

基于直方统计特征的多特征组合航迹关联

徐亚圣^{*①②③} 丁赤飏^{①②④} 任文娟^{②③} 许光奎^{②③}

^①(中国科学院大学 北京 100049)

^②(中国科学院电子学研究所 北京 100190)

^③(中国科学院空间信息处理与应用系统技术重点实验室 北京 100190)

^④(微波成像技术重点实验室 北京 100190)

摘 要: 现有的航迹关联方法主要有基于统计和基于模糊数学两大类方法。基于统计的方法大多依赖阈值的设置, 基于模糊数学的方法参数设置复杂, 且多数方法相关比较时只考虑单个航迹点的信息。针对现有问题, 该文首先从航迹的整体出发, 在传统欧式距离度量的基础上, 提出了一种距离分布直方图的特征并提取了航迹的相似特征, 有效地利用了航迹间的整体特性, 具有较好的抗噪声性能以及关联准确率。其次充分考虑了船舶运动特征以及不同数据源位置精度, 提取了航迹间的速度差分布直方图特征、传感器来源特征。然后将这些特征组合并利用机器学习的方法训练关联模型, 有效地避免了需要人工设定阈值以及参数设置复杂的问题。最后, 该文构建了一个真实的船舶数据集, 实验结果表明距离分布直方图特征相比传统的距离特征总体关联准确率提高了3.23%~11.65%, 组合特征相较于单一的距离分布直方图特征总体关联准确率提高了0.068%, 验证了该文方法的有效性。

关键词: 多传感器航迹关联; 直方统计; 机器学习; 航迹相似度; 多特征组合

中图分类号: TP391

文献标识码: A

文章编号: 2095-283X(2019)01-0025-11

DOI: 10.12000/JR18028

引用格式: 徐亚圣, 丁赤飏, 任文娟, 等. 基于直方统计特征的多特征组合航迹关联[J]. 雷达学报, 2019, 8(1): 25-35. doi: 10.12000/JR18028.

Reference format: XU Yasheng, DING Chibiao, REN Wenjuan, *et al.* Multi-feature combination track-to-track association based on histogram statistics feature[J]. *Journal of Radars*, 2019, 8(1): 25-35. doi: 10.12000/JR18028.

Multi-feature Combination Track-to-track Association Based on Histogram Statistics Feature

XU Yasheng^{*①②③} DING Chibiao^{①②④} REN Wenjuan^{②③} XU Guangluan^{②③}

^①(*University of Chinese Academy of Sciences, Beijing 100049, China*)

^②(*Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China*)

^③(*Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, Chinese Academy of Sciences, Beijing 100190, China*)

^④(*Key Laboratory of Science and Technology on Microwave Imaging, Beijing 100190, China*)

Abstract: Existing track-to-track association methods are mainly based on statistics and fuzzy mathematics. However, most methods based on statistics depend on thresholds, and parameters based on fuzzy mathematics are complex to set. In addition, most methods only consider the information of a single track point in comparison. To solve the existing problems, this paper presents a distance distribution histogram feature to extract the similarity features of a trajectory and measure them using the standardized Euclidean distances; this method effectively utilizes the characteristics of the whole trajectory and has a good anti-noise performance and accuracy. The motion features of ships and the location accuracy of different data sources were fully

收稿日期: 2018-03-29; 改回日期: 2018-05-29; 网络出版: 2018-07-09

*通信作者: 徐亚圣 xuyasheng93@163.com *Corresponding Author: XU Yasheng, xuyasheng93@163.com

基金项目: 国家自然科学基金(61725105, 61331017)

Foundation Items: The National Natural Science Foundation of China (61725105, 61331017)

considered. After obtaining the histogram features of velocity difference and the source features of sensors, the authors combined them and trained association models using machine learning, which effectively avoids the problem of manually setting thresholds and complex parameter settings. Finally, a real ship data set was constructed. The experimental results show that compared with the traditional distance feature, the overall association accuracy was improved by 3.23%~11.65% using the distance distribution histogram feature, and by 0.068% using the combination feature, which verifies the effectiveness of the proposed method.

Key words: Multi-sensor track-to-track association; Histogram statistics; Machine learning; Trajectory similarity; Multi-feature combination

1 引言

数据融合目前在各个领域有着广泛地应用^[1,2], 分布式多传感器融合系统是一种典型的多传感器数据融合系统^[3,4], 具有鲁棒性强、成本低的特点。航迹关联方法的研究一直是分布式多传感器融合系统中的一个热点。航迹关联就是确认多条来自不同传感器的目标轨迹是否属于同一目标, 关联后可以有效地降低数据的冗余, 正确的关联也是实现数据融合的前提。目前航迹关联的方法可以分为两大类, 一类是基于统计的方法, 另一类是基于模糊数学的方法。基于统计的方法主要是利用状态估计的差作为统计量并建立统计假设, 然后以设定的概率阈值来判定航迹是否关联^[4-6]。基于模糊数学的方法主要是选定或设计关联隶属度, 通过计算两两航迹的关联隶属度来判断航迹是否关联^[6-8]。这些方法主要存在以下问题: (1)对目标的航迹进行逐点比较, 没有从航迹整体的角度来考虑, 忽视了航迹的全局性的特征; (2)基于统计的方法需要人工设定阈值和大量的调试, 扩展性不强; (3)只考虑了随机误差, 而忽视了其他误差的影响。

