

A DISCUSSION OF RIVER STRUCTURE KNOWLEDGE DISCOVERY METHOD BASED ON RIVER CHARACTERS

YANG Pin-fu^{a,*}, DU Qing-yun^a

^a School of Resources and Environmental Science, Wuhan University, 129 Luoyu Road, Wuhan, P.R.China, 430079
yangpinfu@telecarto.com

KEY WORDS: Spatial Data Mining, River Structure, Characteristic Analysis, Spatial Analysis

ABSTRACT:

After given a brief introduction on spatial data mining, the knowledge of river structure and the relations between mainstreams and branches were discussed. Some parameters were imported as characteristic variable to distinguish various features from river structure in digital data. Then a method as **Binary Tree Hierarchy** was designed to analysis the digital map using these linear parameters. At last an experiment was given to validate the method to discover the patterns of these test data.

1. INTRODUCTION

Spatial data mining, or knowledge discovery in large spatial databases, is the process of extracting implicit knowledge, spatial relations, or other patterns not explicitly stored in spatial databases. SDM is a synthetical technology that relates to Computer Technology, Artificial Intelligence, Data Collection, Information Processing, Knowledge Engineering, Remote Sensing, GIS etc. There are many methods to discover spatial knowledge, such as Spatial Statistics, Spatial Induce, Spatial Clustering, Spatial Trend Detector, Visualization, etc.

River is a kind of important feature in GIS. How to get the form and characteristics from digital data is a basement to fully understand them. In this article, a method to distinguish the various river structures was proposed, which was integrated Spatial Data Mining with Expert Knowledge to discover patterns river distribution.

2. KNOWN KNOWLEDGE OF RIVER STRUCTURES

2.1 River Structure Forms

River structure shows the relations between mainstreams and branches in an area. The patterns of river distribution can be categorized as follows:

1. Arborization form. An important feature is that there are many branches and they arrange disorderly. The mainstreams and branches mostly conflux as an acute angle.
2. Latticed form. An important feature is that mainstreams and branches conflux as a right angle or almost right angle.
3. Parallel form. The mainstreams almost parallel with each other, and the branches almost parallel with each other too.
4. Radiate form. It exists in the volcano mountain mostly. An important feature is that the river flows around from the mountaintop.
5. Confluent form. The rivers flow to the center of basin as confluent streams from around.
6. Reticulate form. The river branches are connected with some others, and there are many flow network in these branches.

7. Ring form. The mainstreams are formed as a ring type. In the follow of this article, we take an experiment to get the knowledge of some river form such as Arborization form, Latticed form, Parallel form etc., which based on the feature of river's relations.

2.2 Study on River Structure Feature

River Network Structure is a compositive relation form that structured with many rivers in an area. In the digital data processing, it is a useful method to study the river structure based on the spatial relation of the rivers. Following are some main relations between rivers:

1. Confluent relations between mainstreams and branches, which means an angle that a branch flows to a mainstream.
2. Length ratio of mainstreams and branches, which presents the difference of length between branches and their closer grade's mainstreams. The difference of length ratio shows the length distribution of different grade's river.
3. River loop is another feature in river structures, which means the connectivity between different rivers.
4. Elevation distribution, which displays the elevation of headstream and endpoint of rivers.
5. Number of river classes while distinguish the rivers as grades with some character parameters.

There are some character relations that show the river structure in Table 1.

Base on the spatial relation, this article proposed a method to discover the implicit knowledge of river structures. Some other parameters were introduced to division different rivers in digital map environment.

