Fuzzy Self-Organizing Incremental Neural Network for Fuzzy Clustering

Tianyue Zhang, Baile Xu, and Furao Shen(✉)

National Key Laboratory for Novel Software Technology,
Department of Computer Science and Technology,
Collaborative Innovation Center of Novel Software Technology and Industrialization,
Nanjing University, Nanjing, China
njucszty@gmail.com, dg1633021@smail.nju.edu.cn, frshen@nju.edu.cn

Abstract. In this paper, a neural network named fuzzy self-organizing incremental neural network (fuzzy SOINN) is presented for fuzzy clustering with following four characteristics: fuzzy, incremental learning, topological representation and resistance to noise. No predefined structures of clusters is required due to the self-adjusting nodes and edges which fit the learning data incrementally. A removal of nodes and edges promises the robustness of the network to the noisy data. Experiments on artificial and real-world data prove the validity of the clustering method.

Keywords: Fuzzy clustering · Incremental or online learning · Topological representation · Self-organizing incremental neural network (SOINN)

1 Introduction

Clustering is assigning objects to clusters which have higher similarity in the same cluster and dissimilarity between the different clusters with an extensive application in diverse research fields [1]. However, the issue of bridges (or overlaps) between clusters is often encountered in the procedure of clustering according to Nagy [2] which is hard to be solved by hard clustering. The idea that assigning patterns with grades of membership rather than clustering them into disjoint clusters, which was introduced and termed fuzzy set by Zadeh [3], exploited into clustering and termed fuzzy clustering by Ruspini [4], is more appropriate for dealing with the issue. When data clusters have vague boundaries and the fuzzy characteristic of data structures needs to be reserved, fuzzy clustering methods work well. Applications of fuzzy clustering to problems of clustering, feature selection, and classifier design have been reported in biology, medicine, psychology, economics, and many other disciplines.

Previous researches on fuzzy clustering leads to the most popular classical algorithm: fuzzy C-means (FCM, also termed fuzzy K-means) [5] with frequent applications in clustering for its simplicity and computational efficiency. Though classical batch algorithm, FCM has developed its on-line version. For very large data sets, first online fuzzy c-means is presented in [6] and two online fuzzy

© Springer International Publishing AG 2017
D. Liu et al. (Eds.): ICONIP 2017, Part I, LNCS 10634, pp. 24–32, 2017.
https://doi.org/10.1007/978-3-319-70087-8_3

c-medoid based clustering algorithms are presented in this paper [7]. James C. Bezdek presented three new incremental kernel FCM [8] and recommended using the rseKFCM at the highest sample rate. A kernel fuzzy c-means (KFCM) is also proposed in this paper, dealing with linearly non-separable clusters. However, the requisite priori knowledge of cluster structures and inadequate ability to learn incrementally are still the shortcomings of FCM.

Fuzzy self-organizing incremental neural network has the following four characteristics which are not be realized in any one method altogether. It's based on the work of SOINN [9]:

Fuzzy clustering. Assigning input patterns fuzzy membership grades decided by Gaussian membership function and similarity threshold reflects not the probability of the belonging but the degree. Take the human height for an example, a height of 180 cm may be assumed as 'half tall' and 190 cm as 'completely tall', the 'half' and 'completely' are the grades of membership.

Incremental learning. Compared with traditional batch learning, on-line learning processes data in the sequence they come without occupying the memory. Moreover, incremental learning enables the system to learn information from the new arriving patterns without omitting or corrupting old knowledge whether the new patterns belong to the clusters already learned. Although on-line and incremental learning may encounter an issue termed stability-plastic dilemma, the learning rate capable of guaranteeing convergence keeps the process stable and the plasticity are realized by adding new nodes to the network and removing old nodes.

Topological representation. Connections between learned patterns are designed to preserve the topological structures of original clusters. With the ability of incremental learning, fuzzy SOINN needs no prior knowledge of the structures of clusters.

Resistance to noise. Due to noisy data in data sets, the removal of those nodes which have less neighbors than others obviously improves the performance of the network on data sets with slight noises.

