General Entropy Model of Line Segments Uncertainty in GIS

SHI Yufeng SHI Wenzhong

Abstract  Spatial data uncertainty can directly affect the quality of digital products and GIS-based decision making. On the basis of the characteristics of randomness of positional data and fuzziness of attribute data, taking entropy as a measure, the stochastic entropy model of positional data uncertainty and fuzzy entropy model of attribute data uncertainty are proposed. As both randomness and fuzziness usually simultaneously exist in linear segments, their omnibus effects are also investigated and quantified. A novel uncertainty measure, general entropy, is presented. The general entropy can be used as a uniform measure to quantify the total uncertainty caused by stochastic uncertainty and fuzzy uncertainty in GIS.

Keywords  uncertainty; line segment; general entropy; stochastic entropy; fuzzy entropy; uniform measure

Introduction

With the development of GIS, people are more and more interested in spatial data uncertainty. The purpose of spatial data uncertainty research is to discuss the occurrence, propagation and control of uncertainty. Spatial data uncertainty generally refers to error, imprecision, fuzzy and vagueness [1]. Generally, it can be classified as positional uncertainty, attribute uncertainty, temporal uncertainty, logistic inconsistency and incompleteness of data, and the uncertainty of spatial data characteristic is one of the hot issues in GIS research fields [1-3]. At present, research on the positional uncertainty is mainly focused on the model and propagation of line segment and area unit. Based on statistical theory, some researchers have studied the uncertainty models of point, line segment and polygon [2-6]. On the basis of “ε-band” model proposed by Chrisman, Shi developed it and presented a universal model to describe positional uncertainty of GIS data, and he also presented a general statistical description of these uncertainties [2]; based on randomized graph algebra, Shi also studied a probabilistic paradigm for handling uncertain objects [7]. Dai defined the visual indexes of point ellipse, line error band and polygon error donut to assess effect scope of their positional uncertainty according to the probability that the feature points dropping into their error ellipses based on the error ellipse expressing the positional uncertainty in surveying and mapping [8]. Cheung and Shi developed a model and continuous index, the probability value, to indicate the extent of the uncertainty point located inside the uncertain polygon [6]. Based on Shannon information theory, Fan proposed a new uncertainty band of error entropy (H-2 band), H-2 band can be used as an objective index of uncertainty [9]. Li proposed the information entropy models of spatial data positional uncertainty and discussed the error entropy models of point, line segment and polygon respectively [10]. From the viewpoint of pure mathematics, the uncertainty model
based on information entropy is indeed of statistical uncertainty. Because error entropy is educed on the basis of probability density, and error entropy model is a kind of stochastic entropy uncertainty model. But there are many fuzzy geographical entities in GIS, and due to these entities have intrinsic ambiguity, they can not be expressed by statistical theory. On the basis of field’s theory and model, Zhang combined positional uncertainty and attribute uncertainty to describe and analyze the spatial data uncertainty[11,12]. Owing to the procedure of spatial data storing and processing resembles information transmission very much, so based on the characteristics of information entropy and fuzzy entropy, this paper proposes stochastic entropy model of positional uncertainty of spatial data and fuzzy entropy model of attribute uncertainty of spatial data respectively, and takes randomicity and fuzziness into consideration comprehensively, proposes general entropy model of spatial data uncertainty. As for some ambiguous geographical phenomena, due to the randomicity and fuzziness exist in continuous form, so general entropy can better embody their uncertainty.

1 Stochastic entropy and fuzzy entropy

1.1 Definition and properties of stochastic entropy

Entropy is a kind of measurement to a system’s disorder, instability, imbalance, uncertainty, etc. Historically, entropy has three origins: thermodynamics, statistical mechanics, and information theory. Information entropy is an important concept in information theory, and it denotes the average uncertainty of information source[13].

Definition of information entropy (discrete sample space): let $A$ be a probabilistic experiment with sample space $X$ and probability distribution $P$, where $p_i$ is the probability of outcomes $x_i \in X$ and they satisfy $p_i \geq 0$ and $\sum_{i=1}^{n} p_i = 1$. Then Shannon information entropy is given by:

$$H_r (X) = E [-\log p_i] = -\sum_{i=1}^{n} p_i \log p_i$$

Information entropy $H_r (X)$ has the following main properties.

Symmetry: $H_r (X)$ is symmetric. $H_r (p_1, p_2, \cdots, p_n) = H_r (p_2, p_1, \cdots, p_n) = H_r (p_n, p_2, \cdots, p_1)$. That is, the ordering of the probabilities $p_1, p_2, \cdots, p_n$ does not influence the value of $H_r (X)$.

Non-negativity: $H_r (X)$ is non-negative. $H_r (p_1, p_2, \cdots, p_n) = -\sum_{i=1}^{n} p_i \log p_i \geq 0$.

Additivity: $H_r (X)$ is additive. If $X$ and $Y$ are two sample spaces, where outcomes in $X$ are independent of those in $Y$, then the information entropy of joint events $(a_i, b_j)$ is given by:

$$H_r (X, Y) = -\sum_{i=1}^{r} \sum_{j=1}^{s} p(a_i, b_j) \log p(a_i, b_j) =$$

$$= -\sum_{i=1}^{r} p_i \log p_i - \sum_{j=1}^{s} p_j \log p_j = H_r (X) + H_r (Y)$$

where $p_i$ and $p_j$ are the probability of space $X$ and $Y$ respectively, $0 \leq p_i \leq 1$, $\sum_{i=1}^{r} p_i = 1$ and $0 \leq p_j \leq 1$, $\sum_{j=1}^{s} p_j = 1$.

Maximality: $H_r (X)$ is maximum if all probabilities are equal. That is, $H_r (p_1, p_2, \cdots, p_n) = \log r$. This corresponds with the situation where maximum uncertainty exists. $H_r (X)$ is minimum if one outcome has a probability equal to 1.

Definition of information entropy (continuous sample space): for the continuous stochastic variable $x$ with probability density function $p(x)$, the amount of information is equal to

$$H_r (x) = -\int_{-\infty}^{+\infty} p(x) \log p(x) dx$$

Clearly, the definition of the continuous information entropy is based on analogy with the discrete one.

From the definition and characteristics of information entropy, we can see that information entropy is a function of probability density function and can be used as the measure of stochastic uncertainty. In order to differentiate and compare with fuzzy entropy in the following sections, information entropy is called stochastic entropy in this paper.

1.2 Definition and properties of fuzzy entropy

Fuzzy entropy is a proper measure of fuzzy subset