Type-2 Fuzzy Sets Applied to Geodemographic Classification

P.F. Fisher[1], N.J. Tate[1], A.Slingsby[2]

[1]Department of Geography, University of Leicester, Leicester LE1 7RH UK
Email: {pff1; njt9}@le.ac.uk

[2]Department of Computer Science, School of Informatics, City University, London EC1V 0HB UK,
Email: a.slingsby@city.ac.uk

1. Introduction

Fuzzy set theory has been widely and successfully applied to model uncertainty in a variety of geospatial contexts, including the classification of land cover (Foody 1996), and geodemographics (Grekousis and Hazichristos 2012). However, as noted by Fisher et al. (2007) such work predominantly makes use of fuzzy sets as originally specified by Zadeh (1965); characterised by crisp membership functions (Mendel and John, 2002). Zadeh (1975) extended his initial ideas to define fuzzy sets with fuzzy membership functions (Mendel and John 2002) and the nomenclature 'type-1' and 'type-2' is now used to refer to these different forms. Fisher and Tate (2014) employed type-1 fuzzy sets to soften a geodemographic classification (the UK Output Area Classification: OAC) of the City of Leicester UK. In this paper we extend that work to explore the use of type-2 fuzzy sets to the OAC.

2. Methods and data

Conventional geodemographics classifications are hard, assigning a single class to each output zone. However, in the process potentially useful information on variation in both geographic and classification spaces is effectively lost (Longley and Goodchild, 2008) and it is impossible to discriminate between zones which exhibit 'strong' and 'less strong' degrees of belonging to a class. This becomes more critical when exploring the local variation (e.g., by town or city) of a global (national) classification. While usually lost, this information may be very useful in applications (Slingsby et al., 2011). Fortunately various methods exist to soften hard classifications to derive multiclass memberships (Bezdek, 1981), especially fuzzy c-means (FCM) clustering (Bezdek, 1981) and possibilistic c-means (PCM). The latter replaces the constraint (Equation 1) of the FCM with a more inclusive constraint (Equation 2); in other words in FCM memberships for one particular case or object are constrained to sum to 1, but in PCM they are simply constrained to be in the range 0 to 1, the sum being no more than the number of classes.

$$\sum_{i=1}^{c} \mu_{ij} = 1, \quad \forall_j \in \{1, \dots, n\} \tag{1}$$

$$0 < \sum_{i=1}^{c} \mu_{ij} \le c, \quad \forall_j \in \{1, \dots, n\} \tag{2}$$

At the core of both is an iterative process between class centroid calculation/update and distance-based membership calculation for *c* classes and *n* elements (zones) which optimises an objective function (Equation 3; Krishnapuram and Keller, 1993; Kruse et al. 2013; Pal et al, 2004):

$$J_{PCM}(\mathrm{X}, \mathrm{U}_{PCM}, \mathrm{C}) = \sum_{i=1}^{c}\sum_{j=1}^{n} \mu_{ij}^{m} d_{ij}^{2} + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n} \left(1 - \mu_{ij}\right)^{m} \quad (3)$$

where the possibility of zone $\mathbf{x}_j$ being in each class/cluster c_i is given in Equation 4.

$$\mu_{ij} = \frac{1}{1 + \left(\frac{d_{ij}^{2}}{\eta_i}\right)^{\frac{1}{m-1}}} \quad (4)$$

Here m is the fuzzifier and d_{ij} the distances in classification space from each zone to the class centroid. η_i is the distance of the "cross-over" point where μ_{ij} = 0.5 (Krishnapuram and Keller 1993), and is obtained from Equation 5.

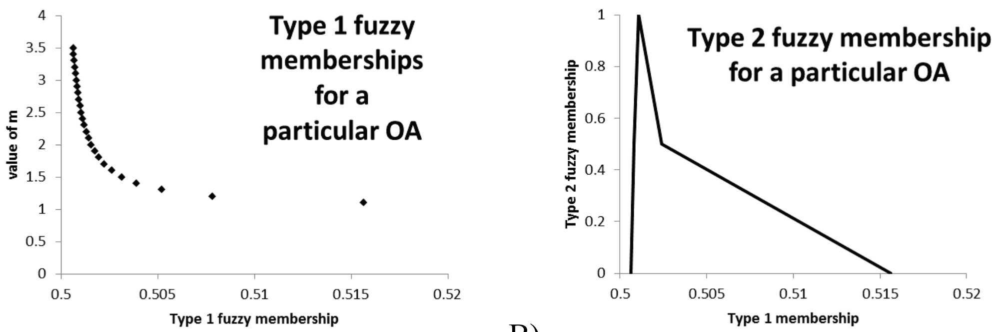
$$\eta_i = K \frac{\sum_{j=1}^{n} \mu_{ij}^{m} d_{ij}^{2}}{\sum_{j=1}^{n} \mu_{ij}^{m}} \quad (5)$$

