QUANTITATIVE MEASUREMENT FOR GEOSPATIAL INFORMATION CONTENT

KYAW Sann Oo
GIS Application Center, Asian Institute of Technology, PO Box 4,
Klong Luang, Pathumthani, Thailand. 12120
Tel: (66)-2-524-6403 Fax: (66)-2-524-6147
E-mail: kyawsannoo@ait.ac.th

Xiaoyong CHEN
Remote Sensing & Geographic Information Systems (RS&GIS)
School of Advanced Technologies, Asian Institute of Technology, PO Box 4,
Klong Luang, Pathumthani, Thailand. 12120
Tel: (66)-2-524-5597 Fax: (66)-2-524-6407
E-mail: xychen@ait.ac.th

KEY WORDS: Geospatial Information, Entropy, Distribution, Size, Shape

ABSTRACT

In this information revolution period, everything has representational medias. GIS technology is one of the best ways to represent the location based geospatial information. In this host technology, the art of information measuring was being needed for representation the information. Prior, entropy is one to represent information content in the communication technology and quantify probabilistic information distribution. But communication theory don't have geospatial component. There is being needed a completely new line of thought. It is a commonplace that a geospatial representation contains geometric information, thematic information and events distribution. Therefore, a set of measures needs to be developed for geospatial representation. This study will apply the information theory to do the native measurement for geospatial information to represent information content. First, the study will evaluate the existing measurement representation based on then evaluation results, a set of new quantitative measure is then introduced. Finally these new measures are experimentally evaluated.

1. INTRODUCTION

Shannon (1948) was the first person to introduce entropy in the quantification of information by employed the probabilistic concept in modeling message communication and he believed that a particular message is one element from a set of all possible messages. If the number of messages in this set is finite, then this number or any monotonic function of this number can be regarded as a measure of the information when one message is chosen from the set, all choices being equally likely. Based upon this assumption, information can be modeled as a probabilistic process. He then introduced the concept of '*entropy*' to measure the information content.

The pioneering work in quantitative measurement of map information was done by Sukhov (1967, 1970), who considered the statistics of different types of symbols represented on a map. That's the entropy of these symbols are computed. Later, Neumann (1987, 1994) did some work on topological information of maps. Quantitative measures for map information has been used for comparing the information contents between maps and images, maps at different scale, evaluation of map design and so on (Knopfli 1983, Bjorke 1996). In the Zhilin LI (2001) measures, Voronoi region of map features play a key role, which not only offer metric information but also some sort of thematic and topological information. This study is followed by an evaluation of existing measures. Based on the evaluation results, a set of geospatial analysis based quantitative measures is then introduced. Finally, some conclusions area made.

2. EVALUATION OF EXISTING MEASURES

2.1 Shannon (1948)

He introduced the concept of 'entropy' to measure the information content. Let X be the random message variable, the probabilities of different message choices are P_1, P_2, … P_i, … P_n. The entropy of X can be computed as follows:

$$H(X) = H(P_1, P_2, \ldots, P_n) = -\sum_{i=1}^{n} P_i \ln(P_i) \quad (1)$$

Statistically speaking, H(X) tells how much uncertainty the variable X has on average. When the value of X is certain, $P_i = 1$, then H(X)=0. H(X) is at its maximum when all messages have equal probability.

2.2 Sukhov (1967, 1970)

He has adopted the entropy concept for cartographic communication. In such a work, only the number of each type of symbols represented on a map is taken into account. According to definitions both map (figure 1) have same amount of information (H=1.5). But symbols distribution are very different.

Figure 1: map (a) and map (b)

2.3 Neumann (1994)

He proposed a method to estimate the topological information of a map. The method consists of two steps: (a) to classify the vertices according to some rules, such as their neighbouring relation and so on, to form a dual graph, and (b) to compute the entropy with Equations (2) and (3). Both below maps have same amount of information (H=1.38). But Figure 2b more complex than figure 2a.

