Research on Neural Network Model Based on Subtraction Clustering and Its Applications

ZHANG Tie-jun, CHEN Duo, SUN Jie

Abstract
This paper takes advantage of the integration of subtraction clustering and fuzzy c-means clustering algorithm to obtain precise number of clusters and the degree of membership, and establishes the neural network model based on the multiple criterion information fusion and the fuzzy technology. Collecting a variety of data in actual underground mining process and integrating them as a whole, the simulations are performed. Experimental results indicate that the proposed method is valid.

Keywords: subtractive clustering; fuzzy c-means clustering; fuzzy neural network

1. Introduction
In fact, most of the systems of actual production process can not be modeled using an existing model; accordingly, to solve this problem is quite difficult. The neural network does not rely on the accurate model and has a series of advantages of parallel computing, distributed information storage and adaptively learning function and son on. It is just because of these advantages, the neural network technology has been gradually applied to many industrial processes. However, in more complex industrial processes, which consist of the quantity of data and multiple dimensions, if the neural network is directly used, we might have the problem of yielding inaccurate results and not being convergence in their training phase. Aimed at these shortcomings, the neural network model based on subtraction clustering and the fuzzy c-means is presented in this paper, in which, the fuzzy clustering is run two times to deal with the initial complex data, and then the neural network are utilized for further operation.
2. Data clustering process

Clustering is a process of classification when only data available are unlabeled and no prior knowledge about it [1]. According to a definition of metric of similarity, a set of objects are partitioned into a certain number of clusters, so that objects in the same clusters are as similar as possible and objects in different clusters are as dissimilar as possible in the sense of the definition.

2.1 Subtractive clustering

In practice, the structure of data distribution cannot be usually known in advance, but in clustering method some parameters, such as the number of cluster, is required to be given. In order to conduct more accurately the inherent characteristic of training data, the subtractive clustering is utilized to obtain the initial cluster centers of the fuzzy c-means algorithm in this paper, in which the subtractive clustering algorithm is based on the mountain function. By the establishment of mountain function that is an index of data density, subtractive clustering algorithm can adaptively determine the number of clusters and cluster center itself, furthermore the initial structure of the objective system. The subtractive clustering is a kind of simple and effective algorithm based on the forming principle of human visual data set and the correlation between data.

Let the number of m-dimension input data set X is equal to n, since each data point is a candidate for cluster centers, the mountain function that expresses the density of X is first constructed as follows:

\[ m_i^k(x_i) = \sum_{j=1}^{n} \exp \left( \frac{\|x_i - x_j\|^2}{(\alpha / 2)^2} \right) \]  

(1)

Where, \( \alpha \) is a positive number, it expresses cluster radius and defines a neighborhood for the point. If there are larger amount of data points near a data point, then the data will has a higher peak, while data points outside the radius has little contribution to the point density of the point. After calculating the density of all points, the point with the maximum density is selected as the first cluster center \( x^c_1 \), where, \( m^c_1 \) is its corresponding index value of density.

To compute the next cluster center, it is required to eliminate the impact of the existing cluster center; therefore the density index of mountain function is modified as follows:

\[ m_i^k(x_i) = m_i^{k-1}(x_i) - m_i^{k-1} \exp \left( \frac{\|x_i - x_c^{k-1}\|^2}{(\beta / 2)^2} \right) \]

(2)

where, \( \beta \) expresses the neighbourhood that the density index is significantly reduced, and to avoiding yielding very near cluster centres, we have \( \beta = 1.5 \alpha \). Mountain function value will be pared to select a new cluster centres, when the new cluster center to meet the appropriate density index \( m_c^k / mlc < \delta \), stop searching the new cluster center. The initial condition of the fuzzy c-means clustering algorithm is the cluster centres computed by the subtractive clustering in advance, therefore avoiding yielding the error by assigned arbitrarily the number of cluster, when only using the fuzzy c-means algorithm to obtain the membership degree of data set.

