Journal of Visual Languages and Computing 23 (2012) 213–222

Contents lists available at SciVerse ScienceDirect

Journal of Visual Languages and Computing

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

Probability issues in locality descriptions based on Voronoi neighbor relationship ☆

Yongxi Gong [a,b], Lun Wu [a], Yaoyu Lin [b,c], Yu Liu [a,*]

[a] Institute of Remote Sensing & Geographical Information Systems, Peking University, Beijing 100871, PR China
[b] School of Urban Planning and Management, Harbin Institute of Technology Shenzhen Graduate School, Shenzhen 518055, PR China
[c] State Key Laboratory of Subtropical Building Science, South China University of Technology, Guangzhou 510641, PR China

ARTICLE INFO

Article history:
Received 29 July 2008
Received in revised form
28 March 2012
Accepted 30 April 2012
Available online 9 May 2012

Keywords:
Voronoi diagram
Probability function
Locality description
Voronoi neighbor relationship

ABSTRACT

Spatial relationships play an important role in spatial knowledge representation, such as in describing localities. However, little attention has been paid to how to describe the position of a target object (TO) with a qualitative referencing system that consists of a set of reference objects (ROs) in the locality description context. We propose a method that accounts for the differences between two scenarios in locality descriptions. This method is probabilistic and is based on the Voronoi neighbor relationship to determine candidate ROs for describing a given TO's position in the second scenario. The Voronoi neighbor relationship is adopted to determine candidate ROs of a TO and to compute the neighboring area of an RO. A probability function is presented to model the uncertainty of selecting appropriate ROs. To build locality descriptions that are consistent with commonsense, four constraints are placed on the probability function. Two probability functions based on Euclidean distance and stolen-area, and a mixed probability function that considers both Euclidean distance and stolen-area, are analyzed and compared. With the mixed probability function, we establish a method to construct the locality description of a given TO. Finally, three examples demonstrate how to select ROs to describe a TO's position.

1. Introduction

In recent decades, a substantial amount of effort has been made to research spatial knowledge representation from a variety of disciplines, such as artificial intelligence (AI), geographical information science (GIScience), psychology, linguistics, and philosophy [1–7]. Spatial relationships, especially binary spatial relationships, play an important role in spatial knowledge representation and are extensively investigated.

It is argued that three components, i.e., the reference object (RO), the target object (TO), and the spatial relationship (SR), should be considered in a spatial assertion for describing a locality [8]. For example, in the statement "A is inside B" for describing the location of A, A and B are the TO and the RO, respectively, and the topological relationship "inside" serves as the SR. A locality description often takes named places as the ROs to depict a TO's location with spatial relationships [9] but not with numerical coordinates. When a locality description is communicated among people verbally or textually, a set of known spatial entities is preferred to be taken as ROs to qualitatively describe a TO's position with specific spatial relationships [10].

When a set of ROs is known, our research aims to provide a method for determining which ROs are selected

☆ This paper has been recommended for acceptance by Shi Kho Chang.

* Corresponding author. Tel./fax: +86 10 62753220.
E-mail address: liuyu@urban.pku.edu.cn (Y. Liu).

1045-926X/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jvlc.2012.04.002

to build locality descriptions that are consistent with commonsense. Two aspects of locality descriptions should be considered. The first aspect is the difference between the spatial knowledge representation context and the locality description context. In general, a locality description is a special case of the spatial knowledge representation, with more requirements. In the spatial knowledge representation context, a spatial assertion should be logically correct, whereas in the locality description context, not only logical correctness but also spatial cognition reasonability should be required [11]. For example, the assertion “California is to the southwest of Montana” shows the relative direction between California and Montana in the spatial knowledge representation context (Fig. 1), but Montana is not the best RO to represent the location of California in the locality description context. Statements such as “California is to the south of Oregon”, “California is to the southwest of Nevada”, or “California is to the northwest to Arizona” are more consistent with commonsense than the Montana statement because people tend to use known spatial entities near the TO to describe the position of the TO.

Second, two scenarios can be found in locality descriptions according to spatial cognition [12,13]. Spatial knowledge of places is acquired in three aspects according to spatial cognition: landmark knowledge, route knowledge, and survey knowledge [14]. Route knowledge consists of ordered sequences of landmarks, while survey knowledge is the spatial configuration of the features in the environment. According to the different aspects of spatial knowledge, two scenarios, i.e., Scenario I and Scenario II, can be addressed in locality descriptions. When one arrives at a new environment, he (or she) will acquire the spatial knowledge of the prominent spatial entities, i.e., landmark knowledge. This is the Scenario I for describing localities. In Scenario I, people organize spatial knowledge with landmarks as anchor points, and any point in the environment will be attached to the nearest landmarks. In this scenario, the reference systems are constructed based on landmarks, and each reference system is separated from the other reference systems.

Fig. 1. Spatial descriptions is various in spatial knowledge representation context and in locality description context.

When one acquires survey knowledge of the environment, the spatial relationships of the landmarks are established, and a spatial entity will be related to multiple landmarks. This is the Scenario II in locality descriptions. In Scenario II, the reference frameworks based on each landmark can be in a unified spatial reference framework.

Human spatial cognition is so complex that the manner in which people select ROs to position a TO cannot be modeled exactly. Uncertainty is, therefore, inevitable and should be addressed in locality descriptions. Probabilistic methods have been widely investigated to address the uncertainty of spatial knowledge representation and reasoning [15–17]. We explore the probabilistic approach based on Voronoi neighbor relationship to build locality descriptions in Scenario II. Four constraints are placed on the probability function to make the corresponding description consistent with spatial commonsense. The method presented here is an attempt to describe a TO’s position in such a way that it is more consistent with human spatial commonsense. We do not claim that this method is a complete solution, but we believe that it is an improvement over conventional methods. To focus on the model, we assume that ROs and TOs are points, and they can be extended into polylines and polygons to apply the method presented in this paper.

