A Recognition Method of Surface-Water Based on RBF Neural Network

CHEN Xue-lian1,2,3,4, HU Jing-tao1,2
1 Dept. of Information Service and Intelligent Control, Shenyang institute of automation, CAS Shenyang 110016, China; 2 Key Laboratory of Industrial Information Technology, CAS Shenyang 110016, China; 3 University of Chinese Academy of Sciences Beijing 110000, China; 4Dept. of Information Engineering, Liaoning Provincial College of Communications Shenyang 110122, China
Xuelian_chen@163.com

Abstract—For the extraction of surface-water data characteristics was not easily, and pseudo-depression was easy to confuse with the surface-water in the recognition of surface-water, surface-water radial basis function neural network recognition model was proposed. According to the features of surface-water and the computational result as input and output neural cells, the surface-water radial basis function neural network recognition model was established. The proposed model solved the problem that the distinguishing of surface-water and pseudo-depression was difficult. As the results, the proposed surface-water recognition model can recognize the surface-water efficiently.

Keywords—RBF neural network; potential outlet; recognition of surface-water

I. INTRODUCTION

With the development of information technology, such as RS, GPS, GIS, etc., digital watershed technology play more important role. Digital Elevation Model is a discrete digital expression on the earth’s surface morphology, which reflects the distribution of the area elevation. Digital watershed and the related information extracted from DEM have an important application in the hydrology analysis.

Surface-water is one type of depressions, which was formed by the nature surface morphology, such as the lakes, reservoirs and so on. It is one important part of digital watershed and plays a major role in preventing flood. The traditional extraction methods of digital watershed don’t contain lakes and reservoirs, which is mainly for the researchers take the depressions as pseudo depressions in DEM processing. The pseudo depressions can affecting the flow direction and should be filled in different method. At present, with the rising status of the surface-water in the field, researchers begin to focus on the study of extraction of surface-water from the DEM data, in order to make the digital watershed drainage containing the information of surface-water.

However, owing to exist resolution ration, elevation interpolation and round off error in the original DEM data in the produce of DEM, pseudo depressions will affect the accuracy of the extraction of surface-water. Because the data features between pseudo depressions and surface-water are extremely similar, which make it different to recognize the surface-water, how to identify the surface-water becomes a difficult problem.

Tribe proposed that depression in DEM should be divided into lakes, reservoirs, etc., and pseudo depressions, which respectively represent the real surface morphology and data error. According to the depression and its depth, size and location, depth of specified threshold and area threshold to identify on lake, reservoir, etc.

Xu analyzed DEM data characteristics of the lakes and pseudo depressions, and considered pseudo depressions which resulted from data error should be on a small scale and presented a scattered distribution; depression caused by the lake is real surface morphology, which should have a certain area and the elevation difference, and presents a large area of patchy distribution.

Cheng et al. analyzed the DEM data feature of the lakes and reservoirs from the original DEM data, and pointed out the difference between DEM data feature of the lakes and reservoirs and data feature of the pseudo depressions, which regarded the potential outlet as to identify the unique characteristics of lakes and reservoirs and had great limitations.

This paper analyses the difference between surface-water and pseudo depressions in the data features, and extracts the feature of surface-water including area feature, depth feature and potential outlet feature, proposes a water feature recognition model. According to the characteristics that RBF neural network can fast to pattern classification, a recognition method of surface-water based on RBF was introduced. The model solved the problem using accurate mathematical model to distinguish surface-water and pseudo depressions and provides a new method and thought to deal with surface-water.

II. AN RECOGNITION MODEL OF SURFACE-WATER CHARACTERISTIC

A. Area characteristic of surface-water

The area characteristic of surface-water is a basic data feature, which represents that the size of surface-water
accounts for a geographical area. Calculation method is based the polygon area which is made of the surface of water that depends on the boundary defined as the border between the surface-water and land. Surface-water presents as irregular polygon areas which are consisted of multiple adjacent grid unit in DEM data(see Fig.1). Therefore, the area of surface-water can be calculated by the statistics of the number of grid units in DEM data. Thus the formula of the area characteristic of surface-water can be written as:

\[ f(Area_i) = nl^2 \]  

(1)

Where \( f(Area_i) \) is the area of the surface-water, \( n \) is the total numbers of grid units contained by surface-water, \( l \) is the length of grid unit.

