Application of Self-organizing Feature Maps Neural Network on Hydrographic Zones Partitioning

Liu Dedong 1, Ruirui Zhu 2, Liang Jinyan 3, Zhang Juan 4

1. Shandong Survey and Design Institute of Water Conservancy, Jinan, China
   east_liu@foxmail.com

2. Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China
   zhurr@igsnrr.ac.cn

3. Bengbu Survey and Design institute of water conservancy, Bengbu, China,
   ljy0552@163.com

4. Yellow River Investigation and Design Institute of Shandong, Jinan, China
   xtvk_2008@yahoo.com.cn

Abstract

Hydrographic zones partitioning is the critical issue in hydrological station network planning and a hydrographic zones partition method that can divide the hydrographic zones objectively is desirable for any region. As a kind of pattern classification problem, the hydrographic zones partitioning of Shandong province is achieved by the self-organizing feature map neural network (SOFM network) which has been successfully applied to pattern classification problems. Based on 50 hydrological stations and 8 basic factors which reflect the land surface and hydroclimate characteristics, Shandong province is automatically divided into 2 hydrographic zones by SOFM network. The average watershed characteristics of each sub-zone are consistent with the local terrain and surface conditions. And based on the partitioning results of SOFM network method, the maximum peak flow and the accuracy are analyzed which are proved reasonable and achieve pass rate of 82%. Which indicate that it is an effective method to use SOFM neural network to divide the hydrographic zones.

Keywords: Hydrographic Zones, SOFM Neural Network, Hydrological Station Network

1. Introduction

The hydrographic zones are the regions divided according to the different regional climate, hydrology and physiographic conditions.

Currently, the general hydrographic zones partitioning methods mainly include geographical landscape method, contour map method, runoff producing characteristics method, storm flood parameter method, hydrological model parameter method and principal component cluster analysis method etc. For some methods, personal experiences have to been relied to determine the hydrographic zones, so different partitioning results could be gotten by different hydrology experts. And some methods are very demanding on the integrality of data, thus they are not suitable for data shortage areas. All of these partitioning methods can be roughly grouped into three categories : (1) physiographic hydrographic zones partitioning method, such as the geographical landscape method; (2) Causes partitioning methods, such as basin hydrological model parameter method. In 1980s and 1990s the Xin’anjiang model has been applied for hydrographic zones partitioning and got initial results in China [1]; (3) Statistical hydrographic zones partitioning method basing on statistical parameters. Hydrographic zones partitioning is the critical issue in hydrological station network planning and gets much attention of hydrologists all over the world. Statistical hydrographic zones partitioning method has been used in America and Canada which can be used to determine the minimum number of hydrological stations in each region and the approximate location of the stations, also, can assess the effectiveness of other sources of information for the station network. Additionally, the method can assess the error of borrowed data in data shortage areas. The statistical hydrographic zones partitioning methods include multiple regression method and grid method, respectively used in America and Canada in common.
In recent years, with the rapid development of information technology, biotechnology and computer technology, some new methods such as the artificial neural network (ANN) have been widely used in various fields. It has become a trend to solve complex hydrologic problems using ANN model, especially to apply the multilayer feedforward neural network to simulate rainfall-runoff process and get reasonable results [3,4,5,6]. Neural network is an important part of human brain and the artificial neural network model is information processing system which simplify and model the structure, function and other basic properties of human brain. Currently there are 40 kinds of ANN models, in which the self-organizing feature map neural network (SOFM network) can solve the pattern recognition and classification problems[7,8]. And hydrographic zones partitioning problem can be regarded as example of pattern recognition and classification. The purpose of this article is to study the SOFM network method of hydrographic zones partitioning, which is applied to divide the hydrographic zones in Shandong of China.