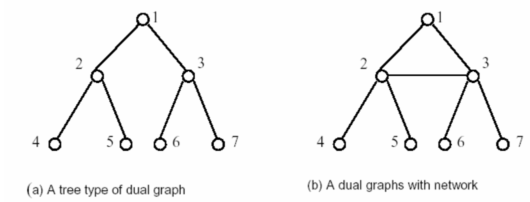
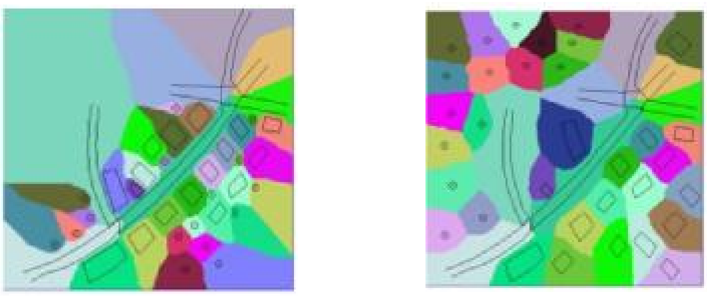


Figure 2: Dual graph

2.4 Zhilin LI (2001)



He developed a set of measures based on voronoi region which represent not only offer metric information but also some sort of thematic and topological information.

Figure 3: Features' Voronoi regions

2.4.1 Metric information

Metric information here considers the space occupied by map symbols only. In this case, an analogy to the entropy of binary image is used. That is, if the space occupies by each symbols is similar, the map has larger amount of information. If the variation is very large, the amount is smaller.

For example, the two maps shown in Figure 4 have different amount of metric information although both are tessellated by 9 polygons. The map in Figure 4(b) has the maximum H(M) for any tessellation into 9 polygons.

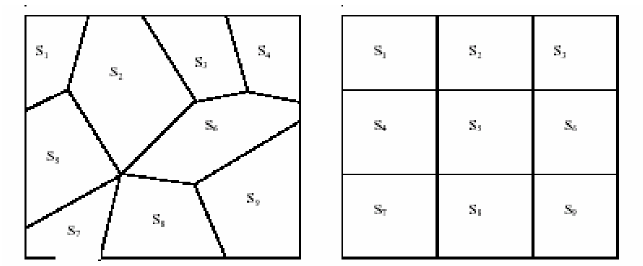


Figure 4: Voronoi's Metric

In the case of map, it is clear that for the same number of feature, the entropy will be larger is the symbols are more evenly distributed. However, it is clear that such entropy is related to the number of map symbols and thus it would be not convenient to compare two maps with different number of symbols.

2.4.2 Topological information

The construction of dual graph for map features is a difficult task because the vast majority of map features are disjoint. However, with the Vonoroi region, all features have been connected together to form a tessellation. The generation of dual graph for map features could be replaced by the dual graph of the Voronoi region of these features. This is illustrated in Figure 5. Figure 5(a) is the Voronoi region and Figure 5(b) is the corresponding dual graph.

