Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 3, Issue 5: 23-30, DOI: 10.25236/AJCIS.030504

Dominant Set Based Fuzzy Clustering Ensemble for Aggregation of Tortuosity Measures

Pan Su

School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, Hebei, China

ABSTRACT. *The tortuosity of corneal nerve fibres is correlated with a number of diseases such as diabetic neuropathy. The assessment of corneal nerve tortuosity in in vivo confocal microscopy images can inform the detection of diseases and complications. Clustering is a typical technique which is often used to discover data distribution. This paper proposes a dominant set based fuzzy clustering ensemble method based on cluster filtering. The proposed method uses similarity relationship to represent the connections among the clusters to form a graph and the dominant sets are extracted to form meta-clusters. Then, each meta-cluster is aggregated to obtain a cluster in the final clustering ensemble outputs. Experimental results on a public corneal nerve tortuosity data set demonstrates the effectiveness of the proposed method in IVCM image tortuosity grading.*

KEYWORDS: *clustering ensemble, fuzzy clustering, dominant set, nerve tortuosity*

1. Introduction

Compared with individual clustering method, clustering ensemble may produce better and more robust result in several applications [1]. Similar to classifier ensemble [2], clustering ensemble maps results provided by multiple base clustering members to a single final one through the consensus function. The performance of clustering ensemble is usually affected by the quality and diversity of the results of base clustering members (BCMs). Usually, there are two key factors in the clustering ensemble, one is the generation of BCMs and the other is implementation of the consensus function. A number of consensus functions are proposed in the literature, including: voting based [3], label assignment matrix based [4], similarity [5] and graph based method [6], and so on. Many of the graph based consensus functions group several clusters to form a meta-cluster. The dominant set [7] provides a formal description and search method for a set of nodes which are more correlated internally than externally on a weighted graph. Therefore, in the setting of clustering ensemble, the dominant set can be employed by the graph based consensus for extracting the hybercluster from all clusters generated by the BCMs.

The tortuosity assessment of vessels and nerves in medical images has drawn substantial attention, for its potentials in assisting various medical diagnoses. Numerous novel tortuosity measures that could potentially exhibit high correlations with the disease of interest are proposed in the literature. Nevertheless, there is no universal agreement as to which standard or measure to apply [8]. With the similar idea to clustering ensemble, the simultaneous use of multiple measures may produce a robust and accurate assessment. Therefore, the clustering ensemble method based on the dominant set is applied to fulfill the aggregation of multiple tortuosity measures in this paper. In the experimentation on a public corneal nerve tortuosity data set, results show that the dominant set based clustering ensemble achieves better performance than conventional c-means and fuzzy c-means.

2. Dominant set based fuzzy clustering ensemble

Fuzzy clustering has been successfully applied to practical problems. Different from the hard boundary clustering, fuzzy clustering allows an instance to belong to different clusters according to the membership degree. Each fuzzy cluster $C_k, k = 1, \cdots, K$ in a fuzzy clustering result π is a fuzzy set, where $C_k(x) \in [0,1]$ represents the degree to which the instance x belongs to the fuzzy set C_k and satisfy that $\sum_{k=1}^{K} C_k(x) = 1$.

Formally, fuzzy (or soft boundary) clustering ensemble can be described as follows: let $X = \{x_1, \cdots, x_N\}$ be a set of N instances and $\Pi = \{\pi_1, \cdots, \pi_m, \cdots, \pi_M\}$ be M fuzzy BCMs (FBCMs). Each FBCM π_m provides a set of fuzzy sets (clusters), indicated as: $\pi_m = \{C_1^m, \cdots, C_{K_m}^m\}$, where K_m is the number of fuzzy sets in the FBCM. A set of fuzzy clusters is formed by the union of clusters generated by all FBCMs: $\{C_1, \cdots, C_n\} = \cup_{m=1}^{M} \pi_m$, $n = \sum_{m=1}^{M} K_m$. For each $x \in X$ and each cluster member $\pi_m \in \Pi$, $C_k^m(x) \in [0,1]$ represents the degree to which cluster C_k^m the instance x belongs to. In fuzzy clustering ensemble, the generation of FBCMs can be fulfilled by sampling attribute set, perturbing parameters of a single fuzzy clustering algorithm, and employing different algorithms [9].

