High-resolution Remote Sensing Clustering Analysis Based on Object-based Fuzzy Data Modeling

Tao Jiang[a], Dan Hu[a,**], Xianchuan Yu[a,*]

[a] *College of Information Science and Technology, Beijing Normal University, Beijing 100875, China*

Abstract

This template helps you to create a properly formatted LATEX manuscript. Keywords: elsarticle.cls, LATEX, Elsevier, template

Keywords: Fuzzy Set Data, High-resolution remote sensing image, New IT2-FCM, Fuzzy clustering

1. Introduction

Clustering analysis is a wide useful tool in remote sensing applications. However, there exists uncertainty in classifications of remote sensing image. For instance, owing to the inherent uncertainty of remote sensing and the many sources of interference, there may be a series of uncertainties in the spectral signatures between classes and spectral variation within classes (Cheng et al., 2004). On this account, conventional, crisp clustering algorithms may not perform well in classifications of remote sensing in most cases. Since the 1980s, fuzzy clustering has been extensively studied and successfully applied in remote sensing classification. The most commonly utilized fuzzy clustering algorithm is fuzzy c-means (FCM) algorithm (Bezdek et al., 1984). Many researchers have applied FCM to remote sensing image analyses (Ibrahim et al., 2005; Schowengerdt, 2006; Ghosh et al., 2011) , and have achieved more satisfactory results than hard classifications such as k-means and maximum likelihood classification. Standard FCM algorithm is applied

*Corresponding authors

**Principal corresponding author

Email addresses: hd@bnu.edu.cn (Dan Hu), chuan.yu@ieee.org (Xianchuan Yu)

to low-resolution images and based on image pixels, but high-resolution remote sensing image have smaller targets and more information. More details in the high-resolution (more than 10m) images mean it more difficult to describe a ground object, which indicates that as a pixel-based method, FCM algorithm cannot obtain the desired land cover classification results of high-resolution remote sensing images. To take advantage of more detailed information of high-resolution images, object-based classification methods for medium to high-resolution images can provide a valid alternative to pixel-based classification methods (Geneletti and Gorte, 2003; Guo et al., 2007; Tenenbaum et al., 2011; Yu et al., 2012). However, it is difficult to extract effective and stable features from the segmentation units, which directly affects the accuracy and stability. For example, the mean spectral signature is typically used to describe a segmentation unit, but this may not appropriately partition two different objects with the same mean value. He et al. (2016) recently proposed an unsupervised classification method that adopts an object-based interval value modeling method fuzzy clustering algorithm. However, with the development of remote sensing technology and the launching of third generation commercial Earth observation satellites such as WorldView-4 satellites, the spatial resolution of remote sensing images can reach to about 0.4m, interval-valued modeling cannot represent a feature's uncertainty in the segmentation units.

To put forward the method for describing the uncertainty and obtain better results for high-resolution,remote sensing image clustering analysis, we proposed an object-based fuzzy data modeling method and a new interval type-2 fuzzy clustering algorithm.

This article is organized as follows. Fuzzy Set Data are defined and constructed in Section 2.

2. Fuzzy set valued data modeling and dissimilarity metric

2.1. Definition of fuzzy set valued data

Since Zadeh (1965) introduced the concept of fuzzy set (FS) whose elements have degrees of membership, we know it can describe the uncertainty of objects. A set of membership degrees can be thought of as membership functions (MF) mapping predicates into fuzzy sets. The widely used and fundamental membership function of fuzzy sets is triangle membership function. So we use triangle MF fuzzy set to define fuzzy set valued data.

Definition 1 Triangle MF Fuzzy Set Valued Data

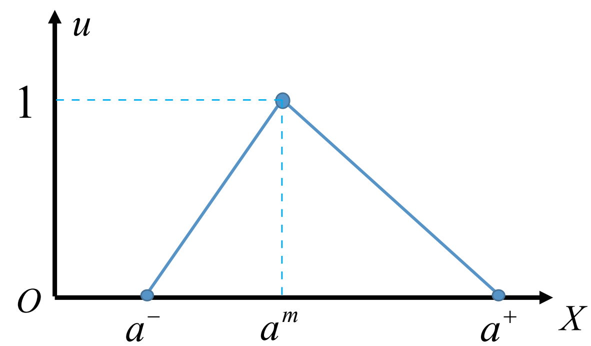

Figure 1: The triangle MF fuzzy set $\tilde{A}$

Let $A = \{a^-, \ldots, a^m, \cdots, a^+\}$ be a set containing some sorted numbers, where a^+ and a^- are real numbers representing the lower and upper bounds of the interval set A, and a^m is the median number of A. Triangle MF fuzzy set valued data modeling denoted by $\tilde{A}$ can be defined by a triangle membership function fuzzy set constructed by these three critical parameters: $(a^-, 0)$,$(a^m, 1)$,$(a^+, 0)$. As shown in Figure 1, $(a^-, 0)$ and $(a^+, 0)$ build the bottom edge of the triangle MF in geometry and form an interval set within a certain range in algebra that ensures the range of variation (Moore, 1966), $(a^m, 1)$ is the apex of triangle MF, and a^m is the median number of set A. As is well-known, the median that is the value separating the higher half of a data sample or a probability distribution from the lower half is about the statistical concept. The basic advantage of the median in describing data compared to the mean is that it is not skewed so much by extremely large or small values, and so it may give a better idea of a 'typical' value (Bissell, 1994). Therefore, we use the median number to construct triangle MF which has good robustness to noise points and outlier.

