods could be used to calculate probabilities on spatial relations. In Section 2, we will provide some examples showing how probabilities on spatial relations can be calculated. We choose some concrete examples in different fields.

In Section 3, we propose a generic database schema to store spatial relations. This database schema can store probabilities in different types of tables.

Section 4 presents our case study. We illustrate the use of our generic database schema in an agri-environmental application. The proposed probabilistic database stores information on (1) the use of pesticide in agriculture and (2) spatial relations between different environmental objects. We also introduce an example of business process modeling that formalizes the analysis of the possible interaction between pesticides and environmental objects. We show that the use of this business process model with an instance of the probabilistic database can help decision-makers to estimate the level of pesticide contaminations for a particular geographical region. This analysis will help to decide if control actions are needed or not (e.g., soil controls).

2. PROBABILISTIC SPATIAL RELATIONS

Spatial relations can provide information on the layout of spatial objects (Freeman, 1975). As shown in (Egenhofer & Herring, 1990), there are different types of spatial relations. For example, topological relations are invariant under topological transformations. Metric information is provided by distance relations. Partial and total orders can be also modeled by spatial objects ("in the front of", "behind", etc.). Numerous spatial relations have been defined between different forms of spatial objects.

Probabilistic spatial relations can be defined between objects. This type of relations is associated to probability. In other words, we are able to assign probabilities on the different spatial relations. We present below three examples to illustrate this concept.

2.1. Exemple 1

Suppose five data sources. Data sources #1, #2, #3 store different spatial representations of the same objects A. Data sources #4, #5 store different spatial representations of the same object B. These different representations are caused by the uses of different measurement techniques. Figure 1 provides an example of the instances A and B in the five data sources; *di* is the different possible minimal distances between A and B. We consider that only one data source stores the correct representation for the object A (#1, #2 or #3) - but we do not know which one is correct. In the same manner, only one data source stores the correct representation for the object B (#4 or #5) - but the correct source is unknown. So, if we make the hypothesis that all the rows of Table 1 are equiprobable events: P(*distance(A,B)=d1*) = 1/6, P(*distance(A,B)=d2*) = 1/2 and P(*distance(A,B)=d3*) = 2/3. In other word, if we do not know which data source stores the correct representation (for A and B), the probabilities that *d1* is the correct distance is 1/6, etc. Note that the distance that minimizes the risk (i.e., the wrong choice) is *d2*. An order can be also provided between these distances (d2 < d1 < d3) and probabilities of other events can be calculated: for example, P(*distance(A,B)<=d1*)= 5/6, i.e., the probability that *distance(A,B)*is less than or equal to *d1* is 5/6.