For a given data set X, the goal of consensus function in fuzzy clustering ensemble is: input the Π of FBCMs and output a partition of X. In graph based clustering ensemble, the weights of nodes and edges of clusters are usually defined by the fuzzy clusters generated by FBCMs and the similarity between clusters, respectively. Since fuzzy clusters in $\{C_1, \cdots, C_n\}$ may share instances, the overlap of those shared instances between a pair of basic clusters can be measured. In this paper, the overlap (similarity) between two fuzzy clusters is defined as:

$$L(C_i, C_j) = \frac{\sum_{l=1}^{N} \min(C_i(x_l), C_j(x_l))}{\sum_{l=1}^{N} \max(C_i(x_l), C_j(x_l))}, \text{if } i \neq j \quad (1)$$

and $L(C_i, C_j) = 0$, if $i = j$.

Conventional graph based consensus functions usually build a partition of $\{C_1, \cdots, C_n\}$, which means that each fuzzy cluster is assigned to a set in the partition.

ISSN 2616-5775 Vol. 3, Issue 5: 23-30, DOI: 10.25236/AJCIS.030504

However, abnormal clusters which have low similarity with other normal clusters can also be generated. In this case, if the abnormal cluster is forced to be assigned to a set of normal clusters, the quality of clustering ensemble outputs will be reduced. Dominant set provides a method to search a set of closely related nodes (named as dominant set) in a weighted graph. Dominant set generalize the concept of maximal clique of unweighted graphs to weighted graphs. The weights of edges between nodes in a dominant set should be large, which indicates the high similarity of those nodes. Let $a_{ij} = L(C_i, C_j)$ represents the weight of edge between node C_i and C_j and satisfy $L(C_i, C_j) = L(C_j, C_i)$, the dominant set can be found by solving a quadratic programming [7]:

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{z}) = \mathbf{z}^{\top} A \mathbf{z} \\ \text{subject to} \quad & \mathbf{z} \in \Delta \end{aligned} \tag{2}$$

where $A = (a_{ij})_{n\times n}$, $i, j = 1, \cdots, n$, and $\Delta = \{\mathbf{z} \in \mathbb{R}^n : \sum_{i=1}^{n} z_i = 1 \text{ and } z_i \geq 0\}$. Equation (2) can be solved by:

$$z_i^{(t+1)} = z_i^{(t)} \frac{(A\mathbf{z}^{(t)})_i}{\mathbf{z}^{(t)\top} A \mathbf{z}^{(t)}}, \tag{3}$$

where t is the number of iteration and $z_i > 0$ indicates C_i is in the dominant set.

It is worth noticing that only one dominant set S can be extracted from the graph at one time. When using this method to search the dominant set and form meta-clusters for fuzzy clustering ensemble, the "peel-off" strategy is employed, once a dominant set $\mathbb{S}$ is found, the node set in the original graph $\{C_1, \cdots, C_n\}$ is replaced by $\{C_1, \cdots, C_n\} \backslash \mathbb{S}$, then the similarity matrix A is calculated and Equation (3) is solved again. This process is repeated till all the dominant sets are extracted, or a number of dominant sets is reached.

If the dominant set is employed by the consensus function of clustering ensemble, each dominant set extracts a meta-cluster, i.e., a set of clusters with high similarity between them. The method of aggregation is then used to transform a meta-cluster into a cluster in the final output of clustering ensemble. Given that the dominant sets (meta-clusters) found in the node set $\{C_1, \cdots, C_n\}$ are $\mathbb{S}_1, \cdots, \mathbb{S}_d$, the membership degree of an instance x_i belongs to the final output cluster $\mathbb{C}_k$, $k = 1, \cdots, d$, of fuzzy clustering ensemble is defined as:

$$\mathbb{C}_k(x_i) = \frac{\sum_{C_j \in \mathbb{S}_k} C_j(x_i) / |\mathbb{S}_k|}{\sum_{k=1}^{d} \sum_{C_j \in \mathbb{S}_k} C_j(x_i) / |\mathbb{S}_k|}.$$

