Fuzzy regions: adding subregions and the impact on surface and distance calculation

Jörg Verstraete

Instytut Badań Systemowych, Polska Akademia Nauk (Systems Research Institute, Polish Academy of Sciences); Ul. Newelska 6, 01-447 Warszawa, Poland
DDCM, Dept. Telecommunications and Information Processing, Ghent University; Sint Pietersnieuwstraat 41, 9000 Ghent, Belgium
`jorg.verstraete@{ibspan.waw.pl,telin.ugent.be}`

`http://www.ibspan.waw.pl,http://telin.ugent.be/ddcm`

Abstract. In the concept of fuzzy regions we introduced before, a region was considered to be a fuzzy set of points, each having its own membership grade. While this allows the modelling of regions in which points only partly belong to the region, it has the downside that all the points are considered independently, which is too loose a restriction for some situations. The model is not able to support the fact that some points may be linked together. In this contribution, we propose an extension to the model, so that points can be made related to one another. It will permit the user to, for instance, specify points or even (sub)regions within the fuzzy region that are linked together: they all belong to the region to the same extent at the same time. By letting the user specify such subregions, the accuracy of the model can be increased: the model can match the real situation better; while at the same time decreasing the fuzziness: if points are known to be related, there is no need to consider them independently. As an example, the use of such a fuzzy region to represent a lake with a variable water level can be considered: as the water level rises, a set of points will become flooded; it is interesting to represent this set of points as a subset of the region, as these points are somewhat related (the same can be done for different water levels). The impact of this extension to the model on both surface area calculation an distance measurement are considered, and new appropriate definitions are introduced.

1 Introduction

The concept of the fuzzy regions was sparked by a lack of models capable of handling imperfect information, built on a solid, theoretical foundation, applicable for a wide range of representations and open to an implementation. In the presented concept, a fuzzy region was defined as a fuzzy set of points over the two dimensional domain. Each point therefore is given a membership grade indicating the extent to which it belongs to the region (i.e. a veristic interpretation

[1]). While this broad definition is useful for many basic applications and allows for unlimited distributions of membership grades over the region, it treats all the elements independently of one another. In some situations however, it is possible a user has more information about a relation between elements of the region. Consider the example of a lake, where there is a variation possible in the level of water. At a random point in time we may not know the water level, but we do know that if a particular point would be under the water, so would other points (i.e. those at the same altitude just around the lake). If the likeliness of that water level is 0.4, then it makes sense to assign all these points that membership grade; however, as these points would be be at the same altitude, it is more natural not to consider them independently but rather as a group the points.

In this paper, a brief overview of the current model for fuzzy regions is given in section 2. Both the concept and the operations that are will be extended are considered here. Section 3 then deals with the proposed extension to fuzzy regions and the impact on the considered operations. The results are summarized in the conclusion.

2   Current fuzzy regions

2.1   Concept

A crisp region is often defined by means of its outline: a polygon or other closed line is defined as the boundary of the region, and the region is said to consist of the points inside this outline ([3]). To define fuzzy regions, a different point of view was adopted: rather than consider a crisp region to be defined by means of an outline, it was considered to be be a set of points. From this point of view, it is a small step to augment the definition to a fuzzy set ([8], [9]) of points. In [5], the fuzzy region was defined over $\mathbb{R}^2$, thus with each element (point) a membership grade was associated.

Definition 1 (Fuzzy region).

$$\tilde{R} = \{(p, \mu_{\tilde{R}}(p)) | p \in \mathbb{R}^2\} \tag{1}$$

A fuzzy region essentially is a fuzzy set defined over a two dimensional domain. As a result, the traditional fuzzy operations for intersection and union (by means of t-norms and t-conorms) are immiately applicable. Functionaly has been added to deal with spatial aspects, some examples include the distance between regions and the (fuzzy) surface area of a region. To determine topological relations however, appropriate definitions for the boundary, interior and exterior had to be derived from the initial given fuzzy set.

2.2   Operations

Surface area: cardinality For the surface area of fuzzy regions, two definitions have been introduced ([6]). The first is an application of fuzzy cardinality,

where each point is said to contribute its own membership grade.

$$\tilde{S}^c(\tilde{A}) = \int_{p(x,y)\in U} \mu_{\tilde{A}}(p(x,y))d(x,y) \tag{2}$$

Where U is the considered universe (usually $\mathbb{R}^2$).