The UK 2001 Output Area Classification (OAC) is a free geodemographic classification which at the highest taxonomic level is comprised of seven classes named 'Supergroups'. This classification employed a variant of hard c-means and critically in addition to the assigned class, distances d_{ij} are also available for all n elements and c classes. Fisher and Tate (in press) employed [Equation 4] to create possibilistic memberships for each of the seven classes for each census reporting zone (Output Area - OA) for the City of Leicester. They compared PCM outcomes with fuzzy memberships from the equivalent FCM calculation favouring the PCM approach because of the constraint change in Equations 1 and 2. Following general practice (Bezdek 1981 among others) Fisher and Tate (in press) selected a crisp value of $m = 2$. However, m may have any value greater than 1, and following the method of Fisher (2010; see also Hwang and Rhee, 2007) by allowing m to vary [1.1 to 3.5 in this instance] we can generate type-2 fuzzy sets for each Output Area.

3. Results

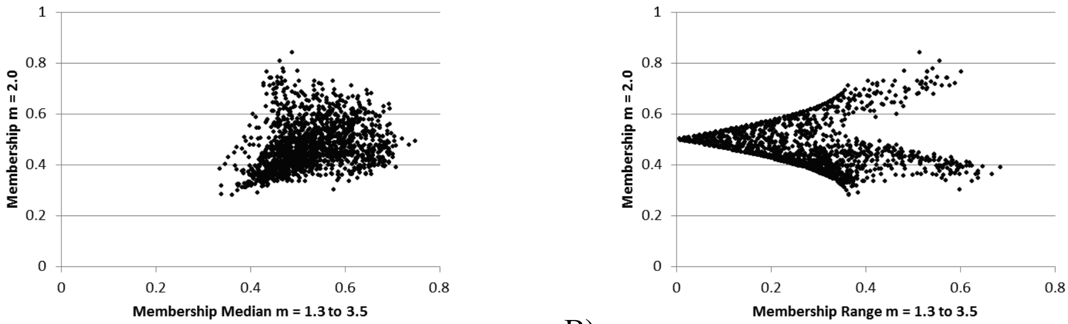
In Figure 1A, for one particular OA within Leicester, the distribution of all type-1 fuzzy memberships are shown plotted against the values of m which yielded them. For convenience type-2 fuzzy sets can be summarised by taking appropriate summary statistics from the distribution of type-1 fuzzy memberships. Thus for the type-1 memberships for one particular OA shown in Figure 1A, the minimum, 1st, 2nd and 3rd quartiles, and maximum are used to summarise the type 2 fuzzy set in Figure 1B. These summary values are assigned memberships of 0, 0.5, 1, 0.5 and 0, respectively, in the type-2 fuzzy set (Figure 1B).

To establish that the type-2 memberships provide new information, the relationship to the type-1 memberships was examined. The relationship where m = 2.0 for the City Living Supergroup is illustrated in Figure 2, and these membership values are seen to not be a good predictor of either the median or the range of the type-2 memberships. Furthermore, Figure 3 shows that neither the median nor the range are predictably related. Figures 2 and 3 are both for the City Living Supergroup only, but the same patterns of poor statistical prediction are repeated for all other Supergroups.



A) B)

Figure 1. For one particular Output Area and the City Living Supergroup A) shows the distribution of type 1 fuzzy memberships plotted against valuations of *m* which yielded them, and B) shows the form of the type 2 fuzzy set from the five summary statistics of that distribution discussed in the text.



A) B)

Figure 2. Scatterplots of membership range and membership median over m valuations from m = 1.1 to 3.5 (horizontal axis) against the type-1 membership (vertical axis) where m = 2.0 (the recommended valuation) for the City Living Supergroup

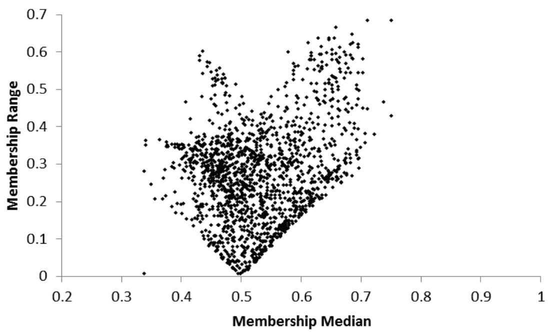


Figure 3. Scatterplot of membership range against membership median over the range of m valuations from m = 1.1 to 3.5 for the City Living Supergroup

