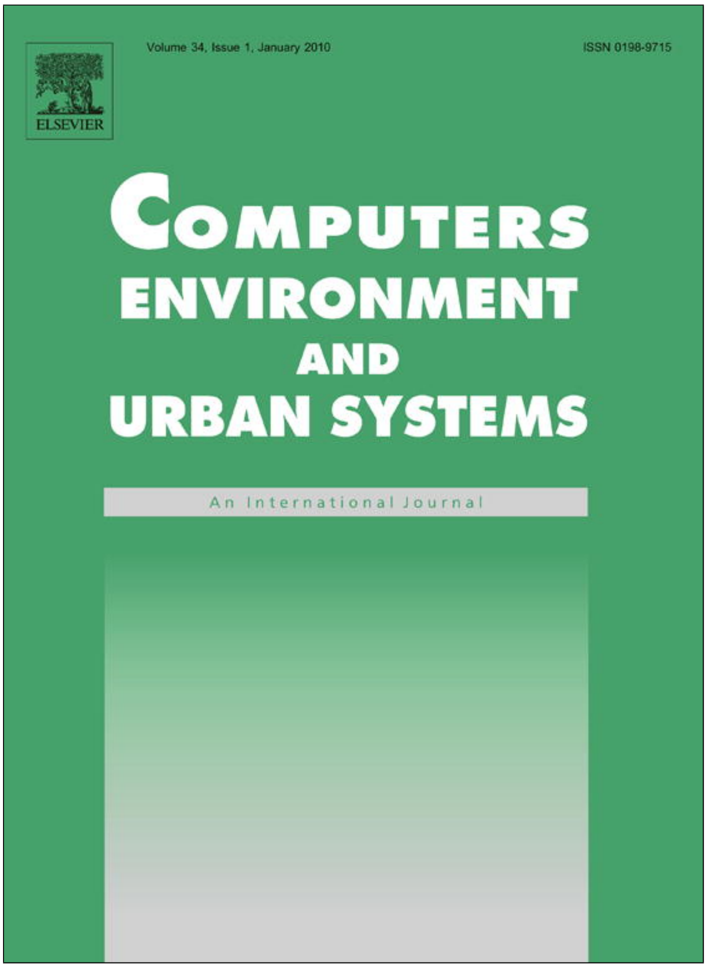

Volume 34, Issue 1, January 2010
ISSN 0198-9715
ELSEVIER
COMPUTERS
ENVIRONMENT
AND
URBAN SYSTEMS
An International Journal

Computers, Environment and Urban Systems 34 (2010) 28–39

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

A point-set-based approximation for areal objects: A case study of representing localities

Yu Liu [a,*], Yihong Yuan [a], Danqing Xiao [b], Yi Zhang [a], Jiangquan Hu [a]

[a] Institute of Remote Sensing and Geographic Information Systems, Peking University, Beijing 100871, China
[b] Department of Spatial Information Science and Engineering, University of Maine, Orono, ME 04469-5711, USA

A R T I C L E I N F O

Article history:
Received 25 May 2008
Received in revised form 11 May 2009
Accepted 11 May 2009

Keywords:
Point-set-based approximation
Vague areal object
Locality
Spatial relationships
Uncertainty

A B S T R A C T

Since partonomy knowledge is an important component of human cognition, we propose a point-set-based region (PSBR) model to approximate areal objects, especially vague areal objects. Two major properties of this model are that it is cognition-accordant and that it can represent vague regions easily. Given a point-set-based model, we can estimate the corresponding region using various methods, including convex hull, minimum bounding box, one-class support vector machine, and point density. We can examine the spatial relationships between two PSBRs using the derived areal objects. Additionally, we present a number of methods to compute relationships directly, based on two PSBRs. In the case study, we use a number of localities in China to demonstrate applications of the PSBR model. The proposed model can be implemented easily in an object-relational database management system. Hence, it provides a reasonable representation for vague objects that takes both manageability and approximation into account, especially now that Web 2.0 is making point data more convenient to collect.

1. Introduction

Vagueness is a ubiquitous phenomenon in the real geographical world (Couclelis, 1996; Fisher, 2000; Varzi, 2001). In urban geography, many areal objects are vague, such as the downtown of a city (Montello, Goodchild, Gottsegen, & Fohl, 2003), an exurban area (Ban & Ahlqvist, 2009), or a megalopolis that contains a number of cities (Gottmann, 1961). It is usually difficult to delineate the boundaries of such objects, although past researchers have applied cognitive experiments to this task (Montello et al., 2003). However, we can relatively easily determine whether a given point object (such as a hotel) is inside a vague locality (such as a downtown), since human beings commonly understand the distinction between a whole object and its parts (Varzi, 1996). Additionally, the geographical position of each point is often determinate. Hence, we can use the point entities inside an areal locality to approximate the geographical range of the locality. In this research, we develop a point-set-based model for vague areal objects from a perspective that incorporates spatial cognition. We call this model with "point-set-based regions" (PSBR), which can represent vague objects and approximate crisp objects.

The boundary of a vague geographical object is generally indeterminate (Burrough, 1996). Two concepts, feature and place (or locality), are used for geographical objects (Goodchild & Hill, 2008). Many geographical features, such as geomorphic units (Fisher, Wood, & Cheng, 2005), are vague; their boundaries cannot be exactly delineated. Moreover, places, which play an important role in representation and communication of geospatial knowledge (Bennett & Agarwal, 2007; Tuan, 1975), are also often vague (Goodchild & Hill, 2008). A typical expression of a place may simply be a feature, such as "Beijing", or a predicate including at least one reference object and spatial relationship, such as "in the south of England". The latter is obviously vague (Worboys & Clementini, 2001), and the vagueness mainly derives from spatial relationships (Yao & Thill, 2006). In most applications, such vague objects are managed using fuzzy sets (Dilo, De By, & Stein, 2007; Hwang & Thill, 2005; Schneider, 1999; Tang, 2004).

In two-dimensional geographical information systems (GIS), we usually distinguish three fundamental types of geometric objects: point, line, and areal objects. However, geographical objects are usually areal, as they are most likely to be of a certain size in a two-dimensional space. Higher-level abstractions could be used to identify point and line objects, depending on the scale adopted. Hence, the literature has focused mainly on vague areal objects (Cohn & Gotts, 1996; Montello, 2003), or fuzzy areal objects if they are being modeled using fuzzy sets (Fonte & Lodwick, 2004). A database can manage a fuzzy region with a raster model or a collection of crisp α-cut level regions. However, both raster models and α-cut level region models have their inconveniences. First, it is relatively hard to manage these types of models in a database management system (DBMS). Raster models often have problems

* Corresponding author. Tel.: +86 10 61751182.
E-mail address: liuyu@urban.pku.edu.cn (Y. Liu).

0198-9715/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compenvurbsys.2009.05.001

related to spatial resolution and data volume, whereas α-cut level region models require interpolation based on lines (Tøssebro & Nygård, 2002). For example, Hwang and Thill (2005) interpolated membership degrees using the triangulated irregular network (TIN) model. Second, and perhaps more importantly, they contradict common sense. Suppose that we have a vague areal object that is modeled with a membership function $m = f(x, y)$. We can ask whether the membership function is unique and correct for this object. For example, a value of $m = 0.7$ at location (x_0, y_0), may lack an interpretation. Also, the membership degree could be another number, say, 0.72. Different individuals may produce similar, but not identical, membership functions for the same vague object. Therefore, we usually consider the pattern that the membership function represents instead of the precise values. Finally, the footprints of places in a digital gazetteer are often managed by means of approximations, such as points, rectangles (Hill, 2006), and circles (Wieczorek, Guo, & Hijmans, 2004), since people do not always find the use of precise actual shapes necessary when they are representing places. For instance, the statement ''I am in Beijing'', without specifying the precise boundary of Beijing, is enough to communicate knowledge, given that the audience knows the rough location of Beijing. When an administrative unit is viewed as a place, it is often vague. Hence, Hwang and Thill (2005) proposed a fuzzy-set-based investigation of the vagueness of the locality ''Buffalo'', even though the city boundary of Buffalo is crisp. Considering the above three issues, an ideal representation for vague objects should satisfy the two requirements of manageability and approximate object representation.

The PSBR model satisfies the above two requirements well. There is some existing literature on establishing two-dimensional membership functions for localities according to point sets. For instance, Jones, Purves, Clough, and Joho (2008) proposed a GIR (geographical information retrieval)-based approach to modeling vague positions, computing the vagueness according to point density. Given a place, we can find a set of point localities of a particular type (e.g., hotel) that are associated with the given place, using a search engine such as Google. The density surface of all the localities obtained represents the vagueness of the target place. Liu, Yuan, and Zhang (2008) created a point landmark-based questionnaire to determine the membership functions of vague places. However, the PSBR model represents vague areal objects directly based on point sets. The philosophical foundation behind this approach is the following: ''represent vague objects approximately rather than precisely''. Compared with fuzzy-set-based approaches, the PSBR model does not emphasize membership degrees, and is easy to manage.

