Advances in theory and applications of fuzzy clustering

GAO Xinbo[1] & XIE Weixin[2]

1. Institute of Electronic Engineering, Xidian University, Xi'an 710071, China;
2. President office, Shenzhen University, Shenzhen 518060, China.
Correspondence should be addressed to Gao Xinbo (e-mail: xbgao@263.net)

Abstract The summarization and evaluation of the advances in fuzzy clustering theory are made in the aspects including the criterion functions, algorithm implementations, validity measurements and applications. Several important directions for a further study and the application prospects are also pointed out.

Keywords: cluster analysis, fuzzy clustering, cluster validity, pattern recognition, image processing.

Clustering is such a procedure that objects are distinguished or classified in accordance with their similarity. During this course there is no teacher to provide guidance, hence it is also called unsupervised classification. And cluster analysis is to study or cope with unsupervised classification of given objects with mathematical methods. As the old saying goes, things of one kind come together[1]. So, clustering is not a new question, and it will be uninterruptedly deepened with the development of human society.

Most traditional cluster analysis methods are crisp partitioning, in which every given object is strictly classified into a certain group. Hereby, the boundaries among classes are sharply in such partition. However, in practice, the class attributes of most objects are not strict but ambiguous, hence it is suitable for soft partitioning. Fortunately, the fuzzy sets theory proposed by Zadeh[2] provides a powerful tool for such soft partitioning. Thus, people began to deal with clustering with fuzzy fashion and named them fuzzy cluster analysis. Since fuzzy clustering obtains the degree of uncertainty of samples belonging to each class and expresses the intermediate property of their memberships, it can more objectively reflect the real world. Thereby, it has become the main content of studies on cluster analysis.

It was Ruspini who proposed the concept of fuzzy partition firstly[3]. With this concept, some typical fuzzy clustering algorithms, such as methods based on the similarity and fuzzy relations[4], the transitive closed package of fuzzy equivalent relation[5], the convex decomposition of data[6], or dynamic programming and indistinguishable relation are developed one after the other. Unfortunately, these methods are unsuitable for the case with large amount data and difficult to meet the requirements of real time realization. Since their practical applications are not very wide-ranged, studies on them are reduced progressively. So far, the objective-function-based method is very popular for its simple designing and wide uses, and for its easy convertion into optimization problems and its implementation with computer. With the development of computer, objective-function-based method has become a studying hotpot in cluster analysis.

Hereinafter, the summarization of the advances in objective-function-based fuzzy clustering is made in the aspects of evolution of objective function, in the ways of algorithm implementation, styles of validity measurements and in practical applications. The systematic summarization about the traditional cluster analysis and other types of fuzzy clustering methods can be referred to refs. [1, 7—10].

1 The evolution of objective function in fuzzy clustering

In mathematical terms, the fuzzy clustering problem can be expressed as classifying a set of given patterns, $O=\{o_1,o_2,\cdots,o_n\}$, into c fuzzy subsets (clusters) $S_1,S_2,\cdots,S_c$. Let m_{ik} ($1\leqslant i\leqslant c$, $1\leqslant k\leqslant n$) denote the membership degree of pattern o_k in the fuzzy subset S_i, then one obtains the fuzzy c-partition of the given patterns, $U=\{m_{ik}\,|\,1\leqslant i\leqslant c, 1\leqslant k\leqslant n\}$. Such an operation as partitioning these unlabeled patterns is called fuzzy cluster analysis. To obtain reasonable classification results, the partitioning criterion, i.e. the criterion of similarity or dissimilarity, $D(\cdot)$ needs to be defined firstly.

Provided that every fuzzy subset, $S_i\,(1\leqslant i\leqslant c)$ has a cluster prototype p_i, the similarity between pattern o_k and subset S_i can be measured by the degree of distortion of this pattern from the prototype p_i, $d_{ik}=D(o_k,p_i)$.

By the distance between the observed value of pattern set O, $X=\{x_1,x_2,\cdots,x_n\}\subset R^s$, and the feature values of the clustering prototype, $B=\{b_i,1\leqslant i\leqslant c\}$, one can easily construct an objective function for the fuzzy clustering problem as eq. (1). Objective-function-based fuzzy clustering algorithm will obtain the optimal fuzzy c-partition by solving such a nonlinear programming problem with constraints.

$$\min_{(U^*,B^*)}\left\{J_m=\sum_{i=1}^{c}\sum_{k=1}^{n}(m_{ik})^m\cdot D(x_k,b_i)+z,\quad s.t.\ f(m_{ik})\in C\right\},\tag{1}$$

where z is a penalty item, $f(m_{ik})\in C$ the constraint condition and m the fuzzy weighting exponent. Obviously, this objective function is determined completely by parameter sets $\{U,D(\cdot),B,m,X\}$. Corresponding to these parameters, the evolution of the fuzzy clustering objective function will be summarized from the following five aspects.