2.2. Definition of distance for fuzzy set valued data

Distance is a numerical description of how far apart objects are, a distance function or metric is a dissimilarity and triangle inequality for different datasets and plays an important role in clustering analysis. There are many distance metrics for fuzzy set data (e.g.,city-block, Euclidean, Mahalanobis, Hausdorff, Wasserstein) can be found in (Wang, 1997; Zwick et al., 1987; Diamond and Kloeden, 1994; Chaudhur and Rosenfeld, 1996; Saha et al., 2002; De Carvalho et al., 2006; Irpino et al., 2014). However, the optimal dissimilarity metric depends on the application.

Hausdorff distance measures how far two subsets of a metric space are from each other, Diamond and Kloeden (1994) introduced a Hausdorff metric distance between fuzzy sets. Let $\tilde{A}$ and $\tilde{B}$ be two fuzzy sets, the Hausdorff distance is as follows:

$$d_\alpha(\tilde{A}, \tilde{B}) = \max\{\sup_{a\in\tilde{A}} \inf_{b\in\tilde{B}} \|a-b\| , \sup_{b\in\tilde{B}} \inf_{a\in\tilde{A}} \|a-b\|\} \tag{1}$$

where $\|\ \|$ means distance between elements in $\tilde{A}$ and $\tilde{B}$.

The location relationship between two fuzzy sets is illustrated in Figure 2. Fuzzy sets differ from real numbers, they may intersect part each other or separate so far away and not intersect at all, there also may be a situation that one set is including in other set. For this reason, to define distance metric of fuzzy sets first of all we need to combine with the characteristics of fuzzy set. Since (Liu, 2008; Mendel et al., 2009) proposed α - plane representation of type-2 FS (comparable to the $\alpha - cut$ representation of a type-1 FS), we know α - plane presentation is useful for both theoretical and computational studies of and for fuzzy sets. Liu (2008) provided examples that show that the mean value(defuzzified value) of the centroid can often be approximated by using the centroid of only 0 and 1 α - planes of a fuzzy set, and Nie and Tan (2015) recently proposed the closed form formulas for the centroid endpoints of the α planes for fuzzy sets, and also indicated that any α plane can respectively be expressed as a generalized linear combination of those of its $\alpha = 0$ and $\alpha = 1$ planes. For all the above results, we define a new Hausdorff distance metric for two type-1 fuzzy sets (T1 FSs) using the $\alpha = 0$ and $\alpha = 1$ cut of the fuzzy set. Let $\tilde{A}_\alpha$ and $\tilde{B}_\alpha$ be the $\alpha - cut$ of FS $\tilde{A}$ and $\tilde{B}$. When $\alpha = 0$, $\tilde{A}_0$ and $\tilde{B}_0$ are the $\alpha = 0$ cut and they are interval sets. After ascending sorted, they can be described as $\tilde{A}_0 = [a_{min}, \ldots, a_{max}]$ and $\tilde{B}_0 = [b_{min}, \cdots, b_{max}]$, For operations 'inf' and 'sup' in Equation 1 can be replcaed by 'min' and 'max' operations respectively. So the new Hausdorff distance metric for $0 - cut$ is as follow:

$$\begin{aligned} d_0(\tilde{A}_0, \tilde{B}_0) &= \max\{\max_a \min_b \|a-b\| , \max_b \min_a \|a-b\|\} \\ &= \max\{\|a_{min} - b_{min}\| , \|a_{max} - b_{max}\|\} \end{aligned} \tag{2}$$

In a similar way, we can get the $1 - cut$ Hausdorff distance metric for two FSs as $d_1(\tilde{A}_1, \tilde{B}_1)$ easily.

Let $\tilde{A}$,$\tilde{B}$,$\tilde{C}$ be three fuzzy sets, Then $d(\tilde{A}, \tilde{B})$ is a distance measure means if it is satisfies:

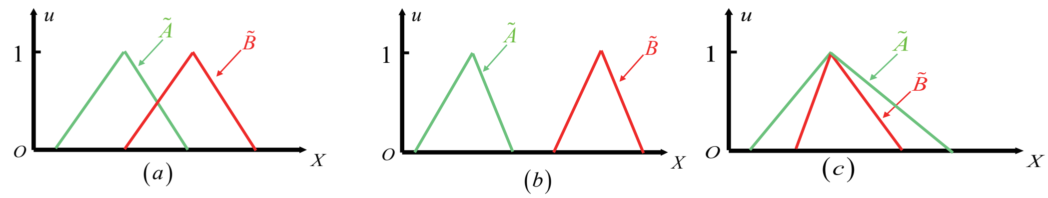

Figure 2: There are three kind of location situations between two fuzzy set $\tilde{A}$ and $\tilde{B}$. (a)$\tilde{A}$ and $\tilde{B}$ intersect; (b)$\tilde{A}$ and $\tilde{B}$ is disjoint; (c)$\tilde{A}$ contains $\tilde{B}$.