The remainder of this paper is structured as follows. After a discussion of the generic nature of PSBR models, Section 2 presents one basic PSBR model and four extended PSBR models, and proposes several ways to compute approximate regions based on PSBRs. Section 3 investigates the binary spatial relationships between two PSBRs. In Section 4, we demonstrate how to adopt the PSBR model to represent localities. Section 5 discusses several issues, including the uncertainty and applications of the PSBR model. We conclude this research in Section 6. For the purpose of demonstration, we provide an implementation to manage Chinese administrative units in an object-relational DBMS (ORDBMS) in the Appendix A.

2. From point sets to regions

2.1. Generic nature of PSBR models

Research on part-whole (or mereological) relationships has a long history (Varzi, 2007). Coad and Yourdon (1999) introduced the part-whole distinction as the philosophical justification of the object-oriented methodology. Mennis, Peuquet, and Qian (2000) viewed partonomy knowledge to be an important component in a geographical database that considers cognition. Winston, Chaffin, and Herrmann (1987) identified the place–area relationship as one type of mereological relationship. In terms of a PSBR, all points have place–area relationships to the target areal object. Hence, we argue that the PSBR model incorporates spatial cognition well.

Depending on the nature of the points inside a PSBR, we can find two types of place–area relationships. For the first category, the points stand for meaningful geographical entities, such as the hotels inside the downtown of a city. Most examples in the case studies fall into this category. This category of PSBRs has the following three characteristics: (1) The inside entities generally have spatial extent and are perceivable, (2) The target region plays the role of a container of the point set, and (3) The target region may be semantically dependent on or independent of the interior entities. In the second category of PSBRs, the points are associated with particular attributes and generally have no spatial extent. According to Goodchild, Yuan, and Cova (2007), a geographical object can be viewed as an aggregation of points with properties. Due to the semantic similarity of these points, we can conceptualize a geographical object. Hence, we can use a set of sampling points to represent the variation inside the geographical object. For example, a number of points with elevation values can represent a plateau. For this type of PSBR, each element is a geometrical point with at least one property, and the target region is semantically dependent on the point set. Obviously, the vagueness of an areal object can be modeled by both the properties and distribution pattern of the points within it. We can further distinguish the two categories of PSBR models from the perspective of object-field dichotomy (Goodchild, 1992). For the first category, points inside a PSBR are geographical objects, while for the second category the points represent a field with the target region as the domain of the field.

In terms of the examples mentioned earlier, the areal objects, such as the downtown of a city, have fiat boundaries, which depend on laws, political decrees, or cognitive phenomena (Smith, 1995). The PSBR model can approximate fiat objects, since it focuses on the points inside rather than the boundary of an object. On the other hand, the boundaries of bona fide objects exist physically and are often measured directly. Hence, we generally need not approximate a bona fide object using the points inside it. In particular contexts, the PSBR model of a bona fide object is acceptable. For example, an island has a bona fide boundary with the water body surrounding it. The residents of the island can use the places inside it to represent the rough shape of this island.

2.2. Basic PSBR models

A point-set-based region R is directly represented by a finite point set $\{P_1, P_2, \ldots, P_n\}$, which is denoted by $\boldsymbol{S}$. The coordinates of each point in $\boldsymbol{S}$ are known and assumed to be certain (Fig. 1a). Generally, the predicate ''inside R'' holds for each element in $\boldsymbol{S}$. In this study, we call R the ''target region''. Given a PSBR, we can compute the corresponding areal approximations, such as its convex hull (Fig. 1b) and minimum bounding box (MBR) (Fig. 1c). These areal approximations can be viewed as substitutes for the actual region.

We notice that if a point, P, is inside the region R, then another point close to P has a high probability of being inside R, according to Tobler's first law of geography (Tobler, 1970). To account for this phenomenon, we introduce the union of buffer zones (i.e., disk areas) of all points to create a region (Fig. 2a). Since a region is often connected in practice, we can achieve this by setting the buffer distance to half of the largest nearest-neighbor distance of the whole point set, that is,

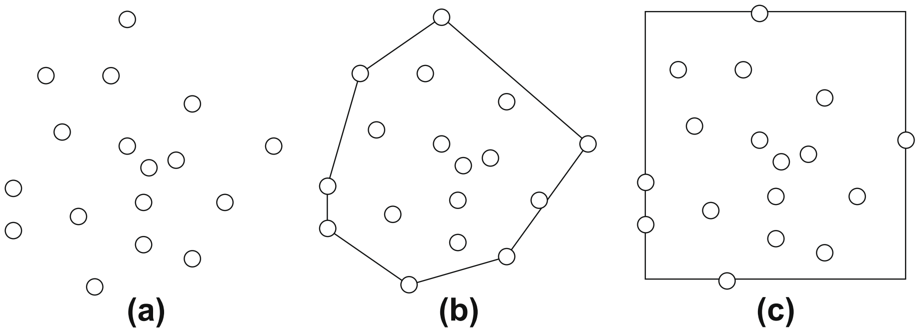


Fig. 1. (a) Point set representation for areal objects; (b) convex hull of a PSBR; (c) MBR of a PSBR.

$$R = \max_{P_i \in R}\left(\max_{P_i \in R} dist(P_i, P_j)\right) \Big/ 2 \tag{1}$$

where $d(P_i, P_j)$ is the Euclidean distance between P_i and P_j. In Eq. (1), we find the closest point to each point, and compute the distance between these two points. Consequently, a set of n distance values is obtained. By setting the buffer distance to half of the maximum distance in this set, we can connect the buffer zones of all points.

Outlining a boundary according to a set of points is actually a one-class classification problem. Machine learning techniques commonly apply support vector machines (SVM) classification or regression (Vapnik, 1995). They can solve one-class problems, such as document classification (Manevitz & Yousef, 2001) and niche modeling (Guo, Kelly, & Grahamb, 2005). One-class SVMs can simulate PSBRs. Fig. 2b–d shows approximations of the same PSBR using different parameters. These three cases use the RBF (radial basis function) kernel function. An RBF kernel can be written as

$$K(x_i, x_j) = e^{-\gamma\|x_i - x_j\|^2}, \quad \gamma > 0 \tag{2}$$

In RBF-kernel one-class SVMs, the parameter γ controls the shape of the target region: a high value of γ will yield a more complex shape. Additionally, in order to train a one-class SVM, an outlier fraction of the training examples should be determined (Scholkopf, Platt, Shawe-Taylor, Smola, & Williamson, 2001). We denote this fraction by μ; it reflects the fact that one-class SVMs can represent a PSBR with errors of commission, that is, some points within ***S*** actually do not belong to the region. In Fig. 2b–d, we set μ to a small value, i.e., 0.001, since all the points are generally assumed to be inside the region.

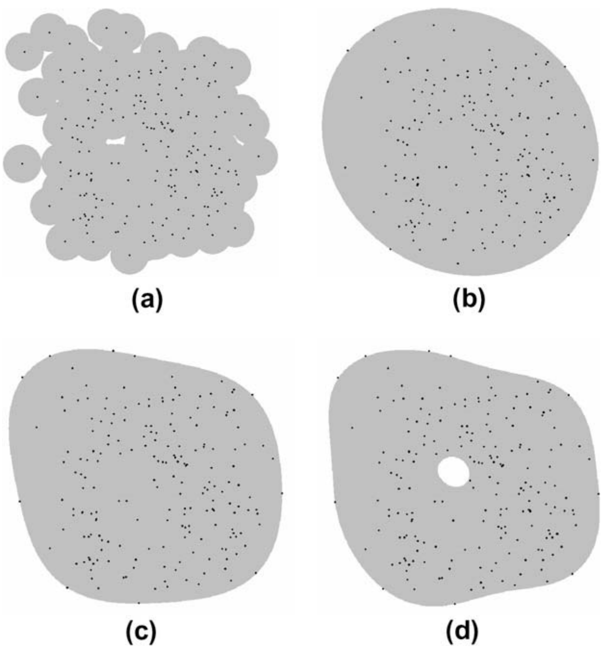


Fig. 2. Generating regions based on point sets using various methods: (a) union of disk areas; (b) one-class SVM (μ = 0.001, γ = 1.0); (c) one-class SVM (μ = 0.001, γ = 5.0); (d) one-class SVM (μ = 0.001, γ = 10.0). During the implementation, LIBSVM developed by Chang and Lin (2001) was adopted.

In addition to crisp regions, we can apply point-density analysis, following the method of Jones et al. (2008), to find a region with a gradual boundary (Fig. 3). The foundation of this approach is that a high point density implies a high level of confidence that the corresponding location belongs to the region.

2.3. Extended PSBR models

The PSBR model can be viewed as a trade-off between precise shapes in conventional geographical information systems and rough approximations in digital gazetteers. In basic PSBR models, we only consider the coordinates of each point. In addition, we can extend a PSBR model from the following four aspects.