(i) Studies on fuzzy partition matrix U. As mentioned above, most traditional cluster analysis methods are hard partitioning, in which $m_i(x_k)\in\{0,1\}$ is the indicator function of the membership of sample x_k, and class-label vector, $m(x_k)=(m_{1k},m_{2k},\cdots,m_{ck})^T$ is the base vector of the Euclidean c-space. To express the similarity information between patterns, Ruspini et al.[3] introduced the concept of fuzzy partition into cluster analysis and generalized $m(x_k)$ from {0,1} to [0,1], which extended label vector $m(x_k)$ to a hyperplane, $\sum_{i=1}^{c}m_i(x_k)=1$, in the Euclidean c-space. Hereby, the label vector can be called a fuzzy or probabilistic label. However, the probabilistic constraint makes the membership function represent only the degrees of sharing of patterns across fuzzy classes but not the typicality. To this end, Krishnapuram et al.[11] proposed possibilistic c-partition by relaxing the probabilistic constraint $\sum_{i=1}^{c}m_i(x_k)=1$. In this way, label vector $m(x_k)$ becomes a unit hypercube without origin. Such a possibilistic clustering algorithm has good noise-proof performance but poor convergence property. Moreover, it is easy to trap into the local optima and results in bad classification. In order to combine the advantages of crisp and fuzzy clustering, Selim et al.[12] presented the concept of semi-fuzzy partitioning only by keeping the fuzzier elements and defuzzifying the other elements in the partitioning matrix. Such a matrix obtained not only has certain distinctness but also keeps the fuzziness of samples distribution. So, the correctness and the convergence property of classification are improved in this way. Later, Kamel et al.[13] and Pei Jihong et al.[14] developed modified version of semi-fuzzy partition from different viewpoints respectively, i.e. threshold fuzzy clustering algorithm and sectional set soft clustering algorithm. The soft partitions above are compared in table 1.

Table 1 Comparison of four space partitioning concepts

Item	Possibilistic	Fuzzy	Traditional (Crisp)	Semi-fuzzy
Label vector sets	$N_{pc}=[0,1]^c-\bar{0}$ $\bar{0}=\{(0,0,\cdots,0)^T\}$	$N_{fc}=:\left\{m_i\in N_{pc}\ \sum_i^c m_i=1\right\}$	$N_{hc}=\{m_i\in N_{fc}: m_i\in\{0,1\},\forall i\}$	$N_{sc}=N_{fc}\cup N_{hc}$
Physical means	typicality of samples to each class	degree of sharing of samples across classes	indicator function of samples membership	either fuzzy or crisp case
Convergence	slower	slow	fast	faster
Sensitivity	higher	low	high	low
Noise-proof	better	good	poor	good

How to improve the convergent speed and decrease the sensitivity to initialization of possibilistic partitioning is still an important task for a further study on cluster analysis from the viewpoint of fuzzy partition. If a breakthrough were made in this aspect, a novel space partitioning method with good noise-proof performance and good convergence property would be obtained, which will not only enrich the theory of the available soft clustering but also shorten their pragmatizing course.

(ii) Studies on the criterion of similarity $D(\cdot)$. It is impossible for a single clustering criterion to solve all the possible problems of unsupervised classification. Therefore, many functions of similarity, such as maximum likelihood function[15], maximum entropy criterion[16], minimum volume criterion[17] and information criterion[18] are proposed. Among the available clustering criterions, the within-group sum of squared error (WGSSE) function is a very popular one.

The classical WGSSE function was originally designed to define the traditional hard *c*-means and ISODATA algorithms. With the emergence of fuzzy sets theory, Dunn[19] firstly generalized WGSSE to square weighting WGSSE function. Later, Bezdek[20] extended it to an infinite family of criterion functions which formed a universal clustering objective function of fuzzy *c*-means (FCM) type algorithms with eq. (1). The studies on criterion functions have mainly been focused on the measurements of similarity or distortion $D(\cdot)$, which are often expressed by the distances between the samples and prototypes. Different distance measurements are used to detect various structural subsets. The distance functions in common use are shown in table 2.

Table 2 Distance functions in common use

Name	Distance functions	Character and purpose
Minkowski	$D_p(a,b)=\left[\sum_{i=1}^{s}\lvert a_i-b_i\rvert^p\right]^{1/p}$	Including a family of distance with $1\leqslant p\leqslant\infty$, it can be used to detect hypercubes structural subsets with the shapes from ◇ to □ in feature space.
Euclidean	$D_2(a,b)=\left[\sum_{i=1}^{s}(a_i-b_i)^2\right]^{1/2}$	Being the Minkowski distance with $p = 2$, it can be used to detect hyperspherical structure with the shape of ○ in feature space.
Hamming	$D_1(a,b)=\sum_{i=1}^{s}\lvert a_i-b_i\rvert$	Being the Minkowski distance with $p = 1$, it can be used to detect hypercubes structural subsets with the shape of ◇ in feature space.
Maximum	$D_\infty(a,b)=\max_{i=1,\cdots,s}\lvert a_i-b_i\rvert$	Being the Minkowski distance with $p = \infty$, it can be used to detect hypercubes structured subsets with the shape of □ in feature space.
Mahalanobis	$D_A(a,b)=(a-b)^T A(a-b)$, A is a positive definitive matrix	It can be used to detect hyperellipsoidal structural subsets in feature space.

Bobrowski et al.[21] discussed fuzzy *c*-means algorithms with L_1 (D_1 in this review) and $L_\infty(D_\infty)$ norms which are the two extreme cases in the class of Minkowski distances. They found that, in some cases, better results could be obtained by using the two norms than the usual Euclidean norm, $L_2(D_2)$. Their work suggests that different distance functions should be explored in cluster analysis. In addition, weighting Euclidean distance, a special case of Mahalanobis distance with A being a diagonal matrix, is also widely used in the cases in which the different contributions of each feature of pattern to classification are emphasized[22].