2. Hydrological factors

The hydrographic zones are the hydrologic regions divided according to the different regional climate, hydrology and physiographic conditions. There are good relationship between hydrological factors and physiographic characteristics in the same hydrographic zone with similar hydrologic characteristics and variation. Therefore the hydrologic station should be laid reasonably so as to calculate hydrological characteristics with certain accuracy at the insertion site. To divide the hydrographic zones reasonably, the hydrological factors chosen as the basis of hydrographic zones partition should be relatively independent and reflect the hydrological characteristics of the region.

For the hydrographic zones partitioning study of China’s Shandong province, 50 stations are selected. And the number of annual maximum peak flow data series is 20 years. The selected hydrological factors which reflect the watershed land surface characteristics include the drainage area(AREA), the main channel length(MSL), weighted average river gradient(WMS), the average basin elevation(ABE), Vegetation Fractional Cover (VFC) and geological characteristic index(GCI). Additionally, the average annual rainfall(AAR) and the average maximum day rainfall(AMDR) which relate to climate factors are chosen. The eight basic factors can reflect the impact of land surface characteristics and hydroclimate factors on watershed runoff to some extent.

3. SOFM network fundamentals and algorithm

Finnish scientists Kohonen T. developed self-organizing map theory. He believes that the neurons in neural network adaptively learn to evolve a special detector competent for classify different signals through lateral interactions and competition of the neurons. Kohonen self-organizing feature map neural network (SOFM network) is a kind of clustering method which can classify the inputs automatically according to set learning rules[9]. In the unsupervised case, SOFM network self-organizing learning is performed. The weights connecting the inputs and outputs are adjusted repeatedly and ultimately reflect the distance between the input samples which would be expressed in the competitive layer. The SOFM network simulates the two-dimensional layer structure of neurons in brain and simulates the clustering, self-organization and self-learning function of brain information processing. The algorithm has been widely used in various pattern recognition and classification problems [10].

3.1. SOFM network structure

SOFM neural network is a two-layer or one-dimensional network which consists of the input layer and competition layer. The Input layer receives the samples and the competition layer classify the samples. All neurons of the two layers connect to each other. The neurons of competition layer arrange in the form of two-dimensional or one-dimension matrix[11]. SOFM network topology is shown in Figure 1.
3.2. SOFM Algorithm

SOFM algorithm is a non-teachers clustering method competent for mapping the input pattern to the output layer with original topology. That is, the classification results will be expressed in the competition layer through self-organizing learning of the input mode. Additionally, the repeated learning of the input pattern makes the spatial distribution of the connection weights vector consistent with the probability distribution of input patterns which means that the spatial distribution of connection weights vector can reflect the statistical characteristics of the input mode. In short, SOFM neural network can map a high-dimensional mode to a plane, while maintaining the same topology. If the distances between input points are short, the mapping points of corresponding input points also cluster.

The basic parameters of SOFM network are given below.

1. Input pattern vector:
   \[ X = \{ x_i; x_i = (x_{i1}, x_{i2}, \ldots, x_{ik}) \in \mathbb{R}^k, i = 1, 2, \ldots, M \} \]
   Where, \( M \) is the number of input vectors; \( k \) is the number of network input nodes, that is the dimension of the input pattern vectors; \( W \) is the network connection weights vector.
   \[ W = \{ w_j; \quad w_j = (w_{j1}, w_{j2}, \ldots, w_{jk}) \in \mathbb{R}^k, j = 1, 2, \ldots, N \} \]
   \( N \) is the number of output neurons.

2. \( \alpha(t) \) is the network learning rate, generally \( 0 < \alpha(t) < 1 \) and decrease monotonically with time;

3. \( \text{NE}_j(t) (j = 1, 2, \ldots, N) \) is the neighborhood of the output layer neuron \( j \) which is related to the number of nodes and gradually decrease with the increasing of \( t \). The so-called neighborhood \( \text{NE}_j(t) \) is the region-wide containing a number of neurons in which the winning neuron node \( j \) is the center. This area can be any shape, but generally is symmetrical, such as square or circular area.

