Spatial data modelling and maximum entropy theory

Modelování prostorových dat a teorie maximální entropie

D. Klimešová[1], E. Ocelíková[2]

[1]Czech University of Agriculture, Prague, Czech Republic & *Institute of Information Theory and Automation, Czech Academy of Sciences, Prague*
[2]Technical University Košice, Slovak Republic

Abstract: Spatial data modelling and consequential error estimation of the distribution function are key points of spatial analysis. For many practical problems, it is impossible to hypothesize distribution function firstly and some distribution models, such as Gaussian distribution, may not suit to complicated distribution in practice. The paper shows the possibility of the approach based on the maximum entropy theory that can optimally describe the spatial data distribution and gives the actual error estimation.

Key words: spatial data classification, distribution function, error distribution, and maximum entropy approach

Abstrakt: Statistické modelování prostorových dat a odhad chyby distribuční funkce jsou klíčové otázky prostorové analýzy. Při řešení praktických aplikací je často obtížné potvrdit či zamítnout určitou hypotézu o distribuční funkci a některé distribuční modely, jako Gaussovo rozdělení, často velice málo odpovídají komplikovanému rozdělení konkrétní úlohy. Příspěvek ukazuje možnosti teorie maximální entropie pro získání dobrého odhadu chyby.

Klíčová slova: klasifikace, distribuční funkce, chyba rozdělení, maximální entropie

INTRODUCTION

Image classification is one of the basic procedures we use to process spatial data. The goal of the classification is to assign pattern to the definite class. We distinguish in general two types of classification*:* supervised and unsupervised approach. It depends on the fact wheiter we have some information about classes and wheiter we have some training data sets in disposal to be able to create sufficient mathematical model of the class etalons in the feature space or wheiter we have no field information. Having no training data, we speak about clusters and we use the methods of cluster analysis. There are two parts to training process, selection of an appropriate class list and an adequately precise mathematical description of each class. The choice of the decision rule, for the discrimination of mentioned feature space, plays key role in supervised classification.

There are two basic approaches:

1) Non-parametric approach – usually uses linear (exclusively non-linear) functions and mathematical or geometrical approaches for subdivision of the feature space. In this case, the Euclidian distances are evaluated. Examples: Minimum distance to means, the method of the nearest neighbor or Parallelepiped box.
2) Parametric approach, where statistical properties of n-dimensional feature space are used – mean vector, covariance matrices, distribution functions and statistical distances are evaluated – Mahalanobis distance, probability, etc. Examples: Maximum likelihood, Bayes method.

The paper is devoted to the problems of the parametric approach. Statistical modeling of spatial data distribution is the key to spatial data pattern recognition. Spatial data error distribution is the base for the spatial data accuracy analysis and quality control. Errors come from the process of data collecting, dealing and methods applying. Spatial data distribution is based on the probability statistical analysis theory and gets parameterised the distribution density function. According to the sample data property, it is hypothesized to suit some distribution model (e.g. Gaussian distribution). Then the model has to be tested under given significance level and according to the testing result, its distribution model is founded (Van Ryzin 1977; Liu, Shi 2000).

The Gaussian model for distributions is very convenient, as the Gaussian density function has many convenient properties and characteristics, both theoretically and practically. The main one among these properties is that its use requires knowledge of only the first two statistical moments; the mean vector and the covariance matrix, both of which can be usually estimated adequately from the reasonable sized training set (Landgrebe 2003; Zhou, Luo 2001).

There are many practical problems of spatial data distributions, it is impossible to hypothesize distribution function firstly and some distribution models such as Gaussian distribution may not suit to the complicated distribution in practice. On the other hand, a purely nonparametric approach is not frequently used, as very large training sets are usually required to provide precise enough estimates of nonparametric class densities. A very practical approach is to model each class density by a linear combination of Gaussian densities. Theoretically, every smooth density function can be approximated within any accuracy by such a mixture of Gaussian densities. And of course, this approach has also its difficulties and limits (Nguyen, Ocelíková 1999; Lin 1997).

