Fuzzy Geodemographics: Application of Fuzzy c-means

Burcin Yazgi Walsh[*1] and Chris Brunsdon[†1]

[1]National Centre for Geocomputation (NCG)

January 29, 2019

Summary

The geodemographics approach tries to provide some classifications based on the socio-economic characteristics of the spatial units. It is an approach applied and discussed in many different countries and fields over the last five decades. This paper discusses one of the main criticisms that have been made for geodemographics classification which focuses on the clustering method by applying a fuzzy clustering technique. The methodology applied in this study is also bringing an alternative aspect for the investigation of the spatial distribution of fuzziness level.

KEYWORDS: Geodemographics, Fuzzy c-means (FCM), Ireland, 2016

1. Introduction

The geodemographics approach tries to provide some classifications based on the socio-economic characteristics of the spatial units. The classifications are organised following the similarities in the dataset. As a result of these classifications varying dynamics of each spatial unit will be represented. Even though the variables and the scale of the spatial units are unique to each study, mostly census datasets are used. The geodemographics approach is applied both in private and public sectors but popularity is mostly gained through the applications in market research for commercial value.

The approach was applied in many different countries and fields over the last five decades. Discussions during this period brought up some criticisms which could be summarised under different topics and one of the main criticisms is related to the clustering method applied. The most commonly used algorithm is focused on a form of crisp clustering in which each data point is forced to belong to only one group.

In this paper in order to deal with this issue, we are applying a fuzzy clustering approach. Fuzzy clustering algorithms which provide the advantage of data points belonging to more than one group with varying membership values will be discussed in more detail in the next sections of the paper. After presenting the outline of the study and the results, the paper will be finalised with some discussion and concluding remarks.

2. Fuzzy approach in geodemographics

In the late 1980s Openshaw was addressing some uncertainties in geodemographic studies. One of them was associated with the nature of the crisp clustering. Spatial units that have only slight differences can end up in different clusters. Besides, not necessarily each geographical unit is exactly belonging to one particular cluster. He was the first to propose fuzziness in geodemographics regarding to these issues (Openshaw, 1989).

After the initial theoretical attempts of Openshaw some researchers started to practice fuzzy approaches in geodemographics. Some of the early examples are Feng and Flowerdew (1998, 1999) and

[*] burcin.yazgiwalsh@mu.ie
[†] christopher.brunsdon@mu.ie

Hatzichristos and Koutsopoulos (1998). In the last decade the number of the studies applying the fuzzy approach accelerated remarkably. Some of these examples are Grekousis and Hatzichristos, 2012, 2013; Son et al., 2012; Fisher et al., 2014.

The Fuzzy c- means (FCM) method, applied in this study is based on Bezdek's (1981) algorithm who followed Zadeh (1965) and Dunn (1973). Instead of classifying each spatial unit under one category, FCM allows a "better profile of each spatial unit" (Grekousis and Hatzichristos, 2012). More information about each spatial unit will be provided through the varying membership values as a result of the algorithm, which then allows some further detailed explanations about the dynamics within the spatial units. Besides, fuzzy approaches can be more robust to noise in data and more efficient in terms of data size (Grekousis and Hatzichristos, 2012).

3. Outline of study and results

The data used in this study are derived from Ireland's 2016 Census Data obtained through Central Statistics Office Ireland open portal. The chosen 40 variables fall under the 10 different themes of this dataset and are organised under the 5 different categories for the analysis. This study follows the variable structure applied in geodemographics studies done by Brunsdon et al. (2011) and Yazgi Walsh et al. (2017) for Ireland for crisp clustering. The geographical representation for this research is the smallest possible scale available in the census dataset for Ireland which is the Small Area (SA) level.

FCM works based on an exponent which controls how fuzzy the outcomes are (1 for no fuzziness - increase of fuzziness with higher values) and there is no right answer for the best 'm' value, as well as the appropriate number of clusters. Some tests are run to validate these values, including PC (Partition co-efficient), CE (classification entropy), XB (Xie-Beni index), and S (Separation index). According to these validation tests, optimal number for clusters is 3 and m is 1.5.