Searching a certain structure from given data sets can be looked as finding an appropriate distance function. It leaves us such a question of what is the criterion for selecting an appropriate distance function. Furthermore, whether or not to develop a fuzzy clustering algorithm does not depend on the distance measurement defined in advance. Few of the existing references involve such questions. So, it needs further efforts.

(iii) Studies on clustering prototype B. An objective-function-based fuzzy clustering is also called a prototype-based method since the construction of objective function depends on the definition of prototypes. Hereby, the types of prototypes should be specified in advance. Studies on prototypes were carried out with the development and requirements of clustering applications. Originally, cluster analysis is only applied to detecting hyperspherical structures from a given data set, so prototypes are "points" in feature space and are also called clustering centroids[20]. In order to detect nonhyperspherical structural clusters, Bezdek et al.[23] defined $r(0 \leqslant r \leqslant p-1)$ dimensional linear cluster prototypes $B_r(v;\{s_i\}) = \{v\} + \mathrm{Span}(\{s_i\})$ through the point $v \in \mathbb{R}^p$, which are shown in table 3.

Table 3 Linear cluster prototypes defined by Bezdek

Dimension	Cluster prototype	Functions and features
$r=0$	$B_0(v;\varnothing)=v$: point	detecting hyperspherical and hyper-ellipsoidal structural subsets.
$r=1$	$B_1(v;s)=L(v;s)$: line	detecting linear structural subsets
$r=2$	$B_2(v;s_1,s_2)=P(v;s_1,s_2)$: plane	detecting plane structural subsets
$2<r\geqslant p$-1	$B_{p-1}(v;\{s_i\})=HP(v;\{s_i\})$: hyperplane	detecting hyperplane structural subsets

In addition, to detect "thin shell" structural pattern subsets, Dave constructed two types of prototypes, i.e. spherical[24] and ellipsoidal[25] shells, which are applied to edge detection with better results. To meet the requirements of application, shell prototypes are generalized to rectangular[26], convex polytopes[27], and even to any shaped shell[28]. Meanwhile, line-shaped prototypes are also extended to curves, such as parabolas[29], quadratic curves and quadratic polynomials[30].

Clustering based on objective functions strongly depends on the prototypes, therefore, it demands that, on the one hand, the priori information should be used sufficiently to select appropriate prototypes; on the other hand, distance measurement should be incorporated to select reasonable similarity criterions.

(ⅳ) Studies on weighting exponent *m*. In the objective function of fuzzy clustering, $\{J_m{:}1<m<\infty\}$, Bezdek[20] introduced a weighting exponent *m* and made Dunn's function a special case (m = 2) of J_m. One may have felt that it is unnatural and unnecessary for the emergence of *m* from the viewpoint of mathematics[16]. However, since J_m is generalized from the WGSSE function, such a generalization will be invalid without weighting the membership function with *m*. Parameter *m* is also called a smooth factor, which controls the membership sharing between fuzzy clusters[20]. Therefore, it is important to select a value of *m* if one implements fuzzy clustering. Unfortunately, it lacks for theoretical basis for optimal choice of *m* at present.

Bezdek had given an empirical range of $1.1 \leqslant m \leqslant 5$. Later, he came to a conclusion from the viewpoint of physical interpretation that FCM algorithm is of specific physical significance in the case of $m=2$ [31]. Cheung and Chen[32] found that *m* should be selected in the range of 1.25—1.75 for the applications of character recognition. From the aspect of convergence property of algorithms, Bezdek et al.[33] considered that the value of *m* depends on the number of samples, says *n*, and suggested that the value of *m* should be greater than *n* (*n*-2). However, Pal et al.[34] obtained the best choice of *m* when *m* being in the interval [1.5, 2.5], and the interval midpoint, m = 2, has often been preferred for many users of FCM algorithm.

All the above ranges of *m* come from experiments or experience, which are heuristic. But it does not provide the straightforward way of optimal choice of *m*. In addition, it lacks for test approaches for the selected optimal *m*. These open problems call for further investigation to establish the theoretical basis for optimal choice of *m*.

(ⅴ) Studies on various data sets *X*. In practical applications, one often meets various types of data sets. For this reason, if one wants to construct an appropriate objective function for fuzzy clustering, he should consider the types of the given data sets firstly. Most of the common data are point-sets in feature space, $X \subset \mathbb{R}^s$. Besides this, relational data[35], directional data[36], interval and fuzzy numbers[37] and other formed data were also studied, and some significant conclusions were drawn. Another type of data sets, symbolic pattern[38] also attracted wide attention in concept clustering, which consists of the usual number valued data, interval valued data, fuzzy numbers and linguistic forms. But the most common data encountered are partially labeled data, data contaminated by noise or mixed data with multi-structural subsets.

