Computers & Geosciences 34 (2008) 939–954

COMPUTERS & GEOSCIENCES

www.elsevier.com/locate/cageo

An algorithm for point cluster generalization based on the Voronoi diagram

Haowen Yan[a,b,*], Robert Weibel[b]

[a]School of Mathematics, Physics and Software Engineering, Lanzhou Jiaotong University, Lanzhou 730070, Gansu Province, PR China
[b]GIS Division, Department of Geography, University of Zurich, Zurich, Switzerland

Received 1 December 2006; received in revised form 17 July 2007; accepted 19 July 2007

Abstract

This paper presents an algorithm for point cluster generalization. Four types of information, i.e. statistical, thematic, topological, and metric information are considered, and measures are selected to describe corresponding types of information quantitatively in the algorithm, i.e. the number of points for statistical information, the importance value for thematic information, the Voronoi neighbors for topological information, and the distribution range and relative local density for metric information. Based on these measures, an algorithm for point cluster generalization is developed. Firstly, point clusters are triangulated and a border polygon of the point clusters is obtained. By the border polygon, some pseudo points are added to the original point clusters to form a new point set and a range polygon that encloses all original points is constructed. Secondly, the Voronoi polygons of the new point set are computed in order to obtain the so-called relative local density of each point. Further, the selection probability of each point is computed using its relative local density and importance value, and then mark those will-be-deleted points as 'deleted' according to their selection probabilities and Voronoi neighboring relations. Thirdly, if the number of retained points does not satisfy that computed by the Radical Law, physically delete the points marked as 'deleted' forming a new point set, and the second step is repeated; else physically deleted pseudo points and the points marked as 'deleted', and the generalized point clusters are achieved. Owing to the use of the Voronoi diagram the algorithm is parameter free and fully automatic. As our experiments show, it can be used in the generalization of point features arranged in clusters such as thematic dot maps and control points on cartographic maps.

Keywords: Point clusters; Map generalization; Measures; Algorithms; Voronoi diagrams

1. Introduction

Scale reduction from source maps to target maps inevitably leads to conflict and congestion of map symbols. To make the maps legible, appropriate operations (e.g. selection, simplification, aggregation, etc.) must be employed to simplify map features. This process is called map generalization in the community of cartography and geographic

*Corresponding author at: GIS Division, School of Mathematics, Physics and Software Engineering, Lanzhou Jiaotong University, Lanzhou 730070, Gansu Province, PR China. Fax: +86 931 4938401.

E-mail addresses: dlkc@mail.lzjtu.cn (H. Yan), weibel@geo.unizh.ch (R. Weibel).

0098-3004/$ - see front matter
doi:10.1016/j.cageo.2007.07.008

information systems (GIS). Nowadays, ''the widespread use of geographic information in computers in the context of GIS has brought with it the demand for automation of map generalization'' (Jones and Ware, 2005, p. 859). Map features may be categorized into three types according to their geometric characteristics of map symbols, i.e. point, linear, and areal, and the algorithms have been developed for the generalization of the three types of features. They include algorithms for point feature generalization (e.g. Langran and Poicker, 1986; van Kreveld et al., 1997; Burghardt et al., 2004; De Berg et al., 2004), line feature generalization (e.g. Mustiere, 2005), and areal feature generalization (e.g. Barrault et al., 2001; Ruas, 2001; Galanda and Weibel, 2002; Sester, 2005).

This paper will focus on approaches for point feature generalization, and aims to propose a new algorithm that may transmit important types of information correctly in the process of map generalization.

It must be clarified before our further discussion that in this paper:

(1) The generalization of an individual point is not of concern. We place emphasis upon the holistic distribution and configuration of point clusters.
(2) Every point has its coordinate pair (x,y) and an importance value (i) for denoting its importance degree.

It is common sense that the main purpose of map generalization is to transmit important types of information from larger scale maps to smaller scale maps (Bjørke, 1996) with the reduction of map features. Hence, what types of information are contained in map features and need to be transmitted in the process of generalization is an essential problem. So, this issue will be discussed in detail in Section 2 based on some pioneering achievements of communication theory. By the answer to this issue, the existing algorithms for point cluster generalization are analyzed so that their advantages and disadvantages are revealed (Section 3). After this, the new algorithm is presented in detail (Section 4), and a method for evaluating the new algorithm is given (Section 5). To illustrate the soundness of the algorithm, some experiments are shown and discussed (Section 6). The article ends with conclusions and an outlook on further research (Section 7).

2. Types of information contained in point clusters

As stated in the introduction, some pioneering research achievements have been made for answering what types of information are contained in map features and how to quantify this information. They will be discussed and summarized in the following section.

2.1. Information contained in point clusters

The concept of information was first used in communication theory (Shannon and Weaver, 1949). 'Entropy' is a quantitative measure for the information content contained in a message. The concept has also been introduced to the cartographic community by Sukhov (1967, 1970), who considered the statistics of different types of symbols represented on a map. The entropy of these symbols is computed using the proportion of each type of symbol to the total number of symbols as the probability. Such a type of information is purely statistical and the spatial distribution of the symbols has not been considered, as pointed out by Li and Huang (2002). Neumann (1994) introduced the concept of topological information by considering the connectivity and adjacency between map features. Bjørke (1996) considered three types of information, i.e. positional, metric, and topological. The positional information of a map considers all the occurrences of the map features as unique events and all map events are equally probable. That is, the positional entropy is simply computed by counting the number of map features. It is obvious that the positional information discussed by Bjørke (1996) is the same as the statistical information by Sukhov (1967, 1970), so we use the term statistical information in the following sections. The metric information considers the variation of the distance between map features. The topological information considers different types of relations between the map features. Li and Huang (2002) identified metric, thematic, and topological information. The metric information considers the size of the area occupied by the Voronoi diagram of each map feature, instead of the distances between map features. The thematic information considers categorical difference of neighboring map features and the topological information considers the connectivity and adjacency relations between neighboring map features, which is similar to the one by Neumann (1994), but with a different mathematical definition.