3.1. Membership plots and entropy

Here an optimal number of fuzzy clusters was found to be 3. This is lower than the amount of 'crisp' clusters found on the same data – this was 6. However this is perhaps not surprising. If some of the crisp clusters were mid-way in characteristics between some of the more polarised, then arguably these could be characterised by mid-level partial membership of the polarised ones. In **Figure 1**, membership functions for the three clusters are shown in map form.

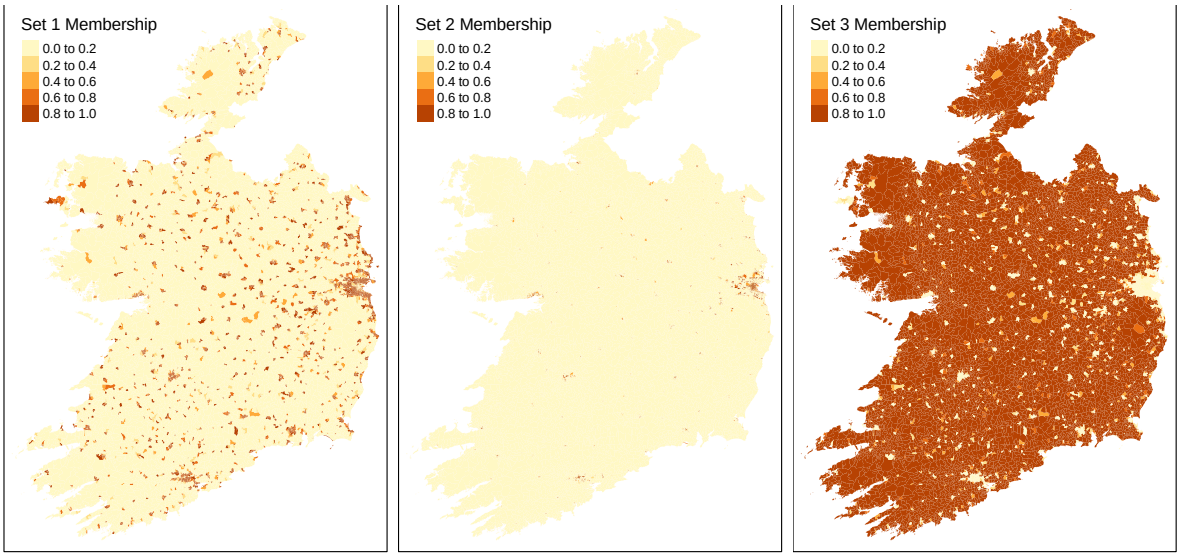

Figure 1 Membership of the three fuzzy clusters.

Cluster 1 identifiers peri-urban and suburban regions generally, whereas cluster 2 focuses on central urban areas, particularly those around Dublin. Cluster 3 identifies more rural areas.

Finally, we can view the *Entropy* of the classifications. This is essentially a measure of the vagueness in classification. If m1, m2 and m3 are the membership levels for each cluster, the entropy is defined as Equation 1.

$$E = log_{10}(m_1) + log_{10}(m_2) + log_{10}(m_{13}) \quad (1)$$

It is maximised when all three memberships are set at 1/3. In **Figure 2** we see that entropy exhibits geographical pattern – with greater levels around urban areas, and particularly in the hinterlands of Dublin.

Figure 2 Entropy of small areas.

3.2. Kernel density ternary plots

The ternary plot is also useful at showing concentration of membership levels across the SAs. Here we see that there is a lot of cluster 1/cluster 2 mixing - and perhaps less so for cluster 1/cluster 3. There are peaks where there is strong membership of one of the individual clusters compared to the other 2 (**Figure 3**).

