Vessel Trajectory Similarity Based on Cubic B-spline Interpolation + Time-constrained Hausdorff Algorithm

Liu Xinyu¹, Wang Zhaoshun¹*, Zhi Ning², Mi Jiping¹, Zhang Xianda², Fang Jifei², Li Bo³

¹ School of Computer & Communication Engineering, University of Science and Technology Beijing, Beijing, 100089, China
² Institute of Electronics, Chinese Academy of Sciences, Beijing, 10019, China
³ Geovis Technology Co., Ltd, Beijing, 101399, China
*Corresponding author’s e-mail: zhswang@sohu.com

Abstract. The spatio-temporal trajectory data sampling period is large, and the general trajectory similarity is not suitable. This paper proposes a new algorithm, cubic B-spline interpolation + time-constrained Hausdorff algorithm to calculate the trajectory distance. This paper compares the time-constrained Hausdorff algorithm, cubic B-spline interpolation algorithm and cubic B-spline interpolation + time-constrained Hausdorff algorithm. The experimental results show that the cubic B-spline interpolation + time-constrained Hausdorff algorithm is significantly higher than other algorithms in the accuracy of vessel trajectory similarity.

1. Introduction
With the rapid development of technology and the wide application of a variety of sensors, massive data is rapidly accumulating. These data play an important role in all aspects of national defense and technological development. In recent years, the navigation has developed rapidly, generating a large amount of spatio-temporal trajectory data. The analysis of vessel trajectory data can dig out the potential useful information of the target. In order to optimize the trajectory when displaying the target trajectory, we need to integrate the trajectory data obtained by different observation sources.

The trajectory similarity is the basis of multi-source data fusion. The accuracy of the similarity directly affects the fusion result. Therefore, it is necessary to study the trajectory similarity specifically. However, most of the current research objects in the trajectory similarity are people or vehicles. These trajectories are carried out on the road network, which is quite different from the vessel trajectory at sea. Wu Cong et al. [1] used the similarity function based on difference and feature point set to analyze the similarity between moving target trajectories; Wang Haiqi et al. [2] used the trajectory similarity method dominated by the motion direction to calculate the similarity between vehicle trajectories; Liu Kun et al. [3] used the improved edit distance to calculate the similarity of human or vehicle trajectories. In this search, the time attribute of trajectory data were not considered; Zhang Chengde et al. [4] proposed a Hausdorff distance calculation method with time constrains for the three-dimensional spatio-temporal trajectory, so as to match the vehicle and achieve the purpose of carpooling; Liu Lei et al. [5] proposed an average distance between the trajectories, combined with the vessel’s trajectory heading and speed characteristics, which constituted the comprehensive distance the
vessel’s trajectories, so as to judge the vessel trajectory similarity; Peng Xiangwen et al. [6] calculated the trajectory similarity by normalizing the weighted summation of vessel position, heading and speed; Zhao Wenwen et al. [7] used the Fréchet distance to calculate the vessel trajectory similarity.

There are many sources of trajectory data, mainly relying on shore-based radar and ship-based radar. Remote sensing satellites, electronic reconnaissance satellites, and Beidou satellites are also important data sources. In the environment where satellites and radars cannot detect, artificial reporting is also used. Different sources of data can cause very different sampling periods. The sampling period of the spatio-temporal trajectory data is quite different, so it is necessary to design an algorithm for this situation.

2. Key Technology

2.1. Time-constrained Hausdorff algorithm

The Hausdorff algorithm calculates the similarity between two sets of unordered points. Suppose the point set \( A = \{ a_1, a_2, \ldots, a_n \} \) and the point set \( B = \{ b_1, b_2, \ldots, b_m \} \), then the distance between the point \( a_i \) of \( A \) and the point \( b_j \) of \( B \) is \( \text{distance}(a_i, b_j) \). The Hausdorff distance can be expressed as:

\[
H(A, B) = \max \left( h(A, B), h(B, A) \right)
\]

\[
h(A, B) = \max a_i \in A \left( \min b_j \in B \text{distance}(a_i, b_j) \right)
\]

\[
h(B, A) = \max b_j \in B \left( \min a_i \in A \text{distance}(a_i, b_j) \right)
\]

The general Hausdorff algorithm does not take into account the time attribute, so it is not suitable for processing spatio-temporal trajectory data. The trajectory data is sorted according to the time attribute, and the distance between the point \( p \) on the trajectory \( A \) and the point \( q \) on the trajectory \( B \) is calculated, and the \( q \) point corresponds to the \( p \) point time. Using a one-to-three method, the distance between \( p \) and \( q \), the distance between \( p \) and a point before \( q \), and the distance between \( p \) and a point behind \( q \) are calculated. The distance between \( p \) and \( q \) takes the minimum value between the three [8].

