Expert Systems with Applications 39 (2012) 9376–9388

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Geographic knowledge discovery from Web Map segmentation through generalized Voronoi diagrams

Ickjai Lee *, Christopher Torpelund-Bruin

School of Business (IT), James Cook University, Cairns, QLD 4870, Australia

ARTICLE INFO

Keywords:
Web map
Segmentation
Topology
Voronoi
Sequential-scan
Complex primitives
Weighted
Obstacles

ABSTRACT

Web maps have become an important decision making tool for our daily lives. We propose a flexible Web Map segmentation method in order to better use them for decision makings. We extend the distance transform algorithm to include complex primitives (point, line and area), Minkowski metrics, different weights and obstacles. The algorithms and proof are explained thoroughly and illustrated. Efficiency and error for the novel algorithms are also detailed. Finally, the usefulness of the algorithms is demonstrated through a series of real-life case studies.

1. Introduction

All the challenges of the modern world, from global warming, economic downturns and spread of disease, to deployment of emergency services are related to human activities, and to confront them, we must understand, analyze and discover the ever changing patterns of human activities. All these challenges have a geospatial dimension, but more importantly, they are leaving more and more trails of data as the world turns digital and the Web takes a more prevalent role as the infrastructure for all sorts of commercial, industrial and political activities. There is a great opportunity to exploit the large traces of georeferenced human behavior that now happen on the Web and our ability to record them. Moreover, the turn to the Web 2.0, the Web as a platform, with massive geospatial content generated by users suggests even far more user-oriented data than ever. These every-increasing georeferenced Web datasets are further fueled by various Web maps from Web Map Service (WMS) such as Google Map (http://maps.google.com), Google Earth (http://earth.google.com), and Open Street Map (http://www.openstreetmap.org). Could this huge repository of geospatial information along with Web maps be analyzed in terms that space becomes informative for gaining insights into the patterns of those users? One of the main obstacles that hinders exploration of these georeferenced Web datasets is the deficiency of topological relationship within datasets. Many interesting topological patterns, such as a hospital having a university in its vicinity tends to attract more patients, cannot be detected with raw georeferenced data. Topology building for georeferenced Web data must be preceded for more enriched geospatial data mining. Segmentation is one approach to build toplogy (Chen, Wang, & Feng, 2010; Hanafizadeh & Mirzazadeh, 2011; Seng & Lai, 2010). The Voronoi diagram (VD) is a popular model to capture natural territories of events (districting, segmentation and tessellation) (Gold, 1991). It tessellates the space and models estimates of the service areas, and captures impact areas of a certain target. It has been widely used in various geoinformatics applications (Okabe, Boots, Sugihara, & Chiu, 2000). The problem with current districting techniques with Voronoi diagrams is their lack of generalizations. Namely, their lack of support representing real world elements which can consist of points, lines, areas and obstacles – all with their own significance. Increased accuracy on geospatial and temporal dimensions on data can lead to increased accuracy on the patterns produced by geospatial data mining techniques. This accuracy to the real world patterns can then possibly lead to new knowledge discoveries. Generalized Voronoi diagrams (GVDs) are generalizations of the ordinary VD to various metrics, different weights, higher order, in the presence of obstacles, and complex data types (point, line and area). Due to the computational difficulty and complexity, vector GVDs have attracted less attention in the literature (Lee & Lee, 2009; Mu, 2004; Okabe et al., 2000).

This paper describes the project dedicated to engineering efficient and effective GVD algorithms for use in geographic knowledge discovery (GKD) from GeoWeb process, which aims to capture new knowledge from spatially referenced data retrieved from distributed Web 2.0 technologies. The GKD from GeoWeb process is summarized in Fig. 1, with the blue highlighted areas describing the contents of this paper. In this paper, we propose novel GVD algorithms that support:

* Corresponding author.
E-mail addresses: Ickjai.Lee@jcu.edu.au (I. Lee), christopher.torpelund@my.jcu.edu.au (C. Torpelund-Bruin).

0957-4174/$ - see front matter Crown Copyright © 2012 Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.02.129