2.3.1. Weighted PSBRs

In the basic PSBR models, all points have equal weight. For epistemic reasons, however, they may in fact have different contributions to the target region. We thus append a weight to each point. Liu, Yuan et al. (2008) obtained weights by doing a cognitive experiment: they asked subjects whether a given point landmark was inside an areal object. The weight of a point is viewed as the membership degree to which this point belongs to the target region. For a weighted PSBR, point interpolation can create a vague region with an indeterminate boundary. A number of point interpolation methods are available, such as triangulated irregular networks (TINs), trend surfaces, and kriging interpolation. For

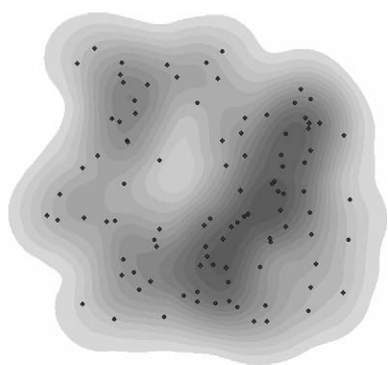


Fig. 3. Generating a region with gradual boundary based on kernel-density analysis.

example, Liu, Yuan et al. (2008) used the support-vector regression method.

If the weights of a PSBR are restricted to be 0 or 1, where a weight of 0 means that the point is definitely outside the target region, then we call it a binary PSBR. For binary PSBRs, an ideal case is that the points with weight 1 are surrounded by the points with weight 0, so that we can compute the target region with a two-class classification method, such as two-class SVMs, or the Voronoi-based region approximation (Alani, Jones, & Tudhope, 2001).

2.3.2. Typed PSBRs

According to Mennis et al. (2000), taxonomy knowledge should also be considered in addition to partonomy knowledge. Goodchild et al. (2007) and Liu, Goodchild, Guo, Tian, and Wu (2008) argued that a feature originates from the semantic similarity of its parts. Therefore, a PSBR can be extended to consider taxonomy knowledge. A region is often represented by points that have semantic relationships to it. For instance, the downtown area of a city may be represented by a set of buildings, while a mountain range can be described by a set of peaks. In other words, the set of points indicates both the spatial distribution and the geographical semantics of the target areal object. Note that the weighted PSBRs and typed PSBRs can be viewed as applications of the second category of PSBRs mentioned in Section 2.1, since they take into account the properties of inside points.

2.3.3. PSBRs limited by other regions

In practice, although the boundary of a vague region cannot be delineated exactly, we can be certain that a region is inside another areal object. For example, we can assert that the Huanghe River delta is inside China. This is also an application of pervasive partonomy knowledge. For a PSBR model, we can thus append a limiting region, which may be crisp or vague, to specify its maximum possible range.

2.3.4. PSBRs incorporating internal cardinal directional relationships

Liu, Wang, Jin, and Wu (2005) proposed models of internal cardinal directional (ICD) relationships, and argued that ICD relationships refine partonomy knowledge. In a PSBR model, we can attach an ICD relationship to each point. Such a model provides more clues for estimating the target region, especially in the presence of errors of omission. As shown in Fig. 4, given two points, P_1 and P_2, and their ICD relationships to the target region R (i.e., P_1 is in the north part of R and P_2 is in the southwest part or R), it is possible to approximate R's location, depicted by the dashed line.

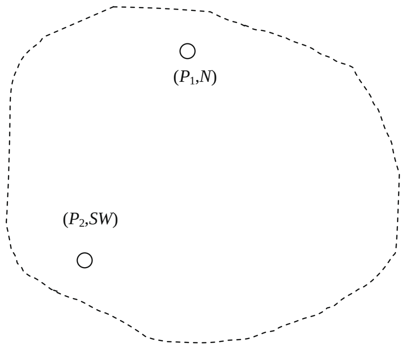


Fig. 4. PSBR model incorporating ICD relationships.

3. Computing spatial relationships between two PSBRs

We can compute the spatial relationships between two PSBRs from their areal approximations. Here we show how this can be done directly based on point sets.

3.1. Topological relationships

Researchers in artificial intelligence (AI) and GIS have extensively investigated topological relationships between two spatial objects. The region connection calculus (RCC) (Randell, Cui, & Cohn, 1992) and the 9-intersection model (9-IM) (Egenhofer & Franzosa, 1991) are two major topological models. Behr and Schneider (2001) extended the 9-IM to model topological relationships between complex objects, which may be multi-points or multi-regions. In terms of complex points, five relationships can be identified: disjoint, overlap, contain, inside, and equal (Fig. 5a–d). Clearly, they can also model the topological relationships between two PSBRs. Such relationships are easy to examine; however, they may be intuitively incorrect. In Fig. 5e, the relationship between R_1 and R_2 is "disjoint" according to the model of Behr and Schneider (2001). However, they are usually perceived to be "R_1 overlaps

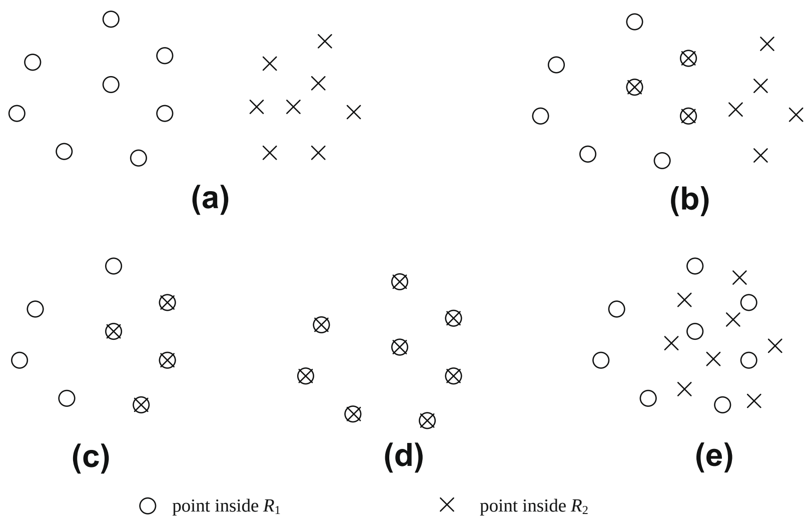


Fig. 5. Topological relationships between two PSBRs: (a) disjoint; (b) overlap; (c) inside/contain; (d) equal and (e) deduced overlap.

R_2" when considering their areal approximations, such as convex hulls. We thus define the relationships that follow Behr and Schneider (2001) to be "direct topological relationships", and the relationships based on areal approximations to be "deduced topological relationships". Direct overlap (or inside/contain) implies stronger geographical semantics than deduced overlap, especially when the coincident points are instances of the same feature type.

3.2. Directional relationships

If the areal approximations of two PSBRs are known, then we can use conventional models, such as the MBR-based model, to represent the cardinal directional relationship (CDR) between the two PSBRs. Moreover, Deng and Li (2008) proposed a statistical method for measuring the directional relationship between two objects with spatial extents. For two PSBRs, we can compute an angle histogram from the angle between any two points in the two objects, and determine the cardinal directional relationship between them. For example, Fig. 6a indicates that "R_2 is northeast of R_1", since the maximum frequency in the histogram corresponds to 40° (Fig. 6b). Additionally, due to the vagueness of CDRs, a fuzzy pattern-compatible approach based on the histogram (Bloch, 2005) can compute the matching degree to which the relationship between two PSBRs belongs to a category of CDRs, such as northeast.

3.3. Metric relationships

For two PSBRs, R_1 and R_2, there are two ways to compute their metric relationship. First, we can find two representative points (e.g., the centroids) of R_1 and R_2, and then compute the distance between the two points. Second, we can compute the distance values of all pairs (P_{1i}, P_{2j}), where $P_{1i} \in R_1$ and $P_{2j} \in R_2$, and then sum the values to get the distance between R_1 and R_2. From a computational perspective, the time complexities of these two ways are $O(n+m)$ and $O(nm)$, where n and m are the numbers of points of R_1 and R_2. In this paper, we provide the following three example metrics.

3.3.1. Distance between centroids

For a PSBR including n points, its centroid can be computed as follows:

$$P_{cent} = (x_{cent}, x_{cent}) = \left(\frac{\sum_{i=1}^{n} x_i}{n}, \frac{\sum_{i=1} y_i n}{n}\right), \tag{3}$$

where (x_i, y_i) are the coordinates of P_i. The above equation can be extended easily to consider the weight of each point. Suppose the centroids of two PSBRs are P_{cent1} and P_{cent2}. Their distance is:

$$D_c(R_1, R_2) = dist(P_{cent1}, P_{cent2}), \tag{4}$$

where *dist* is a function to compute the distance between two points. Note that the function may denote the Euclidean distance or some other available measure, such as Manhattan distance. Moreover, we can adopt other methods to find the representative point, instead of the centroid, of a PSBR. Hence, several variations of this metric can be developed.