The remainder of this paper is structured as follows. Section 2 introduces a method for deciding the candidate ROs to describe a given TO’s position. Section 3 proposes an approach for calculating the neighboring area of an RO. When a TO is in the neighboring area of a spatial entity, the spatial entity is a candidate RO for describing the TO’s position. In Section 4, we placed four constraints on the probability function to ensure that the locality description of the TO will be more consistent with spatial commonsense, and three probability functions are analyzed and compared. In Section 5, we propose a method to build locality descriptions with the mixed probability function based on Euclidean distance and stolen-area. In Section 6, three examples are presented to demonstrate how to describe a TO’s location. We conclude our work and describe future work in Section 7.

2. Candidate ROs for a TO

In this research, both ROs and TOs are spatial entities that are defined as follows.

Definition 0. ***Spatial entity***: Spatial entities are objects embedded in geographical space. A spatial entity is generally perceivable and has a name so that it can be textually represented. For simplicity, all entities are abstracted into points in a two-dimensional space. Note that the number of points may be infinite.

Given a reference system that consists of a set of ROs in two-dimensional space, some literature focuses on selecting a subset of ROs to describe the TO’s location. The qualitative vector space (QVS) method models the qualitative position of a TO by taking into account the distance between the TO and each element in the reference system [18]. Although it is logically correct in spatial reasoning context, it is not the most reasonable method in the locality description context. The capacity of the working memory has restriction [19]; therefore, only a subset of the ROs from the reference system could be considered to position the TO.

In the landmark-based qualitative reference (LBQR) method [20], Voronoi diagrams of all ROs are generated, and the location of a TO is described by the ROs that are nearest to the TO. If the TO falls into the Voronoi region of an RO, then the TO is positioned by the RO, for example T_1 is positioned by R_1 in Fig. 2. If the TO is on an edge or a vertex of the Voronoi diagram, then the TO would be anchored to two or more ROs, for example T_2 positioned by R_1 and R_2, or T_3 positioned by R_1, R_5, and R_6, as in Fig. 2.

The LBQR system is a method for describing the TO's location in Scenario I. In Scenario I, boundaries among the landmarks' Voronoi regions are crisp, so that the position of a TO is presented with only one landmark unless the TO is on the edge or the vertex of Voronoi diagram, following the LBQR model. Locality descriptions in Scenario II, however, are different from those in Scenario I. Human spatial knowledge is qualitative and generally vague, thus the uncertainty of deciding which ROs are selected to describe the position should be taken into account. Not only the relationships between TOs and ROs, but also the uncertainty of selecting landmarks as ROs were not considered in the LBQR method. These issues should be taken into account in Scenario II, where people have much survey knowledge about the environment and the spatial configuration of the landmarks. Suppose that a TO is between R_1 and R_2 and moves from R_1 to R_2 along the line between R_1 and R_2. When the TO exits the Voronoi region of R_1 and enters the Voronoi region of R_2, the RO will change from R_1 to R_2 following the LBQR method. But in Scenario II, we can describe the TO's location using both R_1 and R_2 with different probabilities. The probability describing the TO's location with R_1 increases and the probability with R_2 decreases when the TO moves from R_2 toward R_1. When the TO is very near to R_1, people would like to use only one RO R_1 to describe the TO's location. Therefore, people are accustomed to adopting multiple ROs to locate a TO with a more precise description in Scenario II beyond the situation that the TO is on the edges of the Voronoi diagram in Scenario I. Thus, a new method should be developed to build qualitative locality descriptions that are more consistent with human spatial commonsense in Scenario II.

In Scenario II, the ROs neighboring to a TO are possibly selected to describe the TO's position. The meaning of the neighbor based on Voronoi diagram has been investigated in [21–24]. We present the definition of neighbor relationship among the spatial entities as follows.

Definition 1. ***Voronoi neighbor relationship between two spatial entities***: For a set of spatial entities on the plane, two entities neighbor to each other if they share a common edge of a Voronoi diagram.

If two spatial entities neighbor to each other, they are neighbors. Usually, a spatial entity has several neighbors. For example, in Fig. 3, neighbors of R' are R_1, R_2, R_3, R_4, R_5 and R_6.

Because the position of a TO can be described by its neighbors, R' can be positioned by one or several of its neighbors if R' in Fig. 3 is not a site of the Voronoi diagram but instead is a TO to be positioned.

Definition 2. ***Candidate ROs of a TO***: Let S denote a set of ROs on the plane. The candidate ROs of a TO T are a subset of S that neighbor to T. We denote candidate ROs of a TO with $CandRef(T)$.

When T has been inserted into the Voronoi diagrams, it neighbors to R_1, R_2, R_3, R_4, R_5 and R_6 (Fig. 4). Therefore, $CandRef(T)$ includes R_1, R_2, R_3, R_4, R_5 and R_6, and the position of T can be described by one or several of them.

3. Neighboring area of an RO

We can compute the neighboring area of an RO using Definition 3.

Definition 3. ***Neighboring area of an RO***: The neighboring area of an RO R is the region D that satisfies $\forall T \in D$, $R \in CandRef(T)$, where T is a point. We denote the neighboring area of an RO R with $D(R)$.

If the neighboring areas of all ROs have been computed, whether an RO is in $CandRef(T)$ of a TO can be decided by

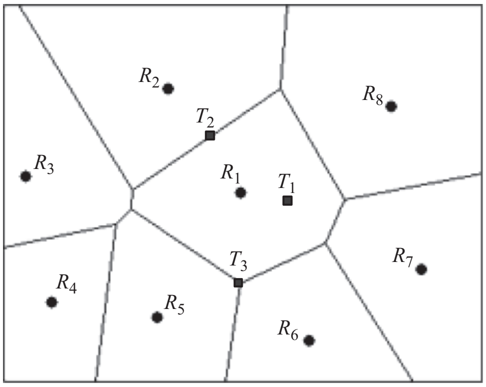

Fig. 2. Locality descriptions based on the LBQR method.

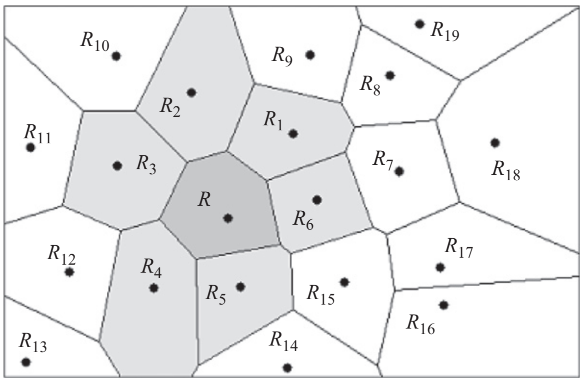

Fig. 3. Voronoi neighbors of R'.

judging whether the TO is in the neighboring areas of the RO.