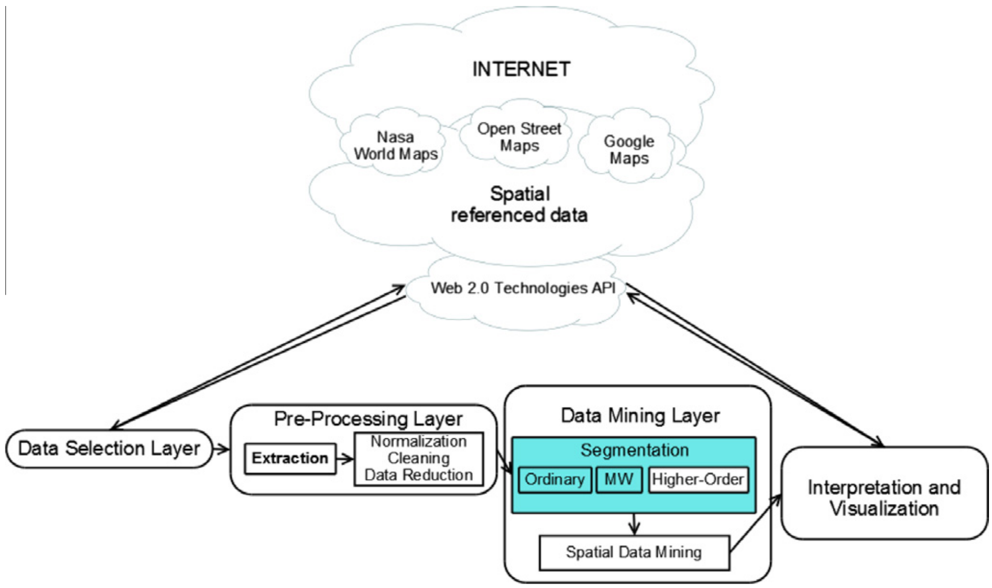

Fig. 1. The GKD from GeoWeb model. The blue highlighted area shows the focus of this paper. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

- Ordinary (modeling normal distriction).
- Weighted (modeling different significance).
- Various metrics (modeling various geographic areas).
- Point, line and polygon primitives of interest.
- Line and polygon obstacles.
- Any combination of these generalized cases.

The proposed algorithms are designed to be as efficient and flexible as possible because of the potential massive geospatial datasets that they will need to generate topological information for. Voronoi diagram algorithms have traditionally been designed as vector-based which have efficient time complexities at the expense of a lack of diversity. This encompasses generators being typically limited to points, the underlying metric limited to the Euclidean metric and weights of generators assumed to be invariant. However, our project aims to accurately produce topological maps and therefore requires a different approach. The need for a robust and versatile method with the ability to accurately model the diverse and various components associated with real world geospatial analysis has been gaining popularity in the scientific community. This has lead to the exploration of efficient raster-based methods for GVD.

This paper introduces a flexible sequential-scan based districting algorithms extended from the distance transform algorithm (Shih & Wu, 2004). It proposes a three-scan algorithm capable of handling generators of complex primitives (point, line and area), Minkowski metrics and obstacles in $O(F)$ time, where F is the number of pixels. Primitives with weights and obstacles can be handled in the plane in $O(F \times G)$ time, where G is the set of generators.

This paper is structured as follows. Section 2 introduces the properties of Voronoi diagrams. Section 3 introduces distance transforms and districting algorithms. Section 4 describes our three-scan propagated relative distance algorithm for weighted primitives. Section 4.2 extends the algorithm for use with obstacles. Section 5.1 defines a proof of the basis of the algorithms. Section 5.2 looks into the average error produced by the algorithms. Section 5.3 shows average performance of the algorithms. Section 6 introduces case studies and possible applications for the algorithms. Finally, Section 7 concludes the paper with an overview of the project and future considerations.

2. Voronoi diagram districting

Let $P = \{p_1, p_2, \ldots, p_k\}$ be a set of *generator* points of interest in the plane in R^m space. For any point p in the plane, $dist(p,p_i)$ denotes the distance from point p to a generator point p_i. The distance metric can be of Manhattan, Chessboard, Euclidean or another metric and the resulting dominance region of p_i over p_j can be defined as:

$$dom(p_i, p_j) = \{p | dist(p, p_i) \leqslant dist(p, p_j)\}. \tag{1}$$

For the generator point p_i, the Voronoi region of p_i can be defined by:

$$V(p_i) = \cap_{j \neq i} dom(p_i, p_j). \tag{2}$$

The partition into subsequent Voronoi regions $V(p_1), V(p_2), \ldots, V(p_k)$ is called the *generalized Voronoi diagram*. The bounding regions of $V(p_i)$ are known Voronoi boundaries and depending on the primitive used as the generator, such as points, lines or polygons and the metric space used, will result in a series of polygons and arcs made up of lines and Bezier curves. These geospatial dominance regions provide natural neighbor relations that are crucial for many topological queries in geospatial modeling and analysis. The most popular distance metric *dist* is the Euclidean distance, which is an instance of the Minkowski metric. However, in urban geography the Manhattan distance metric, another instance of the Minkowski metric, better approximates real world situations (Krause, 1975). The Minkowski distance is described below:

$$dL_p(p, g_i) = \left[\sum_{j=1}^{m} |x_j - x_{ij}|^n \right]^{1/n} \quad (1 \leqslant n \leqslant \infty), \tag{3}$$

where $(x_1, x_2, \ldots, x_m)$ and $(x_{i1}, x_{i2}, \ldots, x_{im})$ are the Cartesian coordinates of p and p_i, respectively. The parameter n can be in the range of $1 \leqslant n \leqslant \infty$. Different values of n give different distance metrics. Different distance metrics influence Voronoi diagrams by changing the Voronoi boundaries that make up each dominance region. When $n = 1$, then $dL_1(p, p_i) = \sum_{j=1}^{m} |x_j - x_{ij}|$ is the Manhattan metric. The Minkowski metric becomes the Euclidean metric when $n = 2$. If $n = \infty$, then the Minkowski metric becomes $dL_\infty(p, p_i) = max_j |x_j - x_{ij}|$, which is called the chessboard metric.