Membership Functions for Spatial Proximity

Jane Brennan[1] and Eric Martin[2]

[1] University of Technology, Sydney, Faculty of Information Technology, PO Box 123, Broadway NSW 2007, Australia; e-mail: janeb@it.uts.edu.au
[2] University of New South Wales, Kensington NSW 2002, Australia

Abstract. Formalising nearness has been the subject of extensive work, resulting in many membership functions based on absolute distance metrics, relative distance metrics, and combinations of those. The possible strengths and weaknesses of these functions have been discussed and argued at length, but strangely enough, no experiment seems to have been conducted to assess the merits and shortcomings of competing approaches. Conducting such experiments can be expected not only to provide an objective evaluation of the various measures that have been proposed, but also to suggest new measures that outperform all those being analysed. This paper fulfills these expectations, and gives further evidence that fuzzy logic provides fruitful and powerful methods to formalise qualitative reasoning and capture fundamental qualitative notions. The proposed fuzzy membership functions can be directly used in qualitative reasoning about spatial proximity in Geographic Information Systems, which are becoming more and more important in software development for diverse purposes such as Tourist Information Systems or property development.

1 Introduction

Love it or loathe it, fuzzy logic is, albeit its unfortunate choice of name, a very practical Artificial Intelligence technique that is used in applications such as control systems in washing machines, elevators, cars, etc. It is quite a valuable method that can generate precise solutions from approximate data, and can be of great use for developing computational models for vague concepts such as those described by natural language expressions. In this paper, we propose and evaluate fuzzy membership functions suitable to implement the notion of spatial proximity, generally represented by linguistic terms such as *near* or *close*, within Geographic Information Systems (GIS).

After a brief review of previous work, this paper will report on studies conducted to evaluate proximity membership functions against a data set of a distance network and predefined nearness information. We will conclude with recommendations for appropriate functions in the context of reachability across a distance network.

2 Factors that influence Spatial Proximity Perception

Several fuzzy approaches have been devised to represent spatial proximity based on distance. In order to determine what exactly influences the perception of spatial nearness within GIS data, Gahegan [3] conducted psycho-metric experiments. The objective of his experiments was to examine when or how people decide whether objects are near a chosen reference object on a GIS map. The pseudo-metric tests were conducted on a group of 50 subjects, who have all had some practical exposure to Geographic Information Systems. The subjects were asked to rate objects in diagrams representing geographic features on a map, according to how close they were to a given reference object. While Gahegan [3] pointed out that his tests were not necessarily conclusive, he could obtain some interesting results and make several observations that could be helpful in modelling spatial proximity. A first observation is that if a scene is devoid of additional objects, namely if only the reference object and the object to be located inhabit the scene, geometrical reasoning is applied. However, in the presence of other objects of the same type, proximity is partially defined by relative distance. Another observation is that proximity perception is impacted by the scale of the scene, which directly depends on the size of the area being considered. This paper is set out to define membership functions that address all of these points.

On the basis of his observations, Gahegan [3] suggested a contextual model of nearness relations in order to account for different influencing factors. Three kinds of metrics are considered in Gahegan's model: an absolute distance metric, a relative distance metric, and a combination of both. Using the absolute distance metric amounts to assuming that proximity is directly proportional to the Euclidian distance between the reference object and the object to be localised. As the scale of the area viewed by the user of a GIS map also seems to have an impact on the perception of proximity, Gahegan [3] further suggested that the bounding boxes of the area in the GIS can be used as scale indicator, the opposite corners of the bounding boxes providing the maximum distance. If the maximum possible distance between objects in the scene is used to normalise the distance between two objects A and B, it is then possible to represent the proximity between A and B by a fuzzy value. Similarly, but not exclusively, we will use the maximal distance between objects on the map of a country or state. This will enable proximity evaluations across several GIS maps if needed. In this paper, the maximal distance will be the largest distance between any two places in the country or state being considered. Absolute distance metrics result in continuous proximity with, for example, *very close* $>$ *close* $>$ *far*, but relations such as *closest* or *farthest* cannot be represented. For this and other reasons, Gahegan [3] proposed to treat proximity in terms of a relative distant metrics, in addition to the absolute distance metrics. More precisely, he suggested an ordinal approach to represent relative distance, by ranking the objects in the scene with respect to their distance to the reference object and the total number of objects. He pointed out that this approach could cause objects to be considered close to a given reference object A, even though such objects might be separated from A by a large distance. As the objects are ranked on the basis of their distance to A, this approach seems to be more absolute than relative in nature. Worboys [6] dealt with this problem in a more efficient way by calculating for each place the mean distance to all other places in the scene.