针对上述问题, 大量文献对其进行了研究。文献^[9]将航迹的整体视为时间序列, 使用动态时间规整(Dynamic Time Warping, DTW)来测量任意两条航迹之间的相似度, 从整体上考虑了航迹形状的相似性。将DTW相似性引入航迹关联当中, 避免了在时域上进行配准, 在一定程度上减小了误差。文献^[10]将系统误差引入到原始的传感器量测中, 构建了一种混合整数非线性规划模型, 并对系统误差进行了估计, 有效地提高了系统的性能。文献^[11]提出了一种具有自适应阈值的最大后验概率(Maximum A Posteriori probability, MAP)关联算法。该算法显示, 具有自适应阈值的算法性能优于固定阈值算法。

针对现有方法存在的需要人工设定阈值、参数设置复杂、只考虑单个航迹点以及航迹信息利用不充分的问题, 本文提出了一种基于多特征组合的航迹关联方法。首先从航迹的整体出发, 在传统欧式距离度量的基础上, 提出了一种距离分布直方图的特征并提取了航迹的相似特征, 有效地利用了航迹

整体的特性, 具有较好的抗噪声性能以及关联准确率。其次挖掘了航迹间的速度差分布直方图特征、传感器来源特征, 然后将这些特征组合, 考虑到目前机器学习方法强大的特征学习能力, 能够自动地学习特征中隐含的知识信息, 且参数设置较为简单, 因此利用机器学习方法对挖掘的特征进行学习得到航迹关联模型, 从而有效地避免了需要人工设定阈值以及参数设置复杂的问题。最后通过实验验证了本文方法的有效性。

2 系统模型描述

船舶的航迹由一系列时空数据点(即经度、纬度、时间)组成, 并且通过多种定位方式获得, 由于同一个目标可能会被多个源观测到, 因此在数据融合中心会出现一个目标的多条航迹观测(对于多条观测的情况, 我们可以对观测中的航迹两两进行关联判断, 最后得到关联结果), 在系统中造成大量的数据冗余, 在未判明目标时则会造成大量虚假信息, 对后续的一些航迹知识挖掘任务, 如目标识别、航迹融合造成很大的影响。

假设分布式多传感系统中有 M 个传感器, 每个传感器输出的航迹数为 m_s , 将其中任意一条的航迹数据即目标的状态估计表示为 \mathbf{X}_k^s , 其中 $s = 1, 2, 3, \dots, M$, $k = 1, 2, 3, \dots, m_s$ 。目标的状态估计 \mathbf{X}_k^s 可以表示成为时间序列(1维或者多维的), 这里定义 \mathbf{X}_k^s 为如下的一个多维的时间序列:

$$\begin{aligned} & x_{11}, x_{12}, x_{13}, \dots, x_{1m} \\ & x_{21}, x_{22}, x_{23}, \dots, x_{2m} \\ & \vdots \\ & x_{n1}, x_{n2}, x_{n3}, \dots, x_{nm} \end{aligned}$$

其中, m 表示航迹的长度, n 表示属性的维数。对于任意的两条航迹数据 $\mathbf{X}_{k_i}^{s_i}$ 和 $\mathbf{X}_{k_j}^{s_j}$ (经过一定预处理后, 在同一时间段里的两条航迹数据)存在以下两种假设 H_0 和 H_1 :

$$H_0: \mathbf{X}_{k_i}^{s_i} \text{ 和 } \mathbf{X}_{k_j}^{s_j} \text{ 源于同一目标}$$

$$H_1: \mathbf{X}_{k_i}^{s_i} \text{ 和 } \mathbf{X}_{k_j}^{s_j} \text{ 源于不同目标}$$

本文对航迹数据 $\mathbf{X}_{k_i}^{s_i}$ 和 $\mathbf{X}_{k_j}^{s_j}$ 提取它们的距离分布, 速度差分布直方图特征、DTW相似度特征以及来源特征, 并将特征组合起来送入训练好的机器

学习关联模型中，根据输出的结果判断是否将两条航迹进行关联。

3 算法描述

3.1 特征挖掘

3.1.1 距离分布直方图特征

距离是航迹间最基础也是最直观的特征，传统的航迹间的距离特征有欧式距离、编辑距离以及其它描述航迹形状相似度的距离等，传统的欧式距离主要是计算航迹对应的距离和或者最大最小距离^[12]，且需要两条航迹在时间上对齐，这些特征虽然简单且能在一定程度上解决关联的问题，但也有着十分明显的缺点：(1)单一数值不能很好地反映航迹间的细节差异；(2)在关联时需要人工设定阈值，依赖于人工经验，且关联准确率敏感于噪声。

