

NATIONAL TOPOGRAPHIC DATABASE INFORMATION MEASURING –CASE STUDY IN GLOBAL MAPPING AND WESTERN MAPPING PROJECT

Hong Wang^{a,b,*}, Tao Wang^b, Shanwu Su^b, Yuxiang Li^b

^a Liaoning Technical University, Liaoning Fuxin 123000, China

^b Chinese Academy of Surveying and Mapping, Beijing 100830, China
- (wanghong,wangtao,sushanwu,liyuxiang)@casm.ac.cn

KEY WORDS: Attribute Information, Joint Entropy, Information Capacity, Topographic Database, Geographic Information Measuring, Global Mapping, Western Mapping Project

ABSTRACT:

The western mapping project is using a series of advanced satellite mapping technology to complete the unpopulated 1:50 000 topographical surveys in China. It provides the related products of 1:50 000 scale topographic maps from year 2009 to year 2010 continuously. These applied the new technical manufacture products urgent to carry on the product information content appraisal and quality analysis, thus they could have the conformity with our country's existing database. At the same time the previously constructed topographic databases including 1:250 000 and 1:1 000 000 scales are needed to be updated from large scale to small scale gradually. Entropy is a kind of information quantity evaluation based on information theory and mostly applied to measure of map information. The authors briefly review the entropy application in geographic information measuring, and then provide the conceptual models of core features information in topographic database. The authors do some research works in the case of 1:50 000 topographic database in western mapping project and 1:1 000 000 topographic database in global mapping project. These research works focus on the geometric information such as measure or form information, topographic information such as connection or order, and attribute information. Two case studies have been accomplished. One of them is about the joint entropy measuring for the small scale map and the other is about the form information measuring for large scale ones between the new and old databases. The studies show that the information theory has effective application in geographic information measurement, and there is some quantity relationship among those different scale terrain maps through the information measuring based on information theory and entropy properties. To some extent these relations provide the basis of information quality and quantity control while updating databases.

1. INTRODUCTION

From Shannon (1948) classical information entropy theory, domestic and foreign scientists have applied it to the map information entropy measuring.

Sukhov(1967) proposed the map statistics information content which started from point of pure statistical view. He took the map symbols as the statistical objects, using each kind of symbol probability which appeared in the map, and calculated the information entropy. Neumann(1994) proposed and calculated the entropy of information content for the contour maps using the dual graphs.

Li(2002) evaluated the existing quantitative measures for map information. He proposed a set of new quantitative measures for metric information, topological information and thematic information. He also pointed out voronoi regions of map features play a key role in information quantitative measurement.

Wang (2007) summarized insufficiency of Li's method and used the spatial division property of weighted voronoi diagrams and a new formula of entropy by adopting the ratio between the voronoi regions over the whole area. On the basis of classical measurement of information, he gave a new way to measure the information quantity in multi level elements maps, at the same time, indicated that people should calculate each object separately.

Liu (2008) proposed measures for contour maps about the geometry information and the topological information. Using the voronoi charts and the contour trees, he calculated those two kinds of information for various landform types including mesa, loess, karsts from 1: 50,000 to 1: 250,000 scale maps.

Wang(2009) completely evaluated statistical information, geometric information, and thematic (attribute) information entropy by selecting 9 sheets from settlement layers in 1:1 000 000 national topographic database, then she researched on the series scales maps from 1:250 000 to 1:1 000000 topographic maps and got entropy relationship between them.

Besides the above methods in geographic information measurement there are some entropy applications in map generalization or map visualization. Tian (2009) proposed a progressive approach to the selection of streets base on information loss. This approach minimized the information loss and can select the streets progressively and dynamically.

Wang(2009) introduced a set of concepts named 'Entropy Centre', 'Double Entropy Centres' and 'Entropy Line' based on (geo)metric information and proposed an algorithm of automatic name placement.

It could be inferred from these studies that the information entropy has been applied gradually from classical map information to digital existing geographic information.

Gradually cartographers or geographic information analysts applied the entropy theory into map generalization or information quantity control.

In this paper the authors summarize the various methods based on information theory and propose the conceptual models of core feature information capacity measurement. And then two case studies including the joint attribute entropy measuring in 1:1 000 000 topographic map and information evaluation of 1:50 000 topographic map in western mapping project are presented.

The authors try to test and verify the measuring methods based on the information entropy, by applying to the topographic map information measuring, and evaluate the quality about the newly surveying maps.