2.2. Cubic B-spline interpolation algorithm

In finance, spline interpolation is used to estimate the dependent variable values of independent variables that are not included in the original observation point. When calculating the distance between the trajectories, in addition to knowing the position value at the sampling time [9], the value at any time is also known, so the data needs to be interpolated. This paper uses the cubic B-spline interpolation algorithm to process the trajectory data.

3. Algorithm

3.1. Algorithm improvement ideas

3.1.1. Improved cubic B-spline interpolation algorithm. The spatio-temporal trajectory data processed in this paper has different sources and different sampling periods. It can not guarantee that the time attributes of the trajectory points can correspond one-to-one. Therefore, the influence of time attributes on the similarity of the entire trajectory may be ignored. The spline curve of the trajectory is obtained by the cubic B-spline interpolation algorithm, and the position of the trajectory at any time is obtained through the B-spline curve. Considering the similarity of the B-spline curves of similar trajectories, when calculating the similarity between trajectory \( A \) and trajectory \( B \), the predicted position of trajectory \( B \) is obtained by using the B-spline curve parameters of trajectory \( A \). Calculate the distance between the predicted position of the trajectory \( B \) and the real position as the distance between
trajectory A and the trajectory B. If the distance is within a certain threshold, the trajectory is considered to be similar.

3.1.2. Cubic B-spline interpolation + time-constrained Hausdorff algorithm. Calculate the B-spline curve parameters of trajectory A and trajectory B, and obtain the position of the two trajectories at any time to solve the problem corresponding to the required trajectory point time, which is required in the time-constrained Hausdorff algorithm. The spline curve parameters of the trajectory B are obtained by the cubic B-spline interpolation algorithm. According to the trajectory point time attribute of the trajectory A, the predicted position of the trajectory B corresponding time is obtained. The time-constrained Hausdorff algorithm is used to calculate the distance between the true position of the trajectory A and the trajectory B corresponding to time predicted position. If the distance is within a certain threshold range, then trajectory A is considered to be similar to trajectory B.

3.2. Improved algorithm in this paper

3.2.1. Data preprocessing. In this paper, we use the marinecadastre.gov/ais/ public data set. However, the format of the data set cannot be directly processed by our algorithm, so it is necessary to preprocess the original data set. The AIS data contains attributes such as MMSI, time, longitude, latitude, speed, direction, etc. MMSI is the unique identifier of a vessel. We first extract the trajectory of each vessel through MMSI, then remove the abnormal points in the trajectory, and then extract the three attributes of time, longitude and latitude we need. When we remove the abnormal point, we consider the drift of the position of the trajectory point, so we simply handle the trajectory point. The theoretical distance travelled by the vessel can be obtained by calculating the product of the current travel speed of the vessel and the time interval. Taking the current position of the vessel as the middle position, the theoretical exercise distance plus a certain threshold is used as the radius. If the neighbour is not in the circle, it is considered to be an abnormal drift point and should be deleted [10].

3.2.2. Cubic B-spline interpolation algorithm for calculating trajectory similarity. When we calculate the similarity between the trajectory A and the trajectory B, the sampling period of the interpolation curve parameter sequence is 15, and the verification data sampling period is 11. For the trajectory A, the one-dimensional curve B-spline representation of the longitude and latitude is obtained from the time t in the interpolation parameter list.

\[ Expression_{lon} = splrep(t) \] (4)
\[ Expression_{lat} = splrep(t) \] (5)

The B-spline curve representation method of attribute time and longitude is obtained in equation (4), and the B-spline curve representation method of attribute time and latitude is obtained in equation (5). Thereby, the position value of the trajectory A at any time can be obtained.

Considering that if the two trajectories are similar, the one-dimensional B-spline curve should be similar. Using the B-spline of the trajectory A, the interpolation result of the trajectory B verification data is calculated. The time attribute \( t_B \), the longitude attribute \( lon_B \), and the latitude attribute \( lat_B \) of the verification data set of the trajectory B are known. Apply the \( t_B \) in trajectory B and the B-spline in trajectory A to obtain the interpolated longitude and latitude of trajectory B.

\[ Estimated_{lon} = splev(t_B, Expression_{lon}) \] (6)
\[ Estimated_{lat} = splev(t_B, Expression_{lat}) \] (7)