针对上述存在的问题，本文在欧式距离算法的基础上，提出了一种基于距离分布直方图特征(图1)。特征描述如下：

经过一定预处理(按照4.1.2节中，第1部分所做的时间对齐和抽样。3.1.2节及3.1.3节的特征提取也经过这样的预处理)，两条时间上对齐(采样点的时间相同)、航迹长度相同、特征维度为经度、纬度、时间的航迹 \mathbf{X} 和 \mathbf{Y} ：

$$\begin{array}{ll} x_{11}, x_{12}, x_{13}, \dots, x_{1m} & y_{11}, y_{12}, y_{13}, \dots, y_{1n} \\ x_{21}, x_{22}, x_{23}, \dots, x_{2m} & y_{21}, y_{22}, y_{23}, \dots, y_{2n} \\ x_{31}, x_{32}, x_{33}, \dots, x_{3m} & y_{31}, y_{32}, y_{33}, \dots, y_{3n} \end{array}$$

其中， m 表示航迹的长度，设航迹 \mathbf{X} 和 \mathbf{Y} 的距离序列为

$$\text{distance}(\mathbf{X}, \mathbf{Y}) = \{d_1, d_2, d_3, \dots, d_m\} \quad (1)$$

其中， d_i 表示 \mathbf{X} 和 \mathbf{Y} 第 i 个航迹点对应的距离。为了规范化处理，本文对距离序列 $\text{distance}(\mathbf{X}, \mathbf{Y})$ 取对数，得到：

$$\begin{aligned} & \log(\text{distance}(\mathbf{X}, \mathbf{Y})) \\ & = \{\log(d_1), \log(d_2), \log(d_3), \dots, \log(d_m)\} \quad (2) \end{aligned}$$

统计 $\text{distance}(\mathbf{X}, \mathbf{Y})$ 的直方图分布，并对直方图分布各个区间的频次进行归一化处理，即将直方图分布在各个区间的频次转换为频率从而得到距离分布直方图特征量。

3.1.2 动态时间规整(DTW)相似度特征

航迹作为一种时间序列，可以考虑将其相似度作为一种特征，而DTW是一种典型的用于计算时间序列相似度的算法，它广泛应用于科学、医学、工业、金融等领域，在人工智能领域使用得更为频繁^[13-22]。DTW定义了时间序列间的最佳匹配，它同时支持不同时间长度的相似度度量，具有更好的鲁棒性。DTW也从一定程度上反映了航迹间的形状相似度。因此本文计算航迹间的DTW相似度，并将其作为一类特征。同上有两条航迹长度分别为 m, n 且 $m, n > 1$ 。特征维度为经度、纬度、时间的航迹 \mathbf{X} 和 \mathbf{Y} ：

$$\begin{array}{ll} x_{11}, x_{12}, x_{13}, \dots, x_{1m} & y_{11}, y_{12}, y_{13}, \dots, y_{1n} \\ x_{21}, x_{22}, x_{23}, \dots, x_{2m} & y_{21}, y_{22}, y_{23}, \dots, y_{2n} \\ x_{31}, x_{32}, x_{33}, \dots, x_{3m} & y_{31}, y_{32}, y_{33}, \dots, y_{3n} \end{array}$$

我们构建一个 $m \times n$ 的矩阵 $\mathbf{D} = (d_{ij})_{m \times n}$ 。其中， d_{ij} 为 $d(x_i, y_j)$ ，是两条航迹两个点 x_i 和 y_j 的距离。本文将距离定义为两个经纬度点的实际距离，因此矩阵中的每个元素 (i, j) 表示的是两个点的匹配程度。在矩阵 \mathbf{D} 中构建一条路径 \mathbf{W} 表示航迹 \mathbf{X} 和 \mathbf{Y} 的匹配程度：

$$\begin{aligned} \mathbf{W} & = \{w_1, w_2, w_3, \dots, w_k\}, \\ & \max(m, n) \leq k \leq m+n-1 \quad (3) \end{aligned}$$