Let R, R_1, and R_2 denote three vertices in a Delaunay triangulation with its circumcircle centering at O, as shown in Fig. 5. Therefore, the circumcircle does not contain other ROs [25]. In the area enclosed by lines RR_1 and RR_2, when a point R' is outside the circumcircle, R_1 neighbors to R_2 and the Voronoi regions of R and R' are separated (Fig. 5a); when R' is on the edge of the circumcircle, then the boundaries of the Voronoi regions of R_1, R_2, R, and R' intersect at point O (Fig. 5b); when R' is inside the circumcircle, the neighbor relationship between R and R' changes, and R' and R share a common edge for the circumcircle of the Delaunay triangle should not contain any other sites (Fig. 5c).

Therefore, in the area enclosed by RR_1 and RR_2, the sub-neighboring area of R is a conic polygon whose boundary consists of line segments and circle arcs, i.e., line segment RR_2, circle arc R_2R_1 in the counterclockwise direction, and line segment R_1R. For every Delaunay triangulation with R as one of its vertices, there is a sub-neighboring area of R, denoted by $D_i(R)$. As a result, the neighboring area $D(R)$ of R can be computed by combining all of the sub-neighboring areas:

$$D(R) = \cup D_i(R) \qquad (1)$$

The neighboring area of R is shown in Fig. 6. If neighboring areas of all ROs have been computed, those ROs whose neighboring areas cover the TO are elements of $CandRef(T)$ to describe the TO's position.

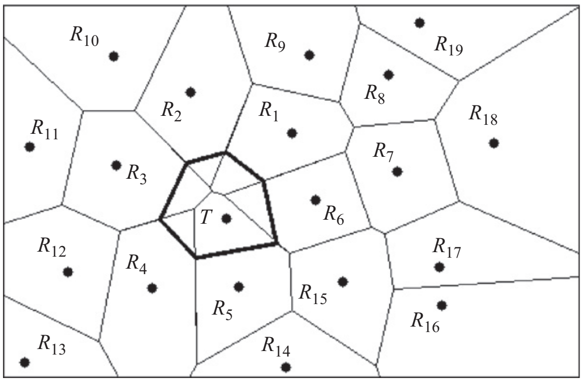

Fig. 4. The TO T can be located with its neighbors R_1, R_2, R_3, R_4, R_5, and R_6.

4. Probability of selecting an RO to describe the position of the TO

4.1. Constraints on the probability function

Neighboring areas of ROs are computed to decide the $CandRef(T)$ of a TO, but the importances of different ROs in a locality description are various. For example, when the TO T is at the position of an RO R, the locality description will select only R to describe the location of T with the highest probability, whereas other ROs neighbor to T would not be likely to be selected to position T. When T is between two ROs, or among several ROs, then multiple neighboring ROs could be selected to describe the location of T. Therefore, we should take the probability of each element in $CandRef(T)$ of a TO into account to build locality descriptions consistent with human spatial commonsense.

The probability of an RO being selected to describe a TO's position is difficult to determine because many factors influence spatial cognition. With some constraints on the probability function, however, we can find a model for building locality descriptions more consistent with human spatial commonsense than conventional method.

Let $P(x,y,R_i)$ denote the probability that an RO $R_i \in CandRef(T)$ is selected to position the TO T with coordinates (x,y). The following four constraints should be maintained.

Constraint 1.

$$\sum_{R_i \in CandRef(T)} P(x,y,R_i) = 1$$

Constraint 1 implies that the position of the TO T should be described by the ROs neighbor to the TO, and

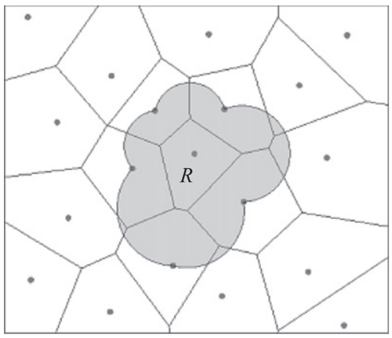

Fig. 6. The neighboring area of R (conic polygon in gray).

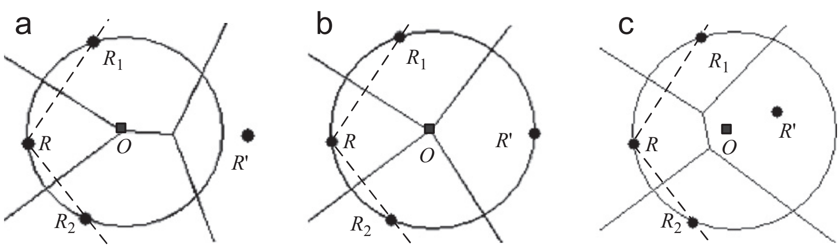

Fig. 5. Change of neighbor relationships among R_1, R_2, R and R'.

the probabilities that ROs are selected to locate the TO are not independent.

Constraint 2. $P(x,y,R_i)=1$, if the TO is at the position of R_i with coordinates (x',y').

Constraint 2 indicates that when a TO is at the position of an RO R_i, the TO can be positioned with only one RO R_i.

Constraint 3.

$P(x,y,R_i)=0,\forall(x,y)\notin D(R_i)$.

Constraint 3 shows that, only if a TO is inside $D(R_i)$, R_i is one of the $CandRef(T)$ of the TO.

Constraint 4.

$$\lim_{(x-x')^2+(y-y')^2\to 0}(P(x,y,R_i)-P(x',y',R_i))\to 0$$

Constraint 4 means that the probability of an RO to be selected to describe the position of the TO on the plane is continuous.

These four constraints are elementary restrictions on the probability function. Additional conditions can be placed on the probability function in different applications contexts. The higher the probability of an RO is, the more likely the RO is selected to position the TO.

Three probability functions are provided and compared in the following sections, and the one that is more consistent with spatial commonsense will be used to build the locality description of a TO.