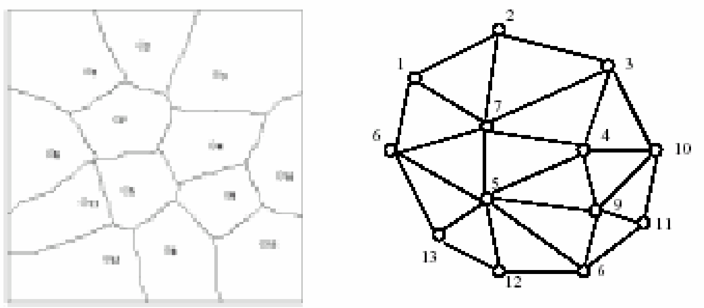


Figure 5: A Voronoi diagram and its dual graph

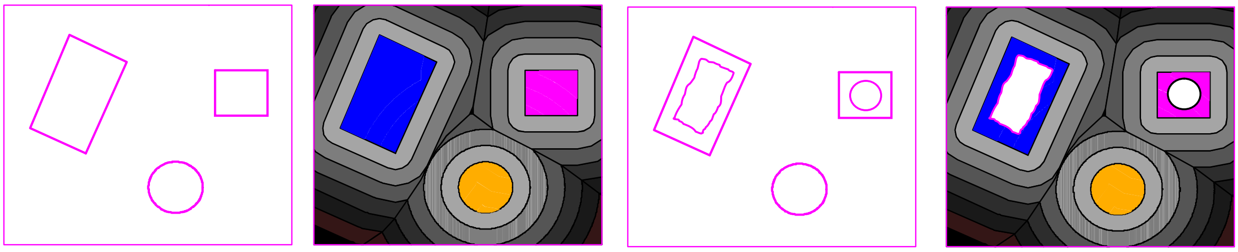
2.4.3 Thematic information

Thematic information related to the thematic types of features. It is understandable that, if a symbol has all neighbours with the same thematic type, then the importance of this symbol is very low, in terms of thematic meaning. In the other hand, if a symbol has neighbours with different thematic types, it should be regarded as having higher thematic information.

3. NEW QUANTITATIVE MEASURES FOR GEOSPATIAL INFORMATION CONTENT

3.1 Sequential Dilating Voronoi Regions

Sequential dilation represent not only offer metric, topological and thematic information but also think about distribution, shape and size of features (see figure 6).



(a) (b) (c) (d)

Figure 6: Sequential dilating on simple features and complex features

4. EXPERIMENTS

The algorithm is tested on the geospatial information for two different scales and different area, 1:1000 for urban area and 1:100000 for regional area to compare the quantitative measurement of geospatial information or “entropy” of that and checking the stabilization of transformation and distribution of geospatial information. The study flow is shown in figure 7.

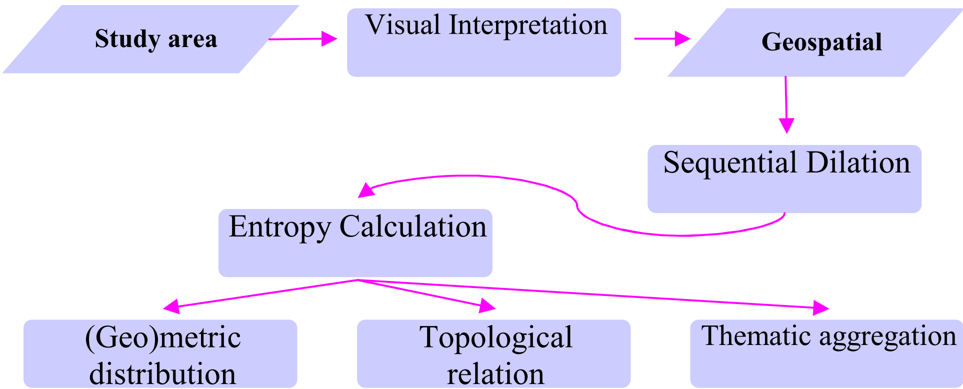


Figure 7: Methodology flow

The basic formula to calculate the entropy of geo(metric) features, thematic features, topology and thematic topology features are as follows:

$$H(X) = H(P_1, P_2, \ldots, P_n) = -\sum_{i=1}^{n} P_i \log_2(P_i) \quad \text{-----------------------------(2)}$$

The choice of a logarithmic base corresponds to the choice of a unit for measuring information. If base 2 is used the resulting units may be called binary digits, or more briefly *bits*, a word suggested by J.W.Tukey.

4.1 Geo(metric) features

In geo(metric) features, we think about distribution, shape and size. The entropy of this information can be calculated as follows:

$$H_{dis}(M) = H(P_1, P_2, \ldots, P_n) = -\sum_{i=1}^{n} \frac{S_i(dist)}{S(dist)} \left(\log_2 S_i(dist) - \log_2 S(dist) \right) \quad \text{-----------------------(3)}$$