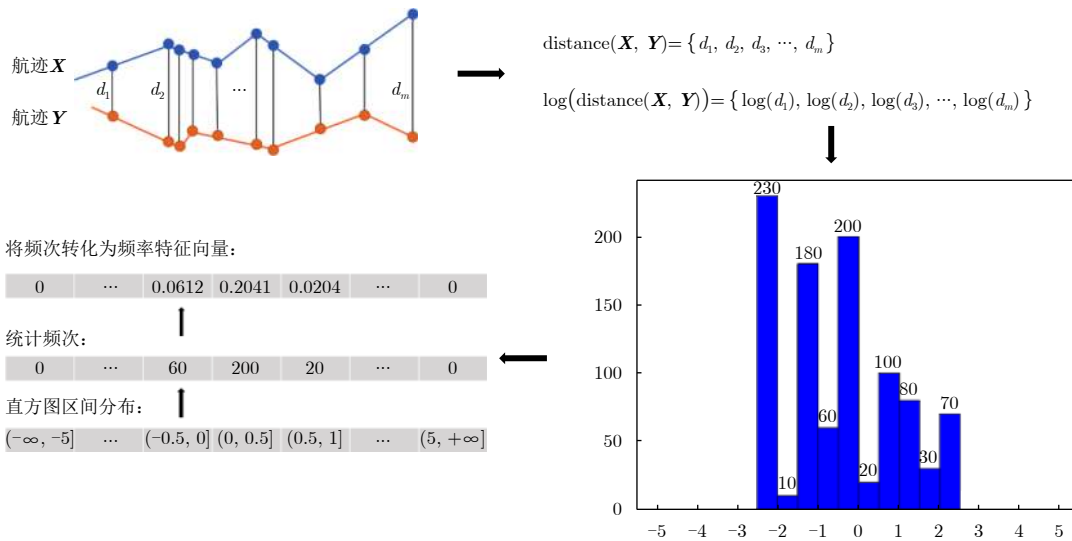


图1 距离分布直方图特征计算示意图

Fig. 1 Schematic diagram of feature calculation of distance distribution histogram

路径 W 必须满足以下 3 个条件:

(1) 边界条件: $w_1 = (1, 1), w_k = (m, n)$;

(2) 连续条件: 给定 $w_p = (i_p, j_p), w_{p+1} = (i_{p+1}, j_{p+1})$, 有 $i_{p+1} - i_p \leq 1, j_{p+1} - j_p \leq 1$;

(3) 单调条件: 给定 $w_p = (i_p, j_p), w_{p+1} = (i_{p+1}, j_{p+1})$, 有 $i_{p+1} - i_p \geq 0, j_{p+1} - j_p \geq 0$ 。

其中有多条路径满足上述的条件, 取满足以下条件的路径:

$$DTW(\mathbf{X}, \mathbf{Y}) = \min_W \left\{ \sqrt{\sum_{l=1}^k w_l} \right\} \quad (4)$$

为了计算式(4), 需要构建一个累加距离矩阵, 并且利用动态规划的思想求解这个矩阵, 递归如下:

$$f(i, j) = d(x_i, y_j) + \min(f(i-1, j), f(i, j-1), f(i-1, j-1)) \quad (5)$$

初始条件为

$$\begin{cases} f(0, 0) = 0, \\ f(i, 0) = \infty, i = 1, 2, \dots, m \\ f(0, j) = \infty, j = 1, 2, \dots, n \end{cases} \quad (6)$$

最终得到的 $f(m, n)$ 为航迹的 DTW 特征。图 2 及图 3 显示了两条航迹在 DTW 算法下的逐点对应关系。

3.1.3 其他特征

航迹间除了距离和 DTW 相似度两种主要特征外, 本文还提取了速度差分布直方图特征以及航迹的来源特征, 考虑到船舶的行驶速度较慢且传感器的采样频率较高, 本文假设任意两个航迹点之间,

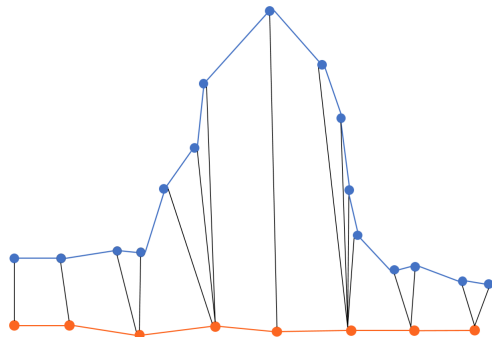


图 2 DTW 航迹点对应关系

Fig. 2 Correspondence of DTW track points

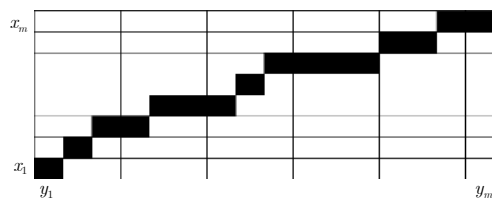


图 3 DTW 算法匹配关系

Fig. 3 DTW algorithm matching relationship

船做的是匀速直线运动, 从而可以得到每一个航迹点的速度, 对于航迹 \mathbf{X} 和 \mathbf{Y} , 记航迹 \mathbf{X} 和 \mathbf{Y} 的速度序列为

$$\begin{aligned} \mathbf{V}_x &= \{vx_1, vx_2, vx_3, \dots, vx_m\}, \\ \mathbf{V}_y &= \{vy_1, vy_2, vy_3, \dots, vy_m\} \end{aligned} \quad (7)$$