2.2 The Fuzzy C-means clustering (FCM)

Since most actual objects are not strictly property, there is the intermediary nature in their property and generic, i.e. they have the nature of both this and that. The degree of samples belonging to various clusters can be determined by the fuzzy clustering, which expresses the intermediary of the sample and well reflects the real world objectively.
The vector \( X_j (j = 1, 2, \ldots, n) \) is partitioned into \( c \) clusters \( G_i (i = 1, 2, \ldots, c) \), the clustering process is described as follows:

Initialization: undertake the cluster centers and the number of clusters \( c \), obtained by the method in section 2.1, set iteration threshold \( \epsilon \), initialize the cluster prototype \( P (0) \), and the iteration counter \( b = 0 \);

Step 1: calculate or update partition matrix \( U^{(b)} = \| c_j - x_j \|, \forall i, k \), if \( \exists d_{ik}^{(b)} > 0 \), we have:

\[
\mu_{ik}^{(b)} = \left( \frac{d_{ik}^{(b)}}{d_{jk}^{(b)}} \right)^{2}^{-1}
\]

(3)

Step 2: update cluster prototype matrix \( P(b+1) \):

\[
p_i^{(b+1)} = \frac{\sum_{k=1}^{n} (\mu_{ik}^{(b+1)})^m \cdot x_k}{\sum_{k=1}^{n} (\mu_{ik}^{(b+1)})^m}, i = 1, 2, \ldots, c
\]

(4)

Step 3: if \( \| p(b) - p(b+1) \| < \epsilon \), then stop and output partition matrix \( U \) and cluster prototype \( P \), else let \( b = b + 1 \), go to the Step 1. The choice of the weighed index \( m \) plays an important role in adjusting fuzzy degree of clustering, and \( m = 2 \) in this paper. Through FCM the fuzzy partition matrix and cluster prototype can be obtained, and then the membership degree of roof stability on different influence factors.

3. Fuzzy neural networks with information fusion

Information Fusion, also known as multi-sensor information fusion technology, or data fusion technology, makes integration and fusion of the data that come from multiple sensors by imitating experts' capacity of comprehensively processing information, so obtaining more accurate and credible conclusions than using each sensor respectively[2].

Artificial neural network is established based on a type of micro-structure and function that is the simulation of the human neural system, with some simulated ability of thinking in images. The information fusion has strong similarities compared with neural networks in structure, just because the information fusion is based on intelligent thought, and its function to be achieved is to imitate the capability of human brain to deal with all kinds of information, which is very close to the idea of neural network. Having the similarity with neural networks in structure, this paper takes full advantage of their superiority and deals with the relationship and interaction among information processing units, consequently, using the five-layer fuzzy neural network shown in Fig. 1.

The first layer: input \( x_j (j = 1, 2, \ldots, n) \) directly, where \( n \) is the number of input variable.

The second layer: fuzzy membership function layer, activation function is the membership function on a fuzzy subset, which represents the input variable membership. Each unit of the layer is actually a small neural network, as shown in Fig. 2.
The third layer, each node represents a fuzzy rule, these nodes conducting "and" operation to integrate the values inputted in the second layer, furthermore matching the antecedent of fuzzy relations and calculating the degree of application of each rule. The fourth level, the number of nodes is still $m$, calculate to achieve normalization.

$$y_i(k) = \sum_{i=1}^{n} a_i^j(k) \cdot \mu_i^j(k)$$

The fifth layer, i.e. output layer, deblurring by the weighted sum method. Where, $a_i^j$ is the connection weight between $i$th node of fuzzy rule layer and $j$th node of output layer, is also activated intensity of $j$th output associated with $i$th rule, which can be adjusted.\[3\].

4. Parameter training

The parameters to be undertaken are $v_{ij}, \delta_j$ and $w_i$, trained by BP algorithm. Aiming the drawbacks of standard BP algorithm, such as slow convergence rate, being easy to form a local minimum, etc. this paper adopts an improved method, in which the activation function are modified as follows:

$$f(x) = \frac{1}{2} \frac{1}{1 + \exp(-x)}$$

The variable-rate learning method is designed, in which, when the network training being the initial stage, learning rate is assigned a larger constant to increase the learning speed; while the output error has been small enough, the $\eta$ is reduced to decrease its speed. In this way, the accuracy of convergence is ensured and the training precision can be improved.