In view of partially labeled data, Pezdrcy[39] proposed partially supervised fuzzy clustering algorithm to utilize the priori information contained in the data sets sufficiently. Bensaid[40] further developed this theory and applied it to image segmentation with better results. For the data contaminated by noise, a lot of robust clustering algorithms were presented to overcome the interference of noise. Dave et al.[41] made a systematic summarization for these robust algorithms, on which we will not give more details here. As to mixed data with multi-structural subsets, Gustafson et al.[42] proposed a clustering algorithm with a fuzzy covariance matrix, which can detect the ellipsoidal and linear structure simultaneously from the given data sets. Meanwhile, Jawahar[43] had some tries for clustering with various geometric structures. Except these, there are few reports involving this aspect.

There remain many problems on clustering for the above three types of data. To make the fuzzy clustering much more effective, one should further analyze the various cases of the practical data and develop clustering algorithms which can utilize priori information, resist the noise and detect multi-type structures simultaneously.

2 Studies on the ways for realizing fuzzy clustering

Having constructed the criterion function for clustering, we will study how to optimize the objective function and obtain optimal clustering results, i.e. studying the ways of algorithm realization. The existing ways were mainly classified into three classes, alternative optimization (AO), neural networks (NN) and evolutionary computing (EC). In what follows, we will summarize the advances of these three aspects.

(ⅰ) Realization based on AO. In the course of optimizing objective function, people had ever tried to use such methods as dynamic programming, branch and bound and convex cut. However, the great amount of storage space and CPU time limited their applications. The most widely used method is fuzzy *c*-means type algorithm, an iterative optimization approach proposed by Dunn[19] and Bezdek[20]. So far, studies in this aspect mainly focus on the proof of convergence property, the choice of stop criterions and on the initialization of the clustering prototypes.

By many times modification, the AO algorithm was proved to converge along a sub-sequence to either a local minimum or a saddle[44,45]. As to the stop criterions, two types of methods are designed[46], i.e. the differences of prototypes or partition matrices between the successive iteration are less than a pre-specific threshold. Unfortunately, since FCM type algorithms are hill climbing with local search strategy in essence, they are very easy to trap into local optima. That is, they are sensitive to initialization. To obtain global optima or satisfied solutions, people place hope on good initialization. Mountain function is one of the famous initialization methods, which was proposed by Yager et al.[47]. However, the amount of calculation increases exponentially with the dimensions of samples. To overcome its drawbacks, Chiu[48] modified this method and made the computing amount only depend on the number of samples, which solved the contradictories between the precision and complexity of computing. In addition, there are still other initialization methods, such as density function estimation[49], morphological method[50], methods with fuzzy measurement and Marr operators.

(ⅱ) Realization based on NN. The application of neural networks in cluster analysis stems from the Kohonen's learning vector quantization (LVQ), self-organization feature mapping (SOFM)[51] and Grossberg's adaptive resonance theory (ART)[52]. The main properties and features of these clustering neural networks are shown in table 4.

Table 4 Comparison of two kinds of clustering neural networks

Item	Kohonen's clustering NN	Grossberg's clustering NN
Input quantity	precise valued quantity	valued or fuzzy linguistic quantity
Output quantity	cluster prototypes (feature vector)	classification results
Learning ways	competitive learning (gradient descent)	fuzzy logic operation
Number of clusters	specified in advance	automatically determining
Utilization	spercial hard clustering, feature mapping	spercial hard clustering, pattern classification

Since neural networks are of capability in parallel processing, people hope to implement clustering at high speed with network structure. However, the classical clustering NN shown in table 4 can only implement spherical hard cluster analysis. So, people made much effort in the integrative research of fuzzy logic and neural networks, which falls into two categories as follows. The first type of studies bases on the fuzzy competitive learning algorithm, in which the methods proposed by Pal et al.[53], Xu[54], and Zhang[55] respectively are representatives of this type of clustering NN. These novel fuzzy clustering NNs have several advantages over the traditional ones. The second type of studies mainly focuses on the fuzzy logic operations, such as the fuzzy ART[56] and fuzzy Min-Max NN[57]. Unfortunately, the researches on the last type are scattered. Moreover, they are not only few in amount, but also immature in theory.

(iii) Realization based on EC. Evolutionary computing (EC) is a random search strategy with the mechanism of natural selection and group inheritance, which is constructed on the basis of biological evolution. For its performance of parallel search, it can obtain the global optima with a high probability. In addition, EC has some advantages such as simple, universal and robust. To achieve clustering results quickly and correctly, evolutionary computing was introduced to fuzzy clustering with a series of novel clustering algorithms based on EC.

This series of algorithms falls into three groups. The first group is simulation-annealing-based approach. Some of them can solve the fuzzy partition matrix U by annealing[58]; the others optimize the clustering prototype gradually[59]. However, only when the temperature decreases slowly enough can the stimulate annealing converge to the global optima. Hereby, the great CPU time limits its applications. The second group is the approach based on genetic algorithm[60] and evolutionary strategy[61], whose studies are focused on such aspects as solution encoding, construction of fitness function, designing of genetic operators and choice of operation parameters[62]. The third group, i.e. the approach based on Tabu search is only explored and tried by AL-Sultan[63], which is very initial and requires further research.

By comparing the three groups of approaches above, we obtained the results shown in table 5. We see that the three group methods have their own advantages and disadvantages, which enlighten us whether we can incorporate their advantages to develop a new clustering algorithm with high speed, high precision and insensitivity to initialization. For instance, both introducing gradient operator to EC and training NN with EC are promising study directions.