3.3.2. Average distance

Assume that R_1 contains n points and R_2 contains m points. The average distance between R_1 and R_2 is

$$D_m = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} dist(P_{1i}, P_{2j})}{nm}, P_{1i} \in R_1, P_{2j} \in R_2. \tag{5}$$

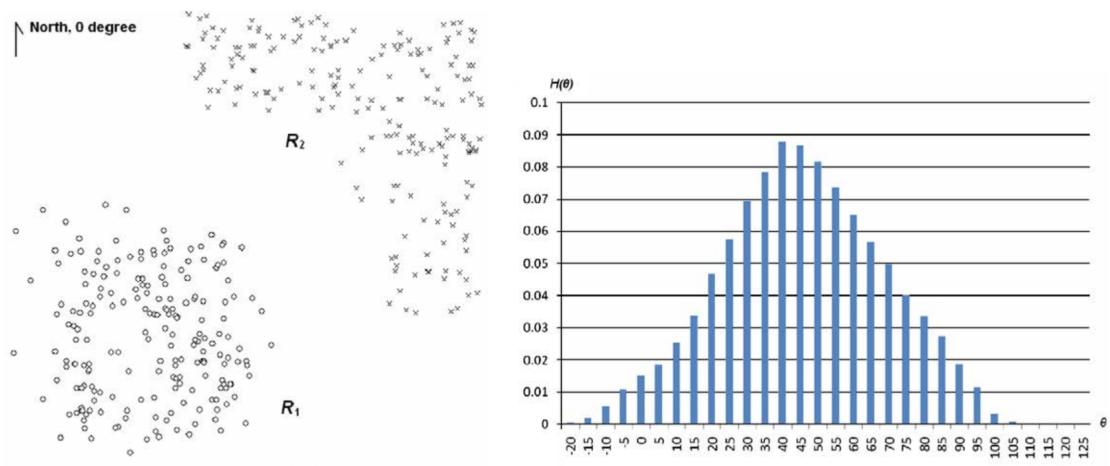
3.3.3. Hausdorff distance

The Hausdorff distance was initially developed to measure how far two compact non-empty subsets of a metric space are from each other. Since it considers object shape, Hausdorff distance is widely used to measure the similarity between two objects. For PSBRs, we can use the maximum and minimum instead of the supremum and infimum, and compute the Hausdorff distance based on point sets:

$$D_H(R_1, R_2) = \max\{\max_{P_1 \in R_{2j}} \min_{P_2 \in R_2}(P_{1i}, P_{2j}) \max_{P_2 \in R_2} \min_{P_1 \in R_1} dist(P_{1i}, P_{2j})\}. \tag{6}$$

The average distance and Hausdorff distance belong to the second category of algorithms with time complexity $O(nm)$, and we can change the summation method to generate new distance metrics. The above three metrics describe the location similarity between two PSBRs. We may further extend them by involving the weights of points, and integrate location similarities and attribution similarities by taking into account the types of points.

In addition to the relationships between two PSBRs, we can analyse the relationships between a PSBR and an ordinary geometric shape, e.g., a polygon. Such relationships can be examined by converting the PSBR to an ordinary region with the methods in Section 2. Moreover, since it is relatively easy to determine whether a point is inside a polygon, we can decide efficiently that a PSBR R_1 is connected to an ordinary region R_2 given that a point within R_1 is inside R_2.



(a) (b)

Fig. 6. Directional relationship between two PSBRs represented by angle histogram.

4. Case studies

In this section, we demonstrate the applications of the PSBR model by using it to represent several localities in China. First, each province in China is approximated by a set of county seats inside it. Although a province has a crisp boundary, a point-set-based approximation is acceptable in many situations. Second, we establish the PSBR model for a vague region, i.e., South China, based on data collected via cognitive questionnaires. The applications demonstrate that the PSBR model can approximate areal localities well.

4.1. Representing provinces in China

China is comprised of 34 provincial-level administrative units (referred to as "provinces" in this paper for simplicity). Each province comprises a number of county-level units. Each county has a county seat, which we abstract to a point object. We thus approximate each province based on a set of cities inside it with the PSBR model. Since Hong Kong and Macau are too small, we merge them into Guangdong Province for simplicity, leaving 32 provinces in our representation (Fig. 7a). Fig. 7b and c show their approximations using convex hulls and Voronoi polygons, respectively. We can see from Fig. 7b and c that a denser point distribution generally leads to a more precise approximation. For example, the convex hulls of the provinces in eastern China are akin to the corresponding actual shapes; while there are noticeable differences between the convex hulls and the actual shapes for the provinces in western China. Note that the Voronoi diagram method has been proposed by Alani et al. (2001). In Fig. 7c, we use the land territory of China as the limiting region for frontier provinces, such as Tibet. According to Fig. 7b and c, the Voronoi diagram based approximation is better than the convex hull based approximation. However, for a PSBR, the Voronoi diagram based approximation should involve additional points that lie outside the target region. In Fig. 7d, we create approximate regions using the union of buffer disks and one-class SVM for an object (Shanxi Province, c.f. the shaded region in Fig. 7a). Fig. 7e depicts the directional relationships between several provinces and Henan Province based on their PSBR models. Note that we use geographical coordinates to compute the angles in Fig. 7e. The above applications indicate that the PSBR model provides an acceptable representation for a region. Note that we can

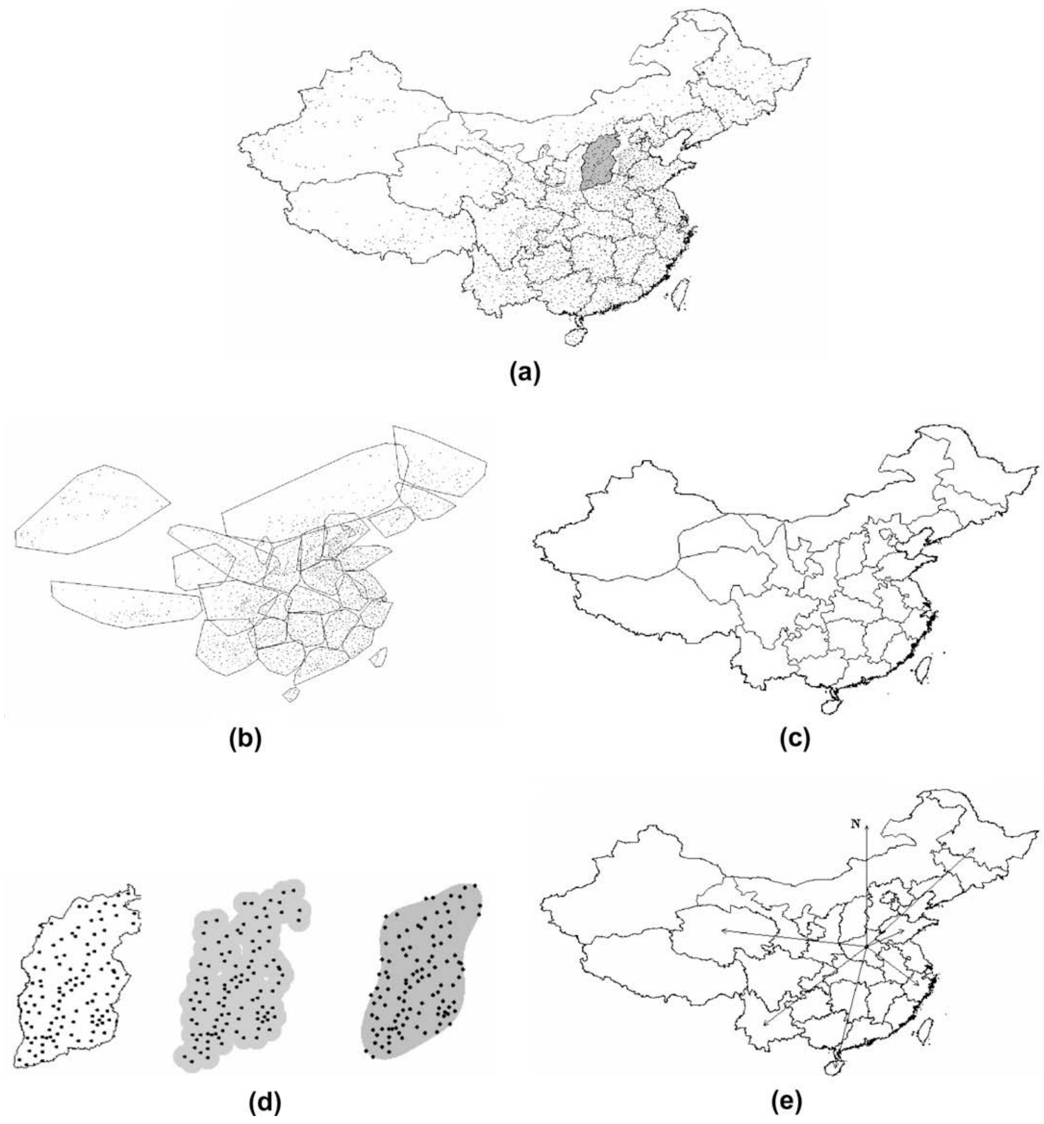


Fig. 7. (a) Actual shapes of China's provinces (without the South China Sea Islands for simplicity); (b) convex hull based approximations of China's provinces; (c) Voronoi diagram based approximations of China's provinces; (d) approximations of an areal object (Shanxi Province) and (e) directional relationships computed based on point sets.

also employ cities to represent megalopolises in addition to provinces. There are differences between these two cases. When we use cities to approximate a megalopolis, the megalopolis is semantically dependent on the cities involved. In terms of the PSBRs for provinces, however, we cannot find such relationships.