Figure 4 shows four type 2 fuzzy sets for the City Living Supergroup derived as in Figure 1B, for those OAs with the largest and smallest median (2nd quartile), and the largest and smallest range (maximum – minimum) of membership values. The median shows the degree to which the OA is typical of the class; the OA with the largest being the most typical and the smallest the least. The range shows the degree to which a particular output area is a good example of that Supergroup. The OA with the smallest range is the most representative; all memberships for different valuations of *m* have similar values. The OA with the largest range is the poorest representative; the memberships are most varied. Shapes and distributions of the seven graphs are remarkably similar showing that the range of memberships in each of the four types of OA are similar for all Supergroups.

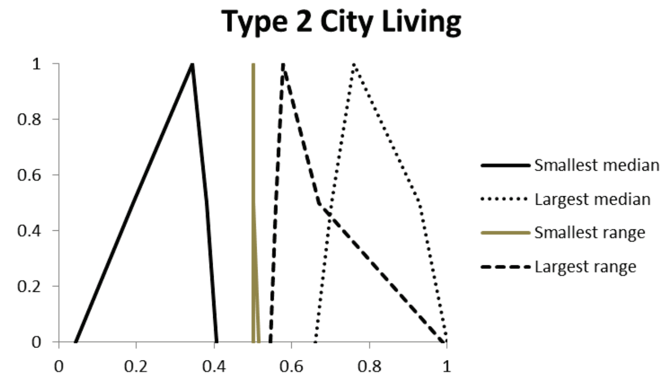


Figure 4. For the City Living Supergroup in the Output Area Classification the distributions of Type 2 fuzzy sets are shown for the four Output Areas having the smallest and largest range and median values.

4. Conclusion

Our research has revealed that for each Supergroup, we can subset OAs into distinct classes on the basis of whether they exhibit sensitivity to *m,* and to differentiate between those zones which display similar type-1 fuzzy memberships.

References

Bezdek, JC, 1981, *Pattern recognition with fuzzy objective function algorithms*. Plenum Press, New York.

Fisher, PF, 2010, Remote sensing of land cover classes as type 2 fuzzy sets. *Remote Sensing of Environment,* 114(2): 309-321.

Fisher PF, and Tate, NJ, in press, Modelling Class Uncertainty in the Geodemographic Output Area Classification, *Environment and Planning B.*

Foody GM, 1996, Fuzzy modelling of vegetation from remotely sensed imagery. *Ecological Modelling*, 17: 3-12.

Grekousis G and Hatzichristos T, 2012, Comparison of two fuzzy algorithms in geodemographic segmentation analysis: The Fuzzy C-Means and Gustafson–Kessel methods. *Applied Geography,* 34:125–136.

Hwang, C and Rhee, FC-H, 2007, Uncertain fuzzy clustering: Interval type-2 Fuzzy approach to c-means. *IEEE Transactions on Fuzzy Systems* **15**: 107-120.

Krishnapuram R and Keller JM 1993, A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems* **1:** 98–110.

Kruse R, Borgelt C, Klawonn F, Moewes C, Steinbrecher M, Held P, 2013, *Computational Intelligence: A Methodological Introduction* (Springer-Verlag, London).

Longley PA, Goodchild MF, 2008, The use of geodemographics to improve public service delivery. In: J Hartley, C Donaldson, C Skelcher, and M Wallace (eds), *Managing to Improve Public Services*. Cambridge University Press, Cambridge 176–194.

Mendel JM and John RI, 2002, Type-2 fuzzy sets made simple. *IEEE Transactions on Fuzzy Systems*, 10(2): 117-127.

Pal NR, Pal K, Keller JM, Bezdek JC, 2004, A new hybrid c-means clustering model in *Proceedings of the IEEE International Conference on Fuzzy Systems, Budapest,* (IEEE) pp 179–184.

Slingsby A, Dykes J. & Wood J, 2011, Exploring Uncertainty in Geodemographics with Interactive Graphics. *IEEE Transactions on Visualization and Computer Graphics*, 17(12): 2545-2554.

Zadeh LA, 1965, Fuzzy sets. *Information and Control,* 8: 338–353.

Zadeh LA, 1975, The concept of a linguistic variable and its application to approximate reasoning—I. *Information Sciences,* 8(3): 199-249.