A density plot is shown in **Figure 4** for each of the six 'crisp' clusters also generated (using PAM). Here we see one of the crisp clusters (crisp cluster 5) focus on fuzzy cluster 1 - but some are a mixture of partial membership of fuzzy clusters 1 and 2 - these are crisp cluster 2 to 4. Crisp cluster 1 generally focusses on fuzzy cluster 3, and crisp cluster 6 focusses on fuzzy cluster 2.

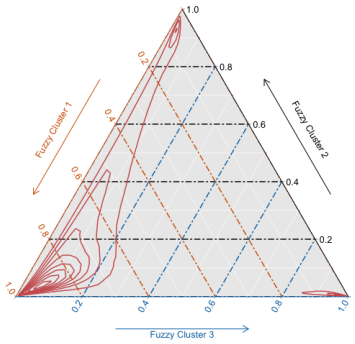

Figure 3 Membership level concentrations.

The image of the last **Figure 5** is also well illustrated when confidence ellipses for the centroid of each of the crisp clusters is shown. Here the nearly-ordered locations of crisp clusters 2-5 may be seen.

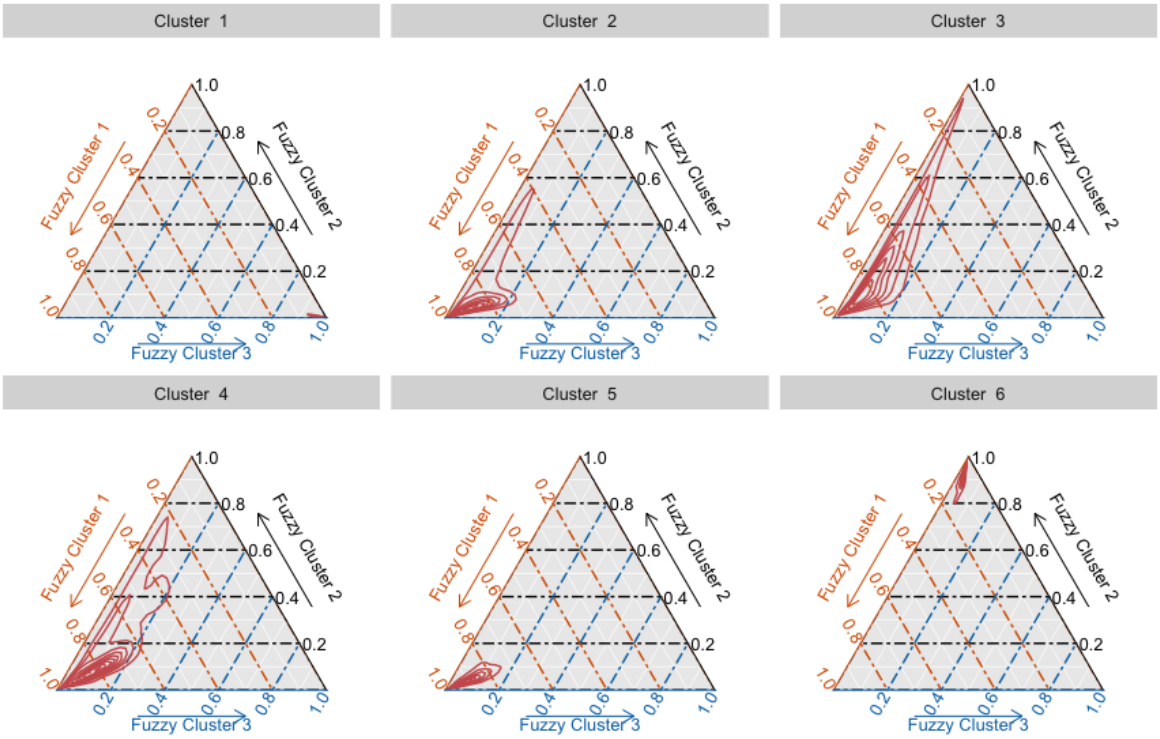

Figure 4 Density plots.