3. Application in tortuosity assessment

The grading of corneal nerve tortuosity level in in vivo confocal microscopy (IVCM) images can be utilized to detect diseases and complications. Usually, IVCM images are graded subjectively with respect to the tortuosity of corneal fibers is 'very high', 'high', and so on by ophthalmologists based on their clinical

ISSN 2616-5775 Vol. 3, Issue 5: 23-30, DOI: 10.25236/AJCIS.030504

experiences [10]. In this paper, the dominant set based fuzzy clustering ensemble is adopted on the Corneal Nerve Tortuosity (CNT) Data Set (available at http://bioimlab.dei.unipd.it/Data%20Sets.htm). In the the CNT data set, 30 IVCM images collected from different normal or pathological subjects are manually graded into three grades ('high', 'medium', and 'low') with respect to the corneal nerve tortuosity and each grade contains 10 images [11]. The segmentation algorithm proposed in [12] is applied on the CNT data set to trace the nerve fibers.

The existing tortuosity measurements are derived from the corresponding geometric concepts (such as length, angle and curvature). Since there is no general method to characterize all types of tortuosity, a variety of methods are usually used to estimate corneal nerve fiber tortuosity. Table 1 lists eight tortuosity measures employed in this paper. Since an IVCM image may contains multiple nerve fibers, for each tortuosity measure, the overall tortuosity value of an IVCM image is obtained by the weighted average of nerve fiber measurements with respect to fiber lengths [8]. After that, each tortuosity measure forms a feature and a FBCM is generated by the fuzzy c-means clustering [13] based on that feature.

Table 1. Employed measures of tortuosity

Name	**Description**	**Reference**
Arc-Chord Length Ratio	the ratio between curve length and the chord length	[14]
Sum of Curvature	the sum of absolute curvature of all points on the fiber nerve centerline	[15]
Sum of Squared Curvature	the sum of squared curvature of all points on the fiber nerve centerline	[15]
Maximum of Squared Curvature	the maximum of absolute curvature of all points on the fiber nerve centerline	[16]
Derivative of Curvature	for all points on the fiber nerve centerline, the squared derivative of curvature is summed	[17]
Inflection Count Metric	the number of changes in sign of the curvature times arc-chord length ratio	[18]
Tortuosity Density	averaged arc-chord length ratio of all portions between two consecutive inflection points	[19]
Slope Chain Code	averaged slope angles between two connected line segments of constant length along the curve	[20]

When calculating the similarity matrix of fuzzy clusters by Eqn (1), if a tortuosity measure is not discriminative, the generated membership degrees of instances to all clusters can be very close. For example, when five instances $\{x_1, x_2, x_3, x_4, x_5\}$ are assigned to 2 fuzzy clusters, C_1 and C_2, with respect to a tortuosity measure, the degree of them belonging to C_1 and C_2 can be $(0.49,0.48,0.51,0.49,0.52)$ and $(0.51,0.52,0.49,0.51,0.48)$, respectively. Since the two clusters are very similar, they are very likely to be considered as the first found dominant set when applying the peel-off strategy. However, such two clusters does not actually provide valuable information for the discrimination of tortuosity. In order to discards such unwanted fuzzy clusters which are similar but ambiguous in

ISSN 2616-5775 Vol. 3, Issue 5: 23-30, DOI: 10.25236/AJCIS.030504

the clustering ensemble, a quantitative index is employed to filter them out from the fuzzy clusters created by FBCMs.

Given that $C_k^m(x) \in [0,1]$ represents the degree of instance x belongs to the fuzzy cluster C_k^m in the FBCM π_m, then the significance index of the fuzzy cluster C_k^m is defined as:

$$sig(C_k^m) = \vee_{x \in X} C_k^m(x) - \wedge_{x \in X} C_k^m(x) - \alpha,$$

where K_m is the number of fuzzy sets in the FBCM π_m, α is a parameter with a theoretical value range of $[0.0, 1.0)$. The fuzzy clusters whose significance index value of is less than or equal to 0 is omitted from the set $\{C_1, \cdots, C_n\}$, before the dominant set based consensus. The larger the value of α is, the more fuzzy sets tend to be omitted, and vice versa.