4.2. Probability function based on Euclidean distance

Let $d(x,y,R_i)$ denote the Euclidean distance between a TO at (x,y) and a RO R_i. The ability of the R_i to describe the position of the TO is the inverse ratio of $d(x,y,R_i)$, for example $1/d(x,y,R_i)^n$. This model indicates that the greater the distance between the TO and the RO R_i, the less the ability of R_i is to describe the TO's position. This model is also known as the Newtonian model or the gravity model in spatial analysis and transportation [26–28].

The most used ability functions are $1/d(x,y,R_i)$ and $1/d(x,y,R_i)^2$. Let $f(x,y,R_i)=d(x,y,R_i)^n$ $(n=1,2)$, and $P(x,y,R_i)$ be the probability of R_i to describe a TO with the coordinates (x,y). The probability of R_i being selected to describe the TO at (x,y) is the ratio of the ability of R_i to the sum of the abilities of all ROs in $CandRef(T)$. This ratio can be computed as follows:

$$P(x,y,R_i)=1/f(x,y,R_i)\Bigg/\sum_{R_k\in CandRef(T)}(1/f(x,y,R_k))$$
$$=\prod_{\substack{R_k\in CandRef(T),\\ R_k\neq R_i}} f(x,y,R_k)\Bigg/\sum_{R_k\in CandRef(T)}\left(\prod_{\substack{R_l\in CandRef(T),\\ R_l\neq R_k}} f(x,y,R_l)\right) \tag{2}$$

Obviously, Eq. (2) makes Constraint 1 hold. Eq. (2) also satisfies Constraint 2 because the denominator of Eq. (2) is $\prod_{\substack{R_l\in CandRef(T),\\ R_l\neq R_k}} f(x,y,R_k)$ when the TO is at the position of the RO R_i. However, Eq. (2) does not satisfy Constraint 3 for $P(x,y,R_i)\neq 0$ when the TO is on the boundary or out of $D(R_i)$. If we enforce $P(x,y,R_i)=0$ when $(x,y)\notin D(R_i)$, there is a discontinuous change of the probability near the boundary of $D(R_i)$. Inside $D(R_i)$, there are several discontinuous changes of the probability because the probability is decided by not only R_i but also other neighbor ROs of the TO. This will lead to a conflict to the fourth constraint. For example, in Fig. 7, when a TO with coordinates (x,y) is on the boundary of $D(R_j)$ and inside $D(R_i)$, the RO R_j should be considered in the computation of $P(x,y,R_i)$, with $\prod_{\substack{R_l\in CandRef(T),\\ R_l\neq R_k}} f(x,y,R_k)\neq 0$. However, when a TO with coordinates (x',y') is not inside $D(R_j)$ but instead is inside $D(R_i)$, and $(x-x')^2+(y-y')^2\to 0$, the proportion of R_j is not considered, which leads to discontinuous changes of $P(x,y,R_i)$ at the position near to the boundary of $D(R_j)$.

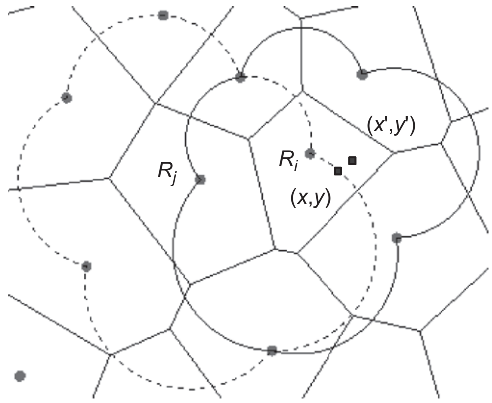

Fig. 7. $P(x,y,R_i)$ discontinuously changes near the boundary of $D(R_j)$.

The reason that Eq. (2) does not satisfy all constrains is that the probability function takes all ROs into account, but $CandRef(T)$ are those who neighbor to the TO. The spatial distribution of $P(x,y,R_i)$ with $f(x,y,R_i)=d(x,y,R_i)$ and $f(x,y,R_i)=d(x,y,R_i)^2$ are shown in Fig. 8a and b, respectively. Obviously, the discontinuity of $P(x,y,R_i)$ is near to the boundary of $D(R_i)$. In $D(R_i)$, $P(x,y,R_i)$ also changes discontinuously near the boundaries of other neighboring areas of the RO.

4.3. Probability function based on stolen-area

An alternative is the model based on stolen-area. "Stolen-area" has been discussed in [21–24,29,30]. When a new site is inserted into the Voronoi diagram, it will steal Voronoi regions from existing sites.

Definition 4. ***Stolen-area***: when a new site is inserted into an existing Voronoi diagram, a stolen-area is the region that is a part of the Voronoi region of the original sites but now belongs to the Voronoi region of the new site after the new site has been inserted into the Voronoi diagram.

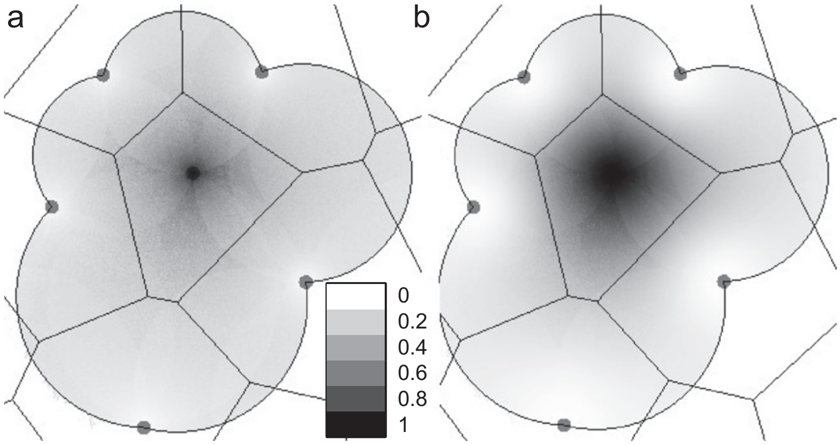

Fig. 8. Distribution of $P(x,y,R_i)$ based on Euclidean distance: (a) $f(x,y,R_i)=d(x,y,R_i)$ and (b) $f(x,y,R_i)=d(x,y,R_i)^2$.