Gahegan [3] suggested to combine both the absolute and the relative distance. As both metrics offer fuzzy representations, he defined the membership function for absolute distance metrics and assumed a distribution function for nearness based on relative distance, and then combined these functions with the fuzzy union operator. This resulted in an object being considered *close* just in case it is geometrically OR relatively close.

Motivated by Gahegan's work, Guesgen and Albrecht [4] suggested to associate spatial binary relations such as *far from* or *close to*, or unary relations such as *downtown*, with fuzzy membership grades that could be calculated from the Euclidian distance between objects on a map. They did not test their suggestions against any data and did not provide any membership function for relative distance metrics. Guesgen [5] proposed to define proximity without any measure of distance by using the notion of fuzzy sets previously defined in Guesgen and Albrecht [4]. These fuzzy sets were used to reason about the relationship between proximity notions by means of transitive closure on ternary proximity relations such as "if B is closer to A than C is to A, and if C is closer to A than D is to A, then B is closer to A than D is to A." This is very similar to van Benthem's [1] approach in his logic of space.

There is no experimental data to give evidence that Gahegan's [3] or Guesgen and Albrecht's [4] fuzyy membership functions can be of practical use. None of their approaches bases fuzzy membership functions on truly relative distance. We therefore find it essential to evaluate Gahegan's [3] absolute distance metrics and Worboys' [6] relative distance metrics before considering whether and how to combine them using fuzzy logic operators.

Worboys [6] did some interesting studies on the qualitative location of cities and the relative distances between them. His definition of relative distance is not based on the comparative concept without Euclidean distance, but it does nonetheless incorporate the context of all places under consideration. He used the road distances between 48 cities in Great Britain, which he called objective distances, and determined their relative distances to each other by first calculating for each centre the mean of the distances to all remaining centres. The relative distance between a centre A and a centre B was then determined by dividing the objective distance between A and B by A's mean. This notion of relative distance is actually asymmetric: this method will most likely produce a different relative distance between A and B than between B and A. The relative distance can then be used to calculate fuzzy nearness values using the following definition: $nearness(x, y) = (relative\_distance(x, y) + 1)^{-1}$.

Places having high nearness values are therefore very close and low ones are not close. The greatest nearness is between a place and itself, with a value of 1. This approach does not suffer from the same restriction as Gahegan's approach. The objects do not need to be fairly evenly distributed. In his more recent work on environmental space, Worboys [7] used the number of subjects and the number of *yes* or *no* votes to calculate fuzzy membership values for nearness. This is a very interesting approach given his experimental data. However it is not practically applicable to do such a kind of data collection for every geographic area that GIS-users might need to work with.

A serious shortcoming of all the approaches that have been described is that none of them does actually evaluate the membership functions against any real data, in order to see how useful these functions are. As has been discussed in this section, we have been conducting experiments with several membership functions and evaluated them against proximity data. The following section introduces the functions we used.

3 Various Distance metrics

As previously mentioned, in order to address all of the observations that Gahegan [3] made in the context of GIS users perceiving proximity, we will evaluate several spatial proximity functions based on absolute distance, relative distance, and combinations of both. Table 1 shows the fuzzy membership functions we evaluated in terms of their usefulness within GIS settings.

Absolute Distance Metrics:	$\mu_{abs}(A,B) = 1 - \frac{Dist(A,B)}{Max}$
Relative Distance Metrics:	$\mu_{rel}(A,B) = \frac{1}{(reldis(A,B)+1)}$
Fuzzy Union:	$\mu_{comb\_u}(A,B) = \mathrm{MAX}(\mu_{abs}(A,B), \mu_{rel}(A,B))$
Fuzzy Intersection:	$\mu_{comb\_i}(A,B) = \mathrm{MIN}(\mu_{abs}(A,B), \mu_{rel}(A,B))$

Table 1. Fuzzy Membership Functions

The fuzzy membership function based on absolute distance metrics is a derivation of Gahegan's [3] function with the maximum value Max being the maximum distance between all of the places in our data set; and $Dist$ being the distance between places A and B. For the fuzzy membership function based on relative distance metrics, we borrowed Worboy's [6] membership function, as we found that Gahegan's ordinal ranking approach is insufficient. Relative distance is calculated using the mean of each place A in the data set, calculated from the n places OP_i, $1 \leq i \leq n$, distinct from A and available in the set: $mean(A) = \frac{1}{n}\Sigma_{i=1}^{n} Dist(A, OP_i)$. The result of this is then used to calculate the relative distance between each two places: $reldis(A,B) = Dist(A,B) * mean(A)^{-1}$. While Gahegan [3] only suggested to combine the membership functions based on absolute and relative distance by applying the fuzzy union, we also investigated the application of the fuzzy intersection operator, which yielded interesting results. The fuzzy union operator will by definition always return the maximum membership function value for each data entry. While the fuzzy intersection operator will by definition always return the minimum membership function value for each data entry. We applied these functions to the data set described in the following section.