SOFM network algorithm steps can be described as follows[12]:

1) Initialization
   - Initialize the weights. Generally the weights from \( k \) neurons of input layer to \( N \) neurons of output layer are given random values from 0 to 1;
   - Define the learning rate function \( \alpha(t) \) and set its initial value \( \alpha(0) \);
③ Define the neighborhood \( NE_j(t) \) of output layer neurons, and set its initial value \( NE_j(0) \). 

\( NE_j(t) \) represents the number of neurons in neighborhood in the learning process of \( t \) times.

2) Normalize the network input vectors: \( x = (x_1, x_2, \cdots, x_k)^T \)

3) Normalize the connection weights \( w_{ij} \) and compute the Euclidean distance \( d_j \) of each weight vector \( w_j \) connecting input mode \( x_i \) with output neuron. \( x_i(t) \) is the output of neuron \( i \) in the input layer at time \( t \). \( w_{ij}(t) \) is the connection weight between input neuron \( i \) and output neuron \( j \). then \( d_j \) is calculated as:

\[
d_j = \left( \sum_{i=1}^{k} \left[ x_i(t) - w_{ij}(t) \right]^2 \right)^{1/2}, \quad 1 \leq j \leq N
\]

4) Select the output node \( j \) with minimum distance as the winning node.

\[
d_j = \min_{1 \leq j \leq N} \{ d_j \}
\]

5) Adjustment all the connection weights vectors between neuron \( j^* \) and neurons in the neighborhood.

\[
w_j(t+1) = w_j(t) + \alpha(t) \left[ x(t) - w_j(t) \right], \quad j \in NE_j(t)
\]

6) Enter the next set of input training vectors and return to step 3) which repeat until all learning modes are provided to the network.

7) New learning rate \( \alpha(t) \) and neighborhood \( NE_j(t) \).

\[
\alpha(t) = \alpha(0) \left( 1 - \frac{t}{T} \right); \quad \alpha(0) \text{ is the initial learning rate, } t \text{ is learning times, } T \text{ is the total learning number.}
\]

\[
NE_j(t) = \text{INT} \left( NE_j(0) \left( 1 - \frac{t}{T} \right) \right); \quad \text{INT}(x) \text{ is rounding function. } \quad NE_j(0) \text{ is the initial value of } \quad NE_j(t).
\]

8) Compute \( t = t + 1 \), then return to step 2) which end until \( t = T \), or achieve a enough study number.

After enough iteration steps or \( t = T \), the weight coefficients will cluster and make the output neurons a topological map of inputs which reflects the structure of the data itself.

### 4. Application

A one-dimensional SOFM network is designed to divide the hydrographic zones of China’s Shandong province using 8 input and 20 output neurons. And the SOFM network is trained by 50 station data with 8 hydrological factors for each station. Before the SOFM network is trained, gave [0,1] random value to the initial weights, and determine the initial learning rate \( n(t) (0 < \eta < 0) \) and the total training number respectively for 0.45 and 6000. The initial values of neighborhood radius are from 12 ~ 15. The results are processed and described by a histogram as showed in Figure 2.
Application of Self-organizing Feature Maps Neural Network on Hydrographic Zones Partitioning
Liu Dedong Ruirui Zhu Liang Jinyan, Zhang Juan

neighborhood radius: 12
station number
output neurons

neighborhood radius: 13
station number
output neurons

neighborhood radius: 14
station number
output neurons
Figure 2. The identification results of hydrological stations under different neighborhood radius

Figure 2 shows there are two areas can be roughly identified but the boundaries between each zone is not obvious when the initial values of neighborhood radius are 12 and 13 respectively. When the initial values of neighborhood radius are 14 and 15, two areas can be clearly identified. Not only the boundaries between each area are significant, but also the corresponding stations in each area are basically uniform which means the partitioning results become stable. So the partition result with initial values of neighborhood radius of 15 is established finally which shows that 2 hydrographic zones with 50 stations are divided.