Men made objects	Natural objects

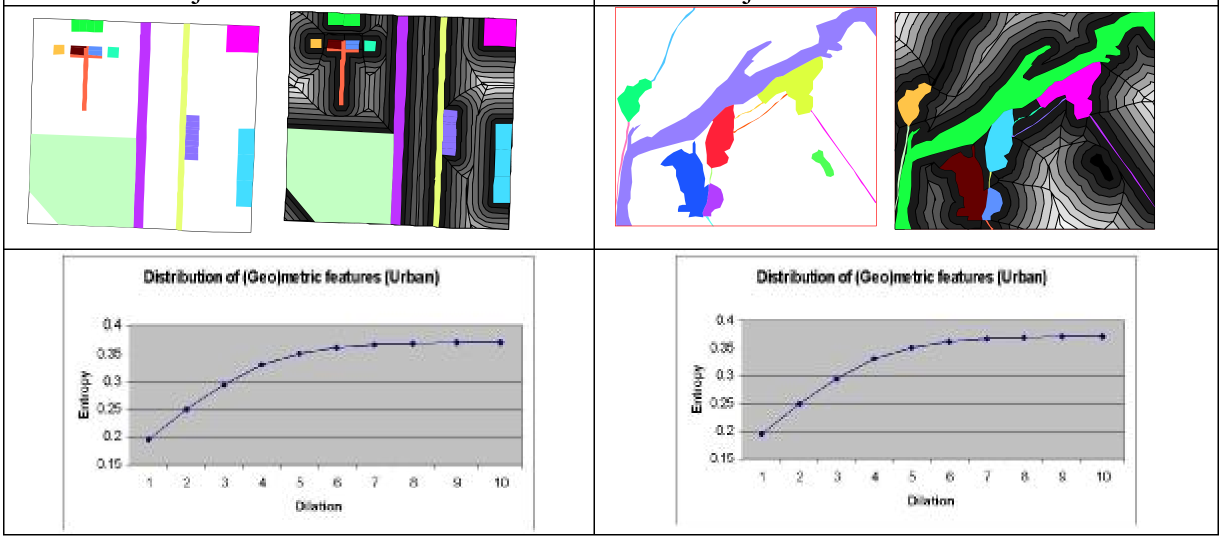


Figure 8: distribution of geo(metric) features

4.2 Thematic features

This geospatial information is based on separate culture and objects scattering. The entropy of thematic information can be calculated as follows:

$$H_i(TM) = H(P_1, P_2, \dots, P_{M_i}) = -\sum_{j=1}^{M_i} \frac{n_j}{N_j} \log_2\left(\frac{n_j}{N_j}\right) \quad \text{------------------------(4)}$$

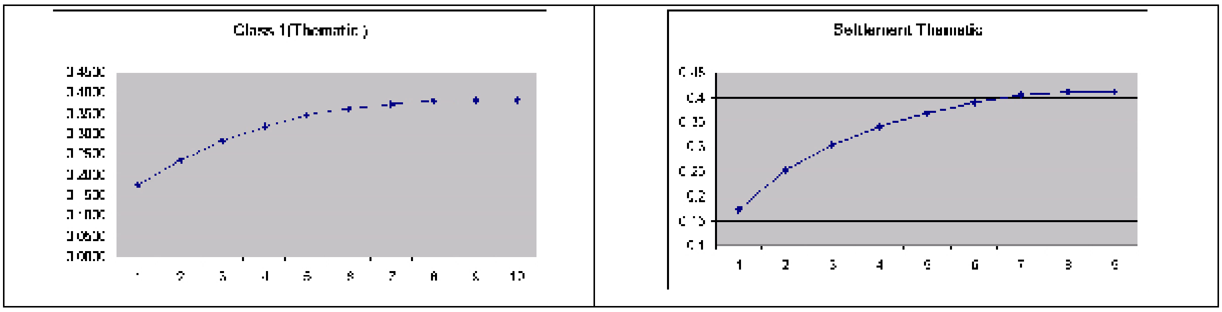


Figure 9: distribution of thematic features

4.3 Topology features

Different cultures and aggregation of objects can be defined as topological features. The entropy of topology information can be calculated as follows:

$$H_i(TP) = H(P_1, P_2, \dots, P_{M_i}) = -\sum_{j=1}^{M_i} \frac{N_s}{N_T} \log_2\left(\frac{N_s}{N_T}\right) \quad \text{-------------(5)}$$