2 The Proposed Approach

2.1 Overview of Fuzzy SOINN

Fuzzy SOINN inherits original SOINN and develops a simplified and fuzzy version which aims to realize the goals mentioned above. Unlike most neural network, it learns without predefined nodes and edges in the net and learning objective function. The model of fuzzy SOINN can be described as $< \{n_i\}, \{e_i\} >$, or $< N, E >$, while N is the neuron set and E is edge set. each node n_i stores the mean coordinates and variance learned by the learning process to represent a hyper-ellipsoid region, which can be denoted as $< c_i, \sigma_i, acc_i >$: $c_i(c_{i1}, c_{i2}, ..., c_{in})$, $\sigma_i(\sigma_{i1}, \sigma_{i2}, ..., \sigma_{in})$ and acc_i are the mean vector, variance of the node region and accumulated active times, and n is the total dimensions

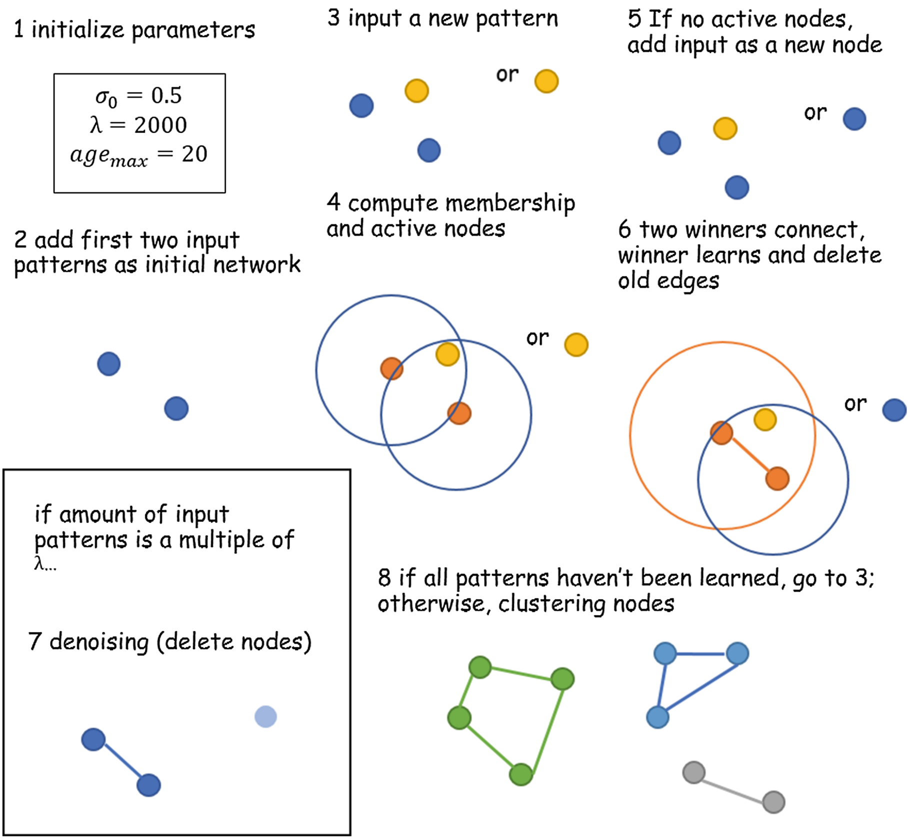


Fig. 1. Overview of fuzzy SOINN.

of n_i. According to Hebbian rule [10], winner and second winner fire together so an edge formed between them to set up the topological representation. Though all active nodes noticed, only winner and his neighbors learn from the input patterns. The learning process is shown in Fig. 1, and in the following patterns, complete algorithm is given.

2.2 Complete Algorithm of Fuzzy SOINN

Combining all the steps mentioned before, we give a complete description of the whole algorithm in Algorithm 1.

2.3 Analyses of Algorithm

Membership Computation. The most significant feature of our network is fuzzy clustering. The computation of membership, or the measure of similarity, which is a critical step, is reflected on the selection of membership function.

Algorithm 1. the complete algorithm of fuzzy SOINN

Input: The set of input patterns $\{x\}$, age_{max}, λ, σ_0.

Output: The neuron set N and edge set E. If needed, the membership matrix of input patterns $M = [m_{ic}]$, i is the number of node, c is the number of cluster.

1: Initialize $N = \phi$, $E = \phi$.

2: Input new pattern x. If it's the first two patterns, directly add it to N according to formula 3.