2. SUMMARY OF PRESENT GEOGRAPHIC INFORMATION MEASUREMENT BASED ON THE CHARACTERISTICS OF CORE FEATURES

2.1 Basic Geographic Information and Features of Classification

According to our country’s standard (GB/T-13923-2006, basic geographic classification and code information), basic geographic information is defined to be the basic information of common framework of spatial location and spatial analysis. Basic geographic information features are including those geographic elements such as the water system, residential areas, facilities, transportation, pipelines, administration boundary and area, geomorphy, vegetation and soil as well as spatial location fundamentals.

The basic geographic information according to the information geometry character is divided into three kinds of feature -point feature, linear feature and area feature. Regardless of the geographic information based on point, line and area, its basic information characteristics include the following three parts:

The first is spatial distribution or location information: entity-related spatial measure information, spatial distribution, and so on spatial information, which is known as a geometric information;

The second is property information: describe the substance and non-spatial information characteristics, such as the administrative class of the residential area, technical level of roads, and so on;

The third is topological information on the relationship between entities: the descriptions of space relationship information between entities, such as adjacency, intersection, and so on.

As a result, the amount of information on the geographical information measuring and calculation including models and ideas are beginning from information theory based on entropy, are defining characteristics and differences in a certain kind of information, at the same time calculating the entropy by converting them to the basis of probability statistically .

2.2 Summary of geographic information measuring models or methods

In the introduction paragraph some studies about map or geographic information measuring have been carried. The

authors summarized the geographic information measuring models or methods in the following table 1.

In the table 1, the author classified the geographic information measuring into three levels. The first level is according to the three characters of the geographic besides Voronoi based thematic information measuring, because it combined the geometric or spatial distributed information and the attributes of the features. The second level contains various measuring methods applied to the special features, for instance, the contour map information measuring especially about topological information according to the present papers. The extended level is mostly extended by the authors except weighted Voronoi graph based geometric information measuring. The double attributes information measuring will be showed in the following case study in 1: 1 000 000 global mapping project.

First Classification	Second Classification	Extended Classification
Geometric information measuring	General geometric measuring	Weighted entropy
		Density map entropy
		...
	Voronoi graph geometric measuring	Weighted Voronoi graph
		...
	Form or shape measuring	Form or shape measuring
Various feature indexes		
Fractal entropy		
...		
Topological information measuring	Dual graph based measuring	...
	Contour based measuring	...
	Voronoi graph measuring	...

Attribute information measuring	Single attribute	Double attributes

Voronoi graph based thematic information measuring		

Table 1. Classification of models about geographic information measuring

3. JOINT ENTROPY MEASURING IN THE CASE STUDY OF 1:1 000 000 GLOBAL MAPPING

Geographic information has certain attributes, these attributes are not entirely independent, such as a settlement with an administrative class, and often its administrative population is classified into a certain category. Combined calculation of its population and administrative level is more practical and meaningful

3.1 Principle of joint entropy

It is well known that the joint of two statistically independent source of information entropy is to sum up their entropies, but how to calculate the two interrelated source of information entropy? The joint entropy is an entropy measure used in information theory. The joint entropy measures how much entropy is contained in a joint system of two random variables.

If the random variables are X and Y, the joint entropy is written $H(XY)$ (Fu, 2007).

The joint entropy can be defined or used in the definitions of the conditional entropy:

$$H(XY) = H(X) + H(Y|X) \tag{1}$$

Where $H(X)$ describes uncertainty about the value of X, $H(Y/X = x)$ is the entropy of the variable Y conditional on the variable X taking a certain value x, then $H(Y/X)$ is the result of averaging $H(Y/X = x)$ over all possible values x that X may take.

$$H(X) = H(p_1, p_2 \dots p_n) = - \sum_{i=1}^n p_i \log p_i \tag{2}$$

$$H(Y) = H(q_1, q_2 \dots q_m) = - \sum_{j=1}^m q_j \log q_j \tag{3}$$

$$H(Y|X) = \sum_{i=1}^n p_i H(Y|X = x_i) = \sum_{i=1}^n \sum_{j=1}^m p_i p_{ij} \log p_{ij} \tag{4}$$

Where $\sum_{i=1}^n p_i = 1$, $\sum_{j=1}^m p_{ij} = 1, (i = 1, 2, \dots, n)$, $p_{ij} = P(Y = y_j | X = x_i)$

