PERGAMON Computers and Mathematics with Applications 37 (1999) 95–100

An International Journal
computers & mathematics with applications

Cluster Validity for Fuzzy Criterion Clustering

A. O. ESOGBUE
School of Industrial and Systems Engineering
Georgia Institute of Technology, Atlanta, GA 30332, U.S.A.

BAODING LIU
Department of Applied Mathematics
Tsinghua University, Beijing 100084, P.R. China

Abstract—In this paper, we define a validity measure for fuzzy criterion clustering which is a novel approach to fuzzy clustering that in addition to being non-distance-based, addresses the cluster validity problem. The model is then recast as a bilevel fuzzy criterion clustering problem. We propose an algorithm for this model that solves both the validity and clustering problems. Our approach is validated via some sample problems. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords—Fuzzy clustering, Fuzzy criterion, Fuzzy prototype, Cluster validity, Multilevel programming.

1. INTRODUCTION

Cluster analysis is an important technique in pattern recognition. The basic problem of cluster analysis is to divide the K data points into N clusters in an optimal fashion with N a preassigned integer. Clustering via fuzzy set theory arises in two cases. One is to group fuzzy data points into some fuzzy subsets. The other is to divide the crisp data points into a specified number of subsets which need not be fuzzy, but utilizing fuzzy set theoretic methods in developing the clusters.

The beginning of fuzzy cluster analysis can be traced to the early works of Bellman *et al.* [1] and Ruspini [2]. According to Yang [3], the studies of cluster analysis via fuzzy set theory can be divided into three categories: fuzzy clustering based on fuzzy relation, fuzzy clustering based on objective function, and the fuzzy generalized k-nearest neighbor rule. The first group, fuzzy clustering based on fuzzy relation, was first proposed by Tamura *et al.* [4]. They presented a multistep procedure by using the composition of fuzzy relations beginning with a reflexive and symmetric relation. Fuzzy clustering based on objective function was proposed by Dunn [5] and generalized by Bezdek [6]. A variety of generalizations of this method has been developed [7].

The fuzzy general k-nearest neighbor rule is a type of nonparametric classifiers. Let a set of n correctly classified samples be $(\mathbf{x}_1, \theta_1), (\mathbf{x}_2, \theta_2), \dots, (\mathbf{x}_n, \theta_n)$, where θ_i represents the labeling variables of N clusters and take values in the set $\{1, 2, \dots, N\}$. A new pair $(\mathbf{x}, \theta)$ is given, where only the measurement $\mathbf{x}$ is observable by the statistician, and it is desired to estimate θ

The contribution of the first author was sponsored in part by the National Science Foundation under the Special Initiative on Intelligent Control, Grant ECS-9216004 and NASA Grant NAG 9-726. The opinions, findings, and conclusions reported here are those of the authors and do not necessarily reflect the views of the sponsoring agencies.

PII: S0898-1221(99)00147-9

by utilizing the information contained in the set of correctly classified points. We shall call $\mathbf{x}' \in \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ a nearest neighbor to $\mathbf{x}$ if $\|\mathbf{x}' - \mathbf{x}\| = \min_{1 \le i \le n} \|\mathbf{x}_i - \mathbf{x}\|$. Then the point $\mathbf{x}$ is assigned to the cluster θ' of its nearest neighbor $\mathbf{x}'$. Most of the foregoing approaches are essentially heuristic.

As a departure, Esogbue [8] introduced fuzzy dynamic programming to the area of fuzzy clustering with optimality as an objective. Application to the evaluation of fuzzy data generated in connection with nonpoint source water pollution control strategies was also reported. Recently, Liu and Esogbue [9] presented a concept of fuzzy prototype as opposed to a crisp prototype and a new kind of fuzzy clustering named fuzzy criterion clustering. Two forms of fuzzy criteria for cluster analysis are proposed. One is fuzzy average criterion which is to maximize the weighted sum of all degrees of membership of given data points, while the other is to maximize the minimum degree of membership of all given data points. Different from the approaches of traditional clustering methods, fuzzy criterion clustering will preassign the membership functions for all possible clusters to form a collection of fuzzy prototypes, and then select a number of clusters from all fuzzy prototypes by fuzzy criterion as the optimal fuzzy partition.

In this paper, we define a validity measure for fuzzy criterion clustering and form a bilevel fuzzy criterion clustering problem which solves both the validity and clustering problems. We introduce a solution algorithm for the model and exercise it on a number of sample problems.

2. FUZZY CRITERION CLUSTERING

Suppose that we have K data points, $\mathbf{x}_k$, $k = 1, 2, \ldots, K$, each $\mathbf{x}_k$ is an m-dimensional vector, i.e., $\mathbf{x}_k = (x_{k1}, x_{k2}, \ldots, x_{km})$. These data points may be crisp or fuzzy. Our problem is to group the set of K (crisp or fuzzy) data points into N clusters, where N is a predetermined integer.