3: Compute the grade of membership of node n_i according to

$$r_i(x) = exp(-\sum_j \frac{(x_j - c_{ij})^2}{2\sigma_{ij}^2}) \tag{1}$$

j means every dimension of patterns. All grades of membership should be normalized that they sum to 1. Get active nodes set N_{active} according to the threshold

$$N_{active} = \{n_i | r_i(x) > r_i(c_i + \sqrt{2}\sigma_i)\} \tag{2}$$

If active node set is empty, add input pattern to N according to:

$$\begin{aligned} n_{new} :< c_{new} &= x, \\ \sigma_{new} &= \begin{cases} \sigma_0, & |N| = 0 \\ c_{nearest} - x, & |N| > 0\,, \end{cases} \\ acc_i &= 0 > \end{aligned} \tag{3}$$

σ_0 is a small initial parameter, and $c_{nearest}$ is the coordinates of the nearest node to x in Euclidean distance. And go back to step 2.

4: Choose winner and second winner according to (second winner is the winner picked without the first winner)

$$n_{winner} = \underset{n_i \in N}{\operatorname{argmax}}\, r_i(x) \tag{4}$$

connect winner and second winner (if second winner exists, and if the edge between the two nodes exists, reset its age to 0). The winner and its neighbors $\{n_{neighbor}\}$learn as

$$\begin{aligned} acc^*_{winner} &= acc_{winner} + 1 \\ c^*_{winner} &= c_{winner} + \frac{1}{acc^*_{winner} + 1}(x - c_{winner}) \\ \sigma^*_{winner} &= \sqrt{\sigma^2_{winner} + \frac{1}{acc^*_{winner} + 1}(x - c_{winner})^2} \\ c^*_{neighbor} &= c_{neighbor} + \frac{r_{neighbor}(x)}{acc^*_{winner} + 1}(x - c_{neighbor}) \end{aligned} \tag{5}$$

5: Delete the edge of which age is larger than age_{max}. If the input number is a multiple of λ, denoise as the Sect. 2.3.

6: If all patterns have been learned, then turn to the clustering step as Sect. 2.3, else go back to step 2 to get a new input pattern.

7: If the membership of all nodes is needed, assign membership to patterns as Sect. 2.3.

8: **return** N and E or M.

The most widely influential method, Fuzzy C-means, selects the mean position (or point) of every cluster to represent a cluster, and the grade of membership of every pattern is determined by the distance, namely Euclidean distance or other distance from the point to the centers, which is suitable to describe spherical distribution. Due to the topological representation feature of SOINN, fuzzy SOINN use nodes connected to describe a cluster, which is more appropriate to describe a non-spherical distribution, namely irregular shape.

Node Learning and Edge Connection. When input patterns which brings new information of distributions of clusters are learned, new neuron is supposed to be added to network incrementally; otherwise, the input pattern helps old neurons to describe the distribution more precisely. Based on the Hebbian rule [10], we connect the winner and second winner if the second winner exists.

Denoising and Clustering. Noise often exists in the real-world data sets, the method should be robust enough to detect them. The accumulated active times of each node classify noisy and normal data based on the assumption that the node density around noisy data is lower than the normal data and more neighbors means more active times. Thus, after a period of learning (predefined by λ), we delete nodes of which accumulated active times is lower than most other nodes, more specifically described as:

$$N_{noise} : \{n_i | acc_i < c * \frac{\sum_j^{|N|} acc_j}{|N|}\} \tag{6}$$

c is a parameter between 0 and 1, and assigned by user. Large c means more nodes will be removed as noise, and small c means less noisy data in the data set.

The edge connection aims to preserve the topological representation of nodes, and the clustering result are also computed on it. For edges will be deleted when the two nodes don't often fire together, we assume that the two nodes connected by an edge belong to one cluster. Therefore, when the learning process stops, the label assigned to the nodes of the same connected subgraph is also the same.

