Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 32 (2007) 47–55
Expert Systems with Applications
www.elsevier.com/locate/eswa

Dempster–Shafer Theory in geographic information systems: A survey

J.A. Malpica [a,*], M.C. Alonso [a], M.A. Sanz [b]

[a] Mathematics Department, Alcala University, Madrid, Spain
[b] Mathematics Department, Politecnica University, Madrid, Spain

Abstract

Since the information used in a Geographic Information System has a certain degree of uncertainly, in general classical mathematics models should not be applied to solve geographical problems computationally. Therefore, probabilistic or fuzzy-related methods should be considered, in order to model the behaviour of real problems that have to be solved by or with a Geographic Information System.

In this paper, a review of the application of Dempster–Shafer Theory of Evidence—also called "belief functions"—in relation to Geographic Information System is given. The review will focus on classification as a way of fusing information in a Geographic Information System. Information fusion, for classification, represents the first step in the abstraction of information and a means of data mining, and both the advantages and limitations of the technique of the Theory of Evidence in comparison to other techniques are analysed.

Keywords: Uncertainty in GIS; Theory of evidence; Dempster–Shafer Theory; Decision making within GIS; Fusion of information

1. Introduction

It is difficult to avoid uncertainty when attempting to make models of the real world. Uncertainty is inherent to natural phenomena, and it is impossible to create a perfect representation of reality. Classic mathematics deals with ideal worlds where perfect geometric figures exist and can verify extraordinary conditions.

The formalisation of fuzzy sets started in the 1960s with the works of Zadeh (1965) in fuzzy sets and Dempster (1968) in belief functions. Belief functions offer a non Bayesian method for quantifying subjective evaluations by using probability. In the 1970s, it was further developed by Shafer, whose book. Mathematical Theory of Evidence (Shafer, 1976) remains a classic in belief functions, or the so-called Theory of Evidence (TE). This theory has been also called the Dempster–Shafer Theory of Evidence. In the 1980s, the scientific community working with Artificial Intelligence got involved in using TE in applications. Parallel formalisations continued in the 1980s with the work of Higashi and Klir (1982, 1983), Höhle (1982) and Yager (1983). In the 1990s, Klir (1991) developed it further and gave it the name of "General Theory of Information."

Today, mathematical formalisations continue to be developed. There are many publications touching upon this subject, and entire congresses are dedicated specifically to uncertainty and its related fields. The increasing computational power has contributed to new possibilities for the solutions of problems in many areas of science and technology—mathematics, medicine, social science, business, and the like—and every aspect of the real world is capable of been tackled by these new computational techniques. The formalisation of uncertainty in order to construct models for the real world is so important that Klir (1995) has even said that we are now experiencing a change of attitude within the scientific community—a change that has all the characteristics of a new framework for understanding natural phenomena; a new concept of knowledge in the sense of a new paradigm, as it has been highlighted in the popular book by Kuhn (1962).

Currently, the use of Geographic Information Systems (GIS) is widespread and used in many applications. Their most important use is as a decision support system, but

* Corresponding author. Tel.: +34 18856761.
E-mail address: josea.malpica@uah.es (J.A. Malpica).

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2005.11.011

to so use it requires that one combine information within the GIS in an optimal manner, in order that the decision-maker can extract the more relevant information in helping him or her make the decision. Moreover, the data store in each layer of the GIS and the method utilised to fuse the information are both inherently rife with uncertainty.

Section 2 of this paper will be dedicated to briefly introducing to the reader the GIS terminology required to grasp the thrust of the paper. In Section 3, the inherent uncertainties are presented, along with the new paradigm of fuzzy methods for modelling natural phenomena; more specifically, uncertainty for geoinformation will be examined, along with the problems with implementing expert systems in GIS. Classification in a GIS allows for the reduction of information, and this classification is carried out in layers within a framework of uncertainty. Section 4 is dedicated to TE, and it has been written in a manner similar to Section 2 (i.e., providing a brief introduction and outlining the terminology involved); this section finishes with the Dempster rule, which is fundamental in fusing information in a GIS. Section 5 presents several case studies where a mixture of information was garnered from layers of a GIS, in order to make decisions. Most of these case studies used TE for combining information.