To do this, fuzzy criterion clustering demands a collection of fuzzy prototypes which will be given by the decision maker. Usually, a crisp prototype, for example, crisp circle, is $(x_1 - a)^2 + (x_2 - b)^2 = r^2$. A point (x_1^*, x_2^*) is on that circle if and only if $(x_1^* - a)^2 + (x_2^* - b)^2 = r^2$. A fuzzy prototype, for example, fuzzy circle, has a prototype center like $(x_1 - a)^2 + (x_2 - b)^2 = r^2$ on which the degree of membership is defined as 1. For any other data point (x_1', x_2'), it is on the fuzzy circle with degree of membership

$$\exp\left(-\left[(x_1' - a)^2 + (x_2' - b)^2 - r^2\right]\right).$$

Certainly, we can define the membership function in other ways. Similarly, we can define fuzzy point, fuzzy line, fuzzy ellipse, fuzzy parabola, etc. Generally, let $U = \{\mathbf{u} \mid \mathbf{u} \subset R^m\}$ be a collection of all fuzzy prototypes given by some decision maker(s). This collection may be finite or infinite (countable or not). Each prototype $\mathbf{u}$ is a fuzzy subset with membership function μ.

After constructing a collection of fuzzy prototypes, we need a clustering measure called fuzzy criterion which will maximize the weighted sum of all memberships of all given data points, that is, select N fuzzy prototypes $\{\mathbf{u}_n,\ n = 1, 2, \ldots, N\}$ from the collection U to maximize

$$\max J(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N) = \sum_{k=1}^{K} \lambda_k \mu_1(\mathbf{x}_k) \vee \cdots \vee \mu_N(\mathbf{x}_k), \tag{1}$$

where $\mathbf{u}_n \in U$ are fuzzy prototypes with membership functions μ_n, $n = 1, 2, \ldots, N$, respectively, λ_k are weighted factors; typically, we can define $\lambda_k = 1/K$.

In practice, we can employ a vector $\mathbf{y}$ to represent the N fuzzy prototypes $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N$. Thus, we can rewrite (1) as

$$\max_{\mathbf{y}} f(\mathbf{y}, N) = \sum_{k=1}^{K} \lambda_k \mu_1(\mathbf{x}_k) \vee \mu_2(\mathbf{x}_k) \vee \cdots \vee \mu_N(\mathbf{x}_k), \tag{2}$$

where $f(\mathbf{y}, N)$ is called a clustering measure of $\mathbf{y}$ when the number of clusters is N.

Let $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N$ be the optimal N fuzzy prototypes optimizing the objective function $J(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N)$ in model (1) or (2). Then $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N\}$ is an optimal fuzzy criterion partition of the given data point set. Each $\mathbf{u}_n$ is a fuzzy cluster with membership function μ_n. A data point $\mathbf{x}$ is considered to be in $\mathbf{u}_n$ if

$$\mu_n(\mathbf{x}) = \max\{\mu_1(\mathbf{x}), \mu_2(\mathbf{x}), \dots, \mu_N(\mathbf{x})\}. \tag{3}$$

That is, a data point belongs to one and only one cluster.

3. VALIDITY MEASURES

Given a number of clusters N, we can obtain N optimal clusters represented by parameter vector $\mathbf{y}$ by employing fuzzy criterion clustering. We denote the N clusters by $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N$. Then, the K data points can be classified into one and only one cluster of $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N$.

Let the length of $\mathbf{u}_n$ be L_n, the number of data points belonging to $\mathbf{u}_n$ be K_n, $n = 1, 2, \dots, N$, respectively; then, we have $K_1 + K_2 + \cdots + K_N = K$. The validity measure $g(\mathbf{y}, N)$ is defined by

$$g(\mathbf{y}, N) = \min\left\{\frac{K_n}{L_n},\ n = 1, 2, \dots, N\right\}. \tag{4}$$

On the other hand, we know that cluster validity should eliminate spurious clusters and merge compatible clusters. According to the partition rule of fuzzy criterion clustering, any point can belong to one and only one cluster, so a cluster is considered spurious if the number of its data points is *too small*; meanwhile, the validity measure $g(\mathbf{y}, N)$ should also be *too small*. In addition, the fact that two clusters are compatible implies that there is at least one cluster such that it contains only a few number of data points, i.e., the validity measure $g(\mathbf{y}, N)$ is very small. Hence $g(\mathbf{y}, N)$ can be regarded as a validity measure.

This validity measure presumes that we have an approximate estimation on the number of potential points of certain prototypes. Meanwhile, the length L_n will be represented by that number. This assumption is motivated by the theory and operation of the digital process of a camera.

