Research on an Improved Adaptive Fitting Algorithm of Trajectory Information

To cite this article: LIANG Futai et al 2019 J. Phys.: Conf. Ser. 1169 012020

View the article online for updates and enhancements.
Research on an Improved Adaptive Fitting Algorithm of Trajectory Information

LIANG Futai1,2,*, LI Hongquan1, ZHENG Mao1 and LI Lingzhi1

1 Air Force Early Warning Academy, Wuhan, China
2 31121 PLA Troops, Nanchang, China

*Corresponding author:417516911@qq.com

Abstract. In order to facilitate the storage, visualization and data mining of large-scale trajectory information, an improved adaptive fitting algorithm of trajectory information is proposed, which can automatically select the optimal fitting interval and generate the key points and coefficients of the fitted interval. The algorithm consists of two steps: Firstly, the adaptive fitting method is used to fit the trajectory points to obtain the most suitable fitted trajectory interval, and the fitting method adopts the least squares method. Secondly, the constrained quadratic programming method is used to optimize the obtained trajectory interval coefficients to make the fitting curve smooth and continuous. The experimental simulation proves that the algorithm has obvious effects on data compression and feature extraction of trajectory information.

1. Introduction

With the development of big data technology, the value of tremendous amount trajectory information obtained by various types of sensors has become more and more important. It is a new topic how to find their laws, tap their potential, and utilize their value. It has great significance in the fields of air traffic control, transportation, and military.

For information transmission, the fitted trajectory information data is greatly compressed, which greatly reduces the amount of data transmitted, improves the efficiency of data transmission and saves the transmission bandwidth. From the perspective of data storage, it can be seen that the fitted trajectory information data reduces the scale of the trajectory information data, which is beneficial to save storage space, increase the storage data scale, and reduce the expenditure of the storage medium. From the perspective of data mining, the fitted trajectory information data can be stored in a large amount, which is conducive to big data analysis, and the fitting extracts and forms new trajectory information data features, which facilitates data mining and increases the potential value of the data.

At present, some of algorithms of trajectory information compression are based on general data compression methods, such as Hoffman coding[1], arithmetic coding, LZS algorithm[2], LZW algorithm[3], and so on. These algorithms only facilitate the transmission of data, and the heavy compression algorithm needs to be decompressed when searching, which is not conducive to big data mining analysis. The others of algorithms are specially designed for trajectory information compression[4, 5]. Although column storage and lightweight compression are suitable for data mining and analysis[6], there are still many shortcomings, such as discontinuous, non-smooth, redundant points and so on. Reference [4] presents a special compression algorithm based on curve fitting of trajectory information. Because the algorithm adopts fixed partition fitting, the error is large.
Reference [5] proposes a trajectory information fitting algorithm which adaptively selects the appropriate step size. Although the algorithm improves the fitting accuracy, it has discontinuous fitting curve and obvious inflection point.

In this paper, an improved adaptive fitting compression algorithm is proposed to overcome the shortcomings of trajectory information fitting algorithm in reference [4, 5], and its practicability is verified by experiments.

2. Algorithm Introduction

2.1 The basic method of trajectory information fitting

Curve fitting is to select the appropriate type curve to fit the observed data[7]. The common fitting functions contain exponential function, logarithmic function, power function and polynomial. For the fitting of trajectory information, considering the polynomial is simple and easy to use, the fitting effect is good, and it is suitable for most curve forms, and becomes the first choice of most trajectory information fitting[4]. Experiments show that the cubic polynomial curve fitting method is effective[8].

\[ y = ax^3 + bx^2 + cx + d \]  

(1)

The least square fitting method is adopted here. The least square method, also known as the least square method, is a mathematical optimization technique. It searches for the best function matching of data by minimizing the sum of squares of errors. By using the least square method, the unknown data can be easily obtained and the optimal value of the objective function can be obtained. It can be used for curve fitting to solve the regression problem. Least square method is one of the famous and effective algorithms in machine learning field.