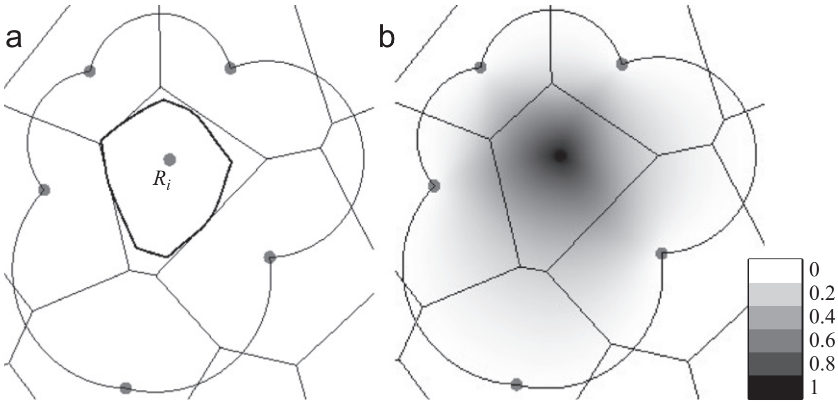

Fig. 9. The model based on stolen-area: (a) dominant area of the RO R_i based on stolen-area (in bold lines) and (b) distribution of $P(x,y,R_i)$ based on stolen-area.

Stolen-areas are associated with spatial neighbor relationships. When a new site is inserted into the Voronoi diagram and steals area from an existing site, the new site obviously neighbors to the original site.

Let A denote the total area of the Voronoi region of the a TO, and let A_i be the area that the TO steals from the RO R_i. The probability of R_i to be selected to describe a TO's position can be measured by the following:

$$P(x,y,R_i)=A_i/A \tag{3}$$

Because stolen-area relates to the neighbor spatial relationship, it is appropriate to be adopted stolen-area to compute the probability of an RO. This model holds all four constraints. Since $\sum A_i=A$, the model satisfies Constraint 1. The model meets Constraint 2 because a TO would steal the whole Voronoi region of R_i but no area from other ROs when the TO is at the position of R_i. Obviously, this model will satisfy Constraint 3. Constraint 4 is maintained because $P(x,y,R_i)$ changes continuously with the area stolen. An RO's dominance of $P(x,y,R_i)$ based on stolen-area is not the same as its Voronoi region (Fig. 9a), but its shape tends to approximate the Voronoi region. An example of the distribution of $P(x,y,R_i)$ is shown in Fig. 9b. The change of $P(x,y,R_i)$ is continuous.

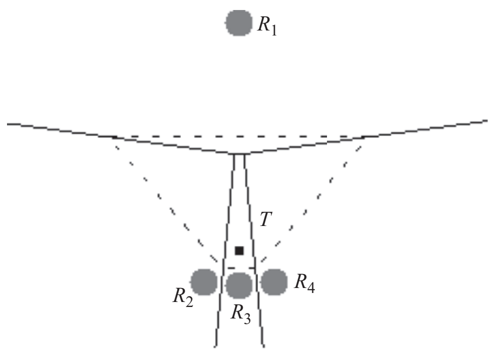

Fig. 10. A case for which the stolen-area-based probability function is inconsistent with commonsense.

Although the probability function based on stolen-area holds four constraints, it could be inconsistent with commonsense in special cases. As shown in Fig. 10, R_1, R_2, R_3 and R_4 are ROs on the plane, and the last three are close to each other. According to intuition, the probabilities of the last three ROs to be selected to locate T are similar and that of R_1 is much lower than the other three. If we follow the method based on stolen-area, however, the probabilities of R_1, R_2, R_3, and R_4 being selected to describe the position of T are 0.12, 0.38, 0.38, and 0.12, respectively. The probability of R_3 is equal to that of R_1, but is much less than that of R_2 and R_4. Obviously, these results are inconsistent with spatial commonsense.

4.4. Mixed probability function based on Euclidean distance and stolen-area

The probability function based on Euclidean distance is original from the global model, which does not satisfy all constraints. The model based on stolen-area meets all constraints but could be inconsistent with commonsense in some special cases. Meanwhile, Euclidean distance is an important factor that should be considered in locality descriptions. Therefore, taking both stolen-area and Euclidean distance into account to build a new model is a better solution.

We conceive the model as follows:

$$P(x,y,R_i)=(A_i/(f(R_i)))\Big/\sum_{R_l\in CanfRef(T)}(A_l/f(R_l))$$

$$=A_i\prod_{\substack{R_k\in CanfRef(T),\\ R_k\neq R_i}} f(R_k)\Bigg/\sum_{R_k\in CanfRef(T)}\left(A_k\prod_{\substack{R_l\in CanfRef(T),\\ R_l\neq R_k}} f(R_l)\right) \tag{4}$$

where $f(x,y,R_i)=d(x,y,R_i)^n$ $(n=1,2)$ and A_i is the area stolen from R_i by the TO.

The mixed model obviously holds Constraints 1 and 2. On the boundary or outside of $D(R_i)$, $A_i \prod_{\substack{R_l\in CanfRef(T),\\ R_l\neq R_i}} f(R_k)$ is 0; therefore, $P(x,y,R_i)$ is 0 when the TO is on the boundary or outside of $D(R_i)$, and this makes Constraint 3 hold. The proportions of R_i go to 0 continuously near the boundaries of $D(R_i)$, which leads to a continuous change of $P(x,y,R_i)$ near the boundaries of $D(R_i)$. The proportions of other ROs whose neighboring areas intersect with $D(R_i)$ also go to 0 continuously near the boundaries of their neighboring areas, which makes $P(x,y,R_i)$ change continuously inside of $D(R_i)$. Therefore, the probability $P(x,y,R_i)$ is continuous on the plane, with the result that Constraint 4 is maintained. Examples of the distribution of $P(x,y,R_i)$ with $f(x,y,R_i)=d(x,y,R_i)$ and $f(x,y,R_i)=d(x,y,R_i)^2$ are shown in Fig. 11.

The dominance of the mixed model is much closer to the Voronoi region than that based on stolen-area because it considers the factor of Euclidean distance. The boundary of the dominance in this model is between the boundary of the Voronoi region and the boundary of the dominance in the model based on stolen-area (Fig. 12).