4. BILEVEL FUZZY CRITERION CLUSTERING

We can design our clustering model as a bilevel programming problem which can be formulated as follows:

$$\begin{aligned} &\max_N g(\mathbf{y}, N) \\ &\text{where } \mathbf{y} \text{ solves} \\ &\quad \max_{\mathbf{y}\|N} f(\mathbf{y}, N), \end{aligned} \tag{5}$$

here, N is a positive integer representing the number of clusters, $\mathbf{y}$ is a vector of parameters describing the selected N clusters.

We note that the model (5) is a bilevel programming problem. However, in this case, since the validity measure $g(\mathbf{y}, N)$ is a decreasing function of N, the optimal solution is clearly $N^* = 1$, always. Consequently, this form is not considered a *good* choice.

To circumvent this difficulty, we reformulate it as follows:

$$\begin{aligned} &\max N \text{ such that} \\ &\quad g(\mathbf{y}, N) \geq \alpha \\ &\text{where } \mathbf{y} \text{ solves} \\ &\quad \max_{\mathbf{y}\|N} f(\mathbf{y}, N), \end{aligned} \tag{6}$$

where α is a predetermined level called the *critical number*. The determination of the critical number is not a difficult proposition because the validity measure will decrease quite rapidly from correct clustering to a clustering with a spurious cluster. Usually, the validity measure takes on a value of about 1 for correct clustering if we have a correct estimation of all of L_n, and a value of less than 0.5 for a clustering with at least one spurious cluster.

The model of problem (2) is solved via a genetic algorithm. The general form of bilevel fuzzy criterion clustering based on a genetic algorithm is shown below.

```
Procedure for Bilevel Fuzzy Criterion Clustering
    Set N = N_0 as an underestimated number;
    FOR number of clusters N DO
        Initialize the fuzzy prototypes
            (chromosomes);
        REPEAT
            Update the fuzzy prototypes by genetic
                operators;
            Select the fuzzy prototypes by sampling
                mechanism;
        UNTIL(termination_condition)
        Report optimal N clusters y*_N;
        IF g(y*_N, N) < α THEN break;
        N = N + 1;
    ENDFOR
    Report the optimal solution (y*_{N-1}, N - 1);
```

5. NUMERICAL EXPERIMENTS

5.1. Example 1

In the sequel, we present the results of numerical studies which exemplify our approach and are used to implement the foregoing algorithm. Let us consider a perfect case in which we produce 63 data points from circle $(x_1 - 5)^2 + (x_2 - 2.5)^2 = 2.5^2$, 50 data points from circle $(x_1 - 2.5)^2 + (x_2 - 7.5)^2 = 2^2$, and 50 data points from circle $(x_1 - 7.5)^2 + (x_2 - 7.5)^2 = 2^2$ on the region 10×10. The total number of data points is 163.

We define the critical number for validity measure as 0.5. We have developed a computer program to implement the algorithm. The computer program starts at $N = 1$ and finds that the optimal cluster is $(x_1 - 5.004)^2 + (x_2 - 2.500)^2 = 2.504^2$, meanwhile, the clustering measure $f(\mathbf{y}^*, 1) = 0.349$ and the validity measure $g(\mathbf{y}^*, 1) = 2.591$, which is greater than the critical number 0.5.

So the number of clusters N is replaced by $N + 1$, i.e., $N = 2$. The optimal two clusters are $(x_1 - 5.003)^2 + (x_2 - 2.500)^2 = 2.504^2$, and $(x_1 - 7.500)^2 + (x_2 - 7.500)^2 = 2.012^2$. The clustering measure $f(\mathbf{y}^*, 2) = 0.630$ and the validity measure $g(\mathbf{y}^*, 2) = 1.542$, which is also greater than the critical number 0.5.

The number of clusters N is replaced by $N + 1$ again, i.e., $N = 3$. The following optimal three clusters are then obtained:

$$\begin{aligned}
&(x_1 - 5.003)^2 + (x_2 - 2.500)^2 = 2.504^2, \\
&(x_1 - 2.495)^2 + (x_2 - 7.499)^2 = 2.014^2, \\
&(x_1 - 7.493)^2 + (x_2 - 7.500)^2 = 2.011^2.
\end{aligned} \tag{7}$$

The clustering measure $f(\mathbf{y}^*, 3) = 0.911$ and the validity measure $g(\mathbf{y}^*, 3) = 0.988$, which shows that the number of clusters N can be enlarged again.