To sum up, the dominant set based fuzzy clustering ensemble for corneal nerve fiber tortuosity analysis is as follows: 1) segment the nerve fibers from the IVCM images; 2) generate FBCM from each tortuosity measure; 3) filter out unwanted fuzzy clusters according to the significance index; 4) generate similarity matrix of fuzzy clusters; 5) peel off dominant sets and form final fuzzy cluster.

4. Experimental results

The number of dominant sets peeled off is set to 3, and the number of fuzzy clusters K_m in each BCM is set to an integer from 2 to 6. The parameter α in the calculation of the significance is set to 0.0, 0.1, 0.2, 0.3 and 0.4. Fuzzy c-means and $\mathbf{z}^{(0)}$in Equation (3) are initialized randomly in each run, but the ensemble results are compared based on the same BCMs. The experimental results of accuracy are shown in Figure 1. Each point in the figure is the average of 30 runs. It can be seen from Figure 1 that in the test data set under the same conditions, the dominant set based clustering ensemble with cluster filtering process can obtain higher accuracy than it without filtering. However, if the value of$\alpha$is set too high, too many clusters will be omitted, which will result in a decrease in the accuracy of cluster ensemble.

The performance of the proposed method is also evaluated based on the weighted accuracy (wAcc), sensitivity (wSe), specificity (wSp), positive predicted value (wPpv), and negative predicted value (wNpv) [8] and the comparison to the classic k-means and fuzzy c-means is shown in Table 2. Each value in Table 2 is based on the average of 30 runs with random initialization. It can be clearly seen from Table 2 that proposed method outperforms the k-means and fuzzy c-means in the task of corneal nerve tortuosity grading.

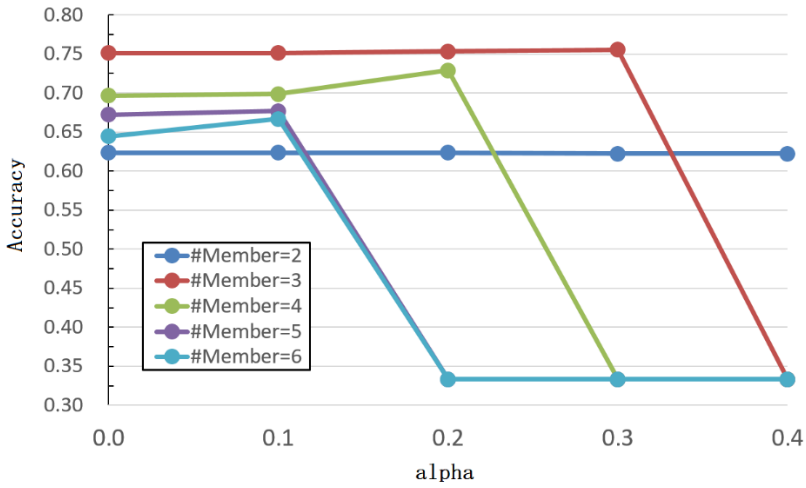

Figure 1. Accuracy of the proposed method with different parameters

Table 2. Corneal nerve tortuosity grading performance evaluation (%)

	Acc	**wAcc**	**wSe**	**wSp**	**wPpv**	**wNpv**
k-means	56.31	70.50	58.32	50.31	44.51	70.23
Fuzzy C-means	60.53	73.69	62.00	58.06	50.96	80.82
Proposed	75.11	83.36	75.03	87.60	77.65	88.80

5. Conclusion and future work

This paper proposes a pipeline for the assessment of curvilinear structure tortuosity based on fuzzy clustering ensemble where the concept of dominant set is adopted to implement consensus function. The proposed work is verified on a real-world data set with superior results achieved over classic clustering algorithms. Whist promising, the proposed method could be extended to deal with a broader range of unsupervised learning tasks and also opens up an avenue for further investigation of unsupervised methods in automated tortuosity grading with multiple measures..