4.2. Representing South China based on cognitive experiments

South China is an approximate region in China. This concept is associated with several factors, including climate, geography, culture, and physical traits, in addition to the internal cardinal directional relationship. It is thus difficult to delineate the exact boundary of South China. To overcome this problem, we designed a cognitive experiment to investigate its borders. In the experiment, we chose 20 cities that could potentially be considered to be in the southern part of China (Fig. 8a). For each city, 30 undergraduate students from the department of geography at Peking University served as subjects, and we asked them whether each city was inside South China. The students had sufficient geographical knowledge of the approximate locations of all cities to provide an educated answer. Based on the collected data, we established a weighted PSBR for the concept of "South China".

4.2.1. City selection

Each chosen city was sufficiently well known in China for participants to make plausible estimates and thus provide reasonable data. Given that the geodetic center of China is in the city of Xi'an, all selected cities were south of Xi'an (except for Xi'an itself).

4.2.2. Methods and results

For each landmark, let μ_i denote the degree of belonging to South China; then

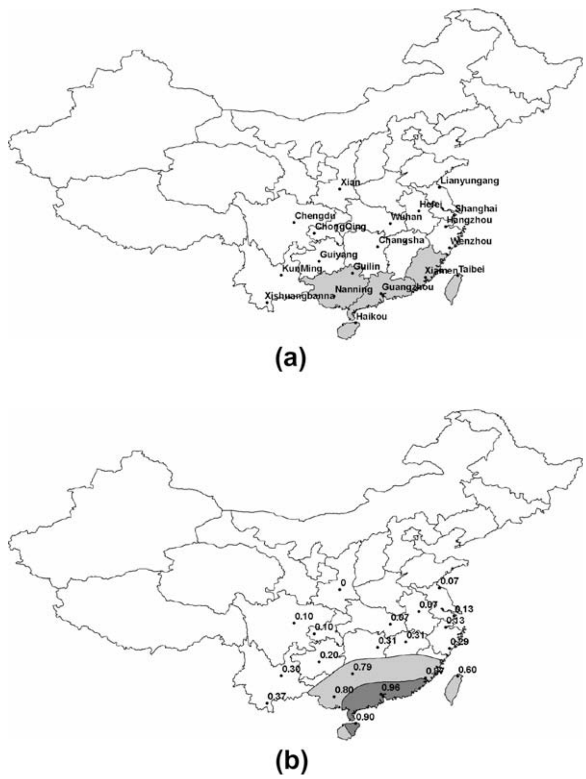


Fig. 8. (a) Cities selected in the cognitive experiment and (b) visualizing the concept of "the South of China" using kriging interpolation for a weighted PSBR.

$$\mu_i = \frac{A_i}{A_i + B_i}, \tag{7}$$

where A_i is the number of subjects agreeing with the assertion that "the ith city is inside South China", and B_i is the number of subjects disagreeing with the assertion.

The set of cities chosen in the experiment can be viewed as an instance of the weighted PSBR model described in Section 2, in which each point (i.e., each city) in this set has a value μ to indicate its membership degree. Such a weighted PSBR serves as a representation of South China. We can visualize the vague region using the kriging interpolation from the weighted PSBR (Fig. 8b).

4.2.3. Brief discussion

According to common geographical knowledge, South China includes the five provinces Guangxi, Guangdong, Taiwan, Hainan, and Fujian (Fig. 8a) in a broad sense, and excludes Taiwan and Fujian in a narrow sense. As shown in Fig. 8b, the gray regions indicate membership values higher than 0.5. According to the experimental design, a point with a degree higher than 0.5 implies that more than half of the participants agree that this point is inside South China. It is thus reasonable to assert that the 0.5-cut set corresponds to the outer boundary of South China. In other words, we can consider the 0.5-cut set to be the "egg", according to the "egg yolk" model (Cohn & Gotts, 1996). It is highly coherent with the broad concept. The membership value is greater than 0.8 in the dark-gray regions. It is consistent in the narrow sense, and can be viewed as the "yolk part" in the "egg yolk" model. Fig. 8b introduces the land territory of China as the limiting region of the PSBR model, so that the concept of South China does not extend into other neighboring countries, such as Vietnam. Note that different interpolation methods will yield different membership functions. These functions will generally have similar patterns, and their differences therefore do not influence our discussion. This case study demonstrates that the PSBR model can be a valuable representation for vague objects by taking human conceptualization into account.

In the Appendix A, we introduce the detailed implementation of the above PSBRs and performing several queries about the objects in an ORDBMS. The implementation indicates that the PSBR model provides a promising way to manage areal objects, especially vague areal objects.

5. Discussion

5.1. Uncertainty associated with PSBR models

The PSBR models provide approximations for areal objects and are thus inherently uncertain. In a PSBR model, an important feature associated with uncertainty is its point pattern. A higher point density implies less uncertainty, whereas a lower point density

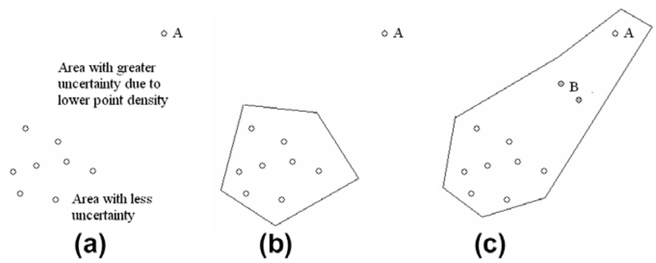


Fig. 9. (a) Relationship between point density and uncertainty of a PSBR model; (b) low point density due to errors of commission and (c) low point density due to errors of omission.

indicates greater uncertainty (Jones et al., 2008, Fig. 9a). From the perspective of data collection, two types of errors may cause a low point density: errors of commission and errors of omission. In Fig. 9b, point *A* is not actually inside the target region. Due to errors of commission, *A* is incorrectly included in the point set. As mentioned earlier, one-class SVMs can handle such situations. On the other hand, as shown in Fig. 9c, if point *A* is inside the target region, but some points (see the points around *B*) between *A* and the clustered points are omitted, then the convex hull or MBR based methods can find the boundary of the target region. Note that the buffer disk based approach may overestimate the target region, since the maximized minimum distance (see Eq. (1)) is greater than that in a point set with random pattern. However, if we cannot distinguish the two types of errors for a given PSBR, we can at least estimate the associated uncertainty using the point-density analysis.

Given a PSBR and an arbitrary point, we can determine whether the point is inside the target region with an uncertainty degree. Interpolation and extrapolation are two ways to make a prediction. As shown in Fig. 10a, the PSBR model contains eight points. For point *A*, which is inside the convex hull, we can interpolate in order to decide whether it is inside the target region. On the other hand, we should extrapolate to make the same decision for point *B*. Since extrapolation is generally less precise than interpolation, the uncertainty at point *B* is greater than that at *A*, even though the point densities at these two points are equal (Fig. 10b). As mentioned earlier, the point density provides an indicator of the uncertainty associated with a PSBR. However, point density estimations depend on spatial scales. For a kernel-density analysis, the kernel radius denotes the scale. Fig. 10b uses a moderate kernel radius. In comparison, Fig. 10c shows a lower radius value and Fig. 10d shows a higher radius value. From Fig. 10d, we can see that the point density is high in the middle of the PSBR. This is consistent with the intuition that the uncertainty at *A* is less than that at *B*. However, Fig. 10b and c do not support this intuition. The above discussions imply that we should estimate the uncertainty of a PSBR model with multi-scale analyses. We plan to investigate this issue in the future.

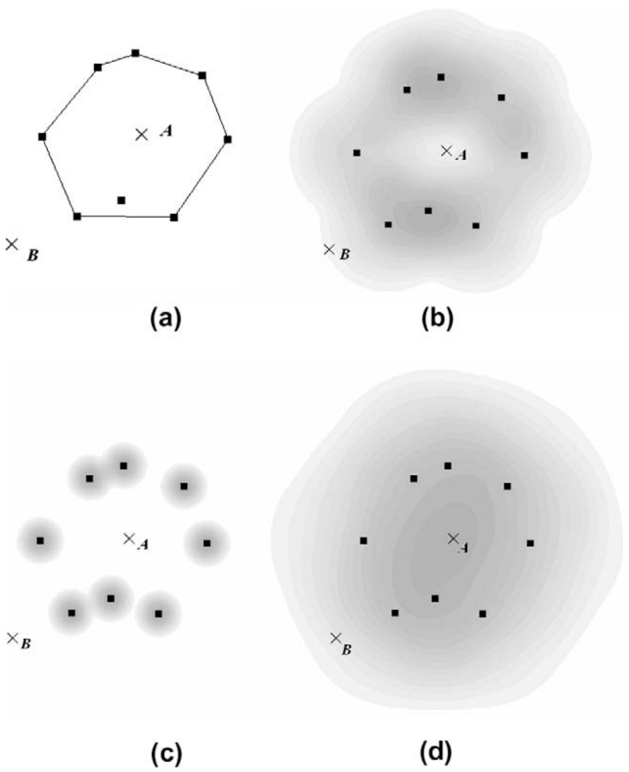


Fig. 10. (a) Predicting whether a point is inside the target region using interpolation and extrapolation; (b) point density estimation with a moderate kernel radius; (c) point density estimation with a small kernel radius and (d) point density estimation with a great kernel radius.