<table>
<thead>
<tr>
<th>25</th>
<th>34</th>
<th>33</th>
<th>29</th>
<th>30</th>
<th>31</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>28</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>31</td>
</tr>
<tr>
<td>28</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>28</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>23</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>28</td>
<td>29</td>
</tr>
<tr>
<td>22</td>
<td>20</td>
<td>24</td>
<td>24</td>
<td>16</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>28</td>
<td>24</td>
<td>23</td>
<td>12</td>
<td>25</td>
<td>26</td>
<td>29</td>
</tr>
</tbody>
</table>

Fig. 1 The area of surface-water

**B. Depth features of surface-water**

Another basic feature of the data is the depth features, which demonstrated the fluctuated level of surface-water. It can be obtained through calculating the difference value between the boundary elevation and the bottom elevation of surface-water, while the boundary is defined as the border between the surface-water and land.

Because the surface-water is usually generated at the bottom of the plane as the center, then spreads around with gradual increase until the border with land. So the depth of surface-water can be obtained by calculating the different value between the maximum of boundary elevation and the minimum of bottom elevation of surface-water. Above all, the computational formula of area feature about surface-water is given as the following form:

\[ f(\text{Depth}) = e_{\text{max}} - e_{\text{min}} \]  

(2)

Where, \( f(\text{Depth}) \), \( e_{\text{max}} \), \( e_{\text{min}} \) donate the depth, the maximum of boundary elevation, the minimum value of bottom elevation of surface-water, respectively.

**C. The potential outlet of the surface-water**

The potential outlet of the surface-water is defined as the boundary outlet which the surface-water flows into the downstream drainage, and the position of the outlet is usually located at the grid cell with the highest flow accumulation in the catchment area of surface-water.

All of the grid cells loaded in the catchment area of surface-water can be the outlet of the surface-water, which is flagged as \( A_1 \) and demonstrated in Fig. 2. In general, the grid cell which gathers up water most would be the outlet position with highest probability and best potentiality. Account to the parameters put forward above, if the grid cell located in the boundary of surface-water meets the conditions that \( \Delta h_1 \) and \( \Delta h_2 \) reach the highest point and \( l_1 \) and \( l_2 \) reach the lowest point, it indicates that this grid cell will gain max flow, which means this grid cell has potential outlet. Then the potential outlet of the surface-water can be obtained by the following formula:

\[ f(\text{outlet}) = \max \{ \psi(e_i) | e_i \in \text{boundary grid cell} \} \]  

(3)

\[ \psi(e_i) = \frac{\Delta h_1 + \Delta h_2}{l_1 + l_2} \]  

(4)

Fig. 2 The potential outlet of surface-water

**D. The model of identification of surface-water features**

In this essay, area features, depth features, potential outlet features are selected as the main parameters which influence the identification of surface-water features. Then the model of identification of surface-water features is established in the following form:

\[ P(i) = \alpha \mu(Area_i) + \beta \mu(Depth_i) + \gamma \mu(\text{Outlet}_i) \]  

(5)

Where \( P(i) \) is the probability of surface-water; \( \alpha, \beta, \gamma \) express the weight coefficients, \( \mu(Area_i) \) is the normalized function of area, \( \mu(Depth_i) \) is the normalized function of depth, \( \mu(\text{Outlet}_i) \) is the normalized function of the potential outlet. When \( P(i) \in [0,0.5] \), the depression type is defined as surface-water, while \( P(i) \in (0,0.5) \), it is defined as pseudo depression type.

This paper statistics area and the depth of characteristic values of 532 lakes, reservoirs and other water area in the national 1:250000 DEM data topographic maps. The statistical data is shown in Tab. 1 and 2. According to the statistical data, the function of \( \mu(Area_i) \) and \( \mu(Depth_i) \) are
confirmed.

Tab. 1 The statistical table of surface-water area

<table>
<thead>
<tr>
<th>Data feature</th>
<th>range</th>
<th>Water quantity</th>
<th>Scale value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area (number of grid cell)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area $\geq 1000$</td>
<td>498</td>
<td>93.6%</td>
<td></td>
</tr>
<tr>
<td>$100 \leq$ Area $&lt; 1000$</td>
<td>32</td>
<td>5.9%</td>
<td></td>
</tr>
<tr>
<td>Area $&lt; 100$</td>
<td>3</td>
<td>0.5%</td>
<td></td>
</tr>
</tbody>
</table>

From Tab. 1, most areas of surface-water are over 10 km² (about 1000 grid cells), accounts for about 93.6% of the total water, drainage areas less than 1 km² account for 0.5%, then the normalization function of areas is deduced as:

$$
\mu(\text{Area}) = \begin{cases} 
1 & \text{Area} \geq 1000 \\
\frac{\text{Area}}{1000} & \text{Area} < 1000 
\end{cases} \quad (6)
$$

