Determining the most proper number of cluster in fuzzy clustering by using artificial neural networks

N. Alp Erilli a,⇑, Ufuk Yolcu a, Erol Eğrioglu a, Ç. Hakan Aladağ b, Yüksel Öner a

a Department of Statistics, Ondokuz Mayis University, Samsun, Turkey
b Department of Statistics, Hacettepe University, Samsun, Turkey

Article info

Keywords:
Artificial neural networks
Cluster validation index
Fuzzy clustering
Number of cluster

Abstract

In a clustering problem, it would be better to use fuzzy clustering if there was an uncertainty in determining clusters or memberships of some units. Determining the number of cluster has an important role on obtaining sensible and sound results in clustering analysis. In many clustering algorithm, it is firstly need to know number of cluster. However, there is no pre-information about the number of cluster in general. The process of determining the most proper number of cluster is called as cluster validation. In the available fuzzy clustering literature, the most proper number of cluster is determined by utilizing cluster validation indices. When the data contain complexity are being analyzed, cluster validation indices can produce conflictive results. Also, there is no criterion point out the best index. In this study, artificial neural networks are employed to determine the number of cluster. The data is taken as input so the output is membership degree. The proposed method is applied some data and obtained results are compared to those obtained from validation indices like PC, XB, and CE. It is shown that the proposed method produce accurate results.

1. Introduction

Clustering analysis is one of the multivariate data analysis methods and it has been being frequently in business and science in recent years. Clustering analysis was developed to provide a detailed explanation to the classification of individuals or objects. Clustering analysis is a method which classifies the units under examination in a study by gathering them in specific groups depending on their similarities, to introduce the common features of the units and to make general definitions about these classifications. The aim here is to classify the ungrouped data according to their similarities and to help the researcher to obtain appropriate, useful and round up data (Rencher, 2002). In other words, the aim is to help gather the similar data in the same group or cluster by taking the similarities between the data into consideration. Clustering analysis classifies the similar individuals or objects in the same cluster according to predetermined selection criteria. The homogeneity of the clusters which are formed as a result of the analysis will be high and the intercluster heterogeneity will be poor (Grabmeier & Rudolph, 2002).

The usage of fuzzy clusters in clustering was first proposed by Bellman, Kalaba, and Zadeh (1966). In fuzzy clustering, fuzzy techniques are used to cluster the data and with these techniques an object can be classified in more than one cluster. Since such algorithms undertake the uncertainty of real numbers, they are helpful in forming clustering types which are appropriate for the experiences of daily life. Determining the most proper number of clusters in fuzzy clustering analysis is an important issue. In the literature, various cluster validation indexes for the determination of the most proper number of clusters have been proposed (Chong, Gedelon, & Koczy, 2002; Fukuyama & Sugeno, 1989; Halkidi, Batistakis, & Vazirgiannis, 2001; Kim, Kim, Lee, & Lee, 2004; Kim & Ramakrishna, 2005; Kwon, 1998; Tang, Sun, & Sun, 2005; Windham, 1981; Wu & Yang, 2005; Zahid, Limouri, & Essaid, 1999; Zhang & Wang, 2007). In this study, for the determination of the most proper cluster number in fuzzy clustering, a criterion based on artificial neural networks has been proposed.

In the second part of the study, there is information available on fuzzy clustering. In Section 3, Fuzzy C-Means method, which is well known in literature, is introduced. Section 4 summarizes fuzzy clustering validation indexes and Section 5 gives brief information about artificial neural networks. In the sixth part, the proposed criterion is explained by being applied to three real-like and one real data. It is shown that the proposed method produce accurate result. In the last part, the obtained data are discussed.

2. Fuzzy clustering

This approach comes into the picture as an appropriate method when the clusters cannot be separated from each other distinctly
or when some units are uncertain about membership. Fuzzy clusters are functions modifying each unit between 0 and 1 which is defined as the membership of the unit in the cluster. The units which are very similar to each other hold their places in the same cluster according to their membership degree.

