

# Curve Generalization Algorithm Based on Area of Bends using Head/tail Breaks

Hangyu Wang<sup>a</sup> and Haowen Yan<sup>a, \*</sup>

<sup>a</sup> Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou, China

**Keywords:** Area of bends, Head/tail breaks, Curve generalization

## Abstract:

For the early curve generalization algorithms, most of them only consider the reduction of the number of vertices, and do not take into consideration the important role of bends, especially the characteristic bends, on the shape of the curve. And the existing generalization methods based on the bends of the curve have complex algorithms and a large amount of calculation, focus on relationship between adjacent bends excessively and ignore the relationship among the overall bends. In addition, the threshold setting for filtering the bends is based on the unreasonable experience. Aiming at the problems above, a generalization algorithm based on the area of bends is proposed to achieve the purpose of simplifying the curve with the head/tail breaks classification method in this paper. Experiment shows that the algorithm is simple and efficient, and can iteratively take account of the overall bends with reasonable threshold, discarding the small bends and retaining the characteristic bends of the original curve to obtain generalization results which conform the natural law and is highly similar to the original graphics at different levels of detail.

Head/tail breaks is a classification method that is always applied to the classification of heavy-tailed data. Heavy-tailed data is universal in nature and human society. For example, there are more small towns than big cities in the world. However, small towns are less important than big cities in the field of economy and politics. Thus, cartographers will mark the big cities on the map and eliminate the small town. Map generalization is a progress of retaining important elements and delete unimportant elements. Head/tail breaks is able to extract significant data which can be retained as a generalization result by arithmetic mean.

Figure 1 shows the algorithm flow chart. First of all, we divide the curve into several bends with oblique-dividing-curve method. Secondly, we calculate the area of each bend, and then use head/tail breaks to complete the classification of the area of bends. If the percentage of bends in the head is less than 40%, it means the data conform to heavy-tailed distribution and can be classified with head/tail breaks. If the percentage of bends in the head is greater than 40%, the head/tail breaks is not applicable to this data. After classification, for the bends which is more important in the head, we reserve them directly. For the bends in the tail, we extract the feature points of each bend by retaining the point farthest from the axis so as to maintain the local shape of the original curve. Finally, we merge the bends in the head and the feature points as a generalization result.

The experimental result is shown in the Fig 2. The data of this experiment is administrative division map of Gansu Province extracted from a China map with a scale of 1:10,000,000. Because algorithm can be executed iteratively, it can generate results at different levels of detail. We can see that from the result in detail to concise result, the graphic changes progressively and there is no oversimplified result. With comparison of three algorithms in the Fig 3, the generalization results of both this paper and bend group division algorithm have better retention of characteristic bends than Douglas-Peucker algorithm. However, the algorithm of this paper has higher compression ratio and less execution time than bend group division algorithm, as shown in Table 1.

The algorithm of this paper is based on nature law rather than empirical threshold, and can generate progressive results at different levels of details by iteration. In addition, it takes overall relationship of bends into consideration, so the generalization result is unique. The experimental result shows this algorithm has not only better retention of characteristic bends than Douglas-Peucker algorithm but also higher compression ratio and less execution time than bend group division algorithm. To further optimize the algorithm, we will study how to evaluate the apparent extent of the curved feature better and how to extract and eliminate the small bend inside of the bend in the head in order to improve compression ratio in the future.

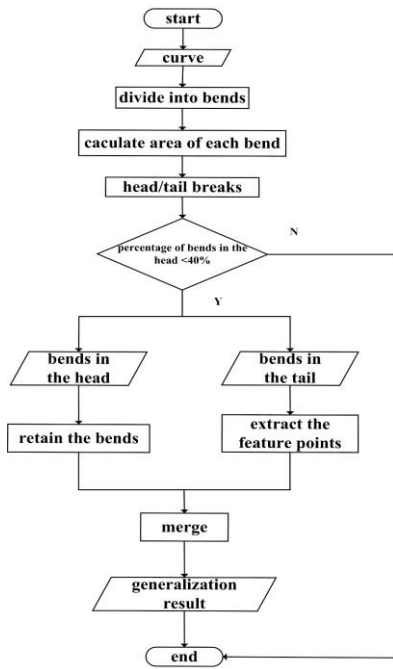


Fig 1. Algorithm flowchart

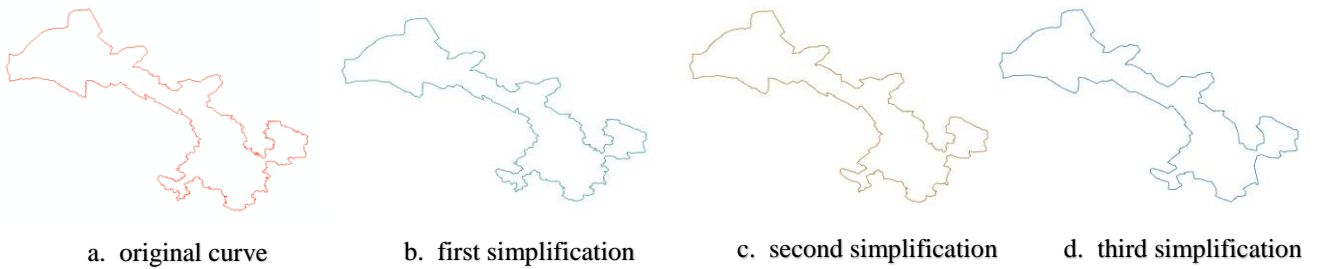


Fig 2. Original curve and generalization results in different levels of details

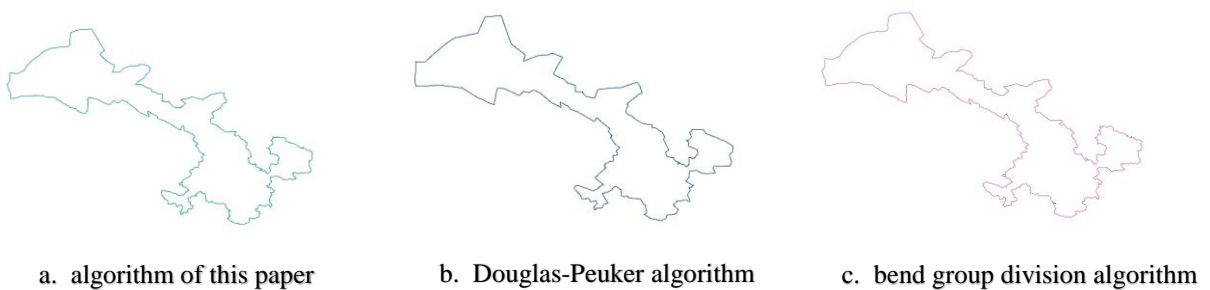


Fig 3. Generalization results of three algorithms

algorithm	origin points	retained points	compression ratio	execution time (s)
This paper	1576	841	53.36%	0.08897
D-P	1576	211	13.39%	0.06018
BGD	1576	1042	66.12%	0.53180

Tab.1 Comparison of the three algorithms