Pattern Recognition 48 (2015) 3652–3672

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Interval type-2 credibilistic clustering for pattern recognition

M. Rostam Niakan Kalhori [a], M.H. Fazel Zarandi [a,b,*]

[a] Department of Industrial Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran
[b] Knowledge Intelligent Systems Laboratory, University of Toronto, Toronto, Canada

ARTICLE INFO

Article history:
Received 18 October 2014
Received in revised form
15 March 2015
Accepted 9 April 2015
Available online 18 April 2015

Keywords:
Credibilistic clustering
Interval type-2 fuzzy clustering
Fuzzy integral
Validity index
Compactness
Separation

ABSTRACT

This paper presents a new approach to interval type-2 fuzzy clustering. In order to consider compactness within the clusters and separation of them simultaneously, the objective function of this paper is designed such that it generates both degrees of membership and non-membership of each data in each cluster, and integrates them using credibility concept. Also, a new approach to separation of clusters is proposed and utilized in designing the objective function. In this approach, the borders of clusters and therefore their compactness contribute in attaining their separation. So, the separation of clusters is not assessed only by the distance of their centers. The credibility degrees are transformed to interval type-2 form to handle different sources of uncertainty. Finally, a new validity index to characterize the number of clusters based on proposed approach to separation of clusters and Choquet integrals are proposed. The advantages of paper contributions are illustrated using several examples.

1. Introduction

Clustering an unlabeled data set $X=\{x_1,x_2,\ldots,x_N\}\subseteq R^P$ into c $(1<c<N)$ subgroups, partitions them so that each data is assigned to one subgroup. Each member of partition matrix U, u_{ik} is degree of belonging x_k to the ith cluster $(1\leq i\leq c)$. Each cluster has following properties [1]:

Homogeneity within the clusters, i.e., data that belong to the same cluster should be as similar as possible; Heterogeneity between clusters, i.e., data that belong to different clusters should be as different as possible.

New objective function-based clustering is proposed in this paper. Generally, the objective function-based clustering models can be classified into two classes: crisp and fuzzy. In hard (crisp) partitioning the degree of belonging x_k to the ith cluster is 0 or 1. In this model each data can be assigned to only one cluster. In fuzzy clustering models, the u_{ik} takes value in [0, 1]. The first fuzzy objective function-based clustering model, Fuzzy C-Means (FCM), was proposed by Bezdek in [2]. This model imposes sum of one for degree of membership of each data to clusters. Although FCM is a very useful clustering method, its memberships do not always correspond well to the degrees of belonging of the data, and it may be inaccurate in a noisy environment [3]. A possibilistic type of objective function has been proposed by Krishnapuram and Keller [3] to solve the problem of FCM (Possibilistic C-Means (PCM)) by relaxing its constraint. Although, PCM is robust to deal with noise and outlier, sometimes it generates coincident clusters because of relaxation of the FCM's constraint. By relaxing this constraint, the degree of belonging of each data to each cluster in PCM is obtained only considering that cluster. So, belonging each data to each cluster is independent of the other data and the other clusters. These two models are the basis of most proposed models.

In this paper a new objective function is designed in which both degrees of membership and non-membership are taken into account. This way of designing objective function comes from considering both compactness within the clusters and their separation in objective function to fulfill the main purpose of clustering. In this regards, a new approach to separation of clusters is introduced and applied in objective function. Also, the credibility concept is utilized to integrate degrees of membership and non-membership. Then, the interval type-2 version of proposed model is applied to deal with uncertainty of fuzzifier m. The other contribution of this paper is developing a new validity index to obtain cluster numbers. The proposed approach to separation in addition to Choquet integral is utilized to define this validity index.

The rest of the paper is organized as following: Literature review is presented in Section 2. In Sections 3 and 4 the proposed interval type-2 credibilistic clustering and validity index are explained, respectively. Experimental results are drawn in Section 5. The concluded points are presented in Section 6.

* Correspondence to: Department of Industrial Engineering, Amirkabir University of Technology, P.O. Box 15875-3144, Tehran, Iran.
Tel./fax: +98 21 64545378.
E-mail addresses: niakan@aut.ac.ir (M. Rostam Niakan Kalhori), zarandi@aut.ac.ir (M.H. Fazel Zarandi).

http://dx.doi.org/10.1016/j.patcog.2015.04.010
0031-3203/

2. Literature review

In this section four subjects are reviewed: first, two popular objective-function based fuzzy clustering models, FCM and PCM, are analyzed. Then, credibility-based clustering, interval type-2 fuzzy clustering and then the validity indices are reviewed, respectively.

2.1. Two basic objective function-based clustering

The most popular models in objective function-based fuzzy clustering are FCM and PCM. Also, some models have utilized the credibility theory in objective function. Many models have modified these three main models.

In FCM, u_{ik} is named as membership grade. The model is as follows [2]:

$$\text{Min}\, J_m(U,V;X)=\sum_{i=1}^{c}\sum_{k=1}^{N}(u_{ik})^m d_{ik}^2 \tag{1}$$

$$<\sum_{k=1}^{N}u_{ik}\le N \quad \text{For all } i=1,2,\ldots,c \text{ and } u_{ik}\in[0,1] \text{ for all } i,\ k, \tag{2}$$

$$\sum_{i=1}^{c}u_{ik}=1 \text{ for all } k=1,\ldots,\ N \tag{3}$$

In (1), $d_{ik}=x_k-v_i$ represents the distance of x_k to the ith cluster center $\{v_i\}$. Also, $m>1$ is known as fuzzifier which in most fuzzy clustering models is applied. With higher values for m the boundaries between the clusters become softer; with lower values they get harder [4]. In formulation of membership grade of x_k to ith cluster center, each cluster is inversely dependent on relative distance of x_k to the all c cluster centers (prototypes).