References

[1] Yu Z, Zhu X, Wong H S, You J, Zhang J, Han G. Distribution-based cluster structure selection. IEEE Transactions on Cybernetics. 2017; 47(11): 3554–3567.
[2] Diao R, Chao F, Peng T, Snooke N, Shen Q. Feature selection inspired classifier ensemble reduction. IEEE Transactions on Cybernetics. 2014; 44(8): 1259–1268.
[3] Sevillano X, Al´ıas F, Socor´o J C. Positional and confidence voting-based

consensus functions for fuzzy cluster ensembles. Fuzzy Sets and Systems. 2012; 193: 1–32, .
[4] Yu Z, Li L, Wong H S, You J, Han G, Gao Y, Yu G. Probabilistic cluster structure ensemble. Information Sciences. 2014; 267: 16–34.
[5] Yan D, Chen A, Jordan M I. Cluster forests. Computational Statistics & Data Analysis. 2013; 66: 178–192.
[6] Iam-On N, Boongoen T, Garrett S, Price C. A link-based approach to the cluster ensemble problem. Pattern Analysis and Machine Intelligence. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2011; 33(12): 2396–2409.
[7] Pavan M, M Pelillo. Dominant sets and pairwise clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2007; 29(1): 167–172.
[8] Annunziata R, Kheirkhah A, Aggarwal S, Hamrah P, Trucco E. A fully automated tortuosity quantification system with application to corneal nerve fibres in confocal microscopy images. Medical Image Analysis. 2016; 32: 216-232.
[9] Su P, Shang C, Chen T, Shen Q. Exploiting data reliability and fuzzy clustering for journal ranking. IEEE Transactions on Fuzzy Systems. 2017; 25(5): 1306–1319.
[10] Oliveira-Soto L, Efron N. Morphology of corneal nerves using confocal microscopy. Cornea. 2001; 20(4): 374-84.
[11] Fabio S, Zheng Xiaodong, Yuichi O, Alfredo R. Automatic evaluation of corneal nerve tortuosity in images from in vivo confocal microscopy. Investigative Ophthalmology & Visual Science. 2011; 52(9): 6404.
[12] Zhao Y, Rada L, Chen K, Harding S P, Zheng Y. Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images. IEEE transactions on medical imaging. 2015; 34(9): 1797–1807.
[13] Bezdek, C James. Pattern recognition with fuzzy objective function algorithms. 1981; 22(1171): 203–239.
[14] Lotmar W, Freiburghaus A, Bracher D. Measurement of vessel tortuosity on fundus photographs. Albrecht Von Graefes Archiv Fr Klinische Und Experimentelle Ophthalmologie. 1979; 211: 49-57.
[15] Hart W E, Goldbaum M H, Kube P, Nelson M. Measurement and classification of retinal vascular tortuosity. International journal of medical informatics. 1999; 53(2-3): 239-52.
[16] Annunziata R, Kheirkhah A, Aggarwal S, Cavalcanti B M, Hamrah P, Trucco E. Two-dimensional plane for multi-scale quantification of corneal subbasal nerve tortuosity. Investigative ophthalmology & visual science. 2016; 57: 1132-1139.
[17] Patasius D J M, Marozas V, Lukosevicius A. Evaluation of tortuosity of eye blood vessels using the integral of square of derivative of curvature. In Proceedings Eur. Med. Biol. Eng. Conf. (EMBEC05); 2005.
[18] Bullitt E, Gerig G, Pizer S M, Lin W, Aylward S R. Measuring tortuosity of the intracerebral vasculature from MRA images. IEEE Transactions on Medical Imaging. 2003; 22: 1163-1171.
[19] Scarpa F, Zheng X, Ohashi Y, Ruggeri A. Automatic evaluation of corneal nerve tortuosity in images from in vivo confocal microscopy. Investigative

ISSN 2616-5775 Vol. 3, Issue 5: 23-30, DOI: 10.25236/AJCIS.030504

ophthalmology visual science. 2011; 52(9): 6404-6408.
[20] Bribiesca E. A measure of tortuosity based on chain coding. Pattern Recognition. 2013; 46(3): 716-724.