A STUDY AND EXPERIMENT ON SPATIAL STATISTICAL CHARACTERISTIC TO DISCOVER RIVER STRUCTURE KNOWLEDGE

YANG Pinfu[a], WU Xiaoyan[b]

[a]Changjiang Waterway Bureau, 16 Jiefang Park Road, Jiang'an District, Wuhan, Hubei, China, 430010, - pinfu.yang@gmail.com
[b]Wuhan Railway Bridge Secondary School, No.2 Qiaoji Xincun, Yingwu Road, Hanyang District, Wuhan, Hubei, China, 430052

KEY WORDS: Spatial Statistics, Data Mining, River Structure, Classification

ABSTRACT:

River structure is a direct concept for the distribution of rivers in nature. In spatial database, river structure is a type of spatial relations composed by river features. This paper proposes the characteristic properties and classification rules to classify the river forms based on spatial statistics. Also a method to discover the classification knowledge of river forms is presented based on the data processing and statistical analysis of river features in spatial databases.

1. INTRODUCTION

Spatial Data Mining (SDM), or Knowledge Discovery from spatial database, refers to the extraction of implicit knowledge, spatial relationships, or other patterns that are not explicitly stored in spatial database (Han, 1999). Spatial Data Mining and Knowledge Discovery is a process to extract the implicit but interesting spatial and none-spatial patterns or common characteristics from spatial database (Di, 2001).

The methods of Spatial Data Mining can be classified into two categories. The first is primarily based on attributed information. This method applies statistical methods to attributed information of geospatial features with traditional data mining methods, and converts the results into spatial rules. The methods usually used include inductive learning, association rule, rough set fuzzy set, cloud theory, etc. The other category deals with spatial features, and analyzes location and spatial relations of features, and directly derives spatial rules. The methods include spatial statistics and spatial analysis, clustering and classification, characterization and trend detection, geo-informatics graphic methodology etc. In this article we investigate an application to extract river structure knowledge with SDM dominated by spatial features.

2. SPATIAL STATISTICS OF RIVER CHARACTERS

Spatial Statistics is a science developed from statistics, which focus the similarity between the adjacent spatial locations. In other words, spatial neighbor relations of the study objects are related to the similarity among the attributes in the locations (Anselin, 1992). In the studies of nature science, some rules are induced from statistical results, e.g., the distribution rule of river class with river length, the distribution rule of river length with river area (such as Horton rules, Hack rules) etc.

Spatial analysis is an important method for spatial data analysis and processing. It is developed from spatial statistics. In the spatial analysis of geographic features, there are two categories of spatial analysis. The first is spatial distribution analysis, which studies point-line relation, line-line relation, point-area relation, line-area relation, area-area relation etc. And the other is spatial relation analysis, which studies distance relation, orientation (direction) relation, topological relation, similarity relation and correlative relation etc.

Rivers in the nature distribute as special forms (that is the river flow) which flow from upper to lower. And the flow direction of river can be replaced as the order of river feature digitized in spatial database. Base on statistical analysis of river flow direction and river angle, Haralick etc. (1985) develop a method to retrieve river direction and the reasoning relations. De Serres and G. Roy（1990）develop the parameters to extract river direction and the reasoning rules based on the study of Haralick. Paiva J. Egenhofer M. and Frank A.（1992）investigate the method to deduce the river direction based on ontology theory. These analyses of river direction and river relation are obtained using spatial statistics methods.

3. A STUDY OF SPATIAL STATISTICAL METHOD FOR RIVER STRUCTURE CLASSIFICATION

3.1 Characteristics of river structures

The structure characteristics of river networks, which are also named as forms of river system, are descriptions of relations of river features in spatial databases, such as scale, connectivity etc. There are many research papers that discuss forms of river networks and the characteristic differences among them. However, very few studies have been conducted in terms of classification and recognition of them in spatial databases. The automated intelligent mechanisms to recognize the forms are scare. The forms of river networks can be classified as *arborization form, latticed form, parallel form, radiate form, confluent form, reticulate form, ring-like form* etc. (Zhu 1992, Guo 1999). There are different statistical characteristics in these forms, which can be calculated in spatial databases.