Calculate the distance between \((lon_B, lat_B)\) and \((Estimated_{lon}, Estimated_{lat})\). If the verification data sequence of trajectory B is \( b_0, b_1, \ldots, b_{n-1} \). From the true position of each point \( pos(i) \) for \( 0 \leq i < n \) and the
position $\text{posEstimated}_i$ obtained by interpolation, the distance between the two points can be calculated. The average of the distance of all points is taken as the distance between the trajectory $A$ and the trajectory $B$. It can be expressed as:

$$\text{Distance} = n^{-1}\left(\sum_{i=0}^{n-1} \left(\text{lon}_B(i) - \text{Estimated}_{\text{lon}}(i)) + \left(\text{lat}_B(i) - \text{Estimated}_{\text{lat}}(i)\right)\right)^{1/2}\right)$$

(8)

According to the experiment and the running characteristics of the vessel, set a certain distance threshold $\text{Threshold}$. If $\text{Distance} < \text{Threshold}$, the two trajectories are considered similar.

3.2.3. Cubic B-spline interpolation + time-constrained Hausdorff algorithm to calculate trajectory similarity. Through the cubic B-spline interpolation, the longitude spline parameter $\text{Expression}_{\text{lon}A}$ of trajectory $A$, the latitude spline parameter $\text{Expression}_{\text{lat}A}$, and the longitude spline parameter $\text{Expression}_{\text{lon}B}$ of trajectory $B$ and the latitude spline parameter $\text{Expression}_{\text{lat}B}$ are obtained. For the verification data sequence $\text{track}_A$ of trajectory $A$ and the verification data sequence $\text{track}_B$ of trajectory $B$, the same sampling interval is taken to obtain the sampling sequence $\text{samplingTra}_A(0 \leq i < \text{len}(\text{track}_A))$ of trajectory $A$, and the sampling sequence $\text{samplingTra}_B(0 \leq j < \text{len}(\text{track}_B))$ of trajectory $B$. Calculate the distance $\text{distance}_{AB}$ from the trajectory $A$ to the trajectory $B$, and take the time attribute $t_A$ of the trajectory $A$ verification data sequence to obtain the approximate position of the trajectory $B$ corresponding time:

$$\text{Estimated}_{\text{lon}B} = \text{splev}(t_A, \text{Expression}_{\text{lon}B})$$

(9)

$$\text{Estimated}_{\text{lat}B} = \text{splev}(t_A, \text{Expression}_{\text{lat}B})$$

(10)

Using the time-constrained Hausdorff algorithm, the distance $\text{distance}_{AB}$ between the real position of the $\text{samplingTra}_A$ and the predicted position at the time corresponding to the $\text{samplingTra}_B$ is calculated. Similarity, the distance $\text{distance}_{BA}$ from trajectory $B$ to the trajectory $A$ is calculated. Then the trajectory distance $\text{Distance} = \max(\text{distance}_{AB}, \text{distance}_{BA})$, if there is $\text{Distance} < \text{Threshold}$, the trajectories are considered similar.

3.2.4. Algorithm flow in this paper. The algorithm flow chart is shown in Figure 1:
4. Experimental results and comparison

In this paper, the time-constrained Hausdorff algorithm, cubic B-spline interpolation algorithm, cubic B-spline + time-constrained Hausdorff algorithm are used to calculate the similarity between two trajectories, and the accuracy of the three algorithms is compared.

4.1. Time-constrained Hausdorff algorithm

The time-constrained Hausdorff algorithm calculates the trajectory spatial distance, sorts the trajectory points according to the time attribute, and calculates the distance between the trajectory points by a one-to-three method.

The time-constrained Hausdorff algorithm takes into account the effect of time attribute on similarity. However, when calculating the distance between the trajectory points, the nearest point of the time distance is first searched, so the calculation speed is slow, and the accuracy of the similarity needs to be further improved.

4.2. Cubic B-spline interpolation algorithm

The time-constrained Hausdorff algorithm takes a while to find the closest point of the time attribute when calculating the distance of the trajectory point. The result is a certain error.

The B-spline curve is obtained by the cubic B-spline interpolation algorithm, and the position of the trajectory at any time is obtained, which can improve the calculation speed. According to the cubic B-spline curve of one track, the predicted value of another trajectory at a certain moment is obtained. The difference between the predicted position and the real position is taken as the distance between the two trajectories, which ensures the accuracy of the similarity to a certain extent.
4.3. Cubic B-spline + time-constrained Hausdorff algorithm
According to the cubic B-spline interpolation algorithm, the distance between the predicted position of the trajectory point and the real position is calculated as the trajectory distance. Although the running speed is increased, the similarity accuracy rate is decreased.