reflexivity: $d(\tilde{A}, \tilde{A}) = 0$, *symmetry*: $d(\tilde{A}, \tilde{B}) = d(\tilde{B}, \tilde{A})$, and *the triangular inequality*: $d(\tilde{A}, \tilde{B}) \leq d(\tilde{A}, \tilde{C}) + d(\tilde{C}, \tilde{B})$.

3. sec 3

4. sec 4

5. sec 5

6. sec 6

Acknowledgments

This research is supported in part by National Natural Science Foundation of China (Grant No. 11001019, 11471045, 41272359, 41672323) and Major scientific research projects of universities in Guangdong(2016KTSCX167).

References

Bezdek, J. C., Ehrlich, R., Full, W., 1984. Fcm: The fuzzy c-means clustering algorithm. Computers & Geosciences 10 (2-3), 191–203.

Bissell, D., 1994. Statistical Methods for SPC and TQM. Vol. 26. CRC Press.

Chaudhur, B., Rosenfeld, A., 1996. On a metric distance between fuzzy sets. Pattern Recognition Letters 17 (11), 1157–1160.

Cheng, J., Guo, H., Shi, W., et al., 2004. The uncertainty of remote sensing data.

De Carvalho, F. d. A., de Souza, R. M., Chavent, M., Lechevallier, Y., 2006. Adaptive hausdorff distances and dynamic clustering of symbolic interval data. Pattern Recognition Letters 27 (3), 167–179.

Diamond, P., Kloeden, P., 1994. Metric spaces of fuzzy sets: theory and applications. World scientific.

Geneletti, D., Gorte, B., 2003. A method for object-oriented land cover classification combining landsat tm data and aerial photographs. International Journal of Remote Sensing 24 (6), 1273–1286.

Ghosh, A., Mishra, N. S., Ghosh, S., 2011. Fuzzy clustering algorithms for unsupervised change detection in remote sensing images. Information Sciences 181 (4), 699–715.

Guo, Q., Kelly, M., Gong, P., Liu, D., 2007. An object-based classification approach in mapping tree mortality using high spatial resolution imagery. GIScience & Remote Sensing 44 (1), 24–47.

He, H., Liang, T., Hu, D., Yu, X., 2016. Remote sensing clustering analysis based on object-based interval modeling. Computers & Geosciences 94, 131–139.

Ibrahim, M., Arora, M., Ghosh, S., 2005. Estimating and accommodating uncertainty through the soft classification of remote sensing data. International Journal of Remote Sensing 26 (14), 2995–3007.

Irpino, A., Verde, R., De Carvalho, F. d. A., 2014. Dynamic clustering of histogram data based on adaptive squared wasserstein distances. Expert Systems with Applications 41 (7), 3351–3366.

Liu, F., 2008. An efficient centroid type-reduction strategy for general type-2 fuzzy logic system. Information Sciences 178 (9), 2224–2236.

Mendel, J. M., Liu, F., Zhai, D., 2009. α-plane representation for type-2 fuzzy sets: Theory and applications. IEEE Transactions on Fuzzy Systems 17 (5), 1189–1207.

Moore, R. E., 1966. Interval analysis. Vol. 4. Prentice-Hall Englewood Cliffs.

Nie, M., Tan, W. W., 2015. Closed form formulas for computing the centroid of a general type-2 fuzzy set. In: Fuzzy Systems (FUZZ-IEEE), 2015 IEEE International Conference on. IEEE, pp. 1–8.

Saha, P. K., Wehrli, F. W., Gomberg, B. R., 2002. Fuzzy distance transform: theory, algorithms, and applications. Computer Vision and Image Understanding 86 (3), 171–190.

Schowengerdt, R. A., 2006. Remote sensing: models and methods for image processing. Academic press.

Tenenbaum, D. E., Yang, Y., Zhou, W., 2011. A comparison of object-oriented image classification and transect sampling methods for obtaining land cover information from digital orthophotography. GIScience & Remote Sensing 48 (1), 112–129.

Wang, W.-J., 1997. New similarity measures on fuzzy sets and on elements. Fuzzy sets and systems 85 (3), 305–309.

Yu, X., An, W., He, H., 2012. A method of auto classification based on object oriented unsupervised classification. Progress in Geophysics 27 (2), 744–749.

Zadeh, L. A., 1965. Fuzzy sets. Information and control 8 (3), 338–353.

Zwick, R., Carlstein, E., Budescu, D. V., 1987. Measures of similarity among fuzzy concepts: A comparative analysis. International Journal of Approximate Reasoning 1 (2), 221–242.