To compare the SOFM network partitioning method with the traditional partitioning methods, such as principal component cluster analysis, the results of two methods of partitioning are listed in Table 1.

Table 1. The average basin characteristics of 2 sub-zones divided by SOFM and The principal component cluster analysis method(PCCAM)

<table>
<thead>
<tr>
<th>Zones</th>
<th>Sub-Zone I</th>
<th>Sub-Zone II</th>
</tr>
</thead>
<tbody>
<tr>
<td>partitioning method</td>
<td>SOFM</td>
<td>PCCAM</td>
</tr>
<tr>
<td>station number</td>
<td>12</td>
<td>19</td>
</tr>
<tr>
<td>AREA</td>
<td>6784</td>
<td>5046</td>
</tr>
<tr>
<td>MSL</td>
<td>195</td>
<td>174</td>
</tr>
<tr>
<td>WMS</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>ABE</td>
<td>29</td>
<td>34</td>
</tr>
<tr>
<td>VFC</td>
<td>0.38</td>
<td>0.42</td>
</tr>
<tr>
<td>GCI</td>
<td>0.443</td>
<td>0.415</td>
</tr>
<tr>
<td>AAR</td>
<td>553</td>
<td>565</td>
</tr>
<tr>
<td>AMDR</td>
<td>97</td>
<td>94</td>
</tr>
</tbody>
</table>

The principal component cluster analysis method is a common method of data compression, and its main idea is orthogonal transform to retain those feature vectors with significant contribution. Applied to the hydrographic zones partition, it is thought that the anterior principal components with a combination of more hydrological factors represent the collective effect of these hydrological factors. At the geographical sites with the similar main components, the collective effect of hydrology should also alike and the sites should be within one zone. Using the same data, China's Shandong Province hydrological zones are divided by principal component cluster analysis method. The results show that
the geographical distribution of the stations are scattered which can’t reflect the land surface status of the basin. Instead, the partitioning results of SOFM network methods pointed in Figure 2 show that each sub-area boundaries are clear, and from Table 1, it can be informed that 2 sub-zones identified by the SOFM network clearly reflect the different land surface characteristics of the basin in Shandong province. For example, the average drainage area of all hydrological stations is larger with relatively short river length, small gradient and the lower vegetation fractional cover in sub-zone I. And all stations fall on the plains (showed in Figure 3). On the contrary, the stations in sub-zone II fall on the valley basin with higher basin elevation, bigger river gradient and larger vegetation coverage. It is obvious that the hydrographic zones partition results reflect well the local terrain and surface conditions.

The accuracy of hydrographic zones partitioning is quantified established in China’s “Technical Guidelines of hydrological station network planning”. That is the peak flow error is less than 20% and the test pass rate is 70% at least. Based on the partitioning results of SOFM network method, the maximum peak flow and the accuracy are analyzed which are proved reasonable and achieve pass rate of 82%. Additionally, the number of sub-zones can be identified automatically by SOFM neural network. Therefore, the hydrographic zones partition method of SOFM neural network is desirable.

5. Conclusions

By SOFM network method, Shandong province can be divided into two hydrographic zones in which the average watershed characteristics are consistent with the local terrain and surface conditions. The statistical test shows that the partition results meet the accuracy requirements.

The selected 8 hydrological factors fully embody the watershed land surface and climate characteristics and reflect the hydrological characteristics to a greater extent. The choice of hydrological factors is appropriate.

All hydrographic zones are divided by the SOFM network mathematical model which model the self-learning and self-organization function of human brain. Compared with the dividing method depending on experience, the SOFM neural network can automatically identify the number of sub-region, so the dividing results are more objective. The hydrographic zones are the direct basis for region representative hydrological station setting, hydrological data transfer and the hydrological
station network planning research. It is the first attempt to using SOFM network method to study the hydrological station network research in Shandong province of China which achieved satisfactory results.

6. Acknowledgement

The author would like to acknowledge the valuable discussions with Jeff Pompeo.

7. References