This definition has the advantage of yielding a crisp number that takes the fuzziness into account. The crisp number can easily be used in systems that cannot deal with fuzzy data.

Surface area: fuzzy number The second definition for surface area yields a fuzzy number, such that all possible surface areas are possible values. This approach has the advantage of yielding a result that holds more information than the cardinality approach above, but it requires working with fuzzy numbers. For each possible surface area, the membership grade matches the possibility of the region being a region of this size. This is achieved by considering the surface area of all α-levels.

$$\tilde{S}^f(\tilde{A}) = \{(x, \mu_{\tilde{S}^f(\tilde{A})}(x)), x \in U\} \tag{3}$$

where U is the considered universe, most likely $\mathbb{R}^2$. For fuzzy regions defined as fuzzy sets over $\mathbb{R}^2$, the membership function was defined as:

$$\mu_{\tilde{S}^f(\tilde{A})}(x) : \mathbb{R} \to [0,1]$$

$$x \mapsto \begin{cases} 1 & \text{if } x = S(\tilde{A}_1) \\ \sup\{\alpha \mid S(\tilde{A}_{\overline{\alpha}}) \le x \le S(\tilde{A}_\alpha)\} & \text{if } S(\tilde{A}_{\overline{\alpha}}) \le x \le S(\tilde{A}_\alpha) \\ 0 & \text{elsewhere} \end{cases}$$

Distance There are various options to define the distance between crisp regions; the most common one is the euclidean distance which represents the shortest distance between both regions. For fuzzy regions, the fuzzy distance $\tilde{d}(\tilde{R}_1, \tilde{R}_2)$ is a fuzzy number representing the possible distances between the α-cuts of R_1 and R_2.

$$\tilde{d}(\tilde{R}_1, \tilde{R}_2) = \{(x, \mu_{\tilde{d}(\tilde{R}_1,\tilde{R}_2)}(x)) \mid x \in \mathbb{R}\} \tag{4}$$

where

$$\mu_{\tilde{d}(\tilde{R}_1,\tilde{R}_2)} : \mathbb{R} \to [0,1]$$

$$x \mapsto \sup\{\alpha \mid d(\tilde{R}_{1\alpha}, \tilde{R}_{2\alpha}) \le x \le d(\tilde{R}_{1\overline{\alpha}}, \tilde{R}_{2\overline{\alpha}})\}$$

3   Extended fuzzy regions

3.1   Concept

The main problem with the above model is that all points are considered individually. In many applications, it would be beneficial to group elements of the

fuzzy set together. In order to do this, the fuzzy set defining the region will be defined over the domain $\wp(\mathbb{R}^2)$; i.e. the powerset of $\mathbb{R}^2$. The powerset of a set is a new set containing all the possible subsets of that particular set. To illustrate this, consider the following example:

$$\wp(\{0, 1, 2\}) = \{\{\}, \{0\}, \{1\}, \{2\}, \{0, 1\}, \{0, 2\}, \{1, 2\}, \{0, 1, 2\}\}$$

For $\mathbb{R}^2$ this becomes:

$$\wp(\mathbb{R}^2) = \left\{ X | X \subseteq \mathbb{R}^2 \right\}$$

At present, we are only considering the powerset over $\mathbb{R}^2$ as a means to further model the distribution of membership grades in fuzzy regions. The use of the fuzzy powerset over $\mathbb{R}^2$ (i.e. $\tilde{\wp}(\mathbb{R}^2)$, the set of all fuzzy sets over $\mathbb{R}^2$) is also under consideration, as it may allow the model to cope with more complex distributions of membership grades, but at an increased complexity. Definition 2 shows the new definitions for fuzzy regions.