The mixed model adopts the advantages and avoids the disadvantages of the method based on Euclidean distance and the method based on stolen-area, making it a better approach to build the locality description of a TO. In the following contents, the examples of locality descriptions adopt the mixed model with $f(x,y,R_i)=d(x,y,R_i)$ to position the TO.

5. Locality descriptions based on the probability function

When the position of a TO is to be described, *CandRef*(T) of the TO should be computed first. If there is more than one RO neighbor to the TO, we should determine when to position the TO with a single RO and when to build the locality description with multiple ROs. When

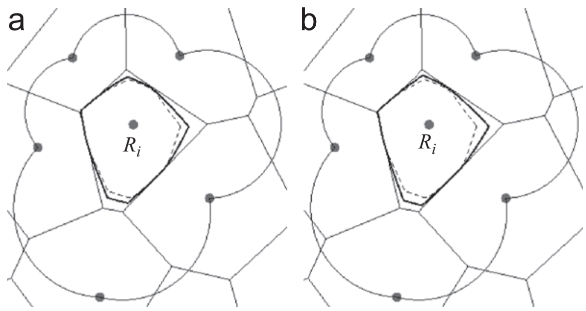

Fig. 12. Dominant areas of the landmark R_i in mixed model when $f(x,y,R_i)=d(x,y,R_i)$ and $f(x,y,R_i)=d(x,y,R_i)^2$, where the dominances in the new model are in bold and the dominance in the method based on stolen-area is in dash lines, for comparison: (a) $f(x,y,R_i)=d(x,y,R_i)$ and (b) $f(x,y,R_i)=d(x,y,R_i)^2$.

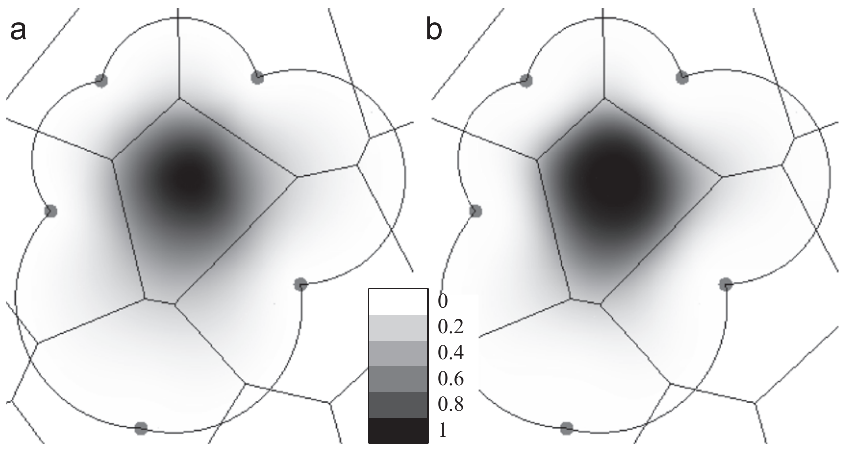

Fig. 11. Distribution of $P(x,y,R_i)$ of the mixed model: (a) $f(x,y,R_i)=d(x,y,R_i)$ and (b) $f(x,y,R_i)=d(x,y,R_i)^2$.

the TO is at the position of an RO, a locality description with a single RO would provide precise positional information of the TO. However, in most cases, the TO is between two ROs or among several ROs, and multiple ROs should be used to position the TO.

Which ROs should be used to describe the position of a TO can be determined by selecting the minimum set of ROs with their probability sum no less than a threshold ε. The threshold is very important in locality descriptions because it can determine whether an RO will be selected to describe the TO's location. If the threshold is low, then fewer ROs will be selected and the locality description will not be precise. If the threshold is high, then taking some ROs, for example, an RO with very low probability, to describe the TO will deviate the TO's position in locality descriptions. The value of ε can be decided according to experience or according to the requirement of the application. Its value should be greater than 0.5.

Let *CandRef*(T) be a set of neighbor ROs of a TO, and S' be an empty set. The detailed steps are as follows:

(1) determine a threshold of the probability value, named ε;
(2) compute the probabilities of all ROs in *CandRef*(T);
(3) remove the RO R with maximum probability from *CandRef*(T);
(4) put R into S' and sum up the probabilities of all ROs in S';
(5) if the sum is no less than ε, then end this algorithm; otherwise, go to step (3).

Those ROs in S' are adopted to describe the TO's location. For example, when a TO T is very near to an RO R, or at the position of R, the probability of R is no less than ε, and R can be used to describe the location of T with the representation "T is near R". When a TO T is between two ROs R_1 and R_2, and the sum of their probabilities is no less than ε, R_1 and R_2 are adequate to describe the position of T with the representation "T is near R_1 and R_2". When T is among several ROs, more ROs could be adopted to describe the position of T, for example, "T is near R_1, R_2, and R_3".

6. Case study

We will introduce some examples to demonstrate how to determine the ROs and how to build the locality description of a TO. Some TOs are marked in Fig. 13. In these examples, we found that 0.95 is a sound value to model locality descriptions according to the experience.

The following example shows how to determine when a single RO and which RO is selected for describing a TO's position in Scenario II.

Example 1. As shown in Fig. 13, describe the position of T_1 with the RO(s).

The TO T_1 neighbors to R_5, R_6, R_8, and R_9, and their probabilities to be selected to position T_1 are as follows:

R_5: 0.0056
R_6: 0.0236
R_8: 0.9633
R_9: 0.0075

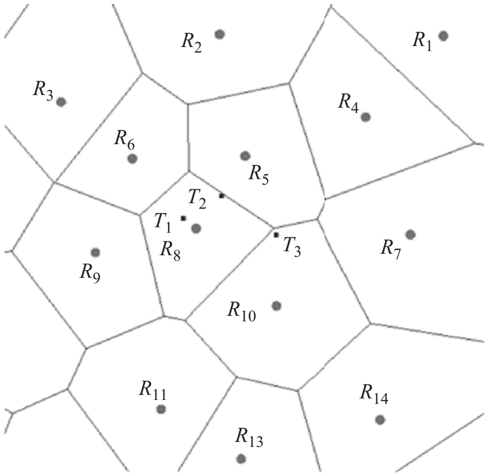

Fig. 13. TOs to be located in the case study.