In many cases, we are not able to test the assumptions of the proposed distribution function. In addition, the probability function distribution is usually one-peak, it means the function has only one maximum, but maybe there are multi-peaks functions in practical problems. Then we meet the difficulties when we use the traditional probability statistical approach for dealing and analysing spatial data information. In this case, we propose to use the approach based on the maximum entropy theory that can optimally describe the spatial data error distribution.

ENTROPY

Shannon successfully introduced the concept of entropy into the information theory. He explained it as the uncertainty of information, and gave us the formula that can measure the amount of information.

Given a random variable X (discrete form), whose value is got randomly, then, information entropy of X can be defined according (Gallager 1968) as:

$$H(X) = -\sum_{i=1}^{n} p\,(a_i)\log p(a_i),$$

where

$X \quad = a_1, a_2, \ldots a_n$

$P(X) \ = p(a_1), p(a_2), \ldots p(a_n)$

and

$0 \le p\ (a_i) \le 1 \quad (i = 1, 2, \ldots)$

$$H(X) = -\sum_{i=1}^{n} p\,(a_i)\,\mathrm{l}$$

and $p(a_i)$ is the probability of a_i

When the value of x is got continuously, and its probability density function *is p(x),* then information entropy of x can be defined as:

$$h(x) = -\int p(x)\log p(x)dx \tag{1}$$

The useful properties of information entropy are listed as follows:

$$H(p_1, p_2, \ldots p_n) \le \log n \tag{2}$$

When and only when

$$p_i = \frac{1}{n} \qquad (i = 1, 2, \ldots n),$$

equation (2) with the sign equal is valid.

It indicates that information entropy of equal probability field is maximal on conditions that the numbers of basic events are equal.

$$H\ (p_1, p_2, \ldots p_n) \ge 0 \tag{3}$$

When and only when the distribution of x is a degenerate distribution, equation (3) is valid with the sign equal. It indicates that determinate field is minimal.

MAXIMUM ENTROPY

On the other hand, we have to be able to determine the probability p_i ($i = 1, 2, \ldots n$) of random variable x_i ($i = 1, 2, \ldots n$) also in case if some numerical characteristics of the random variable can be got from the observed data when the probability distribution function satisfying the observed data may be limitless.

How to approach this case we can find in Janyes (1957). When deducing from part information, we must select such a probability distribution that has maximum entropy and obey to all known information. This is only one unbiased distribution we can do. And using any other distribution means drawing occasional assumptions to information, which may not exist originally.

This principle of statistical deducing is called maximum entropy theory and underlines that the probability distribution should be in accordance with known information. The measurements should suit to the samples and the unknown parts should not be hypothesized at all, because any hypothesis will add some information that may not exist originally. The reason is that the estimation drawing from the maximum entropy principle will approximate to the real distribution best. Maximum entropy theory can be explained by the definition and properties of information entropy. The mathematical formulas of maximum entropy theory are given as follows:

$$\max H = -K\sum p_i \log p_i \qquad (4)$$

$$\sum_{i-1}^{n} p_i = 1, \quad p_i \geq 0 \qquad i = 1, 2, \ldots n$$

$$\sum_{i=i}^{n} p_i g_j(x_i) = E[g_j(x)] \qquad i = 1, 2, \ldots n$$

where $g_j(x)$ represents the observed function $E\lfloor g_j(x)\rfloor$ and corresponding mean for the discrete case of data. When data is got continuously, the following formulas will be valid.

$$\max H = -\int_R f(x)\ln f(x)dx \qquad (5)$$

$$-\int_R f(x)dx = 1$$

$$\int_R f(x).x^n dx = u_n \qquad n = 1, 2, \ldots m$$

u_n denote the n-rank moment of x, which can be calculated from the sample data, and m is the rank of the origin moment and

$$u_n = \frac{1}{N}\sum_{t=1}^{N} x_t^n$$

where N is the number of sample data.

According to the maximum entropy theory, we can formulate such a conclusion: some probability distribution functions in probability theory are actually special cases that can be got from the maximum entropy theory on different conditions. For example, maximum entropy distribution is mean distribution on condition that mean is fixed, and it is Gauss distribution when the variance is fixed, etc.