Definition 2 (Fuzzy region).

$$\tilde{R} = \{(P, \mu_{\tilde{R}}(P)) | P \in \wp(\mathbb{R}^2) \wedge \forall P_1, P_2 \in \tilde{R} : P_1 \cap P_2 = \emptyset\} \tag{5}$$

Note that it is required that no two elements of the fuzzy region share points: the intersection between any two elements should be empty. A point can only be considered to belong to the region once, even if it is to a membership grade less than 1. An extension of this is under investigation, but is not elaborated upon for now. With each element of the fuzzy set, a membership grade is associated; it still has a veristic interpretation to indicate the degree to which this element belongs to the fuzzy region. Before considering the impact of this new definition on operations, it is interesting to verify that the model is a generalization of the previous model, and that it still is capable of representing fuzzy regions as defined before. A fuzzy region $\tilde{R}_1$, defined as:

$$\tilde{R}_1 = \left\{ (p, \mu_{\tilde{R}_1}(p)) | p \in \mathbb{R}^2 \right\}$$

can be defined using this new definition as a fuzzy region $\tilde{R}$, by using an appropriate membership function:

$$\tilde{R}_2 = \{(P, \mu_{\tilde{R}_2}(P)) | P \in \wp(\mathbb{R}^2)\}$$

where

$$\mu_{\tilde{R}_2}(P) = \begin{cases} \mu_{\tilde{R}_1}(p) & \text{if } P = \{p\} \\ 0 & \text{elsewhere} \end{cases}$$

The requirement that two elements need to be disjoint is automatically fulfilled by the given membership function.

3.2 Operators

The change in definition will of course impact the operators that have been defined. In [7], we presented two definitions for the surface area of fuzzy regions: the first is an extension of fuzzy cardinality (each point basically contributes its membership grade to the total area), which yielded a crisp number; whereas the second results in a fuzzy number representing the possible surface areas for the region. Both definitions (2.2) will now be adapted to suit the new definition. For the distance between fuzzy regions, two approaches have been considered: one using the α levels, and one based on topological aspects. In this contribution, only the first approach is considered, as topological aspects are currently under development.

Surface area: cardinality While the meaning of the original definition (each element contributing its membership grade) remains the same, the present formal definition as shown in (2) needs to be changed slightly to deal with the fact that the basic elements of the fuzzy region are now sets rather than points. Each element will once again contribute to the extent of its membership grade, but the elements themselves are now possibly larger than a single point. As a result, the surface areas of each basic element also needs to be taken into consideration:

Definition 3 (surface area (cardinality)).

$$\tilde{S}^c(\tilde{A}) = \int_{X \in \wp(U)} \mu_{\tilde{A}}(X) \int_{p(x,y) \in X} d(x,y) dX \tag{6}$$

This definition yields the same result as the previous definition if all basic elements are singleton sets.

Surface area: fuzzy number The surface area as a fuzzy number is impacted in a far greater way: a given set that is an element of the fuzzy region will contribute its area, or won't contribute at all, with nothing in between. Consequently, there is a smaller number of possible surface areas; as illustrated on figure 1. Figure 1a shows a fuzzy region in the classical definition and its surface area: each point may or may not belong to it, so it may or may not play a part in determining the surface area. The surface area is also shown, and every area between 0 and $2x$ is a possibility. Figure 1b shows a similar fuzzy region, but now the regions consists out of 2 sets: one set which is given membership grade 1, and one given membership grade 0.5. The expected surface area is also shown: both regions either contribute as a whole, or don't contribute at all. As a result, the only possible surface areas are x and $2x$. Figure 1c finally combines both aspects: the points with membership grade 0.5 each are counted individually, whereas the set with membership grade 1 is counted as one whole; this is also reflected in the surface area.

Note that currently, a sub-region is not considered to have 0 as a possible surface area, resulting in the fact that the total surface area for the region cannot

be 0 (e.g. for the region on Figure 1b). This choice is open for discussion when the regions represent imprecision: some argue that there should always be one valid value, which implies that a region must have some elements and thus that 0 is not a possible surface area; whereas others can argue that an empty region is still a region. However if the regions are used to represent uncertainty, each sub-region must also include 0 as a possible surface area. This discussion only has a minor impact on the definition though.
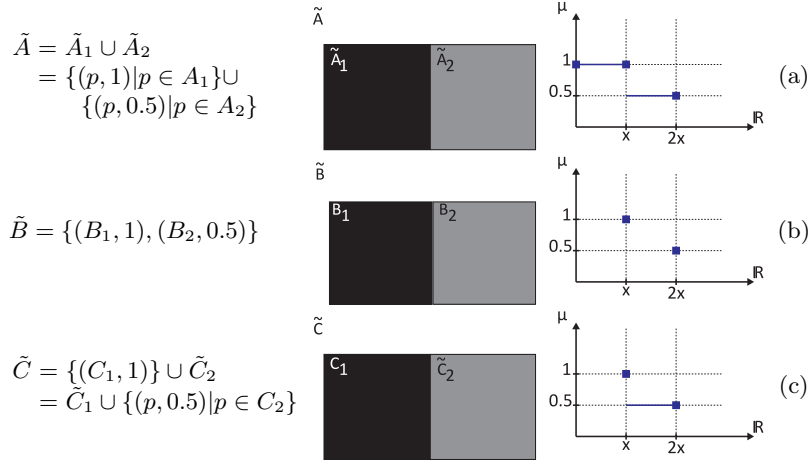
$$\tilde{A} = \tilde{A}_1 \cup \tilde{A}_2$$
$$= \{(p,1)|p \in A_1\} \cup$$
$$\{(p,0.5)|p \in A_2\}$$