5.2. Applications of PSBR models

People perceive many features in geographical space as points, depending on the concrete situation. For instance, residents of a city may represent and talk about a landmark building as a point (Lynch, 1960), while a city is often abstracted to be a point at a nationwide scale. Human activities are tightly coupled with settlements, including cities, towns, and villages. Among them, cities play the most important role, due to their high concentration of activities. Considerable geographical knowledge, especially personalized geographical knowledge, is tightly associated with cities. We thus identify two levels of points, according to their sizes relative to cities: points of interest (POI) inside cities, and cities themselves. These points can represent different levels of vague areal objects. A set of buildings can approximate an area, such as the central business district (CBD), inside a city, while a set of cities can represent geographical concepts with larger spatial scales. For example, the concept of megalopolis refers to a huge metropolitan area (Gottmann, 1961). It is often difficult to draw the exact boundary of a megalopolis, such as the BosWash (Boston-Washington). However, a number of cities, including New York, Philadelphia, etc., can be enlisted to build a PSBR model for the BosWash megalopolis.

At present, a number of POI databases and digital gazetteer services have been developed, such as the GEOnet Names Server[1] of the NGA (National Geospatial-Intelligence Agency), USA, the POI database[2] of the Ordnance Survey, UK, and the Alexandria Digital Library[3] (ADL) of the University of California, Santa Barbara. Most of the records in such databases are points with geographical coordinates, and PSBR models can use them directly. For a given areal locality, however, it is still an open question as to how best determine which points are associated with the locality. Clearly, finding the point set from survey data is costly and thus not practical for a large number of areal objects. Since the World Wide Web (WWW) has become a massive knowledge base, the GIR-based approach (Jones et al., 2008) can extract points associated with a vague areal object. In addition, we can restrict the source web pages to improve the reliability of a PSBR model. With the development of Web 2.0, each individual can contribute his or her own information about a geographical concept (Goodchild, 2007). Wikipedia (http://www.wikipedia.org) is a typical application. In Wikipedia, many entries are about geographical features or places, and each article contains a number of places, which can be abstracted be to point objects inside the corresponding object. For instance, the entry "San Francisco Bay Area" (http://en.wikipedia.org/wiki/Bay_Area) mentions objects like Silicon Valley, the Golden Gate Bridge, and the University of California, Berkeley, which are widely accepted to be inside the "San Francisco Bay Area". These place names can be identified automatically and geo-referenced with help from POI databases and digital gazetteers (Hill, 2006), so that the PSBR model for the "San Francisco Bay Area" is established. In the wikification of GIS (Sui, 2008), each participant creates and maintains the description of a geographical object. Such a public participation mechanism in wiki systems can filter bogus data and guarantee that the obtained PSBR models are consistent with common-sense knowledge.

[1] http://earth-info.nga.mil/gns/html/index.html.
[2] http://www.ordnancesurvey.co.uk/oswebsite/products/pointsofinterest.
[3] http://www.alexandria.ucsb.edu/.

6. Conclusions

Fuzzy sets are widely applied to represent vague geographical objects. However, the fuzzy-set-based approach has its disadvantages, e.g., it is difficult to manage a two-dimensional membership function. In this research, we presented a point-set-based approximation model for areal objects, especially vague areal objects. The model relies on the part-whole distinction, which is pervasive in human knowledge. Compared with fuzzy-set-based representations, it incorporates human cognition well and can be implemented easily.

From a mereological perspective, we find two categories of points inside the target region, namely geographical entities and sampling points. Additionally, the PSBR model can be extended to involve weights, types, limiting regions, and internal cardinal directional relationships in order to represent richer geographical semantics. Depending on the requirements of the application at hand, we can find regions with crisp or gradual boundaries. It should be noted that a PSBR is compatible with a conventional raster model when it consists of an adequate number of points that are regularly distributed.

Spatial relationships, including topological, directional, and metric relationships, can be measured directly based on point sets, in addition to the corresponding areal approximations. Some relationship-examining operations are even more efficient. For example, we can easily decide that the intersection of a PSBR *R* and a polygon is not empty, given that at least one point in *R* is inside the polygon.

The major objective of the PSBR model is to deal with areal objects with uncertain boundaries. We can estimate the uncertainty of PSBRs from the perspective of point density. Generally, a higher point density implies less uncertainty. Other factors, including errors of commission versus errors of omission, interpolation versus extrapolation, and spatial analysis scale, may also influence point density analyses and uncertainty estimations.

The PSBR model can be applied to urban geography research. Two types of points, namely, POIs inside cities and cities themselves, are important elements to represent localities. We can use POIs to model vague districts inside cities and adopt cities to model large-scale regions. In terms of data collection, many localities are associated with human experiences and often described textually. Hence, with the development of Web 2.0, geographical information retrieval, and digital gazetteers, points can be extracted automatically to build the PSBR model for a given areal locality.

Acknowledgements

This research was supported by the National High Technology Development 863 Program of China (Grant No. 2007AA12Z216) and NSFC (Grant No. 40701134). We thank the anonymous reviewers and Dr. J.-C. Thill for their constructive comments, which strengthened this paper. Dr. M.F. Goodchild and Dr. A. Stein provided comments on an earlier version of this paper. Dr. Yongmei Lu helped us to reorganize the manuscript and to polish the English presentation. We thank them all.

Appendix A. Managing PSBRs in spatial databases

A DBMS can manage vague areal objects using the PSBR model. In this study, we use PostgreSQL as a development platform, since it is currently a widely applied ORDBMS, and it can be conveniently extended to implement self-defined types and functions for uses. According to the development guide for PostgreSQL, such an implementation generally includes three steps:

(1) defining and coding types and functions in source files;
(2) compiling source files to DLL (Dynamic Linking Library) files, and embedding the DLL files into the ORDBMS (note that Microsoft Windows operating systems use DLL files, while GNU/Linux operating systems use SO (Shared Object) files; and
(3) registering the corresponding data types and operation functions using SQL (Structured Query Language) commands.

A.1. Definition of PSBR data types

As mentioned in Section 2, PSBR models can be extended to represent richer semantics. For the sake of simplicity, however, this paper only implements basic PSBR models in PostgreSQL.

A.1.1. The logical data model of PSBR

A basic PSBR can be represented directly by a finite point set $\{P_1, P_2, \ldots, P_n\}$, where P_i, $i = 1, \ldots, n$, is an instance of the Point type. The definition of the Point type is consistent with the geometry types specified in the OpenGIS Simple Features Specification (OGC, 1999).

A.1.2. Auxiliary data type: WKB_PSBR

PSBR cannot be applied directly to the transformation between a DBMS and client application, due to possible differences between their storage structures. We thus define an auxiliary data type WKB_PSBR based on the well-known binary (WKB) representation (OGC, 1999), which provides a transplantable representation of a PSBR using a continuous byte stream. It permits PSBR objects to be exchanged between a client application and a DBMS in binary form. The definition of WKB_PSBR in the C programming language is as follows.

```
typedef struct tagWKB_Point

{
  double x;
  double y;
} WKB_Point;
typedef struct tagWKB_PSBR
{
  long numPoints;
  WKBPoint points[numPoints];
} WKB_PSBR;
```

Note that we use an array instead of a pointer in the above definition so as to guarantee the continuity of a storage schema (Fig. 11).

A.1.3. Auxiliary data type: WKT_PSBR

WKB_PSBR stores a PSBR as a stream of binary code. It cannot be used directly in SQL commands, so WKT_PSBR is grounded on the well-known text (WKT) representation to provide a standard textual representation (OGC, 1999).

In EBNF (Extended Backus Naur Form), the grammar of the WKB_PSBR type is as follows:

$$\langle \text{WKT\_PSBR} \rangle ::= \text{PSBR} \; (\langle \text{WKT\_Point} \rangle \{;\langle \text{WKT\_ Point} \rangle\}^{*})$$

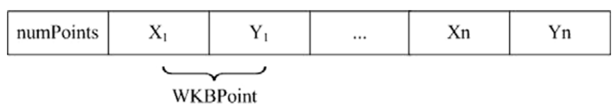


Fig. 11. Storage structure of a WKB_PSBR object.

$\langle$WKT_Point$\rangle$ ::= Point($\langle$Coordinate$\rangle$)
$\langle$Coordinate$\rangle$ ::= $\langle x\rangle\langle y\rangle$
$\langle x\rangle$ ::= double precision literal
$\langle y\rangle$ ::= double precision literal

The extended PSBR models can be defined similarly. In practice, a reasonable solution is to define the PSBR type by taking all extensions into account. The basic PSBR type can thus be viewed as a specialization. In the following sections, we will use the PSBR type in its broad sense, where the ExPoint type substitutes for Point. An ExPoint object includes the weight, type, and corresponding ICD relationship of each point within an extended PSBR instance. Consequently, the WKT associated with the ExPoint point is defined as follows:

<WKT_ExPoint>::= ExPoint(<Coordinate > <Weight> <Type> <ICD_RELATION>),

where "Weight", "Type", and "ICD_RELATION" specify the characteristics of an ExPoint object. The WKB_Point type can be defined similarly.

A.2. Definitions of relevant functions

Based on the definitions of WKB_PSBR and WKT_PSBR, we can introduce a number of functions to manipulate the user-defined PSBR data. The functions can be sorted into the following four categories (Tables 1–4).

In these functions, the Region type is defined to be equivalent to the MultiPolygon type defined by OGC (1999), since an areal object may be disconnected in GIS.

Note that we only provide the interfaces of PSBR functions. Some interfaces may correspond to different implementation strategies. For example, in the DirectionalRel function, one cardinal directional relationship may be decided according to the highest frequency in the histogram. Meanwhile, a degree can be computed by matching a given histogram with a standard histogram associated with the cardinal directional relationship.

A.3. Registration in PostgreSQL

As described in the extension regulation of PostgreSQL, user-defined data types and functions should be registered using the SQL command "create type" or "create function". The source files of types and functions also should be specified (libPSBR.dll in this paper).

Table 1
PSBR operation functions.

Create(npts:integer, pts:ExPoint*, limitation: Region): PSBR	Create a PSBR from a collection of npts ExPoint objects and a limiting region
Clone(pset: PSBR): PSBR	Reproduce a new PSBR object from an existing PSBR
Destroy(pset: PSBR):void	Destroy a PSBR object and release the memory
Subset(pset: PSBR, condition: String): PSBR	Obtain a subset that satisfies given condition, e.g., weight > 0.5

Table 2
Conversion functions.

PSBR2WKB(pset: PSBR): WKB_PSBR	Convert from PSBR to WKB_PSBR
WKB2PSBR(wkb_pset: WKB_PSBR): PSBR	Convert from WKB_PSBR to PSBR
PSBR2WKT(pset: PSBR): WKT_PSBR	Convert from PSBR to WKT_PSBR
WKT2PSBR(wkt_pset: WKT_PSBR): PSBR	Convert from WKT_PSBR to PSBR

Table 3
Functions for generating ordinary regions or vague regions.

MBR(pset: PSBR): Region	Obtain the MBR from a PSBR
ConvexHull(pset: PSBR): Region	Obtain the convex hull from a PSBR
OneClassSVM(pset: PSBR, param: SVM_PARAM): Region	Obtain a region from a PSBR using one-class SVM method
PointDensity(pset: PSBR, threshold: real): Region	Obtain a region from a PSBR using density analysis, and the density valued inside such a region is greater than a threshold
Interpolation(pset: PSBR, threshold: real): Region	Obtain a region from a weighted PSBR using interpolation, and the weight inside such a region is greater than a threshold

Table 4
Spatial relationship examination functions.

Disjoint(pset1: PSBR, pset2: PSBR):bool	Disjoint relationship examination
Equal(pset1: PSBR, pset2: PSBR):bool	Equal relationship examination
Overlap(pset1: PSBR, pset2: PSBR):bool	Overlap relationship examination
Contain(pset1: PSBR, pset2: PSBR):bool	Contain relationship examination
Inside(pset1: PSBR, pset2: PSBR):bool	Inside relationship examination
TopologicalRel(pset1: PSBR, pset2: PSBR):TOPO_REL	Compute the topological relationship between two PSBRs
DirectionalRel(pset1: PSBR, pset2: PSBR):DIR_REL	Compute the directional relationship between two PSBRs
MetricRel(pset1: PSBR, pset2: PSBR):real	Compute the metric relationship between two PSBRs

A.3.1. Type registration

In an ORDBMS, input and output (I/O) functions should be specified for a user-defined type. These functions determine how this type appears in text and how this type is organized in memory. In this research, I/O functions are implemented based on conversion functions defined in Table 2. Conventionally, we create the user-defined I/O functions before creating the type, as follows:

```
CREATE OR REPLACE FUNCTION PSBR_in(cstring)
  RETURNS WKB_PSBR
  AS 'libPSBR','PSBR_in'
  LANGUAGE C IMMUTABLE STRICT;
CREATE OR REPLACE FUNCTION PSBR_out(WKB_PSBR)
  RETURNS cstring
  AS 'libPSBR', 'PSBR _out'
  LANGUAGE C IMMUTABLE STRICT;
```

With registered I/O functions, the PSBR type can be registered in PostgreSQL using the following statement:

```
CREATE TYPE PSBR (
  internallength = VARIABLE,
  input = PSBR_in,
  output = PSBR_out,
  storage = main);
```

A.3.2. Function registration

It is the same as registering I/O functions for users to implement other operation functions into PostgreSQL. For instance, consider the following example:

```
CREATE FUNCTION Disjoint(PSBR, PSBR)
  RETURNS bool
  AS 'libPSBR','Disjoint'
  LANGUAGE C IMMUTABLE STRICT;
```

A.4. Database operations

After building the extension in PostgreSQL, we can perform several database operations with SQL commands, such as data manipulations and spatial queries.

A.4.1. Creating tables

We can create a table "provinces" to record the Chinese provinces represented by the PSBR:

CREATE TABLE provinces (id integer, name text, pset PSBR).

Similarly, the table "vague_region" records regions without a crisp boundary:

CREATE TABLE vague_region (id integer, name text, pset PSBR).

A.4.2. Appending data onto tables

After defining the schema of the table "provinces", it can store the PSBR approximations of all Chinese provinces described in Section 4.1. The SQL command for appending a record (Beijing in this example) is as follows:

```
INSERT INTO provinces
  VALUES (001,'Beijing','PSBR(ExPoint(116.22 40.22 1 City
NULL); . . ., ;
      ExPoint(116.84 40.37 1 City NULL),NULL)').
```

Each point involved in the PSBR object approximating Beijing has weight 1 and type "City", and its ICD relationship to the target region (i.e., Beijing) is not considered (cf. the second to last "NULL"). The last "NULL" in the statement means that the limiting region is unspecified. Likewise, the weighted PSBR representation of South China can be inserted into the table "vague_region" as a record.

A.4.3. Spatial queries

Based on the two tables and user-defined functions, we can make some spatial queries. For example, there are four ways to retrieve the provinces that overlap with or are inside South China R_{sc}. First, we can find a sub-PSBR of R_{sc} with the filter criterion that the weight of each point is greater than 0.5 (or some other reasonable value). Then, we can find the required provinces based on the direct topological relationships defined in Section 3.2. Here is the query statement:

```
SELECT provinces.id, provinces.name FROM provinces
  WHERE Overlap(provinces.pset, Subset(vague_region.pset,'
weight> = 0.5'))
  AND vague_region.name = 'South China'.
```

Note that the relationship cover is not considered, since R_{sc} is represented by a group of large cities and cannot cover the point set of any province that contains many small cities or towns.

The second approach is to convert the PSBRs in both tables to ordinary regions and to use the following query statement:

```
SELECT provinces.id, provinces.name FROM provinces
  WHERE (Overlap_RR(ConvexHull(provinces.pset),
  Interpolation(vague_region.pset,0.5)) OR
  Inside_RR(ConvexHull(provinces.pset),
Interpolation(vague_region.pset,0.5)))
  AND vague_region.name = 'South China',
```

where Overlap_RR and Inside_RR are two functions to decide the topological relationships between two ordinary regions, while Interpolation is a function to generate an ordinary region from a PSBR using the interpolation method. Obviously, this approach is time-consuming. Some alternatives can be implemented to improve the efficiency of queries, such as creating temporary tables to store intermediate variables or calling stored procedures.

The other two ways convert the PSBR objects in one table to ordinary regions while keeping the PSBR objects in the other table unchanged. The corresponding statements follows:

```
SELECT provinces.id, provinces.name FROM provinces
  WHERE Overlap_PR(Subset(vague_region.pset,'weight> =
0.5'),
  ConvexHull(provinces.pset))
  AND vague_region.name = 'South China'
```

and

```
SELECT provinces.id, provinces.name FROM provinces
  WHERE (Overlap_PR(provinces.pset, Interpolation(vague_
region.pset,0.5)) OR
  Inside_PR(provinces.pset,
Interpolation(vague_region.pset,0.5))
  AND vague_region.name= 'South China'.
```

where Overlap_PR and Inside_PR are two functions that decide topological relationships between a PSBR object and an ordinary region. From the above queries, we find five provinces (they are mentioned in Section 4.2, cf. Fig. 8a) from the first and third methods and eight provinces from the other two methods (the three additional provinces are Jiangxi, Hunan, and Guizhou). Note that there are other ways to find a region based on a PSBR object. For instance, we may use ConvexHull(Subset (vague_region.pset,'-weight>= 0.5')) instead of Interpolation(vague_region.pset,0.5). This will yield slightly different results.

References

Alani, H., Jones, C. B., & Tudhope, D. (2001). Voronoi-based region approximation for geographical information retrieval with gazetteers. *International Journal of Geographical Information Science, 15*(4), 287–306.

