IMAGE CLASSIFICATION DIFFICULTIES

POZNÁMKY KE KLASIFIKACI OBRAZOVÝCH DAT

Dana Klimešová, Eva Ocelíková

Abstract:
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 practical. This paper shows the possibility of the approach based on the maximum entropy theory that can optimally describe the spatial data distribution and gives actual error estimation.

Keywords:
Spatial data classification, distribution function, error distribution, maximal 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 aim of the classification is to assign pattern to definite class. We distinguish in general two types of classification*: supervised and unsupervised*. It depends on the fact we have training sets in disposal or we have no field information. Having no training data we speak about clusters and we use the methods of cluster analysis. The choice of decision rule, for the discrimination of feature space, plays the key role in supervised classification. There are two basic approaches:

Non-parametric one – usually uses linear (exclusively non-linear) functions and mathematical or geometrical approach for subdivision of the feature space. Euclidian distances are evaluated. Examples: Minimum distance to means, Nearest neighbour, Parallelepiped box.

Parametric statistical properties are known – mean vector, covariance and distribution function. Statistical distances are evaluated – Mahalanobis distance, etc. Examples: Maximum likelihood, Bayes.

The paper is devoted to the parametric approaches problems. The 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 probability statistical analysis theory and gets parameterised 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 [1,2].

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 complicated distribution in practical. 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. Than 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

$$X : a_1, a_2, \ldots\ldots. a_n \qquad P(X) : p(a_1), p(a_2), \ldots. p(a_n)$$

and $0 \le p(a_i) \le 1 \quad (i = 1,2,\ldots.) \qquad \sum_{i=1}^{n} p(a_i) = 1,$

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 $p_i = 1/n$ $(i = 1,2,\ldots n)$, 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

Assume there is a probability process that can be observed, and the values of random variable x can be a discrete sequence: $x_i (i = 1,2,\ldots n)$. If some numerical characteristics of the random variable can be got from observed result, then how can we determine the probability $p_i (i = 1,2,\ldots n)$ of random variable $x_i (i = 1,2,\ldots n)$? The probability distribution function satisfying the observed data may be limitless, which should be selected?

Simple answer we can find in [3].“*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 maximum entropy principle will approximate to 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:

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

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

$$\sum_{i=1}^{n} p_i g_j(x_i) = E[g_j(x)] \quad j = 1,2,\cdots m$$

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

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

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

$$\int_R f(x)\cdot x^n dx = u_n; \qquad n = 1,2,\cdots m$$

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

$$u_n = \frac{1}{N}\sum_{i=1}^{N} x_i^n$$

where N is the number of sample data.

According to 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 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 maximum entropy theory can be regarded as unified theory base of different probability distribution.

AS PROPOSED IN [4] THE FORMULA OF MAXIMUM ENTROPY DISTRIBUTION FUNCTION CAN BE GIVEN LIKE 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:

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

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 $\lambda_n (n = 1,2,...,m)$.

CONCLUSION

Based on maximum entropy theory, the paper presents a new method to model spatial data error distribution function. Based on information theory, this method uses entropy function as an objective function to reduce the human interference. It doesn't need 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 more 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 beyond the mark.

Reference:

[1] LIU, D.; SHI, W. et al. *Accuracy Analysis and Quality Control of Spatial Data in GIS*, Shanghai, 2000, Science & Technology Literature Press.

[2] GALLAGER R.1968*, Information Theory and Reliable Communication,* New York: John Wiley & Sons, Inc., 1968, 298, ISBN 471 29048 3.

[3] JANYES, E. T Information Theory and Statistical Mechanics, *Physical Review*, 1957, Vol. 106(4), 620-630.

[4] SHI, Y.; JIN, F.; QU, G. Modeling of Spatial Data Error Distribution Based on Maximum Entropy Theory, In.: *Proceedings of 7th South East Asian Survey Congress*, Hong Kong, 7-13 November 2003,

[5] NGUYEN, H. T.; OCELÍKOVÁ E. Maximum Entropy Image Reconstruction with Neural Networks, Proc. of *Digital Signal Processing '99*, SR, 1999, 46-50.

[6] ZHOU, CH.; LUO J. Remote Sensing Statistical Analysis Model and Its Extension, *Journal of Image and Graphics,* 2001, Vol. 6(A.12), 1210-1215.

[7] LIN, H. Modern Survey Error Analysis (6), *Measure Technique,* 1997, (6), 41-43.

Kontaktní adresy autorů:

RNDr. Dana Klimešová, CSc., Czech University of Agriculture, Prague
Faculty of Economocs and Management, Dept. of Information Engineering
Kamýcká 129, 165 21 Praha 6 – Suchdol, Czech Republic, klimesova@pef.czu.cz
and
Czech Academy of Sciences, Institute of Information Theory and Automation,
Dept. of Image Processing, Pod vodárenskou věží 4, 182 00 Prague 8, Czech Republic
klimes@utia.cas.cz
Prof. RNDr.Eva Ocelíková, CSc.
*Technical University Košice, Faculty of Electrical Engineering and Informatics, Department of Cybernetics and Artificial Intelligence, Letná 9, 042 00 Košice, Slovak Republic
ocelike@ccsun.tuke.sk