2. GIS as a geospatial database

A GIS can be defined as a system of hardware and software used for the input, storage, retrieval, mapping, display and analysis of geographic data. A reference system is a common feature to all the information used in a GIS—that is, every element stored in a GIS has a set of coordinates that reference to a concrete place on earth. Such information is said to be "georeferenced."

Geospatial data come in layers, and each layer has two types of information: graphic and attributes (the former is also called geometric data, and the latter thematic data). Information in the layers can be represented in two manners: raster and vector. In Fig. 1, an area of 50×25 km around Alcala de Henares in Madrid is presented as an example; there, land use is represented by a raster layer while hydrology is represented by vector data. One can there see that raster and vector representations are two different spatial conceptions of GIS, used to model the real world. Vector data are graphic (i.e., it can have three elements: points, lines and polygons) and, most importantly, these graphic objects are situated in space according to their coordinates. In the case of raster data, the layer (or space) is considered a grid where each cell (called also a pixel) represents a basic element of information (i.e., the space exists beforehand, and the object is placed in it). This dichotomy—vector and raster—means we have two different schools of thought concerning space: in the former, space exists because of the objects, and without the object there is no space; in the latter, space is an intuitive idea, a philosophical category of Kant where objects are placed. There are many books touching upon the basics of GIS, including Longley, Goodchild, Maguire, and Rhind (2001) and Burrough and Mcdonnell (1998).

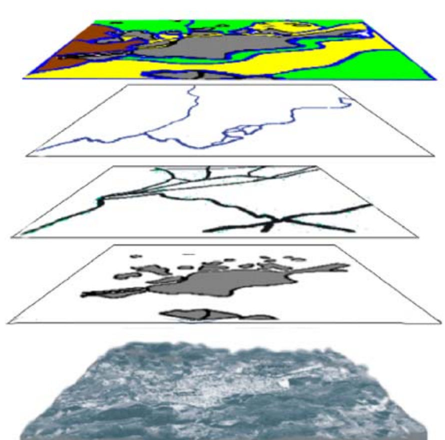

Fig. 1. A GIS representation.

In short, a GIS is a database of geospatial data with some display and analytical tools. Applications use one of the many commercial software packages already on the market, because these software packages are designed specifically for the applications most often at hand: travel, hydrology, utilities, and so forth. They are used as standalone software packages or as modules within general software packages, with many likely working on the Internet and using data from different sites. Today, these packages are user-friendly and can be used by the professional in the specific field of application. This means there is no need to consult an expert on GIS, as was once the case. This specialisation in GIS software for the application at hand has led us to the next step of GIS evolution: expert systems (i.e., systems where all or some of the integral knowledge of the expert is implanted within the system).

In the 1980s, only people with high levels of expertise and access to expensive computers called mainframes were able to use a GIS. However, in the last 15 years, GIS technology has undergone a major revolution, thanks to more advanced technology, a proliferation of data, price reductions on computers, and commercial software having been made ever more user friendly.

One software package with a special module with TE is IDRISI (Idrisi was the name of an Arab-Spanish geographer of the 12th century). It is basically a raster software package and it has been used several times for the applications presented in Section 5. Another raster software package is GRASS (Geographic Resources Analysis Support System), which is open-source and has topological vector and image processing capabilities that could be of use for the implementations of TE, but should be programmed. Recently Benjamin Ducke programmed a module for archaeological applications in GRASS using TE; see the GRASS site for more information http://grass.itc.it. There are many commercial GIS software packages, some of which make use of fuzzy technique but not of TE; Ecognition and RAISON immediately spring to mind.