3.2 Properties for the classification of river structure

a. Network analysis

A network is composed of nodes and edges. Within a specific area, it is a loop network while some paths connect with each other as closed loop. Network in river features is a directed network, which is the closed path starting from two rivers on same node joined together in the flow direction. Characteristics of network are the essential to discover the reticulate form.

b. Analysis of river angle

Angle is the intersect relation of two rivers, showing intersect features of different class river. Serres and Roy (1990) contend that the joint angles in river features of arborization forms are more than 88% acute angle. This rule presents an foundation to classify river forms. The river joint angle shows different statistical values in different river forms, e.g. the acute angle is dominant in arborization forms, but many approximately right angles exist in lattice forms.

c. Analysis of flow direction

Flow direction shows the geologic structure in the area of river flow. The flow of mainstream is dominant in river forms, which express the trend of river features and influence the direction development of other rivers. The river forms can be distinguished by analyzing flow directions, e.g. in the parallel forms of river, the flow direction of mainstream is almost parallel and the flow direction of first branch is also almost parallel.

d. Curvature

Curvature reflects the curve degree of a linear feature. It is a statistical concept of a linear feature distribution. Curvature of river features can be calculated as the ratio of river length to distance from the start point to the end point in a river feature. Generally, the river curvature is a total curvature which means the ratio of river length to the length of its closed line.

e. Length, Average Length and Statistics of Length Frequency

Length is a typical property of a linear feature, which shows its stretch range. It was expressed by a set of coordinates in digital, and the length can be calculated by cumulating the distance of the coordinate points approximately (see formula 1).

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

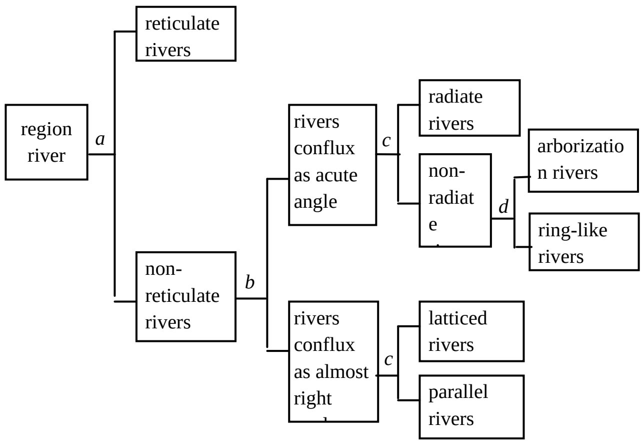
Average length shows the common status of all rivers in an area. It is a statistical variable, which can be calculated as formula 2.

$$\overline{L} = \frac{1}{N} \sum_{i=1}^{N} L_i \quad \text{(Formula 2)}$$

Statistics of length frequency of rivers is a distribution statistics of river length in different classes, which reflects the development of rivers in an area. The statistics of length frequency is usually calculated with the relation of mainstreams and branches.

3.3 A study on the method to discover river structure

The classification rules are given based on the properties of river forms and differences in these properties. We proposed a Hierarchical Method to classify the river forms using these rules (see Figure1). That is a classification process used these properties one by one, until the form of river is classified. Firstly the network property is used to classify the river forms as *reticulate form* and *non-reticulate form*. Secondly the analysis of river angle is conducted to classify as *rivers conflux as acute angle* and *rivers conflux as almost right angle*. Thirdly, the analysis of flow direction is conducted to classify the *rivers conflux as acute angle* as *radiate forms* and *non-radiate rivers*, and the *rivers conflux as almost right angle* as *latticed forms* and *parallel forms*. Lastly, the curvature is used to classify the *non-radiate rivers* as *arborization forms* and *ring-like forms*. The length property and the statistics of length frequency can be used to check the results of classification of river forms.



Properties: a-Network property, b-river angle property, c-flow direction property, d- curvature

Figure1 Hierarchical Method for the classification of River Networks

4. AN EXPERIMENT TO CLASSIFY RIVER FORMS

In the experiment data, river features are organized as direct graph, which is composed of edges and nodes. The direction of edges reflects the flow direction, and in-degree and our-degree of node reflect the distribution of river direction. In-degree is the number of upper rivers which flow to the node, while out-degree is the number of lower rivers which flow from the node. Figure 2 presents a direct graph of rivers, and the corresponding distribution of in-degree and out-degree is given in Table1. The position of a node can be inferred based on its in-degree and out degree. If in-degree is 0 and out-degree is 1, the node may be a start-point of river. If in-degree is 1 and out-degree is 0, the node may be end-point of river. If in-degree and out-degree is all 1, the node must be a medial point. If in-degree is more than 1 and out-degree is 1, the node almost is a conflux point.

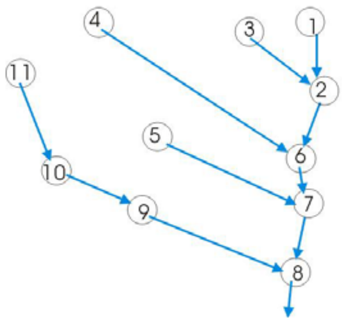


Figure2 A Directed Graph of River Networks

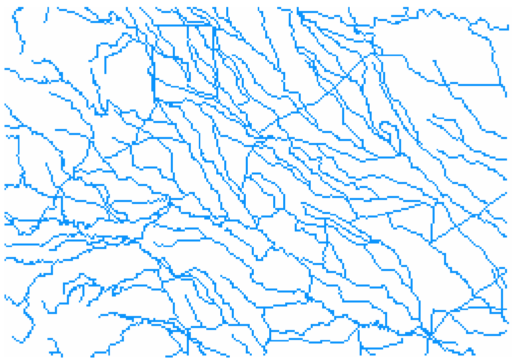
node-num	out-degree	in-degree	node-num	out-degree	in-degree
1	1	0	7	1	2
2	1	2	8	1	2
3	1	0	9	1	1
4	1	0	10	1	1
5	1	0	11	1	0
6	1	2			