The objective function is to minimize the sum of squares of deviations (or sum of squares of residual errors) between the fitting value of trajectory information and the actual value of trajectory information:

\[ \min \hat{\sigma} = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \]  

(2)

Among them, \( \hat{y}_i \) is the fitted value, \( y_i \) is the actual value of the track information.

2.2 Selection of the optimal fitting interval

The trajectory information is complicated, it is difficult to describe all the features with one curve, so the adaptive fitting interval selection becomes the key to solve the problem, and the fitting interval should be carefully selected to achieve the optimal effect. Too long fitted interval will cause the fitting error to increase, and too small fitted interval will affect the compression effect[1].

The optimal fitting interval can be selected by the backtracking method, and the optimal fitting interval can be approached by recursion. If the fitting error is greater than the threshold value, half of distance will be reduced from the judgment point. If the error is less than the threshold value, half of distance will be extended outward. If the error is equal before and after, the loop will be terminated and quit.

The calculation of the fitting error is:

\[ \sigma = \sum_{i=1}^{m} (\hat{y}_i - y_i)^2 \]  

(3)

2.3 Fitted interval of trajectory information fitting

All the adaptive fitting intervals of the whole trajectory information which are obtained by using the above fitting method are recorded as: \( X_m = (x_{m1}, x_{m2}, x_{m3}, \cdots, x_{mn}) \), \( Y_m = (y_{m1}, y_{m2}, y_{m3}, \cdots, y_{mn}) \). \( m \) is the interval number. \( n_m \) is the number of points of \( m \) interval trajectory information. The goal of whole trajectory information fitting is to keep the original trajectory information as close as possible within the allowable range of error.
In order to maintain the continuity of fitted trajectory information, it is necessary to add such a restriction condition to the first fitted interval, that is, the first and last points of the segment are consistent with the fitting value:

\[
\min \sum_{i=1}^{n_i} \left( \hat{f}_1(x_{i_1}) - y_{i_1} \right)^2
\]

s.t. \( \hat{f}_1(x_{i_1}) = y_{i_1} \)

By matrix operation, the objective function and its constraints are transformed into quadratic programming problems. The nonlinear programming problem in which the objective function is a quadratic function and the constraints are given in linear form is called quadratic programming problem. So far, there have been many algorithms for solving quadratic programming problems, such as Lagrange method, Lemke method, interior point method, efficient set method, ellipsoid algorithm and so on\[9, 10\], and there are still many scholars engaged in this research work. In this paper, we don't need to care about the specific solution of the problem. In the python environment, we only need to call specific functions to solve this quadratic programming problem.

But the problem is that not every fitted interval of trajectory information conforms to convex optimization conditions. We converted it into a standard form of quadratic programming.

\[
\min \frac{1}{2} X^T GX + X^T C
\]

s.t. \( AX = B \)

Among them, \( X^T \) is the vector of coefficients for each fitted trajectory interval.

\[
G = \begin{bmatrix}
\sum x_{i_1}^0 & \sum x_{i_1}^1 & \sum x_{i_1}^2 & \ldots & \sum x_{i_1}^{n_m} \\
\sum x_{i_2}^0 & \sum x_{i_2}^1 & \sum x_{i_2}^2 & \ldots & \sum x_{i_2}^{n_m} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\sum x_{i_n}^0 & \sum x_{i_n}^1 & \sum x_{i_n}^2 & \ldots & \sum x_{i_n}^{n_m}
\end{bmatrix}
\]

According to the knowledge of convex optimization theory, the matrix \( G \) in the above equation should be semidefinite. In this paper, damping factor \( u_m \) is used to make \( G = G + u_m I \) positive definite, which solves this problem well. The value of damping factor \( u_m \) is determined according to the error requirement. In each fitted interval of trajectory information, we can take different values for it or the same values for the whole trajectory. In this paper, after overall consideration, the value of damping factor \( u_m \) is set as 0.00001.
By Python calculation, the parameters and key points of each interval fitted curve are obtained. The fitted trajectory interval information can be expressed as: [(head and tail point), (longitude, latitude), (a,b,c,d)] .