其中:

$$v_i = \frac{d(i-1, i)}{t(i-1, i)} \quad (8)$$

$d(i-1, i)$ 表示的是第 i 个点与 $i-1$ 个点的距离, $t(i-1, i)$ 表示的是第 i 个点与 $i-1$ 个点的时间差。

设:

$$\text{dif}V(\mathbf{X}, \mathbf{Y}) = \{\text{dif}v_1, \text{dif}v_2, \text{dif}v_3, \dots, \text{dif}v_m\} \quad (9)$$

为航迹 \mathbf{X} 和 \mathbf{Y} 的速度差序列, 其中 $\text{dif}v_i$ 表示 \mathbf{X} 和 \mathbf{Y} 第 i 个航迹点对应的速度的差。然后对 $\text{dif}V(\mathbf{X}, \mathbf{Y})$ 做与 $\text{distance}(\mathbf{X}, \mathbf{Y})$ 相同的后续处理得到速度差分布直方图特征。

不同来源误差往往不同, 传统的一些方法在进行仿真实验的时候假定误差服从高斯分布, 而实际的误差情况不仅有随机误差还有系统误差的影响, 同时不同来源的精度、特性也不相同, 因此本文将不同数据来源量化为特征, 结合机器学习的方法隐性地学习误差信息从而在一定程度上更好地克服误差影响。假设有 n 个不同的来源:

$$S = \{s_1, s_2, s_3, \dots, s_n\} \quad (10)$$

本文将对其进行一定的编码作为来源特征

$$Cd = \{cd_1, cd_2, cd_3, \dots, cd_n\} \quad (11)$$

3.2 关联流程

对任意两条航迹数据提取距离、速度差分布直方图特征、DTW 相似度特征以及来源特征, 并将这些特征进行组合。然后利用机器学习的模型来进行训练和判断。算法的流程图如图 4 所示。

4 实验验证

4.1 实验数据集构建及预处理

4.1.1 数据集构建

传统的一些方法实验都是仿真数据上进行的, 虽然为了接近真实的情况, 在仿真数据上加入了噪声等干扰, 但大多数都是假定噪声服从高斯分布, 忽视其他误差的影响, 因此与实际的船舶航迹数据仍然有着较大的差距。本文构建了一个真实的船舶航迹数据集, 船舶航迹数据集是 1 年多的船舶航迹数据, 涵盖了较多的航迹运动情况, 数据集中每一个实验样本由两条航迹组成, 两条航迹可能源于同一目标, 也可能源于不同的目标。数据集的航迹样本标注都是经过一定的历史分析以及多个有经验的

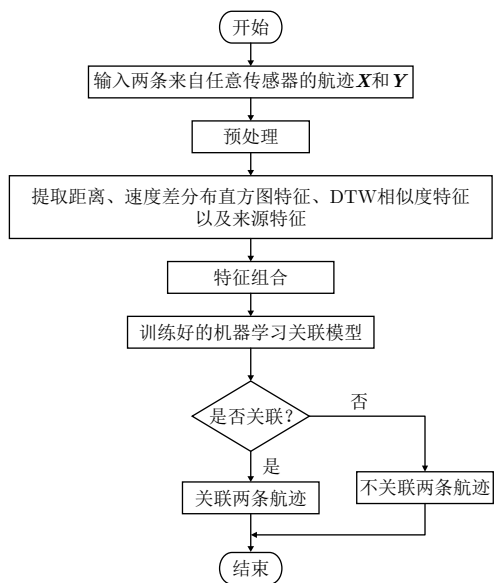


图4 关联流程图

Fig. 4 Association flow chart

判读员检验标注的。因此数据集的标注具有较高的可靠性。本文一共标注了8063个航迹实验样本对，令两条航迹源于同一目标的为正样本，两条航迹源于不同目标的为负样本。其中正样本数量为2522个、负样本为5541个。

4.1.2 数据预处理

(1) 时间对齐和抽样

一般情况下，各个传感器采样频率上报间隔都不尽相同，因此会出现如图5所示的情形。不同传感器的采样频率往往不同，得到的航迹点的时间并不一一对应，如果直接计算距离会有较大的误差，且不满足计算距离分布直方图特征的条件，因此需要对航迹进行插值，再在同一时刻进行采样。

通过对航迹数据集中的数据(在本文构建的数据集里)进行分析后，发现传感器的采样频率大多数在3~300 s，船舶行驶速率大多在11节左右(20 km/h)。因此本文假设船舶的两个采样点之间是做匀速直线运动，从而进行线性插值处理，此假设主要是为了减小直接计算非时间对齐航迹点的欧式距离时的较

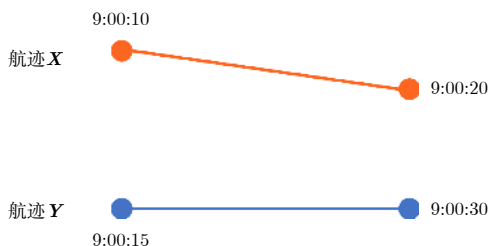


图5 两条航迹对应点时间不同示意图

Fig. 5 Different timings of the corresponding points on the two tracks

大误差同时满足距离分布直方图特征的计算条件。线性插值采样如图6所示。同时为了更符合匀速直线运动的假设，本文将不对采样间隔大于600 s的点做插值处理。

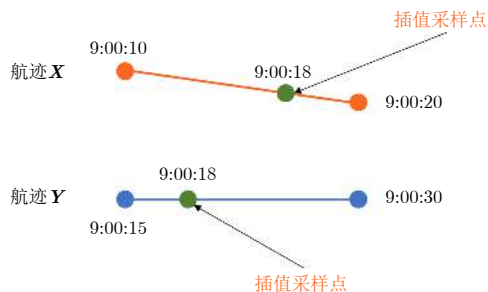


图6 航迹线性插值采样示意图

Fig. 6 Linear interpolation sampling schematic diagram

(2) 样本扩充

原始的可用样本较少，正样本数量仅为2522个、负样本为5541个，通过对样本的观察发现：部分样本的时间跨度较长且样本点数极多(例如有些样本航迹点数有3万多个，时间长度超过20 h)。这样的样本可以对其进行拆分，拆分方式如图7所示，本文将将其拆分为时间长度为2~3 h的多个样本以此来扩充样本。

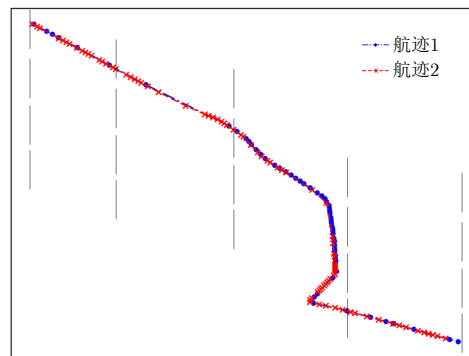


图7 航迹样本拆分示意图

Fig. 7 Schematic diagram of separation of track samples

4.2 实验结果分析

为了验证算法的性能，本文首先比较了传统方法提取的加权距离特征(加权航迹方法计算出的统计量特征)^[23,24]、平均距离特征、最大距离特征在设定阈值与利用机器学习方法时性能的差异，其次比较了距离分布直方特征与传统的平均距离特征、加权航迹特征、最大距离特征，之后对本文提出特征进行组合，比较组合特征和单一特征的性能，最后对比了不同的机器学习方法的性能。实验数据集是经上述预处理后并随机打乱后的船舶样本集，其中80%的样本用来训练，20%的样本用来测试。

4.2.1 评价指标

本文采用常用的机器学习评价指标^[25], 精度、查全率、查准率、F1度量、ROC (Receiver Operating Characteristic)曲线以及AUC (Area Under Curve)值, 令两条航迹源于同一目标的为正样本, 两条航迹源于不同目标的为负样本。则评价指标的具体定义如下:

$$\text{精度} = \frac{\text{总的关联正确样本数}}{\text{总样本数}} \quad (12)$$

$$\text{查全率} = \frac{\text{正确关联的正样本数}}{\text{总的正样本数}} \quad (13)$$

$$\text{查准率} = \frac{\text{正确关联的正样本数}}{\text{关联为正样本的总数}} \quad (14)$$

$$\text{F1} = \frac{2 \times \text{查全率} \times \text{查准率}}{\text{查全率} + \text{查准率}} \quad (15)$$

ROC曲线根据学习器的预测结果对样本进行排序, 并且按照这个顺序把样本预测为正例, 性能越好的模型, ROC曲线越接近于左上角。由于从图形上无法直观比较ROC曲线, 因此可以通过计算ROC曲线下方的面积, 即AUC值, 来比较模型的好坏, 一般来说AUC值越大则表示模型性能越好。

4.2.2 阈值方法与机器学习方法对比实验

传统的方法阈值设置大多根据统计以及设计者经验来决定^[26], 其中一些方法还依赖于系统的参数^[4,27], 本文根据真实数据集的情况, 利用了3种数据的统计值作为阈值来与机器学习方法进行比较, 它们分别是精度最高阈值、平均数阈值以及中位数阈值。精度最高阈值: 也是一般常用的阈值设置方法^[4], 在特征域值范围内的值, 通过一定的步长设置寻找最佳精度阈值; 平均数阈值和中位数阈值主要是利用特征的统计特性^[28], 本文对3种特征分别求得它们平均数与中位数作为阈值, 具体的阈值与实验结果如表1~表3所示。机器学习模型则选较为常用的树模型, 从表1~表3中可以看出, 机器学习的方法则能够更好学习特征表示的含义, 具有更强的泛化能力, 因此关键指标均远高于设定阈值的方法, 且不需要人工设定阈值(精度最高阈值需要大量的实验才能确定), 进而说明了引入机器学习方法的有效性。