However, when $N = 4$, the computer program reports that the optimal four clusters are $(x_1 - 5.000)^2 + (x_2 - 2.500)^2 = 2.504^2$, $(x_1 - 2.495)^2 + (x_2 - 7.501)^2 = 2.014^2$, $(x_1 - 7.501)^2 + (x_2 - 7.499)^2 = 2.014^2$, $(x_1 - 1.613)^2 + (x_2 - 2.498)^2 = 1.543^2$. Meanwhile, the clustering measure $f(\mathbf{y}^*, 4) = 0.914$ and the validity measure $g(\mathbf{y}^*, 4) = 0.052$, which is less than the critical number 0.5. We can then stop the procedure. We mention that the cluster $(x_1 - 1.613)^2 + (x_2 - 2.498)^2 = 1.543^2$ is spurious. This case is shown in Figure 1.

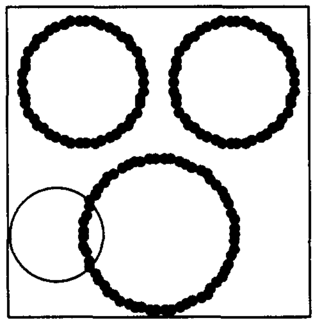

Figure 1. Case of four clusters.

Thus, the optimal number of clusters is three and the optimal three clusters are described by (7). The validity and cluster measures for different number of clusters are shown in Figure 2, in which the decreasing curve represents the validity measure and the increasing curve represents the clustering measure.

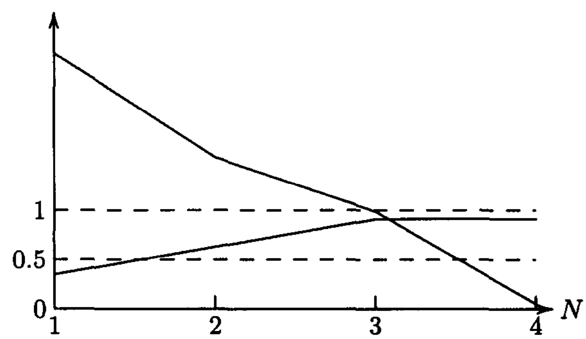

Figure 2. Validity and clustering measures.

6. DISCUSSION

The problem of clustering appears in an array of different and important application areas. Various algorithms for its implementation abound. The efficiency of these algorithms also varies. Often, the particular problem of interest dictates the best one to employ.

The contribution of fuzzy clustering is well documented in the literature. However, the benefits are usually minimized by the absence of reliable validity measures. In a previous effort, we presented a completely different approach to fuzzy clustering. The model combines a fuzzy criterion set approach which we developed for fuzzy dynamic programming with a fuzzified version of clustering based on crisp prototypes which we call fuzzy prototypes.

In this paper, we extended that effort by first advancing a validity measure and then posing the resultant clustering problem as a bilevel fuzzy criterion clustering problem. We developed a computational algorithm based on genetic algorithms for this model. Using this algorithm, we showed, through some sample problems, that both the problem of cluster validity and the clustering problem can be conjunctively solved effectively via this approach.

REFERENCES

1. R. Bellman, R. Kalaba and L.A. Zadeh, Abstraction and pattern classification, *J. Math. Anal. and Appl.* **2**, 581–586 (1966).
2. E.H. Ruspini, A new approach to clustering, *Inform. and Control* **15**, 22–32 (1969).
3. M.-S. Yang, A survey of fuzzy clustering, *Mathl. Comput. Modelling* **18** (11), 1–16 (1993).
4. S. Tamura, S. Higuchi and K. Tanaka, Pattern classification based on fuzzy relations, *IEEE Trans. Syst. Man Cybern.* 1 (1), 61–66 (1971).
5. J.C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact, well-separated clusters, *J. Cybernetics* **3**, 32–57 (1974).
6. J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, (1981).
7. J.C. Bezdek and S.K. Pal, *Fuzzy Models for Pattern Recognition*, IEEE Press, New York, (1992).
8. A.O. Esogbue, Optimal clustering of fuzzy data via fuzzy dynamic programming, *Fuzzy Sets and Systems* **18**, 283–298 (1986).
9. B. Liu and A.O. Esogbue, Optimal fuzzy criterion clustering based on fuzzy prototypes, *J. Math. Anal. Applic.* (1996).
10. A.O. Esogbue and R.E. Bellman, Fuzzy dynamic programming and its extensions, *Fuzzy Sets and Decision Analysis, TIMS Studies in the Management Science* **20**, 147–167 (1984).
11. E.H. Ruspini, Numerical methods for fuzzy clustering, *Information Sciences* **2**, 319–350 (1970).
12. C. Coray, Clustering algorithms with prototype selection, In *Proc. Hawaii Intern. Conf. Syst. Sci.*, pp. 945–955, (1981).
13. R.N. Dave, Fuzzy shell-clustering and application to circle detection in digital images, *Int. J. Gen. Syst.* **16**, 343–355 (1990).
14. R.N. Dave and K. Bhaswan, Adaptive fuzzy *C*-shells clustering and detection of ellipses, *IEEE Trans. Neural Networks* **3** (5), 643–662 (1992).