Tab. 2 The statistical table of surface-water depth

<table>
<thead>
<tr>
<th>Data feature</th>
<th>range</th>
<th>Water quantity</th>
<th>Scale value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth(m)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depth $\geq 3$</td>
<td>394</td>
<td>74.1%</td>
<td></td>
</tr>
<tr>
<td>Depth $= 2$</td>
<td>123</td>
<td>23.1%</td>
<td></td>
</tr>
<tr>
<td>Depth $= 1$</td>
<td>15</td>
<td>2.8%</td>
<td></td>
</tr>
</tbody>
</table>

From the Tab.2, most depths of the surface-water are over 3m. In the western plateau lakes, the depths of the surface-water are generally darker and over 10m and the maximum depth even exceed 100m in some large lakes. While in the eastern plain region, the depths are usually shallow, generally at 3m, some of them are less than 2m. Then the normalization function of depths is deduced as:

$$
\mu(\text{Depth}) = \begin{cases} 
1 & \text{Depth} \geq 3 \\
3 & \text{Depth} < 3 
\end{cases} \quad (7)
$$

According to the rule of min-max normalized method, the data of potential outlet is linear changed, then the value is mapped to $[0, 1]$. So the normalization function of the potential outlet is deduced as:

$$
\mu(\text{Outlet}) = \frac{\text{Outlet}_{i} - \text{Outlet}_{\text{min}}}{\text{Outlet}_{\text{max}} - \text{Outlet}_{\text{min}}} \quad (8)
$$

Where, $\text{Outlet}_{\text{max}}$, $\text{Outlet}_{\text{min}}$ are separately the maximum data and the minimum data of potential outlet in the total surface-water samples, $\text{Outlet}_{i}$ is the data of the current potential outlet.

III. RBF NATURAL NETWORK IDENTIFICATION MODEL OF SURFACE-WATER

A. RBF neural network

Substantially, the identification process of surface-water is the process of classifying depression data, and evaluation of depressions classification can not only reflect the accuracy of the classification model mapping, but also play an important role in selection methods of feature about classification of depression and the comparison between classification methods.

In this paper, the RBF neural network is used as a classifier to identify the surface-water and the subtraction clustering algorithm is applied to learning of the above network center parameters [12]. RBF neural network is composed of input layer, hidden layer and output layer. Input layer nodes only input signals to the hidden layer, hidden layer node is composed of basis functions, output layer node is a simple linear function. The model structure is performed as:

$$
y_k = \sum_{i=1}^{n} w_k R_i(x) \quad (9)
$$

Where, $x$ represents the input layer neurons; $R(x)$ is basis functions of the hidden layer; $R(x)$ is the output of $i_{th}$ nerve cell in the hidden layer; $w_k$ is the output layer weight.

A. The RBF neural network identification model of surface-water

The RBF neural network input is $x=[\text{Area}, \text{Depth}, \text{Outlet}]$, and the gaussian function is made as the basis functions of the hidden layer, which the expression is as follows:

$$
R_i(x) = \exp(-||x - c_i||^2/2\delta_i^2) \quad (10)
$$

Where, $x$ is input layer neuron; $c_i$ is the center of $i_{th}$ basic function; $\delta_i$ is variance of $i_{th}$ basic function; $||x - c_i||$ is the Euclidean norm, demonstrating the distance between the $x$ and $c_i$, then the RBF neural network identification model structure of surface-water can be deduced as:

$$
y_k = \sum_{i=1}^{n} w_{ij} \exp(-||x - c_i||^2/2\delta_i^2) \quad (11)
$$

B. Training algorithm of surface-water RBF neural network recognition model

There are many ways to learn RBF neural network. The most commonly used learning methods include: random selection center method, self-organizing center selected method, supervised select center and orthogonal least squares method. In this paper, self-organizing center selected method is adopted to train the surface-water RBF neural network recognition model. It includes two stages: one is the self-organizing learning stage, namely which implies the centre $c_i$ and variance $\delta_i$ of layer basis functions; the other is supervised learning stage, that is, the learning output layer weights $w_{ij}$.

1) Determine the center of the hidden layer basis function
This paper uses subtractive clustering method \cite{13}, which is a simple and effective clustering algorithm, to determine the center of the hidden layer basis function. It does not need to determine the clustering number in advance, and uses input samples as the candidate set of clustering center to calculate density index of each sample data, which is used to determine the final clustering center. This method can effectively reflect the distribution of data. The density index of sample $x_i$ is calculated by:

$$D_i = \sum_{j=1}^{n} \exp \left( \frac{-\|x_i - x_j\|^2}{(\delta_x/2)^2} \right), \quad (\delta_x > 0). \quad (12)$$

The maximum value of the sample points is selected for the first clustering center, and this sample point is marked as $x_1$, the corresponding density index as $D_1$. The density index of sample point is modified by equation (13), which is given by

$$D_i = D_1 - D_1 \exp \left( \frac{-\|x_i - x_1\|^2}{(\delta_x/2)^2} \right)$$

where, $\delta_x \geq 1.5 \delta_x$ is selected usually to avoid cluster centers in close proximity.