$$\tilde{B} = \{(B_1,1),(B_2,0.5)\}$$

$$\tilde{C} = \{(C_1,1)\} \cup \tilde{C}_2$$
$$= \tilde{C}_1 \cup \{(p,0.5)|p \in C_2\}$$

Fig. 1. Three different examples of fuzzy regions, with their surface areas: (a) the classical model, where each points is treated independently, (b) the new model, showing a region consisting of two subregions each counted as a whole, (c) a region where there is both a subregion where points are treated independently, and a subregion that is counted as a whole. For each region, a mathematical explanation of its elements, a graphical illustration, where the shade of grey is representative of the membership grade, and its surface area are shown.

In the new situation, where fuzzy regions are defined as fuzzy sets over $\wp(\mathbb{R}^2)$, the main problem with the original definition (definition 3) is that it makes all surface areas between every strong and the weak α level possible, which - as illustrated on fig 1b - may not be the case. As before, the surface area of the region should be a (fuzzy) number representing the possible surface areas and their membership grade. Furthermore each subregion is considered either as a whole, or not at all, so its only possible contribution is its surface area; the amount it contributes to the fuzzy number is the same is a the degree to which it belongs to the region.

Consider the example in fig 2a; the surface area is the sum of all the surface areas that make up the region. As each of these sub-regions are possible, their surface areas are considered to be (non-normalized) fuzzy numbers, with just

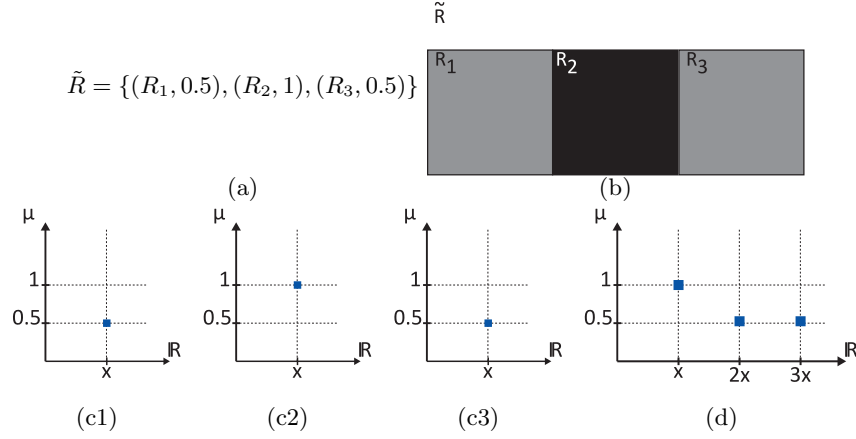one possible value which has a membership grade equal to the possibility of the sub-region in R (fig 2c).

Fig. 2. Example to illustrate the definition: (a) definition of the region, (b) graphical representation, (ci) area of R_i, (d) total surface area of R.

$$\tilde{S}^f(\tilde{R}) = (S(R_1), \mu_{\tilde{R}}(R_1)) \tilde{+} (S(R_2), \mu_{\tilde{R}}(R_2)) \tilde{+} (S(R_3), \mu_{\tilde{R}}(R_3))$$

To add up all these regions, fuzzy arithmetic is used, which implies considering the different possible combinations:

- a surface area of x is only possible by considering either region 1, or region 2 or region 3; the membership grade for x therefore is: $\max(0.5, 1, 0.5) = 1$
- a surface area of $2x$ is possible by considering either region 1 and 2, 1 and 3 or 2 and 3; which yields: $\max(\min(0.5, 1), \min(0.5, 0.5), \min(1, 0.5)) = 0.5$
- a surface area of $3x$ is only possible by considering all three regions; yielding for $3x$ the membership grade: $\max(\min(0.5, 1, 0.5)) = 0.5$

The end result is
$$\tilde{S}^f(\tilde{R}) = \{(x, 1), (2x, 0.5), (3x, 0.5\}$$

as also illustrated on fig 2d. In general, for any number of closed sets, the surface area can be defined as:

$$\tilde{S}^f(\tilde{R}) = \tilde{\sum}_{X | \mu_{\tilde{R}}(X) > 0} \big(S(X), \mu_{\tilde{R}}(X) \big) \tag{7}$$

Generalizing results in the definition:

Definition 4 (surface area (fuzzy number)).

$$\tilde{S}^f(\tilde{R}) = \tilde{\sum}_{X | \mu_{\tilde{R}}(X) > 0} \left(\int_{p(x,y) \in X} d(x, y), \mu_{\tilde{R}}(X) \right) \tag{8}$$

In [2], a fuzzy number X is defined as a fuzzy set over $\mathbb{R}$ that satisfies the properties:

- $\tilde{X}$ is normalized, i.e. there is at least one element x for which $\mu_{\tilde{X}}(x) = 1$
- $\forall \alpha \in\,]0, 1] : \tilde{X}_{\overline{\alpha}}$ is a closed interval
- the support $\tilde{X}_0$ of $\tilde{X}$ must be bounded

For future computations, it is useful to have a fuzzy number as the result of a fuzzy surface area: calculations with different surface areas can be performed using fuzzy arithmetic. The new definition satisfies the first property if the region has at least one element with membership grade 1; and always satisfies the second and third properties if the fuzzy region is bound. While these assumptions are technically not required in the definition of fuzzy regions, they are reasonable assumptions which most likely will be made anyway.

Distance: α level approach For the distance between fuzzy regions, two approaches have been considered [7]: one based on α-levels and one based on topological aspects. In this contribution, only the first approach is considered, in which the distance to a fuzzy region yields a fuzzy number, that models all the possible distances between the corresponding α-levels, not dissimilar to the concept of the surface area as a fuzzy number. To define the distance in the new concept, consider two fuzzy regions $\tilde{A}$ and $\tilde{B}$.

The original definition (definition 4) poses a problem similar to the calculation of the surface area: fewer distances than before need to be taken into consideration, as elements can now be sets that contribute only as a whole. In a similar approach to the surface calculation, the distances between the different elements (subregions) are first considered, and then combined to form the overall distance. As the elements of the newly defined fuzzy regions are basically crisp sets in $\mathbb{R}^2$, there are different options for the distance between any two elements of both fuzzy regions; here the shortest possible Euclidean distance is chosen.

$$d(A, B) = \min_{p_A \in A, p_B \in B} \{d(p_A, p_B)\}$$

where the distance between points is given by the Euclidean distance:

$$d(p_A, p_B) = \sqrt{(x_{p_A} - x_{p_B})^2 + (y_{p_A} - y_{p_B})^2}$$

with $p_i = (x_{p_i}, y_{p_i})$

The likelihood of each distance is given by the intersection of the membership grades of both elements considered; we will use the minimum as the intersection operator. The membership grade associated with $d(A, B)$ therefore is given by: $\min(\mu_{\tilde{R}_1}(A), \mu_{\tilde{R}_2}(B))$. The fuzzy number representing the distances is then obtained as the (fuzzy) union of all possible distances between any element of $\tilde{R}_A$ and any element of $\tilde{R}_B$. This leads us to define the distance between fuzzy regions as follows:

Definition 5 (distance).

$$\tilde{\bigcup}_{A|\mu_{\tilde{R}_1}(A)>0,B|\mu_{\tilde{R}_2}(B)>0} \left\{ (d(A,B), \min(\mu_{\tilde{R}_1}(A), \mu_{\tilde{R}_2}(B))) \right\}$$