* YANG Pin-fu, Ph. D candidate, His major research interests include GIS, Spatial Data Mining

| River Forms | Spatial Relations | | | | |
|-------------------|---------------------------|---------------------------------|----------------|-------------------------------------|--------------------|
| | Confluent relation | Length Ratio | Has River loop | Elevation Distribution | Numbers of Classes |
| Arborization form | conflux as an acute angle | ratio is small, change is great | No | headstream at different elevation | much |
| Latticed form | conflux as a right angle | ratio is great, change is great | No | headstream at different elevation | much |
| Parallel form | conflux as an acute angle | ratio is small, change is small | No | headstream almost at same elevation | little |
| Reticulate form | --- | --- | Yes | developed in plain area | --- |
| Radiate form | conflux as an acute angle | ratio is small, change is small | No | headstream at same elevation | little |

Table 1. Features of river structure

3. PARAMETERS FOR RIVER STRUCTURE ANALYSIS

There are many methods to evaluate river structures in an area including qualitative analysis and quantitative analysis. In this article a method was discussed to evaluate river structures using character parameters and interaction methods in the processing in digital map data. The parameters we used can be recorded as River Length, Average Length and Statistics of Length Frequency, and Curvature, Network analysis and Analysis of Flow Direction and River Angle etc. They will be introduced in follows.

3.1 Curvature

Curvature reflects the curve degree of a linear feature. It is a statistical character of a linear feature distribution. It is difficult to calculate curvature of a linear feature in a strict mathematical method. Usually, the statistical curvature was used instead of mathematical curvature in practice. It was defined as a ratio of the total length(L_t) of a linear feature with the distance(L_d) of start point and end point of a linear feature(see formula 1).

$$w = \frac{\text{TotalLength } h_L_t}{\text{Distance } _L_d} \quad (1)$$

There are three types of curvature such as relative curvature, contour curvature and total curvature. In the experiment of this article, the total curvature was used as curvature of linear feature, which means the ratio of a curve length with the length of its closed side.

3.2 Length, Average Length and Statistics of Length Frequency

Length is a typical character of a linear feature, which shows its stretch range. It was expressed by a set of coordinates in digital. And the length of it can be calculate by cumulate the distance

of two sequential points in the points set approximatively (see formula 2).

$$L = \sum_{i=1}^n ((x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2)^{1/2} \quad (2)$$

Average length shows the common status of all rivers in an area. It is a statistical variable, which was defined by the ratio of total length of all rivers in the study area with the numbers of these rivers (see formula 3).

$$\bar{L} = \frac{1}{n} \sum_{i=1}^n (L_i) \quad (3)$$

Statistics of Length Frequency is the statistical number of different grades of rivers with length parameters. It shows the distribution of length in different river grades in an area. Also it can distinguish the statistical character from different river structure form. Sometimes we used the relation of mainstreams and branches to calculate the statistical variable of rivers.

3.3 Network Analysis

A network was composed of nodes and sides. In an area range, it is a loop network while some paths connect with each other as closed loop. In the natural condition of river developed, the loop network only exists in the plain area where rivers developed irregularly. If there has a closed loop network, it is easy to be recognized as reticulate river structure.

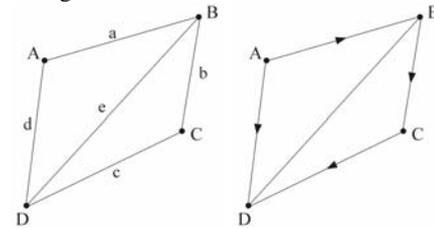


Figure 1. A Simple Network Structure

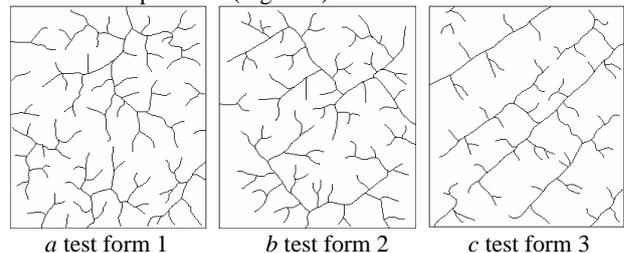
3.4 Analysis of Flow Direction and River Angle

Flow direction is the direction of river developed, where the water flows. River angle means two things. One is the confluent angle of two rivers, especially confluent angle of mainstreams and their branches, which reflects the relations of them. Another means the river's direction angle in a plain coordinate system. Mostly the direction angle of mainstream is the flow direction of the rivers in an area.

4. KNOWLEDGE DISCOVERY IN RIVER STRUCTURE

4.1 Data Organization of This Experiment

In order to try the method to find discover pattern knowledge of river structure, we selected some typical data of river form for our experiment. The purpose of select these data was to improve the results in our experiment and to reduce the complexity in natural river data. Follows is the test data we used in the experiment (Figure 2).



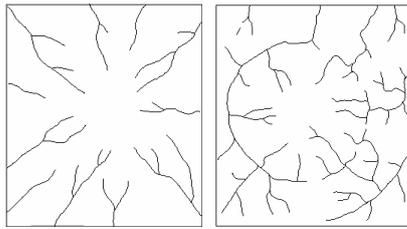


Figure 2. Data Organization of Some Test River Structures

4.2 Method Designed to Discover Knowledge of River Structure

Follow is a method designed to evaluate river structures using parameters of linear feature based on the contribution to distinguish different river structures. These parameters will be used in different steps to principally distinguish river structure. An algorithm flow was designed in Figure3 to discover knowledge of river structure.

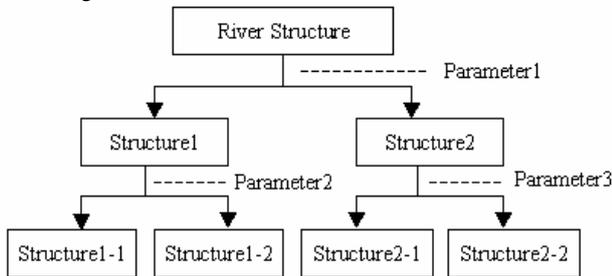


Figure 3. Binary Tree Hierarchy for River Structure Analysis

This method can be described as follow steps:

1. Constructed the statistical feature of rivers used these parameters, and distinguished them as one or more potential structure forms.
2. Analysis the results in previous step with another parameters, and distinguished them as others subsection structure forms.
3. Analysis the results one or more times with other parameters, until the structures were distinguished from others, and knowledge of river structure forms was discovered.

4.3 Processing and Result in Experiments

Using the designed method to discover pattern knowledge river structure, it was analysed with parameters through follow steps, integrated with known knowledge of rivers.

1. Network analysis. After a network analysis with digital river data, if there were some loop networks, a reticulate form can be found out. Otherwise it could be recognized as non-reticulate form.
2. Analysis of flow direction and confluent angle. To distinguish the character of different river structure, confluent angle of mainstream and branches (for example see. Table 2), and flow direction of mainstream (for example see Table 3) were discussed.
3. Curvature analysis. Different rivers curvature was calculated and statistics was given to show the difference of curvature in different river structures (for example see Table 4).
4. Analysis of length and statistics of length frequency. The statistics of river length and river frequency show the different spatial feature in different river

structure developed area. Table 5 is an example of length statistics and frequency statistics in the test data.

| Classification Results | Relation of Confluent Angle | Potential Forms | Ratio of Confluent Angle |
|------------------------|---------------------------------|--------------------------------|--------------------------|
| Class1-1 | conflux as an acute angle | Radiate form (<i>d</i>) | 100% |
| | | Ring form (<i>e</i>) | 75% |
| | | Arborization form (<i>a</i>) | 79% |
| | | Parallel form | --- |
| Class1-2 | conflux as almost a right angle | Latticed form (<i>b</i>) | 45% |
| | | Parallel form (<i>c</i>) | 53% |

Table 2. Classification Based on the Relation of River's Confluent Angle

| Classification Results | River Forms | Potential Forms | Feature of flow direction |
|------------------------|------------------|--------------------------------|---|
| Class 1-1-1 | Radiate form | Radiate form(<i>d</i>) | direction of mainstreams is different, and have same headstream |
| Class 1-1-2 | Non-radiate form | Ring form (<i>e</i>) | direction is disorder |
| | | Arborization form (<i>a</i>) | |
| | | Parallel form (<i>c</i>) | flow direction is same |
| Class 1-2-1 | Latticed Form | Latticed form (<i>b</i>) | flow direction is same |
| Class 1-2-2 | Parallel Form | Parallel form (<i>c</i>) | flow direction is same |