The joint entropy has the following properties:

$$H(XY) \geq H(X), H(XY) \geq H(Y) \tag{5}$$

$$H(XY) \leq H(X) + H(Y) \tag{6}$$

3.2 Data of experiment

Point resident layer is selected from one sheet of maps in the 1: 1 000 000 terrain database. The meaningful attributes or properties of the features in this layer contain the number of people on residential area named GEO_CODE and the administrative level named CLASS. Probabilities of residents are displayed in table 2 according to the population and the administrative class attributes separately. The first left column lists codes for the administration class and the first top row lists the code for each population class

3.3 Results and discussion

Table 3 and Table 4 showed the single entropy for each attribute. In table 5, the greatest value of conditional entropy with gray background displayed that the residents of various administrative classes had the greatest entropy with population classified into code 33030 level, which means the population greater than 100 000 and less than 500 000; in other words, on this sheet the distribution of the residents at this population level spread more averagely or more equally comparing with other residents of different population.

	33010	33020	33030	33040	33050	33060	Total
AD	1						1
AE	2	1					3
AF			5	1			6
AG		3	6				9
AH			6	7	4		17
AI			6	12	117	17	152
AJ			8	4	1		13
AK			2	2	4	1	9
AN			2	11	74	721	808
AO						1	1
BD			2	3	9	1150	1164
CC						18	18
Total	3	4	37	40	209	1908	2201

Table 2. Results of the statistical number of residents in one sheet 1:1 000 000 map

GEO_CODE	Number of residents	P _i	P _i *logP _i
33010	3	0.0014	0.0130
33020	4	0.0018	0.0165
33030	37	0.0168	0.0991
33040	40	0.0182	0.1051
33050	209	0.0950	0.3225
33060	1908	0.8669	0.1787
Total	2201	1.0000	0.7349

Table 3. Entropy for single attribute GEO_CODE

Class	Number of residents	q _j	q _j *logq _j
AD	1	0.0005	0.0050
AE	3	0.0014	0.0130
AF	6	0.0027	0.0232
AG	9	0.0041	0.0324
AH	17	0.0077	0.0542
AI	152	0.0691	0.2663
AJ	13	0.0059	0.0437
AK	9	0.0041	0.0324
AN	808	0.3671	0.5307
AO	1	0.0005	0.0050
BD	1164	0.5289	0.4860
CC	18	0.0082	0.0567
Total	2201	1.0000	1.5489

Table 4. Entropy for single attribute Class

	33010	33020	33030	33040	33050	33060
AD	0.5283					
AE	0.3900	0.5000				
AF			0.3902	0.1330		
AG		0.3113	0.4256			
AH			0.4256	0.4401	0.1092	
AI			0.4256	0.5211	0.4686	0.0607
AJ			0.4777	0.3322	0.0369	
AK			0.2275	0.2161	0.1092	0.0057
AN			0.2275	0.5122	0.5304	0.5305
AO						0.0057
BD			0.2275	0.2803	0.1954	0.4402
CC						0.0635
$H(Y X = x_i)$	0.9183	0.8113	2.8273	2.4349	1.4496	1.1064

Table 5. The conditional entropy

$H(X)=0.7349, H(Y)=1.5489, H(X)+ H(Y)=2.2838$

$H(Y|X)=0.0014*0.9183+0.0018*0.8113+0.0168*2.8273+0.0182*2.4349+0.0950*1.4496+0.8669*1.1064=1.1914$

$H(XY)= H(X)+ H(Y|X)=0.7349+1.1914=1.9263$

Apparently, the experiment results showed that the joint entropy is greater than the single attribute entropy individually and less than the sum of the both attributes entropies. It indicated that there is correlation between the two attributes of the residents point layer. It was coincident with the actual facts. The experiment verified that the joint entropy algorithm was reliable and could reflect the properties information of geographic entities.

We could calculate the other conditional entropy and the result of joint entropy is the same. We also could analyse it and find some more interesting things while doing these joint entropy measuring. The second edition of specifications of Global Map has been introduced the population attributes for population centres.

4. EVALUATION OF THE AMOUNT OF GEOGRAPHIC INFORMATION FOR THE NEW 1:50 000 TOPOGRAPHIC MAPS

The western mapping project is using a series of advanced satellite mapping technology to complete the unpopulated 1:50 000 topographical surveys in China. It provides the related products of 1:50 000 scale topographic maps from year 2009 to year 2010 continuously. The old maps related have 1:100 000

scale. The project signification is doubted while applied to be established. Some people worried about the amount of geographic information in the future maps at the beginning process of the project. The test was designed to verify the effect of some models showed in table 1 and to test the amount of the new maps whether it has been increased than the old one.

