A Heuristic Algorithm for Automatic Classification of Ring River

To cite this article: Chengming Li et al 2018 J. Phys.: Conf. Ser. 960 012012

View the article online for updates and enhancements.

You may also like

- Regional Physiographic Study for the Hydrology of Kali Lamong Watershed Area
  Widya Utama and Rista Fitri Indriani

- River Classification in Line with China’s New Requirements of Water Resources Management
  Huojian Huang, Yong Yuan, Ziqiang Xing et al.

- A classification algorithm based on Cloude decomposition model for fully polarimetric SAR image
  Hongmao Xiang, Shanwei Liu, Ziqi Zhuang et al.
A Heuristic Algorithm for Automatic Classification of Ring River

Chengming Li¹, Yong Yin¹, Zhiwei Hao¹ and Xiaoli Liu¹,*

¹Chinese Academy of Surveying and Mapping, No28 Lianhuachi Xi Road, Haidian district, Beijing, China
Email: *83391860@qq.com

Abstract. In map generalization, a technical difficulty and key to expressing water elements is river classification, which has high requirements for accuracy and calculation speed. In the study of river classification, the current hierarchical algorithm is not quite efficient and lacks classification of ring rivers. Based on this, this paper establishes the attributes of river classification, its geometric rules and ring rules, and classifies different ring types for the existing in river system. When calculating river’s longest path, we use the heuristic algorithm for iterative calculation to automatically conduct river classification. The experimental results show that this algorithm can achieve good results for the ring river system and improve the calculation efficiency and accuracy of river system classification.

1. Introduction
In the era of computer graphing, GIS has become increasingly mature, powerful, and more widely used. In river classification, we no longer have to select and view each part of the river. Instead, we only need to rely on automatic mapping for river classification based on the river’s attributes, geometric features and other information. Compared to manual classification of river, automatic classification not only relieves people from tedious work, but also improves the accuracy of river classification.

At present, there have been several studies of automatic river classification. Hehai Wu[1] establishes a river tree structure for classification and believes that the river’s main stem has the maximum length, so we can find the maximum length of the river to determine the main stem. However, in determining the main stem, merely focusing on the maximum length while ignoring river’s flow and characteristics will lead to inaccurate classification. Paiva et al. [2] propose to use angle information for river classification based on the geological knowledge that at the confluence of the river, the flow of the upper and lower part should almost be on a straight line. Nevertheless, when classifying river only based on angles, there will be large inaccuracy in classification if there is angle deviation at the node of the river. Qingsheng Guo et al.[3] believe that the main stem is often the longest river, and it also meets the 180° assumption at the node. Thus, they propose to combine the aforementioned two methods of length and angle to classify the main stem and tributary of the river, but they fail to consider the attribute information and the practical significance of the river. Wei Wu et al. [4] use the flow, layer, name, and other attributes information in river classification, which strengthens the rationality of the river classification. However, classification based on angles only will lead to some errors. In the above research, the rings in the river system are not treated, and the efficiency in calculating the longest path of the river is relatively low.
In this paper, we use the attribute information and geometric rules to classify the river system. We divide the ring types into single input and single output rings and multi input and multi output ones, and treat each type of rings with methods of the shortest path and the angle-priority rule. Meanwhile, this paper addresses the problem of low efficiency in searching for the longest path in geometric rules. By applying the heuristic algorithm, we reduce the number of searches and thereby improving the efficiency.

2. Rules for river classification

In the existing research results, the basic methods for river classification can be summarized as follows [5-9]:

First, reorganize information of the river system, construct the river tree, and search a river by setting an outflow point from the river as a starting node. Look for the main stem and set it as the first-order stream, then remove it from the river system and continue searching from the tributary node of the main stem, setting it as the second-order stream. Iterate this process and perform “top down” classification until all parts of the river has been searched.

In the searching process, how to accurately determine the main stem is essential, which determines the final result of river classification [10-11]. In determining the main stem of the river, the combination of the attribute information with geometric features is adopted. The attribute information of the river includes: river name, river type, river flow, etc. Geometric features include: the angle of the river, the length of the river, and whether there exists a river ring. According to the above information, the rules for river classification are presented as follows:

2.1. Rule of attribute information

This is to establish a number of rules to determine the main stem from the river’s attribute information. The rules are as follows (priority from top to bottom):

1. The river flow constraint principle: if the flow of one stream is inconsistent with the that of another stream, they cannot constitute the same river.

2. The river type constraint principle: streams of the same type preferentially constitute the same river.

3. The river name constraint principle: streams with the same name preferentially constitute the same river.

2.2. Rule of geometric features

If there is no attribute information of the river or the information is not enough to determine the main stem of the river, we need to determine based on the shape of the river to. The geometric features mainly include the angle and the length of the river. The angle of the river is the angle between the current stream and the next stream, giving priority to the straight stream. The length of the river is to determine whether the next stream is on the longest path of the current stream, giving priority to those on the longest path. The rules are as follows (priority from top to bottom):

1. Length priority principle: If the longest path of the stream is much longer than that of the other ones, then give priority to streams on the longest path to constitute the same river. As shown in Figure 1, according to this principle, we should choose the one with the longest path as the main stem of the river.

2. Angle priority principle. If the longest path of the stream is similar to the length of the candidate stream, but the angle of the candidate stream is much better than the angle on that of the longest path, then the candidate stream should be chosen as the main stem. As shown in Figure 2, according to this principle, we should choose the one with a better angle as the main stem of the river.
2.3. Rule of rings

The basic principles above can be used in river classification for those without rings. However, when rings exist in the river system, the effect of the above rules in river classification is poor. Therefore, this paper has a detailed classification of rings to address this issue by applying different methods to different river rings.

(1) Single input single output ring. A single input single output ring with only one tributary and one distributary. Single input single output rings are the most common in river classification. Since the path of the bulging part of the ring is often longer than those of the straight ones, when applying the previous principles to rings, the stream with longer path will be selected, which will lead to classification error. Therefore, we need to study the single input single output ring in the river. After finding it, by selecting the stream with the shortest path from the tributary to distributary or the stream with the best angle of the arc at the entrance, you can greatly reduce the classification error for single input and single output rings. As shown in Figure 3, the stream with the shortest path in the single input and single output is selected as the main stem.

(2) Multi input and multi output ring. As the name suggests, multi input and multi output ring is the river ring that has multiple tributaries or multiple distributaries. The streams of this type of rings flow to a number of streams, as shown in Figure 4. In this case, we can follow the angle priority principle, i.e., between the tributary stream and the ring, we should select the one with better angles as the main stem. This will avoid most of classification errors in the more into the ring out of the grading error. As shown in Figure 4, for the multi input and multi output ring, we should select the stream with better angle as the main stem of the river.

3. Automatic river classification

Automatic river classification first requires building a river tree. We need to find a distributary point as a starting node to search. And then through the main stem identification, we can determine the order of the stream, and classify each stream through iterative processing. Main stem identification requires the use of river attribute information and geometric features, as introduced above. In the process of classification, the attribute information has higher priority than the geometric features, i.e., we should first classify based on the attribute information. In the absence of attribute information or when the attribute information is not sufficient for classification, then geometric features will be used.

Figure 5 is a river system. N1 is the distributary point, and the attribute information of the river system is listed in Table 1:
Table 1. Attribute information of river network

| Stream Section Identifier          | River Type      | River Name     |
|-----------------------------------|-----------------|---------------|
| (N1, N2), (N2, N3), (N3, N4),     | Perennial river | Banqiao River |
| (N4, N5), (N4, N6), (N3, N7),     | Perennial river | Sanya River   |
| (N7, N17), (N17, N18), (N17, N19) | Perennial river |               |
| (N2, M1), (M1, M4), (M1, M2),     | Perennial river | /             |
| (M2, M3), (N9, M4), (N13, M5),    | Intermittent river | /          |
| (N11, M10)                        | Dry river       | /             |

3.1. Determine main stem with attribute information
According to the rule of attribute information above, we can use the attribute information to determine the main stem of the river. In Figure 6, M2 (M2, S1) is a seasonal stream. Since the former stream is of different type, choose (M2, M3) as the main stem. Another example is when determining the main stem of the next order at N2. Because the two candidate streams are of the same type as (N1, N2), we use name information to determine the main stem. (N2, M1) has the same as (N1, N2), and (N2, N3) has the same name, so (N2, N3) is the main stem. At the N3 node, (N3, N7) and (N3, N4) are two candidate streams, both have the same types and same names. Therefore, it is necessary to determine the main stem using the geometric characteristics.

3.2. Determine main stem with geometric features
The rule of geometric features combines length information with angle information to determine the main stem. When using length information, we need to find the longest path of the stream. This can be considered as finding the longest path of the directed graph. The problem of finding the longest path of the river tree[12] is described as follows:
Let $G=(V, A)$ be a fully connected graph, and $V=\{0,1,2,...,n\}$ be a set of nodes. is the geographical coordinate of node $n$. $A=\{(i,j)\mid i,j \in V\}$ is a set of arcs, where Node 0 is the set $(X_n, Y_n)$ starting node. The goal is to find the longest path starting from the starting node.

In the problem of finding the longest path of the stream, river information is usually stored in a spatial database, which not only stores the attribute data of geographical elements, but also the spatial data. The location data of these geographic features could give heuristics in searching for the path. By introducing the heuristic information into the searching process, we could avoid a large number of invalid searches and improve the searching efficiency.

The heuristic algorithm for calculating the longest path can be described as below:

(1) Determine the queue of target node to be searched. Enter all nodes in the graph of degree 1, i.e., all the leaf nodes. Calculate the distance between all leaf nodes and the starting point, write as $D(n)$, and solve for Max{$D(n)$}. Calculate $D(i)$, the distance between leaf node $i$ and the starting node, and include this leaf node $i$ in the target queue $T$ if $D(i)/\text{Max}\{D(n)\} \geq k$, where $k$ is the threshold value. The smaller the $k$, the more lenient the selection of target nodes, and thus the more nodes to be searched, and the longer the search will last, but the accuracy would be higher.

(2) Search the river map. Using the idea of heuristic search, $f(n)$ is the valuation function at node $n$. We use $f(n)=g(n)+h(n)$ to evaluate, where the cost function $g(n)$ represents the length of the arc from the current node $m$ to the extended node $n$, i.e., $\text{Length}_{mn}$. The evaluation function $h(n)$ represents the Euclidian distance from the extended node $n$ to the target node $t$, i.e., $\text{Dist}_{nt}=\sqrt{(X_n-X_t)^2+(Y_n-Y_t)^2}$. Through calculating $\text{min}\{f(N1),f(N2),f(N3)...\}$, we can select the next node $n$. See if there are other attempted paths at the current node m. If yes, include them into the incomplete search node queue $S$.

(3) Set node $n$ as the current node and continue searching by repeating Step 2 until meeting a leaf node. Check whether this leaf node is in the target node queue $T$ to be searched. If it is in $T$, delete this node from $T$.

(4) If the target node is empty, then end searching. If it is not empty, set the current node as the ending node of the incomplete search node queue $S$, and go back to Step 2 to continue searching.

For example, in Figure 7, (N3, N7) and (N3, N4) are two candidate streams at node N3, and both have the same attribute information, so geometric features are needed to determine the main stem. The heuristic algorithm is used to calculate the longest path from N3 through N7 and from N3 through N4, and then compare to select the main stem.

In Figure 6, first we calculate the longest path from N3 through N7. After screening, the target node set is \{N15, N16\}, and the set of incomplete search node is empty. The current starting node is N7, which has two expandable nodes N17 and N8, and the target node is N16. We can calculate $f(N7,N17)=g(N7,N17)+h(N7,N17)=\text{Length}_{N7N17}+\text{Dist}_{N17N16}$, $f(N7,N8)=g(N7,N8)+h(N8,N17)=\text{Length}_{N7N8}+\text{Dist}_{N8N17}$, and choose $\text{min}\{f(N7,N17), f(N7,N8)\}$. Because $f(N7,N8)$ is smaller, we choose N8 as the next node. At the same time, we include node N7 into the set of incomplete search nodes, and the search continues until reaching the target node N16. Then delete N16 from the target node set, and include node N14 into the set of incomplete search nodes. The next target node is N15. Take the newly added node N14 from the incomplete search set, and continue searching until reaching N15. Delete N15 from the set of target nodes, and set becomes empty, so the program ends.

According to the above algorithm, the longest paths starting from N3 through N7 and starting from N3 through N4 are MaxLengthN3N7 and MaxLengthN3N4, respectively. We get $\text{MaxLength}_{N3N7}/\text{MaxLength}_{N3N4} \geq k$, $k>1$. That is, the longest path starting from N3 through N7 is much longer than that from N3 through N4, so we select (N3, N7) as the main stem.

There are rings at nodes N8 and N10, and most are multi input multi output rings. Thus, we should choose streams with better angles as the main stems. Therefore, (N8, N9), (N9, N10), (N9, N9), (N9, N9), (N9, N9), (N9, N9), (N9, N9) N10, N12), (N12, N13) are selected as the main stems.
3.3. Automatic river classification

Searching results show that the streams from node N1 to node N16 are first-order streams. Then remove the first-order streams from the river system and iterate. Figure 7 shows the results after one iteration.

Continue searching all nodes on the main stem until all streams have been searched and classified. Figure 8 shows the final classification results.

4. Experiment and analysis

The experimental platform is based on the WJ-III stepless map workstation. The development environment is based on Microsoft Visual Studio 2010 and adopts C ++ programming language. The experiment environment is a personal computer with Windows 7 system (64-bit), which has Intel Core i7-6700 (3.40GHz) CPU, and 4GB RAM. The data for experiment is the monitored geographical information of the water system. We conduct the river classification experiment. The information of the water system includes river flow, river type, and river name.

The experiment is conducted using the water system data in two regions. We compare the results with and without optimization using the heuristic algorithm.

Table 2. Comparison of length algorithm before and after optimization

| Comparison Content                  | Region 1 | Region 2 |
|------------------------------------|----------|----------|
| Total number of arcs in river      | 1093     | 13012    |
| Complete ergodic classification time (second) | 1.984   | 15.05    |
| Heuristic classification time (second) | 1.881   | 12.71    |
| Increase rate in time (%)          | 5.3      | 18.4     |
Unoptimized ergodic number of arcs & 3189 & 42532 \\
Optimized ergodic number of arcs & 1961 & 35521 \\
Reduction ratio in ergodic number (%) & 62.6 & 19.7 \\
Difference between heuristic and ergodic classification & 0 (0%) & 0 (0%) \\

Table 2 shows the experimental results. It shows that in river classification in different regions, the optimized algorithm can reduce the ergodic number of arcs and time in different degrees, but the classification results are the same, and the efficiency is improved. Time reduction is more obvious for a larger total number of arcs. The analysis of accuracy and time of the heuristic algorithm are as follows:

1. Algorithm accuracy: The difference between the heuristic algorithm and the original algorithm lies only in the first step of selecting the ending nodes of the river, which might filter out the correct results. But in the experiment, after the classification of attribute information, the streams using this algorithm are often some smaller tributaries, which are usually short and straight, so the correct ending nodes are very rarely screened out.

2. Algorithm time: In the unoptimized case, the algorithm is ergodic, which searches each arc to find the longest path. The optimized algorithm will first find the candidate ending nodes that meet the criteria, and then search for these ending nodes through heuristic information, which reduces the number of searches and thus shortens the searching time of the algorithm. In the best case of optimization, when searching for the longest path, there is only one target node that forms the longest path from the starting node, and each node selected by the heuristic algorithm on this path. In this case, the total number of executions of this algorithm is the number of nodes on the longest path. In the worst case, we need to search all the arcs. It is the same as when not optimized. In general, the time of the optimized algorithm is between the best and the worst case.

![Original river system](a) Original river system ![River classification results](b) River classification results

**Figure 9.** Classification comparison results

Figure 9(a) is the original river system of one of the two experimental regions. The classification results are distinguished by color and thickness, the darker colors and thicker lines represent the higher river hierarchy. From the classification results in Figure 9(b), we obtain good classification results. The results by optimized algorithm are accurate and well-structured, especially for complex river system, such rings and multiple nodes in the river. The optimized classification effect meets the mapping requirements, and beats the existing algorithm.
5. Conclusion
This paper improves the algorithm of river classification. As for accuracy, the heuristic algorithm combines the attribute information with geometric characteristics, and reduces classification error in complex river systems (especially rivers with rings) through treating different types of rings. As for efficiency, it improves the algorithm of finding the longest path of the river system, which reduces the number of searches. The validity and rationality of this method are supported by experiments. In the process of river classification, we can include more geometric features of the river, such as the density of the river and the number of left tributaries and right tributaries of the river, which can further improve the accuracy of river classification.

Acknowledgments
This research was supported by the National Surveying and Mapping Projects (A1713) and the National SciTech Support Plan (2015BAJ06B01).

References
[1] Wu H. Structured Approach to Implementing Automatic Cartographic Generalization. 1996. Journal of Wuhan Technical University of Surveying & Mapping, 21(3), pp 277-285.
[2] Paiva J, Egenhofer MJ, Frank A. Spatial Reasoning About Flow Directions: Towards an Ontology for River Networks. 1992. International Archives of Photogrammetry and Remote Sensing, 24(B3), pp 318-324.
[3] Guo Q, Huang Y. Automatic reasoning on main streams of tree river networks. 2008. Geomatics & Information Science of Wuhan University, 33(9), pp 978 - 981.
[4] Wei WU, Chengming LI, Yin Y, et al. An automatic plotting algorithm of the gradual change of river based on directed topology. 2016. Science of Surveying & Mapping, 41(12), pp 89 - 93.
[5] Zhang Y. Structured Design of Dendritic River Networks Based on Graph[J]. Editorial Board of Geomatics & Information Science of Wuhan University, 2004, 29(6):537-536.
[6] Zhang Q N, Quan H. Construction and Application of River-Tree. 2005. Acta Scientiarum Naturalium Universitatis Sunyatseni, 44(6), pp 101 - 104.
[7] Ai Z. The river net structural model and its construction in geography information system. 1995. Surveying & Mapping of Sichuan, 2, pp 75-79.
[8] Guo Q S. Analysing the Characters of the Networks of Rivers and Structuralizing the Tree-like Network of Rivers Automatically.1999. Surveying & Mapping of Geology & Mineral Resources, 4, pp7-9.
[9] Zhang Q N. Line Generalization Based on Dynamic Segmentation.2004. Acta Scientiarum Naturalium Universitatis Sunyatseni, 43(2), pp 104-107.
[10] Zixing Ai, Niu R F, Hehai Wu, et al. The analyzing method of river network based on dynamic tree topology.2006. Science of Surveying & Mapping, 31(3), pp 80-81.
[11] Xiao T, Fang W U, Qi H, et al. A Multi-criteria Decision Model for Identifying Master River and Its Application in River System Construction.2005. Acta Geodaetica Et Cartographic Sinica, 34(2):154-160.
[12] Wang J X. Algorithms for Longest Path: A Survey.2009. Computer Science, 36(12), pp 1 - 4.