4.2.3 距离分布直方图特征与传统特征对比实验

表4是利用树模型学习不同特征的各项指标, 图8描述了不同特征的ROC曲线图。从表4以及图8可以看出本文提出的距离分布直方图特征的性能优

表 1 平均距离特征阈值方法与机器学习方法指标对比

Tab. 1 Comparison of average distance characteristics threshold method and machine learning method index

	阈值	精度	查准率	查全率	F1
机器学习	-	0.9478	0.9617	0.9590	0.9603
精度最高阈值	4524.7	0.8906	0.9170	0.9170	0.9170
平均数阈值	76231.5	0.5998	1.0000	0.3927	0.5640
中位数阈值	8878.8	0.8142	0.9948	0.7218	0.8366

表 2 加权距离特征阈值方法与机器学习方法指标对比

Tab. 2 Comparison of weighted distance characteristics threshold method and machine learning method index

	阈值	精度	查准率	查全率	F1
机器学习	-	0.9264	0.9538	0.9335	0.9435
精度最高阈值	0.0003	0.8626	0.9822	0.8062	0.8856
平均数阈值	0.2681	0.4528	1.0000	0.1679	0.2902
中位数阈值	0.0008	0.8080	0.9858	0.7190	0.8335

表 3 最大距离特征阈值方法与机器学习方法指标对比

Tab. 3 Comparison of maximum distance characteristics threshold method and machine learning method index

	阈值	精度	查准率	查全率	F1
机器学习	-	0.8636	0.8904	0.9043	0.8973
精度最高阈值	14024.4	0.8036	0.9377	0.7520	0.8346
平均数阈值	108357.6	0.6094	0.9696	0.4026	0.5866
中位数阈值	20853.2	0.7713	0.9471	0.6917	0.7995

表 4 不同特征指标对比
Tab. 4 Comparison of different characteristics

特征	精度	查准率	查全率	F1	AUC
距离分布直方图	0.9801	0.9826	0.9873	0.9849	0.9768
平均距离	0.9478	0.9617	0.9590	0.9390	0.9426
加权距离	0.9264	0.9538	0.9335	0.9435	0.9373
最大距离	0.8636	0.8904	0.9043	0.8973	0.8446

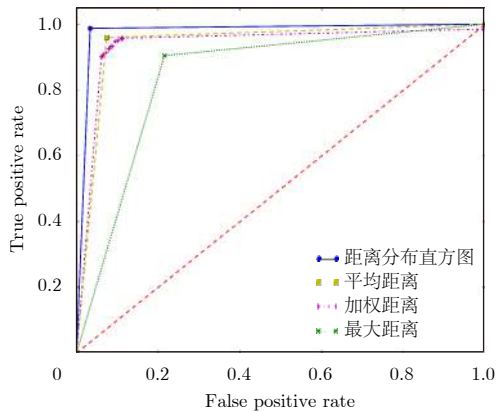


图 8 不同方法ROC曲线

Fig. 8 Different methods of ROC curves

于传统的距离特征。主要原因是距离分布直方图特征从航迹的整体出发，更好地考虑了整体的距离分布。图9是传统距离特征无法识别而距离分布直方图特征能识别的关联目标，图中红色和蓝色分别代表了不同来源观测的同一目标的航迹。从图中可以看出，传统的距离特征由于是单一数值描述，对有噪声干扰的目标不能准确关联，而距离分布直方图特征考虑的是整体的距离分布，因此具有较好的抗噪性能。图10是距离分布直方图特征关联失败的典型情况，图10(a)和图10(b)是两条航迹源于同一目标，图10(c)为两条航迹源于不同目标。从图10(a)和图10(b)中可以直观地看出，其受噪声干扰较为严重，而图10(c)两条航迹十分接近，通过对它们特征的可视化分析发现，图10(a)、图10(b)距离整