4 Data Set and Experiments

We encoded 34 places in the Australian state of New South Wales and the distances between them[1]. For each of the given places, we define the tourist region, the region and the regional area they are located in. For Sydney, a list of regions that are easily accessible for short trips is also supplied, thereby giving some indication of what is perceived and generally accepted as near to Sydney. This data set was collected from the *Tourism New South Wales* site[2]. We will be able to use this information to evaluate our membership functions to see how well they suit the data and "nearness" information for the given places.

[1] as given by "The Official Road Directory of New South Wales" by the Land Information Centre in Bathurst, The New South Wales Government

[2] www.visitnsw.com.au

5 Results and Evaluation

The membership function based on absolute distance as illustrated in the left graph in Figure 1 shows a linear distribution, as expected from the function used. The maximum distance in the dataset is 1710 km. However, the relative membership function as illustrated in the right graph in Figure 1 shows quite a varied distribution, which is very different to Gahegan's more or less linear proposal of ordinal ranking. The issue arising from his kind of approach, that objects could possibly still be considered close to one another even when they are separated by a very large distance, is not a problem for the membership function we used, because ours is a function of the distance between the two places being considered.

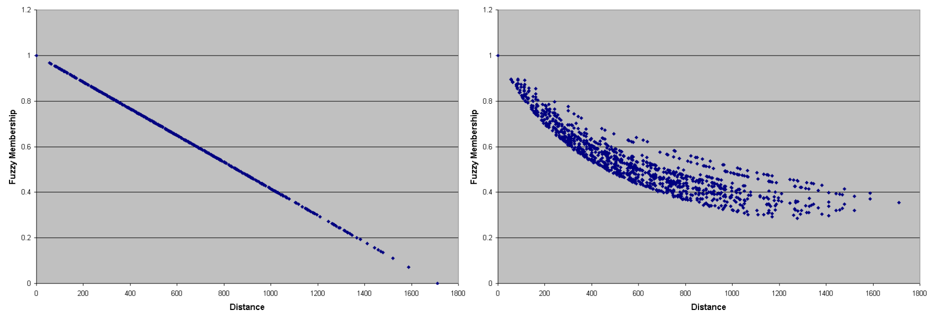

Fig. 1. Fuzzy Distribution Functions for Absolute and Relative Distance Metrics

The two combined membership functions give quite interesting results. On one hand they do support Gahegan's [3] suggestion that absolute measures are more appropriate for non-clustered objects, and relative measures for objects within clusters of same "kinded" objects. On the other hand, the results do contradict Gahegan's suggestion to use the union fuzzy operator for an efficient combined function. Because, when combining absolute and relative distance metrics functions by union we do get a linear distribution for distances until about 800km, where it changes into a more clustered distribution (see left graph in Figure 3).

This is even contradicting Gahegan's own terms that absolute distance metrics i.e., linear distributions in his case, are more suited for proximity assignments between objects that are located in virtually devoid areas. Figure 2 shows that the greater the distances between the places, the fewer places are within the area; which can be explained by the fairly isolated character of the Australian non-metropolitan areas. In order to comply with Gahegan's suggestion to use absolute distance metrics for only lightly and relative distance metrics for heavily populated areas, the membership function distribution should be the reverse of the result we obtained for the combined function using the fuzzy union operator.

When we applied the fuzzy intersection operator to our data set, we attained precisely such a distribution. The fuzzy intersection changes from a clustered to a linear distribution between 1000 and 1200 km. The right graph in Figure 3 shows this clearly. This is even contradicting Gahegan's own terms that absolute distance metrics i.e., linear distributions in his case, are more suited for proximity assignments between objects that are located in virtually devoid areas. Figure 2 shows that the greater the distances

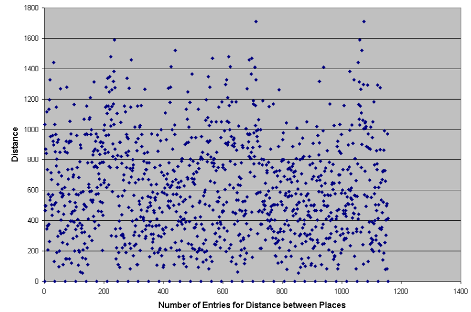

Fig. 2. Distance Distribution in the Data Set

between the places, the fewer places are within the area; which can be explained by the fairly isolated character of the Australian non-metropolitan areas. In order to comply with Gahegan's suggestion to use absolute distance metrics for only lightly and relative distance metrics for heavily populated areas, the membership function distribution should be the reverse of the result we obtained for the combined function using the fuzzy union operator.