Table 5 Comparison of three ways for implementing fuzzy clustering

Item	AO-based approach	NN-based approach	EC-based approach
Search method	gradient descent	gradient descent	random search
Convergent speed	higher	high	low
Precision of computing	high	high	limited by encoding
Structure of algorithms	sequential	parallel on features	parallel on individual
Sensitivity to initialization	high	high	low

3 Studies on fuzzy cluster validity

To analyze the given data sets effectively, one should judge whether there are cluster structures, i.e. one should have a clustering tendency test at first. If there exist such structures, an appropriate algorithm should be employed to discover them. This process is called a cluster analysis. After obtaining these structures, one needs to verify the reasonableness of the result as well, i.e. cluster validity evaluation. In general, the cluster validity problem can be converted into determining the optimal number of clusters[20].

Most of the available studies on cluster validity are designed for the hard c-means and fuzzy c-means algorithms, which were reviewed by Dubes and Jain in 1980[9,64]. The existing validity function can be divided into three classes from the ways of their definitions, i.e. the functions based on fuzzy partition, geometric structure and statistic information of data respectively, which are shown in table 6.

No universal validity function for clustering is the main reason of emergence of new functions in an endless stream. Since single measurement cannot solve all the possible problems of cluster validity, the available validity functions will exist together for a long time. So, it is important to provide a guideline that helps users to choose an appropriate validity function for their application. Because most functions are designed for FCM algorithm, only few effort has been made for fuzzy line, plane and other prototypes of clustering algorithms except for fuzzy c-shell clustering proposed by Dave[72] and Krishnapuram[30]. In addition, with the development of possibilistic and semi-fuzzy clustering, the

Table 6 Comparison of three cluster validity functions

Item	Fuzzy partition based	Geometric structure based	Statistic information based
Theoretical basis	good clustering results corresponding to a more “sharp” partition of data	Every cluster should be compact and well-separated.	Optimal partition provides better statistic information
Advantages	simple, small amount of computing	related closely with the structures of the data set	related closely with the distribution of the given data set
Disadvantages	lack of direct connection with the structures of data	complex, large amount of computing	Performance depends on the consistence of the assuming and practical distribution.
Typical studies	degree of separation[20], partition entropy[20], proportion exponent[65]	partition coefficient[66], separated coefficient[6,7] Xie-Beni’s index[68]	Vogel’s PFS clustering[69], Jain’s Bootstrap method [70], entropy formed functions[71]

validity measurements need further enriching and developing for them to consummate the theory of cluster validity.

The above validity functions are mainly used to determine the reasonable number of clusters to guarantee the more effective clustering result. However, in practical applications, even if the number of clusters takes the proper value, it will not be able to obtain the true structures of data for selecting inappropriate algorithms or inappropriate parameters. All the questions urge us to find more suitable functions to supervise clustering in practical application. Such a job was initiated by Huntsbergery[73]. In order to get better effect of image segmentation, he introduced a supervised function into the FCM algorithm and obtained a great success. Later, Bcnsaid[74] modified Xie-Beni’s index[68] and proposed a new standard one. He pointed out that the study of cluster validity function should not only answer the question of the optimal number of clusters, but also supervise clustering algorithm to obtain classification sufficiently, which will better fit for the practical case.

4 Studies on the applications of fuzzy clustering

The development of fuzzy clustering theory promotes its applications. But in the meantime, the requirement of practical applications accelerates the development of the fuzzy clustering theory. With the development in theory so far, fuzzy clustering has been widely used in numerous fields with satisfied effects and considerable economic benefit. Its applications range from channel equalization in communication systems, code-book design in vector quantization coding, time-series predication, neural networks training and parameter estimation to medical diagnosis, and from weather forecast, food classification to water quality analysis. The above applications are introduced in refs. [1, 7—10, 20, 62]. Here we will not give unnecessary details. For the successful applications of fuzzy clustering in pattern recognition and image processing, we will mainly summarize the two aspects in detail.

(ⅰ) Applications in pattern recognition. The classical pattern recognition theory consists of two branches, i.e. supervised and unsupervised classification, in which unsupervised classification is corresponding to cluster analysis. The natural connection with pattern recognition makes fuzzy clustering successfully applied in this field firstly. The significant advantage of fuzzy clustering lies in its active self-learning and independent of supervisor or training samples.

One of the most important problems in pattern recognition is feature extraction. Fuzzy clustering can not only extract features from raw data directly[75], but also select the optimal feature sets or reduce the dimensionality of obtained features[76]. After extracting the features, the next step is to design classifiers. Fuzzy clustering can provide the nearest-prototype classifier[76] as well as the classifier based on fuzzy IF-THEN rules[77]. Of course, the prototype and fuzzy rule will be obtained by clustering algorithm.

In addition, in some practical applications of pattern recognition, such as character recognition, speech recognition and radar target recognition[81], fuzzy clustering is also playing an important role in the raw data domain[24,25,30] as well as in transform domains[78].

(ⅱ) Applications in image processing. Image processing is a significant component of computer vision. The subjective property of human vision makes image processing suitable for analysis with fuzzy fashion. And the shortness of training images asks unsupervised processing. Because fuzzy clustering just meets these requirements, it becomes a powerful analytic tool in image processing.