Cubic B-spline interpolation obtains the position of the trajectory at any time point and increases the calculation speed. The time-constrained Hausdorff algorithm calculates the trajectory distance to ensure the accuracy of the similarity. It is proved that the trajectory similarity of the cubic B-spline interpolation + time-constrained Hausdorff algorithm used in this paper has certain application value.

4.4. Experimental comparison result
In order to verify the accuracy of the new algorithm proposed in this paper, a different sampling interval of a certain trajectory is sampled as a sub-trajectory. The similarity of the sub-trajectory is calculated by the time-constrained Hausdorff algorithm, the cubic B-spline interpolation algorithm, the cubic B-spline interpolation + time-constrained Hausdorff algorithm. And the results are compared.

In this paper, 11144 trajectories are verified, and different trajectory distance thresholds are set to judge the accuracy of the similarity. The threshold values are: 0.5, 0.6, 0.7, 0.8, 1.0. If the distance between the trajectories is within the threshold range, it is considered similar. The similarity accuracy of each algorithm is shown in Figure 2. The speed comparison of various algorithms to calculate similarity is shown in Figure 3.

![Figure 2](image1.png)
**Figure 2.** Comparison of similarity accuracy.  
**Figure 3.** Calculating time of similarity accuracy.

Algorithm-1 represents the time-constrained Hausdorff algorithm, Algorithm-2 represents the cubic B-spline interpolation algorithm, and Algorithm-3 represents the cubic B-spline interpolation + time-constrained Hausdorff algorithm.

As it can be seen from Figure 2, the accuracy of the similarity of the cubic B-spline interpolation + time-constrained Hausdorff algorithm is significantly higher than other algorithms under different distance thresholds. Figure 3 shows that the cubic B-spline interpolation algorithm and the cubic B-spline + time-constrained Hausdorff algorithm are significantly faster than the time-constrained Hausdorff algorithm.

5. Conclusion
The time-constrained Hausdorff algorithm takes into account the time attribute of the vessel trajectory data, and the similarity has a good accuracy, but the running speed is not high. The B-spline interpolation algorithm significantly improves the speed of similarity calculation, but the accuracy rate decreases. The cubic B-spline interpolation + time-constrained Hausdorff algorithm proposed in this paper improves the accuracy of trajectory similarity and increases the calculation speed.

References
[1] Wu Cong, Cui Jun, et al. An Analysis Method of Moving Object Trajectories Based on Data Dimensionality Reduction[J]. Ship Electronic Engineering, 2015, 35(9): 35-39 (in Chinese).
[2] Wang Haiqi, Zhai Wenlong, et al. Trajectory similarity measure based on motion direction[J/OL]. Computer Engineering.
http://kns.cnki.net/kcms/detail/31.1289.tp.20190123.1605.007.html.

[3] Liu Kun, Yang Jie. Trajectory Distance Metric Based on Edit Distance[J]. Journal of Shanghai Jiaotong University, 2009, 43(11): 1725-1729 (in Chinese).

[4] Zhang Chengde, Bie, Zini. Research on improved carpool algorithm based on 3D space-Time trajectory[J/OL]. Computer Engineering and Applications.
http://kns.cnki.net/kcms/detail/11.2127.TP.20180926.1744.013.html

[5] Liu Lei, Chu Xiumin, et al. Trajectory classification algorithm based on KNN[J]. Journal of Dalian Maritime University, 2018, 44(3): 16-21 (in Chinese).

[6] Peng Xiangwen, Gao Shu, et al. Clustering Method of ship’s Navigation Trajectory Set Based on Spark[J]. Navigation Of China, 2017, 35(10): 112-115 (in Chinese).

[7] Zhao Wenwen, Hu Zhihua, et al. Ship sailing period identification based on AIS[J]. Computer Applications and Software, 2018, 35(10): 112-115 (in Chinese).

[8] Zhang Xiaobin, Yang Dongshan. Hausdorff distance about spatial-temporal trajectory similarity based on time restriction[J]. Application Research of Computers, 2017, 34(7): 2078 – 2079 (in Chinese).

[9] Li Botao, Yang Changchun, et al. Surface fitting and scattered data compressing with B-spline smoothing function[J]. Progress in Geophysics, 2009, 24(3): 936-943 (in Chinese).

[10] Liu Lei, Jiang Zhaoglian, et al. Research on the AIS data restoration and prediction[J/OL]. Journal of Harbin Engineering University.
http://kns.cnki.net/kcms/detail/23.1390.U.20181217.1013.002.html