When we applied the fuzzy intersection operator to our data set, we attained precisely such a distribution. The fuzzy intersection changes from a clustered to a linear distribution between 1000 and 1200 km. The right graph in Figure 3 shows this clearly. The clustered distribution is more appropriate for smaller distances, as there are more

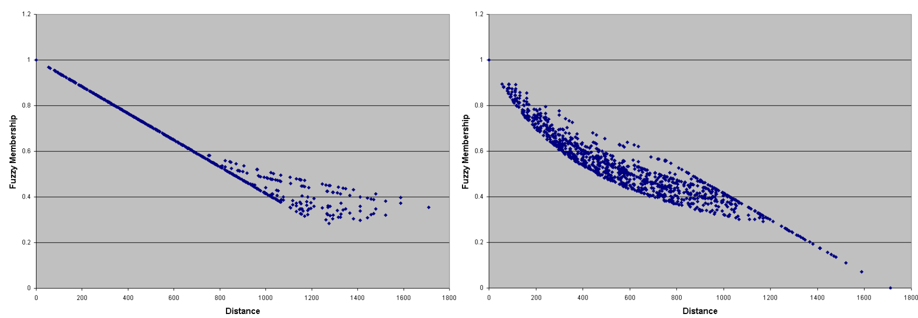

Fig. 3. Combined Fuzzy Distribution Functions using Union and Intersection Operator

places within a smaller area, and the linear distribution will suit areas with fewer objects, which are to be found at greater distances in the given data set. This is perfectly consistent with Gahegan's [3] observations, although it is a different combination operator that gives the desired result.

We evaluated our membership distribution function values against the proximity information, namely *Sydney Surrounds* and regions, to appraise the usefulness of the functions and their combinations in the context of reachability within a road network.

For all the places that are within regions which are generally accepted to be in the Sydney surrounding area, all membership function values were not only well above the usual crossover point of 0.5, but also above 0.7. We tested the distances between all places in our dataset that are in the same region using this value as the crossover point.

As geographic regions are generally defined with the perception of “everything” within a region being close to “everything” within this region. 16 out of 94 matches had two membership values that were below 0.7. These membership values were always the result of the membership function based on relative distance metrics and the combined function that returned the former one. When the threshold was lowered to the normal crossover point (0.5), all matches complied. This is a good indication that the investigated membership functions are useful in the context of a road distance network and the associated reachability of the places within it.

6 Conclusions and Future Work

In this paper, we have shown that while the membership functions based on absolute distance metrics and relative distance metrics, as proposed by Gahegan [3] and Worboys [6] respectively, do evaluate well against distance data, the combined membership function that proves useful in this context is the fuzzy intersection of the two former, and not their union as Gahegan [3] suggested. We will implement these proposed membership functions as part of our already existing qualitative reasoning extension to the Geographic Information System *AccuGlobe*, to further assess the benefits of fuzzy logic as a descriptor for spatial proximity notions.

We also plan to extend our data by including more information about regions and places within the regions that surround major places in the data set. This will allow to analyse further membership functions and to get more conclusive results for asymmetric relative distance relations. Moreover, we will use the expanded data set to automatically learn appropriate membership functions and then evaluate them against a data collection of another geographic area, such as another Australian state or a state within another country of a similar spatial distribution of places.

References

1. van Benthem, J.F.: The logic of time – a model theoretic investigation into the varieties of temporal ontology and temporal discourse. Kluwer Academic Publishing, Dordrecht, 1991.
2. Gahegan, M.: Support for the Contextual Interpretation of Data within an Object-Oriented GIS. Proceedings of Spatial Data Handling ’94, T.C. Waugh and R.G. Healey (eds), Edinburgh, Scotland, pp. 988-1001, 1994.
3. Gahegan, M.: Proximity Operators for Qualitative Spatial Reasoning. Spatial Information Theory - A Theoretical Basis for GIS (COSIT’95). A. U. Frank and W. Kuhn (eds), Springer-Verlag, Berlin, Heidelberg, pp. 31-44, 1995.
4. Guesgen, H.-W. and Albrecht, J.: Imprecise reasoning in geographic information systems, In *Fuzzy Sets and Systems*, vol. 113(1), pp. 121-131, 2002.
5. Guesgen, H.-W.: Reasoning About Distance Based on Fuzzy Sets. In *Applied Intelligence*, vol. 17, pp. 265-270, 2002.
6. Worboys, M.F.: Metrics and topologies for geographic space. Advances in GIS Research II. Taylor and Francis, pp. 365-376, 1996.
7. Worboys M.F.: Nearness relations in environmental space. In *International Journal of Geographical Information Science*, vol. 12(7), pp. 633-651, 2001.