In image processing, one of the widest uses of fuzzy clustering is image segmentation. Since the image segmentation can be equivalent to the unsupervised classification of pixels, Coleman et al.[82] proposed a clustering-algorithm-based image segmentation method in 1979. Later, incorporating a series of advanced techniques, such as two-dimensional histogram[83], pyramid structure[84], wavelet analysis[85] and fractal[86], people presented many novel methods to segment gray-level images with fuzzy clustering. It is also fruitful in research of texture image segmentation[86], color image segmentation[87], sequential images segmentation[87] and segmentation of remotely sensed images[88] based on fuzzy clustering. Furthermore, fuzzy clustering is widely applied in boundary detection[24,25,30], image enhancement[89], image compression[90], curve fitting[62] and in other branches of image processing.

With the development of applications, there occur some new requirements for fuzzy clustering theory. Firstly, the fast implementation of clustering algorithms is an urgent demand in image processing. Secondly, in order to obtain better results, the fuzzy clustering needs to incorporate some new techniques. Thirdly, some priori knowledge in practical problems should be used to supervise clustering for improving the speed and performance of processing. In addition, since most of the existing fuzzy clustering algorithms are proposed for static data sets, it also needs to study the case of dynamic data sets. The above requirements call for further innovations in theory of fuzzy clustering.

5 Concluding remarks

This review systematically summarizes the advances of fuzzy clustering from four aspects, such as the evolution of objective functions, implementation of algorithms, measurements of cluster validity and practical applications. The characteristics of the existing algorithms and their application prospects are also analyzed. In addition, several important directions for further study are pointed out from different viewpoints. Studies on these directions will positively promote the further improvement in theory and development of application of fuzzy clustering.

References

1. He Qing, Advance in fuzzy clustering theory and application, Fuzzy Systems and Mathematics, 1998, 12(2): 89.
2. Zadeh, L. A., Fuzzy sets, Inf. Cont., 1965, 8: 338.
3. Ruspini, E. H., A new approach to clustering, Inf. Cont., 1969, 15: 22.
4. Tamra, S. et al., Pattern classification based on fuzzy relations, IEEE SMC, 1971, 1(1): 217.
5. Zkim, L., Fuzzy relation compositions and pattern recognition, Inf. Sci., 1996, 89: 107.
6. Wu, Z., Leathy, R., An optimal graph theoretic approach to data clustering: theory and its application to image segmentation, IEEE PAMI, 1993, 15(11): 1101.
7. Anderberg, M. R., Cluster Analysis for Applications, New York: Academic Press, 1973.
8. Ryzin, J. var, Classification and Clustering, New York: Academic Press, 1977.
9. Dubes, R. C., Jain, A. K., Algorithms for Clustering Data, NJ: Englewood Cliffs, Prentice Hall, 1988.
10. Li Xianghao et al., Fuzzy Clustering Analysis and Its Applications, Guiyang: Guizhou Press of Science and Technology, 1994.
11. Krishnapuram, R., Kill, J. M., A possibilistic approach to clustering, IEEE FS, 1993, 1(2): 98.
12. Selim, S. Z., Ismail, M. A., Soft clustering of multidimensional data: a semi-fuzzy approach, Pattern Recognition, 1984, 17(5): 559.
13. Kamel, M. S., Selim, S. Z., A threshold fuzzy c-means algorithm for semi-fuzzy clustering, Pattern Recognition, 1991, 24(9): 825.
14. Pei Jihong, Fan Jiulun, Xie Weixin, A new effective soft clustering method: sectional set fuzzy c-means clustering, Acta Electronica Sinica, 1998, 26(2): 83.
15. Trauwaert, E., Kaufman, L., Rousseeuw, P., Fuzzy clustering algorithms based on the maximum likelihood principle, Fuzzy Sets and System, 1991, 85(42): 213.
16. Li, R. P., Mukaidino M., A maximum entropy approach to fuzzy clustering, IEEE-FUZZ’95, 1995: 2227.
17. Krishnapuram, R., Kim J. W., A clustering algorithm based on minimum volume, IEEE-FUZZ’96, 1996: 1387.
18. Liaw, J. N., Kashyap R. L., A new sequential classifier using information criterion window, Pattern Recognition, 1994, 27(10): 1423.
19. Dunn, J. C., A fuzzy relative of the ISODATA process and its use in detecting compact well separated cluster, J. Cybernet, 1974, 3: 32.
20. Bezdek, J. C., Pattern Recognition with Fuzzy Objective Function Algorithms, New York: Plenum Press, 1981.
21. Bobrowski, L., Bezdek, J. C., *c*-means clustering with the l_1 and l_∞ norms, IEEE SMC, 1991, 21(3): 545.
22. Yuan, B. et al., Evolutionary fuzzy c-means clustering algorithm, IEEE-FUZZ’95, 1995: 2221.
23. Bezdek, J. C., Anderson I., An application of the *c*-varieties clustering algorithm to polygonal curve fitting, IEEE SMC, 1985, 15(5), 1985: 637.