Similar to other clustering methods, fuzzy clustering is based on distance measurements as well. The structure of the cluster and the algorithm used to specify which of these distance criteria will be used. Some of the convenient characteristics of fuzzy clustering can be given as follows (Naes & Mevik, 1999):

i. It provides membership values which are convenient to comment on.
ii. It is flexible on the usage of distance.
iii. When some of the membership values are known, they can be combined with numeric optimization.

The advantage of fuzzy clustering over classical clustering methods is that it provides more detailed information on the data. On the other hand, it has disadvantages as well. Since there will be too much output when there are too many individuals and clusters, it is difficult to summarize and classify the data. Moreover, fuzzy clustering algorithms, which are used when there is uncertainty, are generally complicated (Oliveira & Pedrycz, 2007).

3. Fuzzy C-Means (FCM) algorithm

Fuzzy C-Means algorithm forms the basis of all clustering techniques that depend on objective function. It was developed by Bezdek (1974a, 1974b). When the FCM algorithm comes to a conclusion, the dots in the p dimension space become a sphere-shaped figure. It is assumed that these clusters are approximately the same size. Cluster centers represent each cluster and they are called prototypes. Euclidean distance $d_{jk}$ between the data and the cluster center is used as the distance measurement and can be calculated by

$$d_{jk} = d(x_i, v_k) = \left( \sum_{i=1}^{p} (x_{ij} - v_{kj})^2 \right)^{1/2}$$

where $x_i$ represents the position observation value in the coordinated system, and $v_k$ represents the cluster center.

It is necessary to know the number of clusters and the membership degrees of the individuals beforehand to be able to put this technique into practice. Since it is difficult to know these parameters before the application, it is possible to find these values through the method of trial and error or through some techniques developed.

The objective function used for this clustering method is as follows:

$$J(u, v) = \sum_{j=1}^{c} \sum_{k=1}^{m} u_{jk}^m \|x_j - v_k\|^2$$

This function is the weighted least square function. $n$ parameter represents the number of observations, and $c$ represents the number of clusters. $u_{jk}^m$ is the membership of $x_j$ in $k$ cluster. $J(u, v)$ value is a measure of the total of all weighted error sum of squares.

If the $J(u, v)$ function is minimized for each value of $c$, in other words if it is derived from the 1st degree according to $n$’s and made equal to 0, the prototype of FCM algorithm can be given as follows:

$$v_{jk} = \frac{\sum_{i=1}^{n} u_{jk}^m \cdot x_{ik}}{\sum_{j=1}^{n} u_{jk}^m}$$

The required steps for the FCM algorithm are as follows:

**Step 1:** Set the initial values: the number of values $c$, fuzzy index $m$, end process criterion $\varepsilon$ and membership degrees matrix $U$ or $V$ cluster prototypes are generated randomly.

**Step 2:** Considering that the $U$ cluster prototypes are randomly generated, the membership degrees matrix is calculated by using $u_{jk}$ values which is given below.

$$u_{jk} = \left[ \sum_{i=1}^{c} \left( \frac{d_{ik}}{d_{jk}} \right)^{2/m} \right]^{-1}$$

**Step 3:** $U$ cluster prototypes are updated following the Step 2 equation.

**Step 4:** In case of $||U^{(t)} - U^{(t-1)}|| < \varepsilon$ stop, otherwise go back to Step 2.

After the FCM algorithm is applied, membership degrees are used to decide upon which individual will enter in which cluster. The individuals are included in the clusters by considering in which of these clusters they have the biggest membership. However, each individual can enter the other clusters with a certain degree of membership.

The results of the FCM algorithm depend on the randomly generated values to some extent. Thus, various algorithms have been developed and still being developed to solve the problems resulting from randomness (Halkidi et al., 2001).

FCM updates the cluster centers and the membership degrees for each data point through iteration and carries the clusters centers to where they should be inside the data set.

Since the initial places (values) of the cluster centers are generated by using the $U$ matrix, the value of which is given randomly in the first place, FCM will not be able to guarantee getting close to the optimal result (Sintas, Cadenas, & Martin, 1999).