In table 1 the authors summarized the geometric information measuring and shape or form measuring has the most important signification for linear or area features. With the topographic maps the contours which describe the relief or landscape need to be measured the geometric information apparently. On the other hand the western area of China is unpopulated region and the core features on topographic map are certainly terrain features.

4.1 Experiment data of sample area

The sample data are selected from the production of the first dataset. In order to contrast the new map with the old one, the same area covered with 1:100 000 topographic maps are also selected. At the same time, those areas with satellite mapping besides aero mapping are selected. Some sketch maps about contours from these sample areas are showed in the following figure 1.

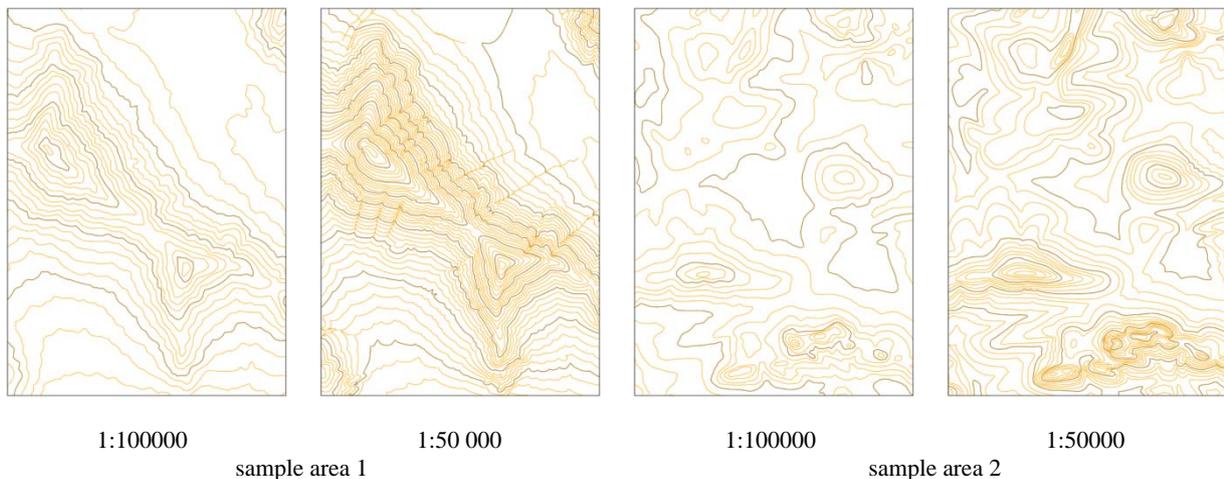


Figure 1. Maps of two sample areas selected for information measuring

4.2 Principle and approaches of information measuring

While researching on the geographic information measuring, the experiment indicated that establishing topological relationships and measuring the amount of information for single layer has little practical significance (Wang, 2009). Therefore in this experiment the topological information measuring for contours haven't been done. The experiment focused on the geometric information measuring because it is the most important and concerned information.

Geometric information measuring

Geometric information is the most important factor to decide the amount of all kinds of information including topologic or attribute information. The measuring models can be classified into three levels including 1.general geometric information measuring such as length or area, 2.voronoi graph based measuring and 3.shape or form information measuring.

Length information measuring: for contours, length for each contour can be gotten automatically through GIS and then calculate the length entropy by considering the percentage of each contour in the total length of map as the probability.

Voronoi graph based geometric information measuring: At first establishing the voronoi graph for the contour map and considering the percentage of voronoi area of each contour in

the total areas of map as the probability and then calculating the entropy.

Shape or form information measuring: The more complex form of linear features, the greater the degree of bending, the greater the amount of information it provides. Measuring the shape complexity of linear elements starts in two ways; one is a simplified method, based on the characteristic points of linear elements. The more characteristic points, the more complex of the shape patterns can provide the larger amount of information. Another is the introduction of fractal geometry in the concept of fractal dimension. Calculating fractal dimension for each linear shape and calculating the entropy of the obtained fractal dimension may be a model. The experiment selected the previous method through Douglas_Peucker algorithm.

Attribute Information measuring

The experiment selected the classification and elevation for contour attributes information measuring because both of them are the main attributes. The approach for attribute information measuring is the same as statistical information measuring. Count the contours by single attribute category and using the entropy algorithm showed in equation 2 or 3. Then the joint entropy could be calculated by equation 1 and 4.