Ban, H., & Ahlqvist, O. (2009). Representing and negotiating uncertain geospatial concepts – Where are the exurban areas? *Computers, Environment and Urban Systems, 33*(4), 233–246.

Behr, T., & Schneider, M. (2001). Topological relationships of complex points and complex regions. In H. S. Kunii, S. Jajodia, & A. Sølvberg (Eds.), *ER 2001, lecture notes in computer sciences 2224* (pp. 56–69). Berlin: Springer-Verlag.

Bennett, B., & Agarwal, P. (2007). Semantic categories underlying the meaning of 'place'. In S. Winter, M. Duckham, L. Kulik, & B. Kuipers (Eds.), *COSIT 2007, lecture notes in computer sciences 4736* (pp. 78–95). Berlin: Springer-Verlag.

Bloch, I. (2005). Fuzzy spatial relationships for image processing and interpretation: A review. *Image and Vision Computing, 23*, 89–110.

Burrough, P. A. (1996). Natural objects with indeterminate boundaries. In P. A. Burrough & A. U. Frank (Eds.), *Geographic objects with indeterminate boundaries* (pp. 3–28). London: Taylor & Francis.

Chang, C. C., & Lin, C. J. (2001). LIBSVM: A library for support vector machines, software available at <http://www.csie.ntu.edu.tw/cjlin/libsvm>.

Coad, P., & Yourdon, E. (1999). *Object oriented analysis* (2nd ed.). NJ: Prentice Hall Press.

Cohn, A. G., & Gotts, N. M. (1996). The 'egg-yolk' representation of regions with indeterminate boundaries. In P. A. Burrough & A. U. Frank (Eds.), *Geographic objects with indeterminate boundaries* (pp. 171–188). London: Taylor & Francis.

Couclelis, H. (1996). Towards an operational typology of geographic entities with ill-defined boundaries. In P. A. Burrough & A. U. Frank (Eds.), *Geographic objects with indeterminate boundaries* (pp. 45–55). London: Taylor & Francis.

Deng, M., & Li, Z. (2008). A statistical model for directional relations between spatial objects. *GeoInformatica, 12*, 193–217.

Dilo, A., De By, R. A., & Stein, A. (2007). A system of types and operators for handling vague spatial objects. *International Journal of Geographical Information Science, 21*(4), 397–426.

Egenhofer, M. J., & Franzosa, R. (1991). Point-set topological spatial relations. *International Journal of Geographical Information Systems, 5*(2), 161–174.

Fisher, PF. (2000). Sorites paradox and vague geographies. *Fuzzy Sets and Systems, 113*, 7–18.

Fisher, P. F., Wood, J., & Cheng, T. (2005). Fuzziness and ambiguity in multi-scale analysis of landscape morphometry. In F. Petry, V. Robinson, & M. Cobb (Eds.), *Fuzzy modeling with spatial information for geographic problems* (pp. 209–232). Berlin: Springer-Verlag.

Fonte, C. C., & Lodwick, W. A. (2004). Areas of fuzzy geographical entities. *International Journal of Geographical Information Science, 18*(2), 127–150.

Goodchild, M. F. (1992). Geographical data modeling. *Computers & Geosciences, 18*(4), 401–408.

Goodchild, M. F. (2007). Citizens as voluntary sensors: Spatial data infrastructure in the world of Web 2.0. *International Journal of Spatial Data Infrastructures Research, 2*(2), 4–32.

Goodchild, M. F., & Hill, L. L. (2008). Introduction to digital gazetteer research. *International Journal of Geographical Information Science, 22*(10), 1039–1044.

Goodchild, M. F., Yuan, M., & Cova, T. J. (2007). Towards a general theory of geographic representation in GIS. *International Journal of Geographical Information Science, 21*(3), 239–260.

Gottmann, J. (1961). *Megalopolis: The urbanized northeastern seaboard of the United States*. New York: The Twentieth Century Fund.

Guo, Q., Kelly, M., & Grahamb, C. H. (2005). Support vector machines for predicting distribution of Sudden oak death in California. *Ecological Modelling, 182*, 75–90.

Hill, L. L. (2006). *Georeferencing – The geographic associations of information*. Cambridge: MIT Press.

Hwang, S., & Thill, J.-C. (2005). Modeling localities with fuzzy sets and GIS. In M. Cobb, F. Petry, & V. Robinson (Eds.), *Fuzzy modeling with spatial information for geographic problems* (pp. 71–104). Berlin: Springer-Verlag.

Jones, C. B., Purves, R. S., Clough, P. D., & Joho, H. (2008). Modelling vague places with knowledge from the web. *International Journal of Geographical Information Science, 22*(10), 1045–1065.

Liu, Y., Goodchild, M. F., Guo, Q., Tian, Y., & Wu, L. (2008). Towards a general field model and its order in GIS. *International Journal of Geographical Information Science, 22*(6), 623–643.

Liu, Y., Wang, X., Jin, X., & Wu, L. (2005). On internal cardinal direction relations. In A. G. Cohn & D. M. Mark (Eds.), *COSIT 2005, lecture notes in computer science 3693* (pp. 283–299). Berlin: Springer-Verlag.

Liu, Y., Yuan, Y., & Zhang, Y. (2008). A cognitive approach to modeling vague geographical features – A case study of Zhongguancun. *Journal of Remote Sensing, 12*(2), 370–377 (in Chinese).

Lynch, K. (1960). *The image of the city*. Cambridge: MIT Press.

Manevitz, L. M., & Yousef, M. (2001). One-class SVMs for document classification. *Journal of Machine Learning Research, 2*, 139–154.

Mennis, J. L., Peuquet, D. J., & Qian, L. (2000). A conceptual framework for incorporating cognitive principles into geographical database representation. *International Journal of Geographical Information Science, 14*(6), 501–520.

Montello, D. R., Goodchild, M. F., Gottsegen, J., & Fohl, P. (2003). Where's downtown? Behavioral methods for determining referents of vague spatial queries. *Spatial Cognition and Computation, 3*, 185–204.

Montello, D. R. (2003). Regions in geography: Process and content. In M. Duckham, M. F. Goodchild, & M. F. Worboys (Eds.), *Foundations of geographic information science* (pp. 173–189). London: Taylor & Francis.

Open GIS Consortium Inc. (1999). *OpenGIS Simple Features Specification*.

Randell, D.A., Cui, Z., & Cohn, A.G. (1992). A spatial logic based on regions and connection, In: *Proceedings of the third international conference of knowledge representation and reasoning*, San Mateo, CA, 1992. pp. 165–176.

Schneider, M. (1999). Uncertainty management for spatial data in databases: Fuzzy spatial data types. In R. H. Güting, D. Papadias, & F. Lochovsky (Eds.), *SSD'99, lecture notes in computer science 1651* (pp. 330–351). Berlin: Springer-Verlag.

Scholkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation, 13*, 1443–1471.

Smith, B. (1995). On drawing lines on a map. In A. U. Frank, W. Kuhn, & D. M. Mark (Eds.), *COSIT 1995, lecture notes in computer science 988* (pp. 475–484). Berlin: Springer-Verlag.

Sui, D. (2008). The wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS. *Computers, Environment and Urban Systems, 32*, 1–5.

Tang, X. (2004). *Spatial Object Modeling in Fuzzy Topological Spaces with Applications to Land Cover Change*. Ph.D. Dissertation. Ensched, Netherlands: International Institute for Geo-Information Science and Earth Observation.

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*, 234–240.

Tøssebro, E., & Nygård, M. (2002). An advanced discrete model for uncertain spatial data. In X. Meng, J. Su, & Y. Wang (Eds.), *WAIM 2002, lecture notes in computer science 2419* (pp. 37–51). Berlin: Springer-Verlag.

Tuan, Y. (1975). Place: An experiential perspective. *Geographical Review, 65*(2), 151–165.

Vapnik, V. (1995). *The Nature of statistical learning theory*. Berlin: Springer-Verlag.

Varzi, A. C. (1996). Parts, wholes, and part-whole relations: The prospects of mereotopology. *Data and Knowledge Engineering, 20*(3), 259–286.

Varzi, A. C. (2007). Spatial reasoning and ontology: Parts, wholes, and locations. In M. Aiello, I. Pratt-Hartmann, & J. van Benthem (Eds.), *Handbook of spatial logics* (pp. 945–1038). Berlin: Springer-Verlag.

Varzi, A. C. (2001). Vagueness in geography. *Philosophy & Geography, 4*(1), 49–65.

Wieczorek, J., Guo, Q., & Hijmans, R. J. (2004). The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science, 18*(8), 745–767.

Worboys, M. F., & Clementini, E. (2001). Integration of imperfect spatial information. *Journal of Visual Languages and Computing, 12*, 61–80.

Winston, M. E., Chaffin, R., & Herrmann, D. (1987). A taxonomy of part-whole relations. *Cognitive Science, 11*(4), 417–444.

Yao, X., & Thill, J.-C. (2006). Qualitative queries with qualitative locations in spatial information systems. *Computers, Environment and Urban Systems, 30*(4), 485–502.