Performance of cluster centers depends on the starting points of the centers (Dave, 1996). The following are two ways defined for a stronger approach:

i. Using an algorithm to define all the centers.
ii. Restarting FCM with different starting centers (Strategy for restart).

4. Fuzzy clustering validity index

Clustering analysis aims to place similar objects in the same groups. The purpose is to get an idea about the sample dispersions and about the correlations between variables in the samples which include huge data. However, many clustering algorithms necessitate pre-knowledge of the number of clusters. The fact that the researchers do not have pre-knowledge of the number of clusters in many studies makes it impossible to know whether the end number of clusters is more or less than the actual number of clusters. If the end number of clusters turn out to be less than the actual number of clusters, then one or more of the present clusters will have to unite; if it turns out to be more, then one or more of the present clusters will be divided. The process of determining the optimal cluster number is called cluster validity in general. Thus, the accuracy of the end cluster number can be determined.

When the data are in the two dimensional space, the number of clusters can be decided upon by commenting on the cluster results visually. However, as the number of dimensions increase in space, visuality gets harder and there becomes a need for validity indexes.

As a result, two criteria can be mentioned for value clusters and the most suitable cluster planning.

1. Density: It measures how close the group members are. The best example to this is variance.
2. Separation: It shows how close two clusters are separated. It measures the distance between two different clusters.
4.1. Partition coefficient (PC)

This method proposed by Bezdek (1974a) and holds a value between $1/c$ and 1. Here, $c$ is the number of clusters. If all membership values turn out to be equal as a result of fuzzy partition, $u_{ik} = 1/c$. This is the smallest degree of the PC. It is desirable that the value of the PC in the appropriate clustering process has a value close to 1. As the PC value gets closer to 1, the value of the PC in the appropriate clustering process has a value between 2 and 0. The best number of clusters will be between the 2 and 6 range.

\[
V_{PC} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}^2
\]

(5)

4.2. Classification entropy (CE)

This method has been proposed by Bezdek (1974b) as well.

\[
V_{CE} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij} \log_2 u_{ij}
\]

(6)

In this equation, a logarithm is e base. CE value needs to be close to 0. The best number of clusters will be between the 2 and $(n-1)$ range.

4.3. Xie–Beni index (XB)

This index developed by Xie and Beni (1991) is also known as the density and secession validity function and it is as follows:

\[
V_{XB} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{c} u_{ijk}^2 \| x_i - v_j \|^2}{\max_{j=1}^{c} \| x_i - v_j \|^2}
\]

(7)

5. Artificial neural networks

Artificial neural networks method is an efficient analysis tool which has been brought forward with the imitation of biological neural network. Artificial neural networks have been applied in many scientific areas because of its simplicity and efficiency. Different artificial neural networks components have been used in application depending on the purpose. With prevision problems, feed forward artificial neural network architectures have been preferred in general and these have had successful results. The components of the feed forward artificial neural networks can be given as follows (Egrioglu, Aladag, & Gunay, 2008):

5.1. Architecture structure

A multi-layer feed forward artificial neural network architecture structure in its simplest form has been given in Fig. 1. As is seen in the figure, a multi-layer feed forward artificial neural network architecture has three parts. These are the input layer, the hidden layer (or layers) and the output layer. Layers are composed of units which are called neurons. In artificial neural networks, neurons are attached to each other with weights. In feed forward networks these connections are unidirectional and it is forward. There is no connection between the units of the same layer.

5.2. Learning algorithm

There are many learning algorithms used in determining the weights in artificial neural networks. One of the most widely used learning algorithms is back propagation algorithm. Back propagation algorithm updates the weights depending on the difference between the available data and the network output. The learning parameter used in this algorithm plays an important role on getting close enough to the optimal results. The learning parameter can be either considered fixed or can be updated dynamically in the algorithm.

5.3. Activation function

Activation function provides the curvy match between the input and output units. Choosing the activation function correctly affects the network’s performance considerably. Activation function can be chosen as unipolar, double pole or as linear. When the chosen activation function is not linear, the slope parameter should be determined. The slope parameter is also a factor which plays an important role on getting close enough to the optimal results.