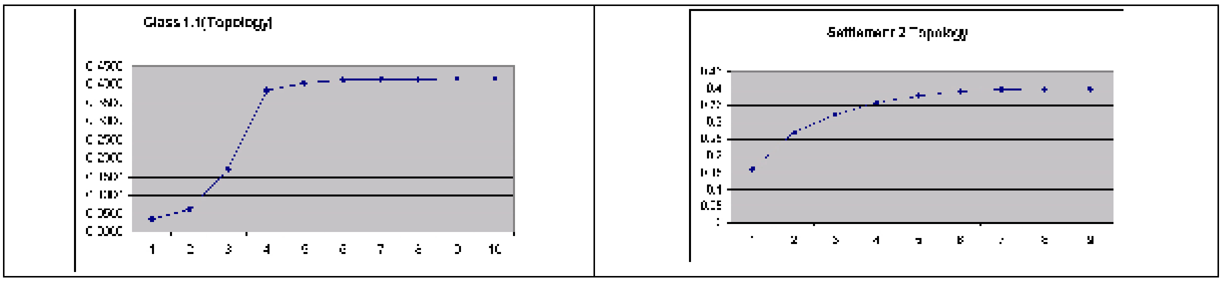


Figure 10: distribution of topology features

4.4 Thematic Topology features

Thematic topology information here considers as the topology information based on the thematic features. The entropy of the thematic topology information, donated as H(TT), can then be defined as follows:

$$H_i(TT) = H(P_1, P_2, \ldots, P_{M_i}) = -\sum_{j=1}^{M_i} \frac{n_{j\,H(TP)}}{N_j} \log_2\left(\frac{n_{j\,H(TP)}}{N_j}\right) \quad \text{----------------------(6)}$$

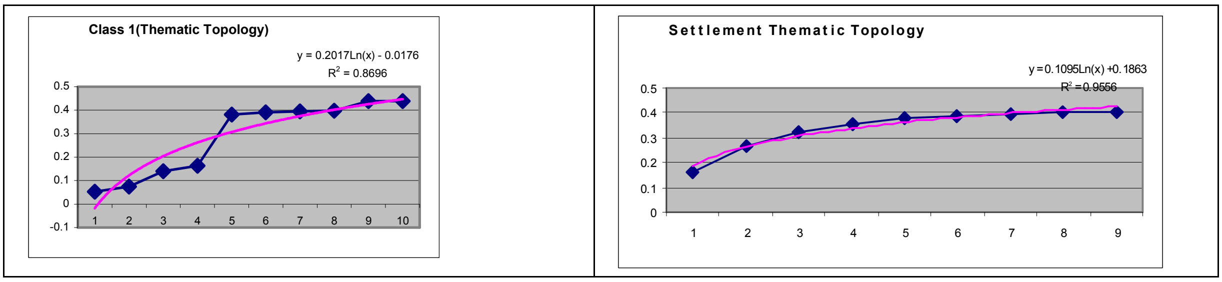


Figure 11: distribution of thematic topology features

5. QUANTITATIVE AND QUALITATIVE DATA ANALYSIS

The distribution of geospatial data for defined area can easily calculate by using new formula as new measurement. To get the comparative outcome for natural objects and urban men made objects. In particular, the suggestion is geospatial information are totally base on the geo(metric), thematic, topology and thematic topology features entropy.

Existing quantitative measures for map information have been pointed out that voronoi regions offer metric information, topological information and thematic information. A set of new quantitative measures is proposed, i.e for distribution, shape and size of sequential dilated voronoi region.

Result show that geo(metric), thematic, topology and thematic topology of sequential voronoi region information representation is carrying great weight than statistical information.

REFRENCE

Bjørke, J.T. (1996). ***'Framework for entropy-based map evaluation'***. Cartography and Geographical Information Systems, 23(2): 78-95.

Knopfli, R. (1983). ***'Communication theory and generalization. Graphic communication and design in contemporary cartography'***, edited by D.R.F. Taylor. John Wiley & Sons Ltd., 177-218.

Li, Z., Huang, P. (2002). ***'Quantitative measures for spatial information of maps'***. *International Geographical Information Science. 16/7, 699-709.*

Serra , J. (1992). ***'Image Analysis and Mathematical Morphology'***. Academic Press, 271-296.

Weaver, W. and Shannon, C.E. (1949). Republished in paperback 1963. ***'The Mathematical Theory of Communication'***. Urbana, Illinois: University of Illinois Press.

Chantanee Chansawang, ***'Fractal dimension base measurement of spatial information for multi-scale representation'***, Asian Institute of Technology, Bangkok, 2002.