Based on the above principle, clustering center $x_{s2}$ is chosen orderly, which is used to modify density index of data points again. Clustering is terminated when the condition $D_{\max} / D_1 < \varepsilon$ is fulfilled, and the final cluster center is obtained.

2) Determine the variance of the hidden layer basis function

The variance $\delta_i$ of the hidden layer basis function represents the distribution width of sample. The average distance between basis function centers and the concentrated sample mode of sub-sample is used to calculate the variance.

$$\delta_i = \frac{c_{\max}}{\sqrt{2h}} \quad i = 1, 2, \ldots, h \quad (14)$$

where $c_{\max}$ is the maximum distance between the selected center, $h$ the number of hidden layer basis function.

3) Determined the connection weights $w_j$ from the hidden layer to the output layer

If the output of the basis function is $A = (a_j)_{n \times n}$, the output of the training sample is $y = [y_1, y_2, \ldots, y_n]$ and the weighting coefficient is $\omega = [\omega_1, \omega_2, \ldots, \omega_n]$.

The connective weights $w_j$ between hidden layer and output layer are got by Least Mean Square Algorithm. In this paper, Moody-Darken algorithm \cite{14} is adopted to solve the connective weights. The equation is given as follow

$$\omega = \exp(-h \frac{c_{\max}}{c_{\max} - \|x_p - c_i\|^2})$$

where, $c_{\max}$ refers to the maximum distance between the selected center, $h$ the number of hidden layer basis function.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

To verify the classification capacity of the proposed surface-water RBF neural network recognition model, 150 depression data are trained and tested, and the input and output data are acquired according to the formula 6, among which the variable of input neural cells is 3d variable, that is the area, depth and potential outlet feature, and the variable of output neural cells is 1d variable, that is the category of depression, 1 for surface-water and 2 for pseudo-depression. Each type of samples are divided into training sample set and test sample set, they are not overlapped and all do not include repeated samples.

After determining the experimental sample data, simulation experiments about the proposed surface-water RBF neural network recognition model will be proceed. Among 150 group experimental data, 100 group experimental data were chosen as training samples to train the neural network model and other 50 group experimental data to verify the precision of the model. Then, using subtractive clustering method to train the neural network model and the minimum average relative error of samples during the training is 2.75%, the final classification results are shown in figure 5, in which horizontal coordinates are sample number, vertical coordinates are classification results and ‘*’ stands for depression category, ‘o’ for classification results of the neural network model, overlapping points of ‘*’ and ‘o’ for the points having the same classification results. In the classification results, 49 samples have the same classification results, so the identification accuracy is 98%. Recognition error statistics of test samples are shown in table 4. There is one simple whose classification result is different from actual classification result, mistaken for pseudo-depression not surface-water and the main reason is that the area of this surface-water is larger, but depth is smaller and potential outlet feature is not apparent. This type of surface-water is usually small lake or reservoir, etc. and will not influence the recognition of typical surface-water. In a word, the surface-water RBF neural network recognition model can accurately identify the information of surface-water.

<table>
<thead>
<tr>
<th>Depression category</th>
<th>sample number</th>
<th>identification number</th>
<th>error</th>
</tr>
</thead>
<tbody>
<tr>
<td>surface-water</td>
<td>35</td>
<td>34</td>
<td>2.9%</td>
</tr>
<tr>
<td>pseudo-depression</td>
<td>15</td>
<td>15</td>
<td>0%</td>
</tr>
</tbody>
</table>

Tab.4 Recognition error statistics of test samples
V. CONCLUSIONS

A surface-water RBF neural network recognition model was proposed in this paper for the extraction of surface-water data characteristics based on DEM digital river network was not easily. Considering the difference of data characteristics between pseudo-depression and the surface-water, the surface-water recognition model was established by statistic and analysis of the area, depth and potential outlet feature of the surface-water. Basing on this, regarding the RBF neural network as the surface-water classifier, the area, depth and potential outlet feature of the surface-water as input neural cells, the computational results based on the surface-water model as output neural cells, the surface-water RBF neural network recognition model was established. Then by using the self-organization select center method to train the established neural network model and subtractive clustering method to direct the learning of basis function center.

To verify the classification accuracy of the surface-water RBF neural network recognition model, 100 group experimental data were chosen as training samples to train the neural network model and 50 group experimental data as verification samples to verify the neural network model. The experimental results show that the identification accuracy of the proposed model is 98% and the method proposed in this paper can recognize the surface-water efficiently. In the following research, improvement and optimization of the surface-water recognition model and subtractive clustering method to direct the learning of basis function center will be the focus of our study.

VI. REFERENCES