6. A new approach and application to establish the best cluster number based on artificial neural network

Determining the best cluster number in fuzzy clustering becomes more important especially if the clusters are not separated from each other significantly. In case of uncertainty, cluster validity indexes help the researcher in making definite decisions. Many cluster validity index in the literature give conflicting results about the cluster numbers with data in complicated form (Cho & Yoo, 2005; Rezaee, Lelieveldt, & Reiber, 1998; Rhee & Oh, 1996). After the application of fuzzy clustering method, each data is appointed to the cluster in which it has the highest membership degree. As a result of a classification done with these results any classification technique is expected to have high percentage of classification. If artificial neural networks method is used as a classification method, the input of the neural network will be the data matrix and the output will be the cluster number where each data is appointed as a result of fuzzy clustering.

When any classification technique is used, considering the high percentage of classification expectation, the most appropriate number of clusters can be determined in the fuzzy clustering with the following algorithm based on artificial neural networks.

Algorithm

Step 1: The highest and lowest number of clusters in accordance with the data is decided upon. The most appropriate cluster number to will be determined will be in this range. If the most appropriate number of clusters is $c_{opt}$, the minimum number of clusters is $c_{min}$ and the maximum number of clusters is $c_{max}$. Thus, the range for $c_{opt}$ can be given as follows:

$c_{min} \leq c_{opt} \leq c_{max}$
Table 1
The results for the simulation data with three clusters.

<table>
<thead>
<tr>
<th>Number of clusters</th>
<th>Cluster validity indexes</th>
<th>PC</th>
<th>CE</th>
<th>XB</th>
<th>ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td></td>
<td>0.8167</td>
<td>0.2765</td>
<td>1.4454</td>
<td>0.000093</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.9992</td>
<td>0.0034</td>
<td>49.2682</td>
<td>0.000198</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.8996</td>
<td>0.1562</td>
<td>43.2621</td>
<td>0.221739</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0.8808</td>
<td>0.1954</td>
<td>24.8372</td>
<td>0.396981</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>0.7817</td>
<td>0.3464</td>
<td>19.2458</td>
<td>0.814937</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>0.8215</td>
<td>0.3078</td>
<td>43.596</td>
<td>0.658898</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>0.742</td>
<td>0.4512</td>
<td>13.6646</td>
<td>1.255811</td>
</tr>
</tbody>
</table>

Fig. 2. The graph of the values obtained from the proposed approach for data with three clusters according to various numbers of clusters.

Step 2: The FCM method is applied for the number of clusters in the designated range. As a result, the FCM method is applied for \( c_{\text{max}} + c_{\text{min}} - 1 \) times.

Step 3: The feed forward artificial neural networks in which input is the data matrix and the target value is the cluster number to which every data is appointed will be operated depending on the number of layers for each of the possible cluster numbers.

Step 4: For every cluster number, the median of the RMSE (Root Mean Square Error) values are calculated. These values are obtained from artificial neural networks according to used various numbers of layer. Median is used here instead of mean since artificial neural networks produce RMSE values that are extreme for some number of layers.

Step 5: Drawing a graph of the median values for each cluster number, the value preceding the number of cluster with the first leap (where the RMSE median rises extremely for the first time) is appointed as the most appropriate number of cluster.

The proposed algorithm has been applied to three simulation and one real life data. First, the proposed method has been applied to the data which has three clusters in real. Moreover, the values of the other cluster validity indexes are obtained and the results are presented in Table 1. When Table 1 is studied, the most appropriate number of clusters for PC, CE and XB criteria is 3. When the results obtained by the proposed method in ANN (artificial neural network) column are studied, it can be seen that the first leap is at 4 and the appropriate number of clusters is 3. The graph of the proposed approach for various numbers of clusters has been presented in Fig. 2.