In this paper the weight adjusting function with momentum term is utilized as follows:

$$a(k+1) = a(k) - \eta \frac{\partial \epsilon_k}{\partial a} |_{a=a(k)} + \beta \Delta a(k)$$

$$\epsilon_k = \frac{1}{2} (y_k - \hat{y}_k)^2$$

where, $a(k)$ is current value of membership function parameters and $a(k+1)$ is its modified values, $\eta$ is the control parameter of speed correction, $\beta$ is the momentum coefficient, $\beta$ is the momentum coefficient, $\epsilon_k$ is the error objective function.
5. Simulation

The above mentioned model is applied to coal mining process with a special complex production conditions, so that the prediction is carried out through the neural network. The eight main factors related closely to the roof stability are conducted to carry out the simulation training, they are mining depth, roof rock feature, structure complexity, angle of coal seam, coal seam thickness, mining method, with or without pillar and mining mode, shown as Table 1.

<table>
<thead>
<tr>
<th>Sample</th>
<th>mining depth</th>
<th>roof rock feature</th>
<th>coal seam thickness</th>
<th>angle of coal seam</th>
<th>with or without pillar</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>521</td>
<td>sandstone</td>
<td>0.6</td>
<td>52</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>752</td>
<td>sandstone</td>
<td>2.0</td>
<td>48</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>855</td>
<td>mudstone</td>
<td>3.1</td>
<td>34</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>560</td>
<td>sandstone</td>
<td>1.2</td>
<td>45</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>530</td>
<td>mudstone</td>
<td>1.4</td>
<td>47</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>490</td>
<td>mudstone</td>
<td>0.8</td>
<td>25</td>
<td>Yes</td>
</tr>
<tr>
<td>7</td>
<td>486</td>
<td>sandstone</td>
<td>1.5</td>
<td>27</td>
<td>Yes</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

5.1 Parameter Selection

The membership degree is used as the network input, which is obtained by clustering the experimental sample at two times.

According to the "Roof classification scheme for gently inclined coal face" suggested by Ministry of Coal Industry of China, the stability of coal mine roof divided into 4 grades: unstable roof, Medium stable roof, stable roof and strong roof. Therefore, the number of output nodes is 4, corresponding to the 4 grades respectively.

The number of cluster is 5 by the hierarchical clustering. In the fuzzy neural network in this paper, the first layer has 8 nodes, which is the same as the dimension; the second layer is the membership function layer and has 40 nodes; The third layer is fuzzy inference layer, has 5 rules that are obtained by the fuzzy c means clustering, thus the layer nodes is also equal to 5; the fourth layer is the de-fuzzy and output layer, has 4 nodes that is correspond to the 4 kinds of stability. Therefore, the neural network has the (8-40-5-4) structure. Through simulation, the initial values of $\eta =0.015$ and $\beta =0.7$, of which, the value of $\eta$ varies with training process.

5.2 Results

In order to avoid overflowing, the samples of mining depth, roof rock feature and coal seam thickness are normalized at first, as shown in Table 2. The discretization processing is made for the other samples, that is, 0 and 1 are used to represent different states, for example, mudstone is expressed as 1 and sandstone as 0, with coal pillar is expressed as 0 and without coal pillar as 1, etc. After normalization process, the samples are partitioned into 5 clusters by the hierarchical clustering algorithm based on field potential topology, and then the fuzzy c-means clustering is run and the results are shown as Fig. 3.

<table>
<thead>
<tr>
<th>Sample</th>
<th>mining depth</th>
<th>angle of coal seam</th>
<th>coal seam thickness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1966</td>
<td>0.6250</td>
<td>0.3263</td>
</tr>
</tbody>
</table>
The fuzzy neural network's parameters derived from clustering results are trained by the self-adaptive learning rate so that the algorithm results converge faster and have less error shown in Fig. 4. The results show that the trained fuzzy neural network showed better fusion effect, which receives more information and optimized parameters.

The prediction system that the fuzzy neural network is adopted to fuse data is superior to the single criterion system.

6. Conclusions

Aiming at the special nature of coal mine roof stability, this paper collects a variety of data and integrate them as a whole, and has carried out the prediction using the fuzzy neural network based on the information fusion after the subtractive clustering and fuzzy c-means clustering. Trained neural network model has satisfied the given demand and can be applied to the simulation prediction. Meanwhile the background database can be used to store data so that the performance of the network model can be improved and its predictability is enhanced continuously. So the system has the real-time property and better usability.

References