This means that the maximum entropy theory can be regarded as unified theory base of different probability distribution. As proposed in Shi et al. (2003), the formula of maximum entropy distribution function can be given as follows:

$$\ln f(x) = \lambda_0 + \sum_{n=1}^{m} \lambda_n x^n = 0$$

It means

$$f(x) = \exp(\lambda_0 + \sum_{n=1}^{m} \lambda_n x^n)$$

where λ denote Lagrange indefinite operator. The error estimation can be computed using the formula:

$$\varepsilon_n = 1 - \frac{\int_R x^n \exp(\sum_{j=1}^{m} \lambda_j x^j)dx}{\mu_n \int_R \exp(\sum_{j=1}^{m} \lambda_j x^j)dx} \qquad n = 1, 2, \ldots m \qquad (6)$$

and the optimal object function can be defined as:

$$\min \varepsilon = \sum_{n=1}^{m} \varepsilon_n^2 \qquad (7)$$

ε_n is the remainder error and ε is optimal value, which may meet the need of formula for ε_n by adjusting the value of λ_n (n = 1, 2, ... m).

CONCLUSION

Based on the maximum entropy theory, the paper presents a new method to model spatial data error distribution function. There is no doubt about the usefulness of given estimations namely when the spatial modelling task results support significant decisions and are used to select optimal scenarios or strategic plans. Based on the information theory, this method uses entropy function as an objective function to reduce the human interference. It does not neccessary to hypothesize sample data to suit some common distribution and then test the hypothesis. But this method also has some disadvantage. To ensure high suited accuracy, much bigger sample is required. And the upper limit and lower limit of integral interval must be selected carefully, or the rear of maximum entropy distribution will be beyond the mark.

The problems of maximum entropy design have been discussed at the conference Agrarian Prospectives 2004 in the session of Applied Informatics.

REFERENCES

Liu D., Shi W. et al. (2000): Accuracy Analysis and Quality Control of Spatial Data in GIS. Science & Technology Literature Press, Shanghai.

Gallager R. (1968): Information Theory and Reliable Communication. John Wiley & Sons, Inc., New York; ISBN 471 29048 3.

Janyes E. T (1957): Information Theory and Statistical Mechanics. Physical Review, *106* (4): 620–630.

Landgrebe D.A. (2003): Signal Theory Methods in Multispectral Remote Sensing. Wiley-Interscience (wiley.com); ISBN 0-471-42028-X.

Liu D., Shi W. et al. (2000): Accuracy Analysis and Quality Control of Spatial Data in GIS. Science & Technology Literature Press, Shanghai.

Lin H. (1997): Modern Survey Error Analysis (6). Measure Technique: 41–43.

Nguyen, H. T., Ocelíková E. (1999): Maximum Entropy Image Reconstruction with Neural Networks. Proceedings of Digital Signal Processing '99: 46–50.

Ryzin Van J. (1977): Classification and clustering. Academic Press, Inc., New York.

Shi Y., Jin F., Qu G. (2003): Modelling of Spatial Data Error Distribution Based on Maximum Entropy Theory. In.: Proceedings of 7th South East Asian Survey Congress, Hong Kong, 7–13 November.

Zhou Ch., Luo J.(2001): Remote Sensing Statistical Analysis Model and Its Extension. Journal of Image and Graphics, 6 (A.12): 1210–1215.

Arrived on 3rd January 2005

Contact address:

RNDr. Dana Klimešová, CSc., Česká zemědělská univerzita v Praze, Kamýcká 129, 165 21 Praha 6-Suchdol,
Česká republika
tel.: +420 224 382 272, fax: +420 224 382 274, e-mail: klimesova@pef.czu.cz, klimes@utia.cas.cz
Prof. RNDr. Eva Ocelíková, CSc., Technická univerzita Košice, Letná 9, 042 00 Košice, Slovenská republika
e-mail: ocelike@ccsun.tuke.sk