Second, the proposed method for the simulation value with four clusters in real is applied. Moreover, the values of the other cluster validity indexes are obtained and the results are presented in Table 2. According to Table 2, the most appropriate number of clusters for PC, CE and XB criteria is 4. When the results obtained by the proposed method in ANN column are examined, it can be seen that the first leap is at 5 and the appropriate number of clusters is 4. The graph of the proposed approach for various numbers of clusters has been presented in Fig. 3.

Lastly, the proposed method is applied to the synthetic data which is a real life data. Besides, the values of the other cluster validity indexes are obtained and the results are presented in Table 3. When Table 3 is examined, the most appropriate number of clusters for PC, CE and XB criteria is 5. When the results obtained by the proposed method in ANN column are studied, it can be seen that the first leap is at 6 and the appropriate number of clusters is 5. The graph of the proposed approach for various numbers of clusters has been presented in Fig. 4.

Third, the proposed method for the simulation value with five clusters in real is applied. Besides, the values of the other cluster validity indexes are obtained and the results are presented in Table 4. According to Table 4, the most appropriate number of clusters is 3 for PC criterion, two for CE criterion, and two for XB criterion. When the results obtained by the proposed method in ANN column are examined, it can be seen that the first leap is at 6 and the appropriate number of clusters is 3. The graph of the proposed approach for various numbers of clusters is presented in Fig. 5. It can be seen that the appropriate number of clusters for the synthetic data graph should be 3. When the PC criterion appoints the number of clusters correctly, CE and XB criteria make the wrong choice. The proposed method appoints the most appropriate cluster number correctly.

Table 2
The results for the simulation data with four clusters.

<table>
<thead>
<tr>
<th>Number of clusters</th>
<th>Cluster validity indexes</th>
<th>PC</th>
<th>CE</th>
<th>XB</th>
<th>ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td></td>
<td>0.7121</td>
<td>0.4626</td>
<td>0.08065</td>
<td>0.000082</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.8034</td>
<td>0.3597</td>
<td>1.9668</td>
<td>0.064163</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.9988</td>
<td>0.0053</td>
<td>7.5563</td>
<td>0.000287</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0.931</td>
<td>0.1129</td>
<td>6.7512</td>
<td>0.132227</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>0.8713</td>
<td>0.2073</td>
<td>4.2561</td>
<td>0.13305</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>0.839</td>
<td>0.274</td>
<td>6.8055</td>
<td>0.398881</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>0.8415</td>
<td>0.2863</td>
<td>6.0238</td>
<td>0.478669</td>
</tr>
</tbody>
</table>

Table 3
The results for the simulation data with five clusters.

<table>
<thead>
<tr>
<th>Number of clusters</th>
<th>Cluster validity indexes</th>
<th>PC</th>
<th>CE</th>
<th>XB</th>
<th>ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td></td>
<td>0.6949</td>
<td>0.471</td>
<td>0.99</td>
<td>0.000094</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.9</td>
<td>0.2286</td>
<td>1.1935</td>
<td>0.0002</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.9265</td>
<td>0.1702</td>
<td>5.9476</td>
<td>0.000325</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0.9999</td>
<td>0.00049</td>
<td>26.0438</td>
<td>0.000483</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>0.966</td>
<td>0.0555</td>
<td>14.7612</td>
<td>0.144221</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>0.9262</td>
<td>0.131</td>
<td>25.4019</td>
<td>1.293433</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>0.8677</td>
<td>0.2123</td>
<td>10.6709</td>
<td>1.210771</td>
</tr>
</tbody>
</table>
The proposed methods are used, and are compared with the results obtained from the number of clusters. In this study, feed forward artificial neural networks can give conflicting results in determining the most appropriate uncertainty of some cluster members. Cluster validity indexes are important problem. In some complicated data, because of the number of clusters in order to reach accurate and sound results is an important problem. In clustering analysis, determining the most appropriate number of clusters correctly. For the real life data called synthetic data, it is shown that only the PC criterion and the proposed methods appoint the most appropriate number of clusters correctly. As a result of the applications, it can be seen that the most appropriate number of clusters can be appointed in fuzzy clustering with the proposed approach based on ANN.

### References