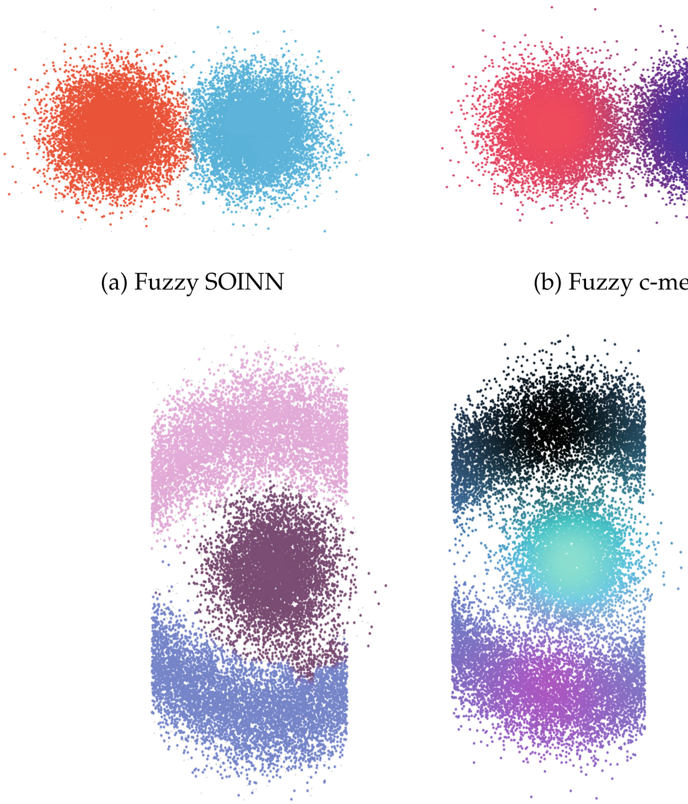
The grades of input patterns of all clusters can be calculated by calculating grades of membership of all nodes according to the formula 1, selecting the largest for every cluster, and normalized them to 1.

3 Experiments

In this section, experiments on artificial and real-world data sets are conducted to show the performance of our approach. Artificial data sets are designed to illustrate the result and process of our method, and the real-world data sets are selected to evaluate the performance of fuzzy SOINN with the comparison with other classical and state-of-art algorithms.

3.1 Artificial Data

Artificial data is designed to show the fuzzy result of fuzzy SOINN directly in figures and compare it with fuzzy c-means. 20000 patterns are randomly generated from two distributions, one of which is linearly separable and the other is not. The clustering result is shown in Fig. 2. Two colors represent different clusters, and in the overlapped region, the gradient colors illustrate the change of input pattern membership of different clusters.



(a) Fuzzy SOINN (b) Fuzzy c-means

(c) Fuzzy SOINN (d) Fuzzy c-means

Fig. 2. The fuzzy results, input patterns are all colored and one color represents one cluster.

The grade of fuzziness of fuzzy SOINN seems less than fuzzy c-means since the color is more "pure" on the whole and only in the overlap region the color is gradient. Due to the multiple neuron nodes which represent one cluster, only the input patterns in the overlap regions have similar grades of membership of two adjacent clusters. The purity of colors in non-spherical distribution which is clustered by fuzzy c-means means that only centers of clusters is insufficient to describe a cluster, especially for clusters which are distributed not so regular. To sum up, the topological-preserving neurons decrease the fuzziness of fuzzy sets of input patterns, but they can describe a cluster more precisely than only centers of clusters.

3.2 Real-World Data

For the characteristics of fuzzy SOINN is fuzzy learning and incremental learning, several related algorithms are picked up for comparing on the real-world data sets, including Fuzzy c-means (FCM) [5], online fuzzy c-means (OFCM) [6] and the original algorithms of SOINN family, namely one-layer SOINN [9], Enhanced SOINN [11]. Six data sets are picked up in this experiment, including iris, glass, Letter, segmentation, wine and isolet, which are taken from UCI Machine Learning Repository, and image data sets, USPS and COIL-100. They are summarized in Table 1:

Table 1. Real-word data sets for test clustering

Data set	Class	Dimension	Instance amount
Iris	3	4	150
Glass	6	10	214
Segment	7	19	2310
Wine	3	13	178
Letter	26	16	20000
USPS	10	256	9298
Isolet	26	617	7797
COIL	100	1024	7200

Normalized Mutual Information (NMI) [12] is taken as the criteria of clustering results, which is able to evaluate the agreement of ground truth and clustering results without the influence of the number of clusters. It's always a number between 0 and 1, and a higher number shows a better result. The parameters of fuzzy SOINN are set as σ_0= 0.1 for letter, isolet and COIL, 1.0 for segmentation and 0.5 for the rest, age_{max} of edges is set to 20, and λ is set according to the size of data sets, namely $\lambda = 1000$ for USPS, letter, isolet and COIL, 100 for iris and wine, 200 for the rest. The final NMI results of fuzzy SOINN are concerned with the input order of data, so we conduct 10 times on random input order and record the average NMI. Results of SOINN and ESOINN is taken from [13]. The result is shown in the Table 2:

Table 2. NMI on real-world data sets compared with other method. The best result is in bold, and the first four results are conducted by FCM and the rest by OFCM

Data set	Iris	Glass	Segment	Wine	Letter	USPS	Isolet	COIL
FCM/OFCM	**0.750**/-	0.360/-	0.515/-	0.417/-	-/0.234	-/0.197	-/0.356	-/0.368
SOINN	0.554	**0.584**	0.382	0.276	0.229	0.507	0.635	**0.607**
ESOINN	0.633	0.515	0406	0.260	**0.376**	**0.607**	0.662	0.513
Fuzzy SOINN	0.701	0.425	**0.597**	**0.456**	0.291	0.325	**0.675**	0.522

We can see fuzzy SOINN outperforms other methods on some data sets like Segment, Wine and isolet (better than FCM/OFCM on most data sets), while performs not so well on other data sets. Without considering the problem of curse of dimensionality, the bad performance on USPS compared with other SOINN methods reveals the difficulty in dealing with data in high dimensions. Since the membership function is a power function and decreases sharply when the node is apart from the mean vector, two nodes of the same class are hard to be active in the same time and hundreds of classed will be generated, which results in a unsatisfied NMI result. It points out that the following research on fuzzy SOINN is supposed to focus on the problem of high dimensions. It also shows an improved FCM-based algorithm should be used for complex structures of data sets such as kernel-FCM.

4 Conclusion

Although the widespread application of methods based on fuzzy c-means has proved it's success, some limits, such as the requirement of prior knowledge of class amount, non-incremental learning (online fuzzy c-means only process online samples without the ability of incremental learning), sensitive to noise, are all improved by fuzzy SOINN. With the addition of new nodes and edges, movement and removal of old, fuzzy SOINN can learn new knowledge infinitely without casting off the topological structure the network has already learned. The experiment results guarantee the performance of fuzzy SOINN in most environments. Though not performing outstandingly on some data sets, the fuzzy SOINN with the ability to fuzzy and incremental learning will play a role in some specific applications.

Acknowledgments. This work is supported in part by the National Science Foundation of China under Grant Nos. (61373130, 61375064, 61373001), and Jiangsu NSF grant (BK20141319).

References

1. Xu, R., Wunsch, D.: Survey of clustering algorithms. IEEE Trans. Neural Networks **16**(3), 645–678 (2005)
2. Nagy, G.: State of the art in pattern recognition. Proc. IEEE **56**(5), 836–862 (1968)
3. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**(3), 338–353 (1965)
4. Ruspini, E.H.: A new approach to clustering. Inf. Control **15**(1), 22–32 (1969)
5. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Springer Science & Business Media, New York (2013)
6. Hore, P., Hall, L.O., Goldgof, D.B., Cheng, W.: Online fuzzy C means. In: Annual Meeting of the North American Fuzzy Information Processing Society, NAFIPS 2008, pp. 1–5. IEEE (2008)
7. Labroche, N.: Online fuzzy medoid based clustering algorithms. Neurocomputing **126**, 141–150 (2014)
8. Havens, T.C., Bezdek, J.C., Palaniswami, M.: Incremental kernel fuzzy c-means. In: Madani, K., Dourado Correia, A., Rosa, A., Filipe, J. (eds.) Computational Intelligence. Studies in Computational Intelligence, vol. 399, pp. 3–18. Springer, Heidelberg (2012). doi:10.1007/978-3-642-27534-0_1
9. Shen, F., Hasegawa, O.: An incremental network for on-line unsupervised classification and topology learning. Neural Netw. **19**(1), 90–106 (2006)
10. Martinetz, T., Schulten, K.: Topology representing networks. Neural Netw. **7**(3), 507–552 (1994)
11. Shen, F., Ogurab, T., Hasegawa, O.: An enhanced self-organizing incremental neural network for online unsupervised learning. Neural Netw. **20**(8), 893–903 (2007)
12. Manning, C.D., Raghavan, P., Schtze, H.: Introduction to Information Retrieval, vol. 1, no. 1, p. 496. Cambridge University Press, Cambridge (2008)
13. Xing, Y., Shi, X., Shen, F., Zhou, K., Zhao, J.: A self- organizing incremental neural network based on local distribution learning. Neural Netw. **84**, 143–160 (2016)