Table 3. Classification Based on the Flow Direction of River

| River Classification | Potential form | Curvature of Mainstreams | Curvature of Branches |
|----------------------|--------------------------------|--------------------------|-----------------------|
| little curve river | Parallel form (<i>c</i>) | 1.055 | 1.186 |
| | Radiate form(<i>d</i>) | 1.032 | 1.035 |
| | Arborization form (<i>a</i>) | 1.154 | 1.150 |
| obtuse curve river | Arborization form | --- | --- |
| big curve river | Latticed form (<i>b</i>) | 1.695 | 1.160 |
| | Ring form (<i>e</i>) | 1.421 | 1.106 |

Table 4. Classification Based on the Curvature of River

| Classification | mainstream | | Branches of Grade 1 | | Branches of Grade 2 | |
|--------------------------------|------------|--------|---------------------|--------|---------------------|--------|
| | Num. | Length | Num. | Length | Num. | Length |
| Parallel form (<i>c</i>) | 4 | 0.7773 | 22 | 0.2188 | 37 | 0.1207 |
| Radiate form(<i>d</i>) | 11 | 0.4927 | 12 | 0.2151 | --- | --- |
| Arborization form (<i>a</i>) | 6 | 0.5973 | 34 | 0.2033 | 36 | 0.1240 |
| Latticed form (<i>b</i>) | 5 | 0.8009 | 31 | 0.1949 | 36 | 0.1105 |
| Ring form (<i>e</i>) | 4 | 0.9791 | 24 | 0.2108 | 20 | 0.1430 |

Table 5. Statistics of River Length

After previous analysis with parameters of linear feature, such as Network analysis, flow direction and confluent angle analysis, Curvature analysis, and statistics of river length and frequency, we can conclude that the *test form 1* is an arborization form, and *form 2* is a latticed form, *form 3* is a parallel form, *form 4* is a radiate form, and *form 5* is a ring form. There was the knowledge we find out in this experiment.

5. CONCLUSION AND FEATURE WORK

In this article, a knowledge discovery process to discover knowledge of river structures was discussed with the hypothetical digital data of river features. And a method as Binary Tree Hierarchy was proposed. After a list of analysis dealt to the test data, it was confirmed as a special river structure as knowledge. It was proved as a useful method to discover pattern knowledge of river structures with these parameters.

There are many works to do in the future studies, such as using more parameters and improving the efficiency of them. Also it needs to study that how to effectively use the known knowledge in the process. Furthermore, it would be developed and improved to suit to analyse natural river data to discover knowledge of river structure, and to build an operable SDM system.

REFERENCE

- Di Kaichang, 2001. *Spatial Data Mining and Knowledge Discovery*, Wuhan University Press
- Han J. and Kamber M., 2000. *Data Mining: Concepts and Techniques*. San matel, CA: Morgan Kaufmann.
- Guorui Zhu, Genshou Zhang, 1994. *Map Analysis*, Mapping Publishing Press
- LI Deren, WANG Shuliang and SHI Wenzhong etc., 2001. *On Spatial Data Mining and Knowledge Discovery (SDMKD)*, Geomatics and Information Science of Wuhan University, Vol. 26
- Di Kaichang, Li Deren and Li Deyi, 1997. *A Framework of Spatial Data Mining and Knowledge Discovery*, Journal of Wuhan Technical University of Surveying and Mapping, Vol.22
- Wu Hehai, 1995, Automation Construct of river tree-like network, Journal of Wuhan Technical University of Surveying and Mapping
- Guo Qingsheng, 1999. Analysing the Characters of the Networks of Rivers and Structuralizing the Tree-like Network of Rivers Automatically, Surveying and Mapping of Geology and Mineral Resources
- Ai Zixing, 1995. The River Net Structural Model And Its Construction in Geography Information System, Surveying and Mapping of Sichuan
- Miller H. J., Han J. 2001. *Geographic Data Mining and Knowledge Discovery: An Overview*. In H. J. Miller and J. Han (eds.) *Geographic Data Mining and Knowledge Discovery*. London: Taylor and Francis.