24. Dave, R. N., Fuzzy shell clustering and applications to circle detection in digital image, Inter. J. General System, 1990, 16(4): 343.
25. Dave, R. N., Generalized fuzzy *c*-shells clustering and detection of circular and elliptical boundaries, Pattern Recognition, 1992, 25(7): 713.
26. Hoeppner, F., Fuzzy shell clustering algorithms in image processing: fuzzy *c*-rectangular and 2-rectangular shells, IEEE FS, 1997, 5(4): 599.
27. Suh, I. H., Kim, J. H., Rhee, F. C., Fuzzy clustering involving convex polytopes, IEEE-FUZZ'96, 1996: 1013.
28. Gao Xinbo et al., Template based fuzzy clustering algorithm and its fast implementation, in Proc. of ICSP'96, 1996: 1269.
29. Hathaway, R. J., Bezdek, J. C., Switching regression models and fuzzy clustering, IEEE FS, 1993, 1(3): 195.
30. Krishnapuram, R., Frigui, H., Nasraoni, O., Fuzzy and possiblistic shell clustering algorithms and their application to boundary detection and surface approximation, Part Ⅰ, Ⅱ, IEEE FS, 1995, 3(1): 29.
31. Bezdek, J. C., A physical interpretation of fuzzy ISODATA, IEEE SMC, 1976, SMC-6: 387.
32. Cheung, Y. S., Chan, K. P., Modified fuzzy ISODATA for the classification of handwritten Chinese characters, in Proc. Int. Conf. Chinese Comput., Singapore, 1986: 361.
33. Bezdek, J. C., Hathaway, R. J., Sabin, M. J., Tucker, W. T., Convergence theory for fuzzy *c*-means: counter-examples and repairs, IEEE SMC, 1987, 17(5): 873.
34. Pal, N. R., Bezdek, J. C., On cluster validity for the fuzzy *c*-means model, IEEE FS, 1995, 3(3): 370.
35. Hathaway, R. J., Davenport, J. W., Bezdek, J. C., Relational duals of the *c*-means clustering algorithms, Pattern Recognition, 1989, 22(2): 205.
36. Yang, M. S., Pan, J. A., On fuzzy clustering of directional data, Fuzzy Sets and Systems, 1997, 91(3): 319.
37. Yang, M. S., Ko, C. H., On a class of fuzzy *c*-numbers clustering problems for fuzzy data, Fuzzy Sets and Systems, 1996, 84:49.
38. Sonbaty, Y. E., Ismail, M. A., Fuzzy clustering for symbolic data, IEEE FS, 1998, 6(2): 195.
39. Pezdrcy, P., Algorithms of fuzzy clustering with partial supervision, Pattern Recognition Letters, 1985, 3: 13.
40. Bensaid, A. M., Hall, L. O., Bezdek, J. C., Clarke, L. P., Partially supervised clustering for image segmentation, Pattern Recognition, 1996, 29(5): 859.
41. Dave, R. N., Krishnapuram, R., Robust clustering methods: a unified view, IEEE FS, 1997, 5(2): 270.
42. Gustafson, E. E., Kessel, W. C., Fuzzy clustering with a fuzzy covariance matrix, in Proc. IEEE CDC, San Diego, CA, 1979: 761.
43. Jawahar, C. V., Biiswas, P. K., Ray, A. K. et al., Detection of distinct geometry: a step towards generalized fuzzy clustering, Pattern Recognition Letters, 1995, 16: 1119.
44. Bezdek, J. C., A convergence theorem for the fuzzy ISODATA clustering algorithm, IEEE PAMI, 1980, 1(2): 1.
45. Bezdek, J. C., Hathaway, R. et al., Convergence and theory for fuzzy *c*-means clustering: counter-examples and repairs, IEEE PAMI, 1987, 17(5): 873.
46. Ismail, M. A., Selim, S. A., Fuzzy *c*-means: optimality of solutions and effective termination of the algorithm, Pattern Recognition, 1986, 19(6): 481.
47. Yang, R. R., Filev, D. P., Approximate clustering via the mountain method, IEEE SMC, 1994, 24(8): 1279.
48. Chiu, S. L., Fuzzy model identification based on cluster estimation, J. Intelligent and Fuzzy Systems, 1994, 2: 267.
49. Chaudhuri, D., Chaudhuri, B. B., A novel multi-seed nonhierarchical data clustering technique, IEEE SMC, 1997, 27(5): 871.
50. Postairs, J., Zhuang, R. D., Lecocq-Botte, C. G. et al., Cluster analysis by binary morphology, IEEE PAMI, 15(2): 170.
51. Kohonen, T., Self-Organization and Associative Memory, Berlin: Springer-Verlag, 1984.
52. Grossberg, S., Nonlinear neural networks: principles, mechanisms and architectures, Neural Networks, 1988, 1: 17.
53. Pal, N. R., Bezdek, J. C., Tsao, E. C. K., Generalized clustering networks and Kohonen's self-organization scheme, IEEE NN, 1993, 4(4): 549.
54. Lei Xu, Krrzyzak, A., Oja, E. et al., Rival penalized competitive learning for clustering analysis, RBF net and curve detection, IEEE NN, 1993, 4(4): 636.
55. Zhang, D., Kamel, M., Elmasry, M. T. et al., Fuzzy clustering neural network (FCNN): competitive learning and parallel architecture, J. of Intelligent and Fuzzy Systems, 1994, 2: 289.
56. Carpenter, C. A., Grossberg, S. et al., Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps, IEEE NN, 1992, 3(5): 698.
57. Simpson, P. K., Fuzzy min-max neural networks, Part Ⅱ: clustering, IEEE FS, 1993, 1(1): 32.
58. Asultan, K. S., Selim, S., A global algorithm for the fuzzy clustering problem, Pattern Recognition, 1993, 26(9): 1357.
59. Rose, K., Gurewitz, E., Fox, G. C. et al., A deterministic annealing approach to clustering, Pattern Recognition Letters, 1990, 11: 589.
60. Buckles, B. P. et al., Fuzzy clustering with genetic search, IEEE-FUZZ'94, 1994: 46.
61. Babu, G. P., Murty, M. N., Clustering with evolution strategies, Pattern Recognition, 1994, 2(27): 321.
62. Gao Xinbo, Study of fuzzy clustering with evolutionary computing and neural networks, Thesis for Master, Xi'an: Xidian University Press, 1996.
63. AL-Sultan, K. S., Fediji, C. A., A tabu search-based algorithm for the fuzzy clustering problem, Pattern Recognition, 1997, 12(30): 12023.
64. Dubes, R. C., Jain, A. K., Validity studies in clustering methodologies, Pattern Recognition, 1979, 11: 235.
65. Windham, M. P., Cluster validity for fuzzy c-means clustering algorithms, IEEE PAMI, 1982, 4(4): 357.
66. Dunn, J. C., Well-separated clusters and the optimal fuzzy partitions, J. Cybernet, 1974, 4: 95.
67. Gunderson, R., Applications of fuzzy ISODATA algorithms star tracker printing systems, in Proc. 7th Triennial World IFAC