Because the probability of R_8 is 96.33% and greater than ε, R_8 is adopted to describe the location of T_1. The position of T_1 can be described as follows:

"T_1 is near R_8".
or
"T_1 is northwest to R_8".

If more ROs, for example, R_8 and R_6, are selected to describe T_1 with the description "T_1 is near R_8 and R_6", audiences of the description may interpret that the TO is more possible in the middle of R_8 and R_6. This representation will not be as precise and concise as the description that uses only a single RO, R_8.

If the possibilities of all ROs in *CandRef*(T) are less than the threshold, however, one RO is not sufficient to describe the position of the TO precisely and concisely, and multiple ROs should be adopted.

Example 2. Describe the location of T_2 in Fig. 13.

The TO T_2 is between R_5 and R_8. Therefore, a single RO could not provide precise positional information of the TO, and multiple ROs should be considered in the locality description. The neighbor ROs of T_2 are R_5, R_6, R_8, and R_{10}, with probabilities as follows:

R_5: 0.4741
R_6: 0.0141
R_8: 0.4923
R_{10}: 0.0195

Because no RO's probability is greater than the threshold, a single RO is not sufficient to describe the location of T_2. Probabilities of R_5 and R_8 sum up to 0.9664, which is greater than the threshold. Therefore, R_5 and R_8 are selected to position T_2, and the representation is as follows:

"T_2 is near R_8 and R_5".

Example 3. Describe the position of T_3 in Fig. 13.

Compared with T_2, the RO T_3 is among several ROs. Thus, three or more ROs are required to describe the position of T_3. The neighbor ROs of T_3 are R_4, R_5, R_7, R_8, and R_{10}, with probabilities as follows:

R_4: 0.0165
R_5: 0.2942
R_7: 0.0800
R_8: 0.1437
R_{10}: 0.4656

Although the probability of R_{10} is much higher than the others, the differences of the probabilities of neighbor ROs are not great. If only one or two ROs are selected to describe the TO, it is difficult to describe the position of T_3 very well. R_{10}, R_5, R_8, and R_7 are the minimum set of ROs in which the probabilities sum to no less than the threshold. Therefore, they are adopted to represent the qualitative position of T_3 as

"T_3 is near R_{10}, R_5, R_8, and R_7".

7. Conclusions and future work

People typically use named places as the ROs to qualitatively describe the position of a TO in everyday life. Given a set of ROs on the plane and a TO, two scenarios are available in locality descriptions. In Scenario II, those ROs neighboring to a given TO are usually selected to textually describe the TO's location. The probabilities of the neighbor ROs for a TO, however, are different. We present a probabilistic method based on Voronoi diagram to build locality descriptions of TOs in Scenario II. The main contributions of this paper are as follows:

(1) Neighbor relationship based on Voronoi diagram is adopted to determine candidate ROs associated with a given TO.
(2) The neighboring area of an RO is computed to decide whether the RO is one of the candidate ROs of a given TO.
(3) We use probability functions to model the uncertainty in selecting ROs, and four constraints are placed on the probability function to build locality descriptions more consistent with human spatial commonsense.
(4) Three probability functions, i.e., the probability function based on Euclidean distance, the probability function based on stolen-area, and the mixed probability functions based on Euclidean distance and stolen-area, are provided and analyzed according to the four constraints, and the mixed probability function based on Euclidean distance and stolen-area holds all constraints and is more consistent with commonsense than the others.
(5) Based on the mixed probability function, a minimum set of ROs in which the sum of probabilities is greater than a threshold is used to build the qualitative locality description of a given TO, and several examples are presented to demonstrate how to textually describe the location of a given TO.

This research focuses mainly on the computational perspective of locality descriptions in Scenario II. In qualitative spatial representations and reasoning, two aspects are emphasized [2,10,31]. One aspect is the formalization of spatial knowledge so that qualitative spatial knowledge can be modeled and manipulated within a computational system, whereas the other aspect is spatial cognition, with which the computation of the spatial knowledge is in accord with the perception of human beings. Although we placed four constraints on the probability function to make the model work in the way in which people do in everyday life, it would be more useful if cognition experiments are conducted to validate and improve the model.

The method in this paper considers only the neighbor relationship to decide a subset of ROs to position a TO. The cardinal direction relationship, however, is important in locality descriptions. Many models, such as the cone-based method [32], the projection-based method [33], and the double-cross calculus method [34], have been proposed. For a fuzzy cardinal direction relationship, the membership degrees of different positions are various. When a cardinal direction relationship is adopted for describing the TO with given ROs, not only the probability but also the membership degree associated with each RO should be considered. Therefore, how to qualitatively describe the TO with cardinal direction is another research task in the future.

Acknowledgments

The authors would like to thank the reviewers for their comments. This work was supported by the National High Technology Development 863 Program of China (Grant 2011AA120301), the National Natural Science Foundation of China (Grants 40928001, 40901080, 51008002, and 41171296), the Research Fund for the State Key Laboratory of Subtropical Building Science (Grant 2011KB20), and the Research Fund for the Harbin Institute of Technology (Grants HIT.NSRIF.2013100 and HIT.NSFIR.2011126).

References

[1] A.G. Cohn, J. Renz, Qualitative spatial representation and reasoning, in: F. van Harmelen, V. Lifschitz, B. Porter (Eds.), Handbook of Knowledge Representation, Elsevier, 2007, pp. 1–47.
[2] M.F. Goodchild, A geographer looks at spatial information theory, in: D.R. Montello (Ed.), Proceedings of the COSIT 2001, Lecture Notes in Computer Science, vol. 2205, Springer, Berlin, 2001, pp. 1–13.
[3] B. Tversky, P.U. Lee, How space structures language, in: C. Freksa, C. Habel, K.F. Wender (Eds.), Spatial Cognition: An Interdisciplinary Approach to Representing and Processing Spatial Knowledge, Lecture Notes in Artificial Intelligence, vol. 1404, Springer, Berlin, 1998, pp. 157–175.
[4] T. Barkowsky, Mental processing of geographic knowledge, in: D.R. Montello (Ed.), Proceedings of the COSIT2001, Lecture Notes in Computer Science, vol. 2205, Springer, Berlin, 2001, pp. 371–386.
[5] J. Renz, Qualitative Spatial Reasoning with Topological Information, Springer, Berlin, 2003 (p. 190).
[6] B. Tversky, Cognitive maps, cognitive collages, and spatial mental models, in: U.F. Andrew, I. Campari (Eds.), Proceedings of the COSIT 1993, Lecture Notes in Computer Science, vol. 716, Springer, Berlin, 1993, pp. 14–24.