Table1 In-degrees and Out-degrees of Nodes

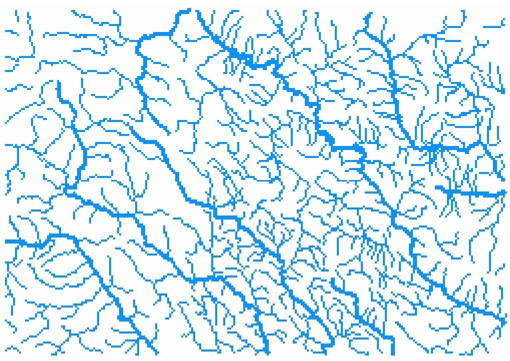
Rules are foundation to classify the river forms. The Expert Rules for classification of river forms (see Table2), which are experience rules, are built based on existing research results and some advices from experts in our study.

	out-degree of notes more than 2 (%)	net-work rings	joint angles		flow directions	curvature	length
			acute angle	almost right angle			
reticulate rivers	>20	Yes	--	--	--	--	--
arborization rivers	<20	No	>70	<30	Flow direction is disorder, but goes to mainstream	<1.5	The ratio of length between different orders is small, but the change is great
latticed rivers	<20	No	<60	>40	Flow direction maybe break	>1.5	The ratio and the change are all great
parallel rivers	<20	No	<60	>40	Flow direction of mainstreams are similar, and braches' are also	1.2~1.5	The ratio and the change are all small
ring-like rivers	<20	No	>60	<40	Flow direction is disorder, but goes to mainstream	>1.5	The ratio and the change are all great

Table2 Rules for River Structure Classification



(a) Reticulate rivers



(b) Arborization rivers

Figure3 Two Basins of Experiment Data

Figure 3 shows experiment data extracted from the 1:1000000 database of china. Data processing was taken and data was organized as direct graph firstly. And then the properties was calculated and the river forms was discovered used our Hierarchical Method. The result is given in Table3. Compare with Experts Rules proposed in our study, we can find that rivers in *data a* should be reticulate form, but rivers in *data b* should be arborization form.

	out-degree of notes more than 2 (%)	network rings	angle analysis	curvature
data a	57	Yes	-	-
data b	-	No	85% right angle	1.374 (mainstream) 1.208 (all river)

Table3 Calculated property in experiment data

5. CONCLUSION AND FUTURE WORK

Spatial data mining is an important direction in the research of artificial intelligence, which is a method to promote the value of spatial database. River features are fundamental features in spatial database, which stored as features lack of semantics of river structure. To discover the semantics of river structure automatically is important to improve the capability of spatial analysis for river features. This paper proposed a data mining method for river structure based on spatial statistical method, which is useful to classification problem for small scale data as demonstrated in the experiment.

There are many tasks to do in future research, such as using more properties and improving the efficiency. Also it needs to study that how to effectively use the known knowledge in the process. Furthermore, it should be developed and improved to be applicable to analyze natural river data for river structure knowledge, and to build an operable SDM system.

REFERENCES

Di Kaichang, 2001. *Spatial Data Mining and Knowledge Discovery*. Wuhan University Press

GUO Qing-sheng, 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.*

Han J. and Kamber M., 2000. *Data Mining: Concepts and Techniques.* San matel, CA: Morgan Kaufmann.

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.

Paiva J. and Egenhofer Max J., 2002. Robust Inference of the Flow Direction in River Networks. *Algorithmica*

Peter Sheridan Dodds and Daniel H. Rothman, 2000. Geometry of river networks I: Scaling, fluctuations, and deviations. *PHYSICAL REVIEW E*, Vol.63

Peter Sheridan Dodds and Daniel H. Rothman, 2000. Geometry of river networks II: Distributions of component size and number. *PHYSICAL REVIEW E*, Vol.63

Peter Sheridan Dodds and Daniel H. Rothman, 2000. Geometry of river networks III: Characterization of component connectivity. *PHYSICAL REVIEW E*, Vol.63

Serres B.de and Roy A., 1990. Flow Direction and Branching Geometry at Junctions in Dendritic River Networks. *The Professional Geographer*, 42(2): 149~201

Wu Hehai, 1995. Automation Construct of river tree-like network. *Journal of Wuhan Technical University of Surveying and Mapping*

ZHU Guorui, ZHANG Genshou, 1994. *Map Analysis*, Mapping Publishing Press