Although FCM is a very useful clustering method, its memberships do not always correspond well to the degrees of belonging of the data, and it may be inaccurate in a noisy environment [3].

Optimal partitions U^* of X are taken from pairs (U^*,V^*) that are local minimizers of J_m. Approximate optimization of J_m by the FCM algorithm is based on iteration through the necessary conditions for its local extrema [5]. It was proved that if $d_{ik}>0$ for all i and k, $m>1$, and X contains at least c distinct point, then $(U,\ V)$ may minimize J_m only if [2]:

$$u_{ik}=\frac{1}{\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^{2/(m-1)}};\quad 1\le i\le c;\quad 1\le k\le N \tag{4}$$

$$v_i=\frac{\sum_{k=1}^{N}u_{ik}^m x_k}{\sum_{k=1}^{N}u_{ik}^m};\quad 1\le i\le c \tag{5}$$

In order to satisfy constraint 3, the noises and outliers should get high membership degrees. This property of FCM sometimes results in a distinct cluster which includes the noises and outliers. Since the problem is minimization of objective function, with relaxation of constraint 3, the algorithm returns 0 for all u_{ik} s. Krishnapuram and Keller in [3] relaxed the constraint 3 to solve this disadvantage of FCM in possibilistic C-Means (PCM). They substituted this constraint with a new constraint $0<\sum_{i=1}^{c}u_{ik}\le 1$. Also, they changed the objective function using the concept of possibility theory. Krishnapuram and Keller in [3] showed that in case of noisy datasets, the PCM approach creates more robust clustering results rather than FCM-based methods. This model is [3]

$$\text{Min}\left\{F_m(U,V;X,\eta)=\sum_{i=1}^{c}\sum_{k=1}^{N}(u_{ik})^m d_{ik}^2+\sum_{i=1}^{c}\eta_i\sum_{k=1}^{N}(1-u_{ik})^m\right\} \tag{6}$$

$$0<\sum_{j=1}^{N}u_{ij}\le N \text{ for all } i, \text{ and } u_{ij}\in[0,1] \text{for all } i \text{ and } j, \tag{7}$$

$$\max_i u_{ij}>0 \text{ for all } j \tag{8}$$

In (6) η_i is a pre-specified constant which determines the distance at which the membership becomes 0.5. Thus, η_i determines the "zone of influence" of a point. A point x_j will have little influence on the estimates of the prototype parameters of a cluster if $d^2(x_k,v_i)$ is large when compared with η_i [3]. In [3] the authors proved that

$$u_{ik}=\frac{1}{1+\left(\frac{d_{ik}^2}{\eta_i}\right)^{1/(m-1)}};\quad 1\le i\le c;\quad 1\le k\le N \tag{9}$$

$$v_i=\frac{\sum_{k=1}^{N}u_{ik}^m x_k}{\sum_{k=1}^{N}u_{ik}^m};\quad 1\le i\le c \tag{10}$$

$$\eta_i=K\frac{\sum_{k=1}^{N}u_{ik}^m d_{ik}^2}{\sum_{k=1}^{N}u_{ik}^m},\quad K>0 \tag{11}$$

where in [3] K=1 has been proposed. The FCM is primarily a partitioning algorithm. It will find a fuzzy C-partition of a given data set, regardless of how many "clusters" are actually present in the data set. In other words, each component of the partition may or may not correspond to a "cluster." In contrast, the PCM is a mode-seeking algorithm, i.e., each component generated by the PCM corresponds to a dense region in the data set. In the PCM, the prototypes are automatically attracted to dense regions in feature space as iterations proceed [6]. However, PCM can be sensitive to the values of the initial cluster prototypes [7]. Since the membership grade of each data in each cluster only depends on the distance of it from that cluster, there is not any information sharing for construction of the clusters. Sometimes this property of PCM leads to create coincident clusters for data sets with close clusters. There are various researches based on PCM and FCM. In these researches the authors tried to solve the problems of these models with modification of objective functions. A significant point which did not consider in these papers is attending both compactness within clusters and separation of them in objective function. In these researches, the compactness of clusters has been pointed out in objective function, while their separation has been considered in validity index. Since in each iterations in order to evaluate the number of clusters, the validity index is evaluated after clustering, it cannot affect the position of centers, the compactness and separation of clusters efficiently. The format of proposed objective function of this paper is similar to FCM and PCM, but it used credibility measure to combine degrees of membership and non-membership.

2.2. Credibility-based clustering

Here the concept of credibility theory is applied to design a new objective function which includes compactness within clusters and separation of them, simultaneously. In order to study the behavior of fuzzy phenomena, credibility theory was introduced by Liu in [8]. Li and Liu presented a sufficient and necessary condition for credibility measure in [9].

Definition 1. Credibility measure

Let θ be a nonempty set, and let $p(\theta)$ be the power set of θ. Each element in $p(\theta)$ is called an event. In order to present an axiomatic definition of credibility, it is necessary to assign to each event A a number $Cr\{A\}$ which indicates the credibility that A will occur. In order to ensure that the number $Cr\{A\}$ has certain mathematical properties of credibility, there are following axioms [9]: