2010 International Conference on Web Information Systems and Mining

Selecting Optimal Parameter Value of Single Parameter Line Simplification Algorithm Based on Maximum Curvature

WANG Xiao-li[1, 2]
[1]Institute of Surveying and Mapping,
[2]Institute of Information Engineering,
Information Engineering University,
Zhengzhou, China, 450052
e-mail: xdwxl@163.com

JIANG Cheng-shun
Institute of Information Engineering,
Information Engineering University,
Zhengzhou, China, 450002
e-mail: csjiang@371.net

SUN Qing-hui
Institute of Surveying and Mapping
Information Engineering University,
Zhengzhou, China, 450052
e-mail: irisqh@163.com

***Abstract*—Line simplification algorithm is an important method in geographic information processing, and point number in result data determines simplification ratio. Based on sample data, curve function is established between parameter and point number using method of curve fit, then curve point with maximum curvature is found out in parameter value range, and the parameter value corresponding with this point acts as optimal parameter value. Law between parameter with simplification algorithm is revealed qualitatively and quantitatively, and maximum curvature method determining optimal simplification parameter value is also put forward. As a conclusion, it's feasible for simplifying large amount of lines data by analyzing factors affected by parameter and for confirming optimal parameter value at maximum curvature point.**

Keywords- line simplification; single parameter simplification algorithm; curve fit; maximum curvature; optimization

I. INTRODUCTION

Vector linear simplifying methods are ofen used for balancing between information quantity and data amount in GIS. Up to now, many linear simplifying methods have been advanced by many scholars, which can be classified into none parameter methods, single parameter methods and multi parameters methods. Besides none parameter method like interval fetching points methods and multi parameters methods like Lang method and objective function method [1], many linear simplifying methods belong to single parameter methods, such as light hurdle method, Circle method [2], Li-Openshaw method, Douglas-Peucker method, and etc, which determine simplifying result by only one parameter value, and which have two representative methods: Li-Openshaw method and Douglas-Peucker method. The former method simplifies lines based on nature principle [3], and also has capacity of re-sampling and smoothing features [1]; Douglas-Peucker method belongs to Global routines, can be called as "mathematically superior", is "overwhelmingly deemed the best perceptual representations of the original lines" [4], and becomes one of the greatest classics of linear simplifying methods applying to many fields like multi scale process, image process, calculating geometry, and et al [5].

Experiment method is often used for determining parameter value of simplifying method, that is, optimal parameter value is selected from many values corresponding to many simplifying results. Key factor effecting parameter value lies evaluating to simplification results. People from the perspective of visual perception or past experience of qualitative analysis and evaluation of quality of the results of simple, balanced line, said elements of the quality and the relationship between the amount of data to determine the best line features of simplification results, then determine the simplification of algorithm parameters. The evaluation process itself has a lot of uncertainty, different people may come to the same difference in simplifying the process of simplification of the larger results. From the geometric simplification process, art, transport, technology, representation, and other aspects to be evaluated, is a flexible, multi-level intelligent processing, while simplifying whether the results meet the requirement, also need to geometric constraints, topology constraints, structural constraints, Gestalt constraints be measured in many areas such as [6]. From the mathematical point of view, it has raised a lot of complexity for the line elements of the measurement method is applied to a single line of some elements of the "single-attribute measurement" is used in some other line of geometric elements of comparison before and after simplification, "Comparison of displacement or measurement "[6], according to the geometric properties can also be subdivided into the length measurement, density measurement, angle measurement and radian measurement [4]. It can be said, the line element of the evaluation results of simple "no best, only better", using as little as possible the amount of data that the line element of the highest possible quality of information is a simplified method to judge good and bad of the highest standards, we only criteria can be as close as possible to this ideal.

In this paper, a representative of the one-parameter class of line elements of Douglas-Peucker simplification algorithm as the research object, specific test data reduction algorithm from the analysis of the choice of parameter values in the simplification of the evaluation of the results of several important indicators of the geometry set of effect, using curve fitting, to establish parameters and geometric simplification function index, by analyzing the function

978-0-7695-4224-9/10 $26.00 © 2010 IEEE
DOI 10.1109/WISM.2010.19

curves to determine the optimal simplification algorithm parameters.

II. Mathematical Description

The use of simple algorithms generates samples according to specific requirements of data to generate sample data to be determined based on curve fitting parameters, fitting curves for the maximum curvature value within a reasonable range of good parameter values.

A. Curve fitting function

In this paper, the sample data were generated using least squares curve fitting. Curve fitting based on least squares means [7]: for given data $\left(x_i, f(x_i)\right)(i = 0,1,\dots,m)$, function $y = s*(x)$ is specified in the function set $\varphi = span\{\varphi_1,\dots,\varphi_n\}$, so that the error sum of squares is minimized:

$$\|\delta\|_2^2 = \sum_{i=0}^{m} \delta_i^2 = \sum\nolimits_{i=0}^{m} (s*(x_i) - f(x_i))^2$$
$$= min_{se\varphi} \sum\nolimits_{i=0}^{m} (s(x_i) - f(x_i))^2,$$

in which,

$$s(x) = a_0\varphi_0(x) + a_1\varphi_1(x) + \cdots + a_n\varphi_n(x)(n < m).$$

For the function set φ, according to the actual situation, you can choose different function types. The most common functions are the exponential function $y = ce^{bx}$, linear function $y = \mathrm{a}x + b$, logarithmic function $y = c\ln x + b$, polynomial function $y = b + c_0x + c_1x^1 + \cdots + c_ix^i$, power function $y = cx^b$,and et al. For specific fit, the first sample data obtained in the single-parameter line simplification algorithm for two groups of factors (including parameter values) data were logarithmic, exponential, power function, (second to sixth) polynomial function fitting the discrete data, and a relatively simple and the best curve goodness of fit (goodness of fit value of the nearest curve 1) function is selected as fitting function.

Depending on circumstances, fitting functions of parameter - length, parameter – point number, as well as parameter - running time can be established.

B. Determine the optimal parameter value

For a given function $y = f(x)$, the curvature k is calculated as: $k = \left|y''/(1 + y'^2)^{3/2}\right|$.

From the application perspective, the significance of the maximum curvature point representing rapid decline curve fitting is: the parameter value of this point is representative of the best balance between data compression rate and features' quality. When parameter value is greater than this point, parameter's influence on compression ratio is smaller, and the quality of line feature is seriously worse.

III. Sample Experiment

Test data are simple elements about China's major roads with Shapefile format using WGS84 global geographic coordinates, and with decimal degrees. All data are line features amounting to 22,629 lines, about 502,741 km total length of roads, totally constituting of 366,813 points. Each line feature has the average length of about 22 km, and averaging 16.2 points as its component elements. National data of all major roads in the state diagram shows the scale of about 1:35,000,000.

A. Sample Data Collection

Simplification process and the statistical parameters related items include: total length of roads, all roads constitute the total points, the average path length, constitute the average points of each road, all roads of a simple time required. Parameter value range is [0.0001, 0.0700], 0.0003 degree increments, thus a total of 235 group interval statistics, part of the sample data shown in Table 1.

B. Determine the optimal threshold

Based on sample data, 3 fitting curves can be obtained between threshold with point number, line length, and running time. Figure 1 and Figure 2 shows respectively the sample scatter and fitting function curve of threshold – point number and threshold - average length.

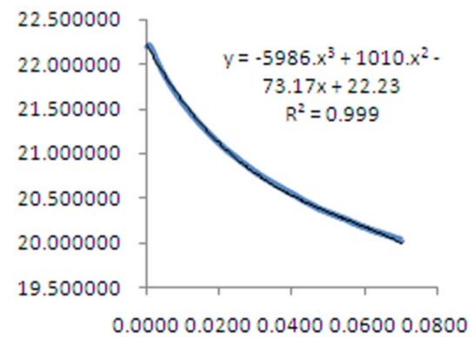

Figure 1. Threshold-length cubic polynomial function fitting results.

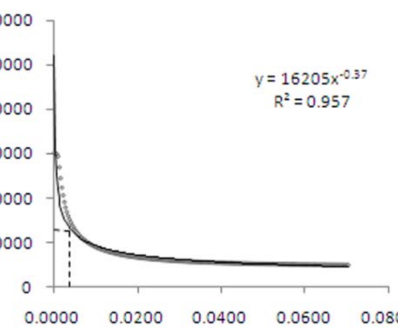

Figure 2. Threshold-length power function fitting results.

In effect diagram, fitting function and curve goodness of fit R^2 value are also marked besides fitting curve.

The threshold – point number fitting function incorporates into the curvature formula available:

$$k = \frac{8214.3145x^{-2.37}}{(1 + 3595017.2225x^{-1.8769})^{\frac{3}{2}}}.$$

Among them, the scope of the threshold is $x \in (0,0.06]$.

Let the derivative of the curvature of fitting function to 0, the corresponding equation's solution-- x value – is function's approximation point of maximum curvature. In the range of x solution, the x value corresponding to point of maximum curvature approximately is 0.00103, which corresponds to the value of fitting function $y = 206490$ (as shown in Figure 2, dashed line), and the large difference among ample data (see Table 1). Through analyzing the reasons of max fitting curvature $x = 0.00103$, there has certain bias between fitting curve with the sample data at the point of max fitting curvature, and the value of fitting curve is smaller than corresponding sample data value at this point. Therefore, taking the maximum curvature Department $y = 206490$ value on fitting curve as criterion, we can get sample with y values of the data closest to the threshold as the optimal threshold value, which can be obtained: $x_{optimize} = 0.0022$. With this threshold, the point data after simplification can be 56.03% of the original data, the average length of line features reduces about 0.44%. Local effects after simplification are seen in the Figure 3 and Figure 4.

IV. RESULT ANALYSIS

Simplification quality evaluation can result from qualitative analysis of some attributes (such as length, feature point displacement, etc.) or other property (such as the algorithm running time) of geometry features before and after simplification.

This paper simply analysis optimal parameters from length value of the sample data generated by algorithm and running time of the algorithm.

A. Visual analysis and evaluation

After simplifying line features with the best threshold using Douglas-Peucker algorithm, the simplified results were obtained from original data in the small scale and large-scale comparison of conditions (Figure 3 and Figure 4). In the small scale (1:500,000) comparison chart (Figure 3), simplification features and original features are almost completely overlapped, the two difference in that effect is almost negligible, and simplifying result can replace the use and processing analysis of simple line features of the original data representation; From large scale (1:100,000) contrast (Figure 4) we can see that, some of the small bends of the original line features is disappeared in the simplified features. Line features are even more concise, in addition to small details of certain features losing, large amount important points of line features are still retained, and important characteristics and key messages of line features are still kept by simplification.

Despite points with acute angle on line features slightly stiff and not very beautiful at state of the equivalent observed 350 times enlarge visual effect to the whole map, very little information of line features is lost using simplifying algorithm with the best threshold.

B. Quantitative analysis of optimal threshold

From the established three fitting curves among threshold with line length, point number, and running time, when the threshold is less than the maximum curvature value, as simplifying threshold increases, point number of line features declines rapidly, the amount of data decreases significantly, simplification running time decreases rapidly too, which indicates that point number and running time are very sensitive to change of threshold. At place of maximum curvature point, the amount of data is compressed to 56.03% of original, the maximum running time is reduced to 54.44 % (see Table 1). Lengths of line features of increases with decreasing of threshold, but this change is relatively slow and flat. At point of the optimal threshold, line length reduced about 0.44%, for a total reduction of 2,209 km (see Table 1). When threshold greater than the optimal threshold, point number and running time on the sensitivity threshold changes gradually decreased, and gradually close, in theory, to a constant value (if the threshold is large enough, all line features would be simplified to lines with only two points). Experiment showed that when threshold x = 0.0064, the data compressibility is 31.3%, while information elements of simplified lines obviously vary at condition of whole map visualization, and the length shortened about 1.96%, total 9,726 km.

Through the analysis of the optimal threshold, in addition to certain loss of line length, with the optimal threshold method simplifying line features, nearly half amount of data can be reduced, running time can be effectively shortened, quality loss of line features can be little at certain extent, which is the optimal balance between performance and processing efficiency about line simplification.

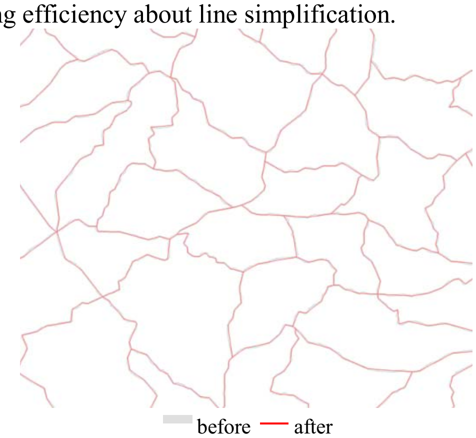

Figure 3. Simplification map before and after using best threshold(1:500,000).

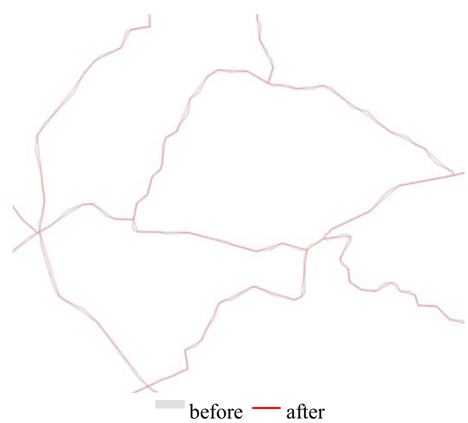

Figure 4. Simplification map before and after using best threshold(1:100,000).

V. CONCLUSIONS

In this paper, using a single parameter line simplification algorithm, a representative of Douglas-Peucker algorithm, for example, based on simple line features of the sample data set, fitting functions between simplifying parameters with line length, point number, and running time were established and analyzed, so that parameters impacts on algorithm performance and results were introduced with qualitative, graphical, more intuitive understanding. By analyzing parameter –point number fitting function, we get optimal parameter value with which balanced between efficiency and quality of data compression simplification algorithm. Through evaluating to the optimal parameter value and correspondingly analyzing, we can effectively decrease data size and on the same time ensure data quality within reasonable extent with optimal parameter value.

The proposed method is suitable for simplifying mass line features, this method reduces influence of human factors, so that the results is more objective and reasonable simplification, and also increase comprehensive analysis and evaluation to simplifying results in a more intuitive way.

As the best way to determine values of practice parameters according to test data is mathematical hypothesis testing and univariate regression analysis, this method can be further extended to analysis of other various algorithms with single-parameter and even multi-parameter to get optimal parameter value.

REFERENCES

[1] Wang Jinbao, Liu Zhenggang. Implementation and analysis of curve vector data compression algorithm. Geomatics&spatial information technology. 2006, 29(2),pp.122-124(in Chinese).

[2] Wu Fang, Qian Haizhong, Deng Hongyan, Wang Huilian. Spatial information intelligence processe facing cartographic automatic aggregation. Beijing: Science Press, 2008, pp. 288-288(in Chinese).

[3] LiZhilin, OpenshawStan. A Natural Principle for the Objective Generalization of Digital Maps. Cartography and Geographic Information Systems. 1993, 20(1),pp. 373-389.

[4] HershbergerJohn, Speeding up the douglas-peuker line-simplification algorithm. SnoeyinkJack. Proc. 5th Intl. Symp. Spatial Data Handling. 1992,pp. 134-143.

[5] McMasterRB. A statistical analysis of mathematical measures of linear simplification. The American Cartographer. 1986, 13,pp.103-116.

[6] McMasterRB. Automatic Line Generalization. Cartographica. 1987, 24,pp.74-111.

[7] Cai Qianqian, Xu Lvhu, Liang Zaizhong, Shen Yonghuan. Practical Math handbook. Beijing: Science Press, 2003: 717-717(in Chinese).

TABLE I. SAMPLE DATA OF DOUGLAS-PEUCKER ALGORITHM (PART)

Length(km)	Point num	Average Length(km)	Average Length(km)	Length/Original Length	Point num/Original Point num	Threshold	Time(msec)
502741.087260	301105	22.216673	13.306156	0.999999	0.820868	0.0001	6515
502738.826486	299709	22.216573	13.244465	0.999995	0.817062	0.0004	5969
502733.477838	298724	22.216336	13.200937	0.999984	0.814377	0.0007	9672
502661.347673	293184	22.213149	12.956118	0.999841	0.799274	0.0010	7781
502285.399764	268662	22.196535	11.872465	0.999093	0.732422	0.0013	5687
501765.681407	244260	22.173568	10.794114	0.998059	0.665898	0.0016	6172
501161.428904	223015	22.146866	9.855274	0.996857	0.607980	0.0019	5969
500532.670900	205536	22.119080	9.082858	0.995607	0.560329	0.0022	5265