4.3 Results and discussion

Sample	Map scale	Threshold (10 metres)		Threshold (5 metres)	
		Characteristic points(N)	Entropy=log(N)	Characteristic points(N)	Entropy=log(N)
sample 1	1:100 000	30692	14.91	42616	15.38
	1:50 000	73011	16.16	106477	16.70
	Increment	137.88%	8.39%	149.85%	8.59%
sample 2	1:100 000	18073	14.14	27918	14.77
	1:50 000	32508	14.99	32513	14.99
	Increment	79.87%	5.99%	16.46%	1.49%

Table 6. Shape information measuring by characteristic points of contour lines

	Map scale	Length entropy	Classification entropy	Elevation entropy
sample 1	1:100 000	7.75	1.22	5.30
	1:50 000	8.19	1.65	5.73
	Increment	5.68%	35.25%	8.11%
sample 2	1:100 000	7.93	1.26	4.70
	1:50 000	7.22	1.28	5.20
	Increment	-8.95%	1.59%	10.64%

Table 7. Contrast table of geometric (length) and two attributes for contour lines

Table 6 shows the shape or form information measuring through characteristic point's entropy. There is a difference between the amounts of information for the same range of maps with different scales. It indicates that in the sample1 the two contrast maps have more differences in shape information. In fact it is easily deduced from figure 1, and the new 1:50 000 topographic map has greater information than the old 1: 100 000 map. More differences from the geometric information, in particular the morphological differences are more obvious. The classification entropy is different from table 7. In sample 1, the contours are more complicated and the attributes entropy is greater. The length entropy for sample 2 seems have some problems. The new 1:50 000 scale map has less length entropy than the old 1:100 000 scale map. Its form entropy changed little comparing

with old scale map. New map therefore with this kind of doubt should be checked back and examined. Maybe the generalization while surveying and mapping is without measure.

5. CONCLUSIONS AND FUTUREWORKS

In this paper entropy theory and applications in geographic information measuring are briefly reviewed. Some extended information measurements are proposed and tested by two case studies in practical projects. The studies show that the information theory has effective application in geographic information measurement, and there is some quantity relationship among those different scale terrain maps through the information measuring based on information theory and

entropy properties. In the future the application of the entropy theory need to be expanded, for example, automatically calculating the entropy or entropy density for each layer while zoom in or out and deciding how many layers or features should be reserved or abandoned. Up to now the information theory and entropy application mostly focus on the syntax information of geographic features and the semantic information should be considered in the further work.

References

Fu, Z. Y. ,2007. Information Theory- Principles and Application, Publishing House of Electronics Industry, pp.33-35.

Li, Z. L., 2002. Quantitative measures for spatial information of maps. International Journal of Geographical Information Science, Vol.16 No.7, pp.699–709.

Liu , W. K., etc., 2008. Quantitative Measures for Spatial Information of Contour Maps. Geometrics and Information Science of Wuhan University, Vol.33 No.2, pp.157-159.

Neumann J., 1994. The topological information content of a map/ an attempt at a rehabilitation of information theory in cartography. Cartographica, Vol.31,pp. 26-34.

Shannon C. E.,1948. A mathematical theory of communication. The Bell System Technical Journal, Vol.27,pp.379-423&623-656.

Sukhov V. I., 1967. Information capacity of a map entropy. Geodesy and Aero photography, X, pp.212-215.

Tian J.,etc.,2009. Progressive Selection Approach of Streets Based on Information Loss. Geomatics and Information Science of Wuhan University, Vol.34 No.3,pp.363-365.

Wang,H.,etc.,2009. Comparing the amount of information for multi-scale national topographic database. Science of Surveying and Mapping, Vol-34 No-6,pp.113-114.

Wang, S. Y., 2007. A measurement method of geographical information considering multi-level map feature. Science of Surveying and Mapping, Vol.32 No.4, pp.60-62.

Wang Zh.,etc.,2009 . Automatic Name Placement of Area Feature:A Metric Information Approach. Acta Geodaetica et Cartographica Sinica, Vol.38, No.2,pp.183-188.

Acknowledgements

The study is supported by the Western Mapping Project. The authors would like to give special thanks to Dr. Q. Yan and Dr.Q.Y.Li for many helpful suggestions, to Prof. T.H. Ai for his ideas and advices on research work about entropy application and future direction. Mr.R.Zhao helped the collection of sample data.