3. algorithm implementation
In summary, the algorithm flow based on Python language is designed as follows:
Firstly select the optimal fitting interval, and use the backtracking method to return the fitting interval key points (the first and last points) by recursive calling. The pseudo code is shown in Table 1.

| Table 1. Pseudo code of fitting interval. |
|------------------------------------------|
| **Input:** fitting interval trajectory information \((X,Y)\) and fitting error threshold \(d\) |
| **Output:** the key points of the fitted interval (each piece of track information) |
| Fit \(X\) and calculate the fitting error |
| \[ \text{while } wch > d \text{ do:} \] |
| \[ \text{Fit } X[l/2\text{len}(X)] \text{ and calculate the fitting error} \] |
| \[ \text{if } wch < d : \text{ break} \] |
| \[ \text{else: recursive the first half of the piece of trajectory information} \] |
| \[ \text{while } wch < d \text{ do:} \] |
| \[ \text{Fit } X[1-l/2\text{len}(X)] \text{ and calculate the fitting error} \] |
| \[ \text{if } wch > d : \text{ break} \] |
| \[ \text{else: recursive the first half of the piece of trajectory information} \] |
| give the latter half of the piece of trajectory information to \(X\) |
| return: the key points of the fitted interval (head and tail points) |

After the adaptive fitting interval of the whole trajectory information is obtained, the constrained fitting operation is carried out on each interval, and finally the fitting parameter list of all intervals is obtained. The pseudo code is shown in Table 2.

| Table 2. Pseudo code of fitting trajectory |
|------------------------------------------|
| **Input:** optimal fitting interval trajectory information \((X_m,Y_m)\) |
| **Output:** the best fitting interval fitting coefficient \((a_m,b_m,c_m,d_m)\) |
| Solve the quadratic optimization for \((X_i,Y_i)\) (Equation 4) |
| \[ \text{while } i < m \text{ do:} \] |
| \[ \text{Solve the quadratic optimization for } (X_m, Y_m) \text{ (Equation 5)} \] |
| \[ \text{Assign } n_i \text{ points value of } i-1 \text{ interval to the head points of } i \text{ interval} \] |
| \[ i+ = 1 \] |
| return: fitting parameter list of all intervals |

4. experiment simulation
The purpose of fitting is to compress data as far as possible under the premise of motion of trajectory information was characterized as perfectly as possible. In the python language environment, the improved fitting algorithm is used to simulate a section of air target trajectory information, and the comparison schematic diagram of fitted and no-fitted trajectory information is obtained. As ‘Figure 1’ showed, this trajectory information contains 300 points, and the position varies obviously, which reflects the characteristics of various air target trajectory information and has a certain representativeness.
Figure 1. Initial trajectory figure

Figure 2. Fixedly divided trajectory figure

In [4], the trajectory information fitting interval is fixedly divided. Although the problem of unmatched fitting is solved to some extent, there is still a problem that the deviation from the real trajectory information is too large when the interval is less divided. If the partition is too dense, the amount of compressed data would be large. In 'Figure 2', the red curve is the fitted curve. The number of intervals fixedly divided is 15. When the number is less than 15, the trajectory curve is not well fitted, the fitting effect is very poor, and the fitted curve is inconsistent. When the number is greater than 15, the effect of the fitting is significantly improved, but the effect of trajectory data compression is not satisfactory very much. When the number becomes larger, the effect of trajectory data compression will be greatly reduced, and the significance of adopting this method is lost.

Figure 3. Figure generated by adaptive trajectory information fitting algorithm

Figure 4. Figure generated by the algorithm in this paper

Experiment with the adaptive trajectory information fitting algorithm of the literature [5], when fitting error $\sigma = 0.00001$, the red points represents the key points of fitting and the amount is 9, and the red curve is the fitted curve. It can be seen from ‘Figure 3’ that, due to the absence of additional constraints, the trajectory information appears discontinuous and unsmooth at the connection of the fitted interval, so that there is a certain distortion in simulating the real trajectory information.