Congr., 1978: 1319.
68. Xie, X. L., Beni, G., A validity measure for fuzzy clustering, IEEE PAMI, 1991, 13: 841.
69. Vogel, M. A., Wong, A. C., PFS clustering method, IEEE PAMI, 1979, 3: 237.
70. Jain, A. K., Moreau, J. V., Bootstrap techniques in cluster analysis, Pattern Recognition, 1987, 20(5): 547.
71. Beni, C., Liu, X. M., A least biased fuzzy clustering method, IEEE PAMI, 1992, 16(9): 954.
72. Dave, R. N., Validating fuzzy partitions obtained through *c*-shells clustering, Pattern Recognition Letters, 1996, 17: 613.
73. Huntsbergery, T. L., Jacobs, C. L., Cannon, R., L. et al., Iterative fuzzy image segmentation, Pattern Recognition, 1985, 2(18): 131.
74. Bensaid, A. M., Hall, L. O., Bezdek, J. C. et al., Validity-guided (re)clustering with applications to image segmentation, IEEE FS, 1996, 4(2): 112.
75. Ramdas, V., Sridhar, V., Krishna, G., An effective technique for feature extraction, Pattern Recognition Letters, 1994, 15: 885.
76. Bezdek, J. C., Castelaz, P. F., Prototype classification and feature selection with fuzzy sets, IEEE SMC, 1977, 2(7): 87.
77. Gao Xinbo, Xu Chunguang, Xie Weixin, Fuzzy partitioning of feature space for pattern classification based on supervised clustering, IEEE ISPACS'98, Melbourne, 1998: 387.
78. Jolion, J.M., Meer, P., Bataouche, S., Robust clustering with applications in computer vision, IEEE PAMI, 1991, 8(13): 791.
79. Wu Youshou, Ding Xiaoqing, A new clustering method for Chinese character recognition system using artificial neural networks, Chinese J. of Electronics, 1993, 2(3): 1.
80. Huang, Z., Kuth, A., A combined self-organizing feature map and multi-layer perception for isolated word recognition, IEEE SP, 1992, 40(11): 2651.
81. Stewart, C., Lu, Y. C., Larson, V., A neural clustering approach for high resolution radar target classification, 1994, 27(4): 503.
82. Coleman, G. B., Andrews, H. C., Image segmentation by clustering, Proc. IEEE, 1979, 5(67): 773.
83. Liu Jianzhuang, 2-D based image segmentation method with fuzzy clustering, Acta Electronca Sinica, 1992, 20(9): 40.
84. Trivedi, M. M., Bezdek, J. C., Low-level segmentation of aerial image with fuzzy clustering, IEEE SMC, 1986, 16(4): 589.
85. Porter, R., Canagarajah, N., A robust automatic clustering scheme for image segmentation using wavelets, IEEE IP, 5(4): 662.
86. Chaudhuri, B. B., Sarkar, N., Texture segmentation using fractal dimension, IEEE PAMI, 17(1): 72.
87. Pei Jihong, Study of image segmentation methods based on fuzzy information processing, Doctorial Thesis, Xi'an: Xidian University Press, 1998.
88. Chen, S. W., Chen, C. F., Chen, M. S. et al., Neural-fuzzy classification for segmentation of remotely sensed images, IEEE SP, 1997, 45(11): 2639.
89. Shih, E. Y., Moh Jenlong, Chang Fuchun, A new art-based neural architecture for pattern classification and image enhancement without prior knowledge, Pattern Recognition, 1992, 25(5): 533.
90. Wai-Chi Lai et al., A VLSI neural processor for image data compression using self-organization networks, IEEE NN, 1993, 3(3): 506.

(Received July 23, 1999, accepted December 24, 1999)