In the case where all element of both $\mu_{\tilde{R}_1}(A)$ and $\mu_{\tilde{R}_2}(B)$ are singleton sets, i.e. $A = \{p_A\}, B = \{p_B\}$, this definition yields the same result at the previous one:

$$\tilde{\bigcup}_{A|\mu_{\tilde{R}_1}(A)>0,B|\mu_{\tilde{R}_2}(B)>0} \left\{ (d(A,B), \min(\mu_{\tilde{R}_1}(A), \mu_{\tilde{R}_2}(B))) \right\}$$
$$\Leftrightarrow \mu_{\tilde{d}(\tilde{R}_1,\tilde{R}_2)}(x) = \sup\{\alpha \mid \alpha = \min(\mu_{\tilde{R}_1}(A), \mu_{\tilde{R}_2}(B)) \wedge d(A,B) = x\}$$
$$\Leftrightarrow \mu_{\tilde{d}(\tilde{R}_1,\tilde{R}_2)}(x) = \sup\{\alpha \mid \alpha = \min(\mu_{\tilde{R}_1}(\{p_A\}), \mu_{\tilde{R}_2}(\{p_B\})) \wedge d(p_A,p_B) = x\}$$
$$\Leftrightarrow \mu_{\tilde{d}(\tilde{R}_1,\tilde{R}_2)}(x) = \sup\{\alpha \mid d(\tilde{R}_{1\alpha}, \tilde{R}_{2\alpha}) \le x \le d(\tilde{R}_{1\overline{\alpha}}, \tilde{R}_{2\overline{\alpha}})\}$$

In the previous definition, the membership was defined on a per-element basis: for each distance, its membership was the highest value so that the distance between the weak α-cuts of both regions at this level is less or equal, and the distance of the strong alpha cuts is greater or equal to it. The membership grade for each distance in the new definition is defined as the greatest value (due to the use of max as union operator) of the smallest membership grades of points in both regions that are this distance apart. This smallest membership grade is equivalent to the α-level in the previous definition: at a higher α level, the distance between these two points will no longer be taken into consideration as one point will not be present in the α-cut. As a result, these two definitions are equivalent if all elements of the newly defined regions are singleton sets.

4   Conclusion

In this contribution, the concept of fuzzy regions was extended with the ability to model groups of points belonging to some extent to the fuzzy region. This enriches the theoretical model of fuzzy regions, but at the same time opens the door for improvements to the derived practical models (both the bitmap based an the triangular network model that have been developed before [4],[6]). The new model can be used to incorporate possible subregions, and by carefully chosing them also possible boundaries, that need to be considered as a whole. Along with the changes to the model, some operators (namely surface calculation and distance) have been considered. The adaptation of the model is such that regions from the previous model can easily be represented with it; as such it is a generalization of the previous model (which in turn was a generalization of the crisp model). The next step is to consider the change that will occur in topological relations, and to optimize three derived models (contour lines, bitmaps, triangular networks) to handle the change in the theoretical model in an efficient manner.

References

1. Dubois D., Prade H.: Fundamentals of Fuzzy Sets. Kluwer Academic Publishers (2000)
2. Klir G. J., Yuan B.: Fuzzy sets and fuzzy logic: Theory and applications. New Jersey: Prentice Hall, (1995)
3. Rigaux P., Scholl M., Voisard A.: Spatial Databases with Applications to GIS. Morgan Kaufman Publishers (2002)
4. Verstraete J., De Tré G., Hallez A.: Adapting TIN-layers to Represent Fuzzy Geographic Information. In: The Seventh Meeting of the EURO Working Group on Fuzzy Sets, pp. 57–62 (2002)
5. Verstraete J., De Tré G., De Caluwe R., Hallez A.: Field Based Methods for the modelling of Fuzzy Spatial Data. Fuzzy modeling with Spatial Information for Geographic Problems, eds. Petry F., Robinson v., Cobb M., 41–69, Springer-Verlag, (2005)
6. Verstraete J., Hallez A., De Tré G.: Bitmap Based Structures for the modelling of Fuzzy Entities. Special issue of Control & Cybernetics vol. 35 no. 1, 147–164, (2006)
7. Verstraete J., : Fuzzy regions: interpretations of surface area and distance. Control and Cybernetics, 38, 509–528, (2009).
8. Zadeh L.A.: Fuzzy Sets. Information and Control 1 3, 338–353 (1965)
9. Zimmerman H-J.: Practical Applications of Fuzzy Technologies. Kluwer Academic Publishers (1999)