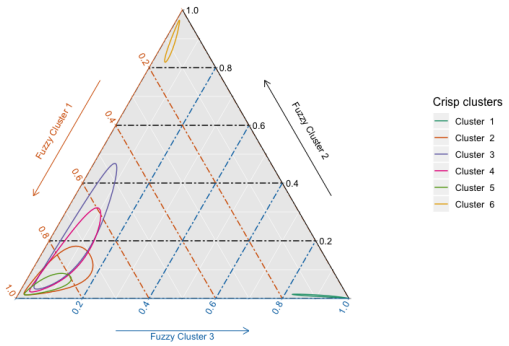

Figure 5 Crisp cluster locations.

4. Discussion and conclusion

This paper was organised to cover one of the main criticisms in geodemographics studies mostly focusing on the method aspect by applying a fuzzy clustering algorithm. Since the classifications as an outcome of the fuzzy clustering are represented through membership values instead of a one specific group number, it is possible to investigate the patterns in the classifications as well as the spatial distribution of the fuzziness. In Ireland context, urban areas especially around Dublin need a further attention in terms of clustering. Outcomes of this study are highly beneficial for further studies in geodemographics for Ireland and also for the development of the approach.

Even though there is no argument on the advantages and flexibility that fuzzy clustering bring to your classifications, it is still a challenge to present them visually, especially the ones based on big datasets. How to visualise these soft classifications is another aspect that needs some further attention.

It is also important to mention that every step of this research is supporting an open approach in order to eliminate the 'black box' processes in geodemographics studies.

5. Acknowledgements

This paper has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant Number 15/IA/3090.

References

Bezdek J C (1981). Pattern recognition with fuzzy objective function algorithms. Plenum Press.

Brunsdon C, Rigby J and Charlton M (2011). Ireland Census of Population 2011: A classification of Small Areas, RPubs. https://rpubs.com/chrisbrunsdon/14998.

CSO, Central Statistics Office Ireland (2016). Ireland Population Census Results, Small Areas, Accessed in September 2017.

Dunn J C (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3, 32-57.

Feng Z and Flowerdew R (1998). Fuzzy geodemographics: a contribution from fuzzy clustering methods, *Innovations in GIS 5*, edited by Carver S, London: Taylor and Francis, 119-127.

Feng Z and Flowerdew R (1999). The use of fuzzy classification to improve geodemographic targeting. In B. Gittings (Ed.), *Innovations in GIS 6*, London: Taylor and Francis, 133–143.

Fisher P, Tate N and Slingsby A (2014). Type-2 Fuzzy Sets Applied to Geodemographic Classification. GIScience 2014, Vienna, Austria.

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.

Grekousis G and Hatzichristos T (2013). Fuzzy clustering analysis in geomarketing research. Environment and Planning B, 40, 95-116.

Hatzichristos T and Koutsopoulos K (1998). Delineation of Demographic Regions in Athens Municipality using Fuzzy Classification. GIS Planet 98 Conference, Lisbon, Portugal.

Openshaw S (1989). Making geodemographics more sophisticated, *Journal of the Market Research Society*, 31, 111–131.

Son L H, Cuong B C, Lanzi P L and Thong N T (2012). A novel intuitionistic fuzzy clustering method for geo-demographic analysis. Expert Systems with applications, 39, 10, 9848-9859.

Yazgi Walsh B, Brunsdon C and Charlton M (2017). Ireland Census of Population 2016: A classification of Small Areas, RPubs. https://rpubs.com/burcinwalsh/343141.

Zadeh L A (1965). Fuzzy sets. *Information and Control*, 8, 338-353.

Biographies

Burcin Yazgi Walsh

Burcin Yazgi Walsh is a researcher at Maynooth University, Ireland as part of "Building City Dashboards Project". She has interests in spatial analysis in urban studies. Her recent works are mostly focusing on geodemographics in Ireland and open research.

Chris Brunsdon

Chris Brunsdon is Director of the National Centre for Geocomputation at Maynooth University, Ireland, and Professor of Geocomputation. He has interests in spatial data science, statistical computing, quantitative modelling and geographical data visualisation.