Using the algorithm shown in this paper, the original trajectory information curve agrees well with the fitted curve when fitting error $\sigma = 0.00001$, and the continuity and smoothness are very good as ‘Figure 4’ shown. We can see that the number of intervals generated automatically is the same as the algorithm in [5], and it also is 9. Using this algorithm, as long as the fitting error range is set, the selection of the interval is automatic, without manual input in advance, and the fitting effect of the algorithm is far better than the previous several methods. The algorithm has higher data compression...
rate than the method in [5], which preserves the trajectory information features well, and demonstrates the effectiveness of this algorithm in feature extraction of trajectory information. Effects of all the above algorithms can be shown in Table 3. By comparison, we can see the advantages of the algorithm in this paper.

Table 3. Comparison table.

| Performance       | The algorithm in [4] | The algorithm in [5] | The algorithm in the paper |
|-------------------|----------------------|----------------------|---------------------------|
| Data compression rate | Low                  | High                 | High                      |
| Self-adaption     | No                   | Yes                  | Yes                       |
| Continuity        | No                   | No                   | Yes                       |
| Smoothness        | No                   | No                   | Yes                       |

5. Conclusion

Based on the actual requirements, the thesis improves the deficiencies of the previous trajectory information fitting algorithm, and proposes an improved trajectory information adaptive fitting algorithm, which well solved the problem of adaptive selection of interval and smooth connection of fitted interval in trajectory information fitting. This algorithm has high compression rate for trajectory data and can solve the problem of high-speed transmission and storage of trajectory data. The algorithm can obtain smooth and continuous fitted curves, and is conducive to the visualization of trajectory information. In addition, it has a unique advantage in the feature extraction of trajectory information, which overcomes the traditional trajectory data feature extraction methods in big data technology and data mining[11,12]. The algorithm improves the speed and efficiency of trajectory data analysis and mining, and is conducive to the large data application of trajectory information. To sum up, the algorithm has certain practical significance.

References
[1] Chen C H, Song H L and Tang X M 2016 Different Compression Algorithms for Data Compression Radar Ship Electronic Engineering 36 pp 57-61
[2] Chen L, Ma K and Li C Y 2017 A New Hybrid Data Compression Algorithm for Weather Radar Computer & Digital Engineering 45 pp 444-448
[3] Song H L, Tang X M and Z T 2016 LZW Algorithm Optimization for Real-time Radar Data Transmission Radar Science and Technology 14 pp 397-402
[4] Song Y, Liu L M and Han Z Z 2017 A Track Data Compression Method for Use on Radar Electronics Optics & Control 24 pp 89-92+98
[5] Liu F Z, Li H Q and Xiao B 2017 An Adaptive Track Fitting Algorithm Journal of Air Force Early Warning Academy 31 pp 424-426+435
[6] Li C, Zhang M B and Xing C X 2010 Survey and Review on Key Technologies of Column Oriented Database Systems Computer Science 37 pp 1-7+17
[7] Gang X X 2015 Research on the Fitting Algorithm of GLONASS Broadcast Ephemeris Parameters. In: 2015 Joint International Mechanical,Electronic and Information Technology Conference p 4
[8] Xiong J 2013 Track Fitting and Fuzzy Optimal Scheduling for Inland Vessels: Wuhan University of Technology
[9] Han J Y, Liu G S and Liu S S 2000 A New Descent Algorithm for Solving Quadratic Bilevel Programming Problems Acta Mathematicae Applicatae Sinica 03 pp 235-244
[10] Rui W J, Cao D X and Song X W 2005 An Interval Maximum Entropy Method for Quadratic Programming Problem Journal of China University of Mining and Technology 04 379-383
[11] Chen Y B, Yu H C and Chen L 2015 A Spatiotemporal Cluster Method for Trajectory Data Springer International Publishing p 3-10
[12] Wu Q 2018 Online GPS Trajectory Data Compression Algorithm Based on Relative Synchronous Euclidean Distance Filtering Computer Applications & Software