[7] B. Smith, D.M. Mark, Ontology and geographic kinds, in: Proceedings of the International Symposium on Spatial Data Handling (SDH'98), Vancouver, Canada, 1998, pp. 12–15.
[8] Q. Guo, Y. Liu, J. Wieczorek, Georeferencing locality descriptions and computing associated uncertainty in a probabilistic approach, International Journal of Geographical Information Science 22 (10) (2008) 1067–1090.
[9] B. Bennett, P. Agarwal, Semantic categories underlying the meaning of 'Place', in: S. Winter, M. Duckham, L. Kulik, B. Kuipers (Eds.), Proceedings of the COSIT 2007, Lecture Notes in Computer Science, vol. 4736, Springer, Berlin, 2007, pp. 78–95.
[10] A.U. Frank, Formal models for cognition—taxonomy of spatial location description and frames of reference, in: Spatial Cognition: An Interdisciplinary Approach to Representing and Processing Spatial Knowledge, Lecture Notes in Artificial Intelligence, vol. 1404, Springer, Berlin, 1998, pp. 293–312.
[11] Y. Liu, Q. Guo, J. Wieczorek, M.F. Goodchild, Positioning localities based on spatial assertions, International Journal of Geographical Information Science 23 (11) (2009) 1471–1501.
[12] Y. Gong, G. Li, Y. Liu, J. Yang, Positioning localities from spatial assertions based on Voronoi neighboring, Science China: Technological Sciences 53 (S1) (2010) 143–149.
[13] Y. Gong, Y. Liu, G. Li, J. Yang, Structural hierarchy of spatial knowledge based on landmarks and its application in locality descriptions, in: Y. Liu, A. Chen (Eds.), Proceedings of the Geoinformatics'2010, Beijing, China, 2010, pp. 1–5.
[14] D.R. Montello, Spatial cognition, in: N.J. Smelser, P.B. Baltes (Eds.), International Encyclopedia of the Social & Behavioral Sciences, Pergamon Press, Oxford, 2001, pp. 14771–14775.
[15] S. Mohammed, R. Dehak, I. Bloch, H.A. Maitre, Spatial reasoning with incomplete information on relative positioning, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (9) (2005) 1472–1484.
[16] Y. Liu, Y. Tian, J. Weng, Probabilistic composition of cone-based cardinal direction relations, Science in China Series E: Technological Sciences 51 (S1) (2008) 81–90.
[17] S. Vasudevan, V. Nguyen, R. Siegwart, Towards a cognitive probabilistic representation of space for mobile robots, in: Proceedings of the 2006 IEEE International Conference on Information Acquisition, Weihai, Shandong, China, 2006, pp. 353–359.
[18] M. Duckham, M. Worboys, Computational structure in three-valued nearness relations, in: D.R. Montello (Ed.), Proceedings of the COSIT 2001, Lecture Notes in Computer Science, vol. 1661, Springer, Berlin, 2001, pp. 76–91.
[19] R.W. Kulhavy, W.A. Stock, How cognitive maps are learned and remembered, Annals of the Association of American Geographers 86 (1) (1996) 123–145.
[20] X.M. Wang, Y. Liu, Z.J. Gao, L. Wu, Landmark-based qualitative reference system, in: Proceedings of the IGARSS 2005, Seoul, Korea, 2005, pp. 932–935.
[21] C.M. Gold, Spatial adjacency—a general approach, in: Proceedings of the Auto-Carto 9, Baltimore, MD, USA, 1989, pp. 298–312.
[22] C.M. Gold, Neighbours, adjacency and theft—the Voronoi process for spatial analysis, in: Proceedings of the First European Conference on Geographic Information Systems, Amsterdam, The Netherlands, 1990, pp. 382–391.
[23] C.M. Gold, The meaning of 'neighbour', in: Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, Lecture Notes in Computing Science, vol. 639, Springer, Berlin, 1992, pp. 220–235.
[24] C.M. Gold, Surface interpolation as a Voronoi spatial adjacency problem, in: Proceedings of the Canadian Conference on GIS, Ottawa, Canada, 1992, pp. 419–431.
[25] F. Aurenhammer, Voronoi diagrams—a survey of a fundamental geometric data structure, ACM Computing Surveys 23 (3) (1991) 345–406.
[26] T. Grosche, F. Rothlauf, A. Heinzl, Gravity models for airline passenger volume estimation, Journal of Air Transport Management 13 (4) (2007) 175–183.
[27] J.J. Lewera, H. Van den Bergb, A gravity model of immigration, Economics Letters 99 (1) (2008) 164–167.
[28] G. Bruno, G. Improta, Using gravitymodels for the evaluation of new university site locations: a case study, Computers & Operations Research 35 (2) (2008) 436–444.
[29] M.A. Mostafavi, C.M. Gold, M. Dakowicz, Delete and insert operations in Voronoi/Delaunay methods and applications, Computers & Geosciences 29 (4) (2003) 523–530.
[30] G. Edwards, The Voronoi model and cultural space: applications to the social sciences and humanities, in: Proceedings of the COSIT 1993, Lecture Notes in Computer Science, vol. 716, Springer, Berlin, 1993, pp. 202–214.
[31] D.M. Mark, A.U. Frank, Experiential and formal models of geographic space, Environment and Planning B 23 (1996) 3–24.
[32] D.J. Peuquet, C.X. Zhan, An algorithm to determine the directional relationship between arbitrarily-shaped polygons in the plane, Pattern Recognition 20 (1) (1987) 65–74.
[33] A.U. Frank, Qualitative spatial reasoning about cardinal directions, in: Proceedings of the Seventh Austrian Conference on Artificial Intelligence, 1991, pp. 157–167.
[34] C. Freksa, Using orientation information for qualitative spatial reasoning, in: Proceedings of the International Conference GIS—From Space to Territory: Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, Springer, Berlin, 1992, pp. 162–178.