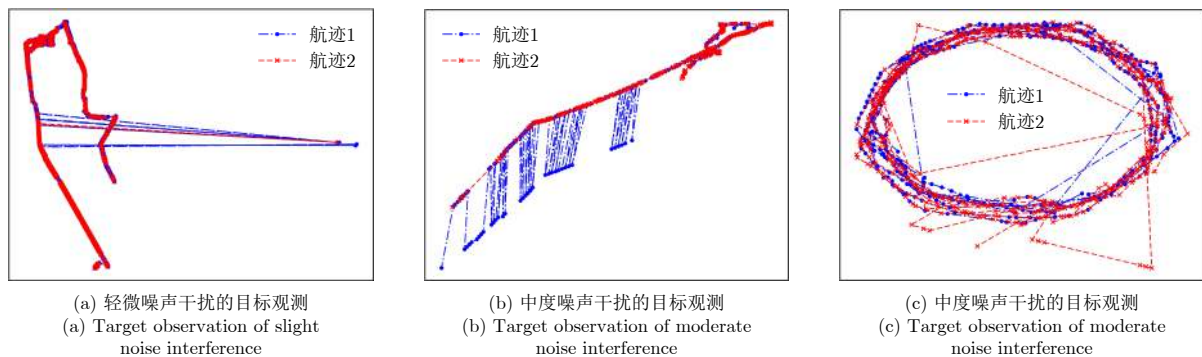


图 9 传统距离特征无法识别而距离分布直方图能够识别的目标

Fig. 9 The targets that can't be identified by traditional distance feature but can be identified by distance distribution histogram

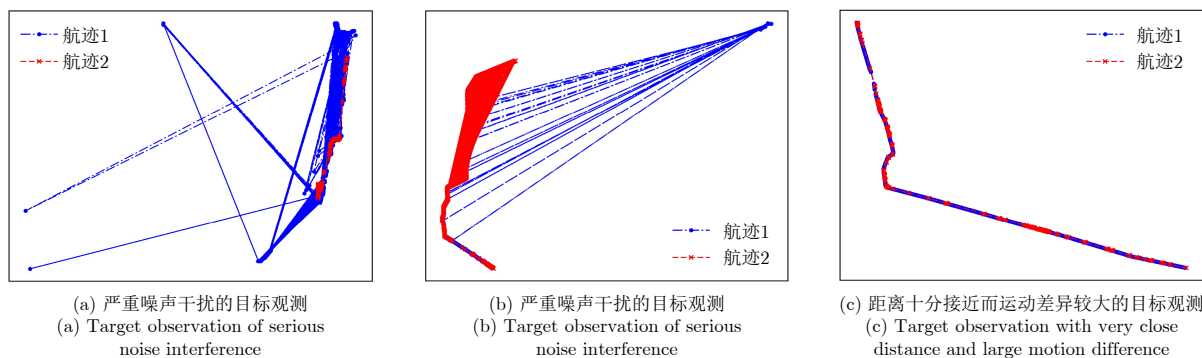


图 10 距离分布直方图特征关联失败的目标

Fig. 10 The targets of distance distribution histogram feature error association

体分布呈现较大差异,此时距离分布直方图特征难以有效地关联图10(a)、图10(b)所示的情况。图10(c)的距离整体分布十分接近,但速度整体分布差异较大,单一的距离分布直方图特征在这种情况下难以正确完成关联判断。

4.2.4 组合特征对比实验

表5是利用树模型学习不同组合特征的各项指标,图11描述了不同组合特征的ROC曲线图,表5以及图11中,DDH (Distance Distribution Histogram)表示的是距离分布直方图特征,SDDH (Speed Difference Distribution Histogram)表示的速度差分布直方图特征。图12给出了单一特征错误关联而

组合特征正确识别情况(图10(c)在加入速度差分布直方图特征后能够准确地判断其为不同的目标),图12(a)是DDH特征错误关联的目标而加入DTW相似度特征准确识别的情形,主要原因是两条航迹的距离较近,在距离维度上很难区分,而其DTW相似度差异较大(DTW相似度在一定程度上描述了航迹形状上的相似度)。图12(b)是DDH+DTW特征错误关联的目标而加入速度差分布直方图特征准确识别的情形,主要原因是两条航迹的距离较近,DTW相似度也较高,而共有12.5%的航迹点速度相差5节航速,速度差分布差异较大。图12(c)是DDH+DTW+SDDH特征错误关联而加入来源特

表 5 组合特征指标对比
Tab. 5 Comparison of composite features

组合特征	精度	查准率	查全率	F1	AUC
DDH	0.9801	0.9858	0.9840	0.9849	0.9783
DDH+DTW	0.9829	0.9859	0.9882	0.9870	0.9804
DDH+DTW+SDDH	0.9842	0.9873	0.9887	0.9880	0.9820
DDH+DTW+SDDH+数据来源特征	0.9869	0.9901	0.9901	0.9901	0.9855

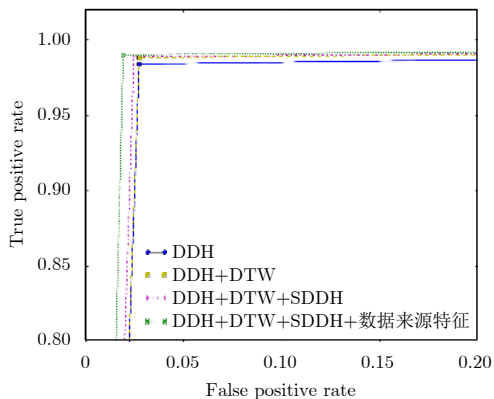
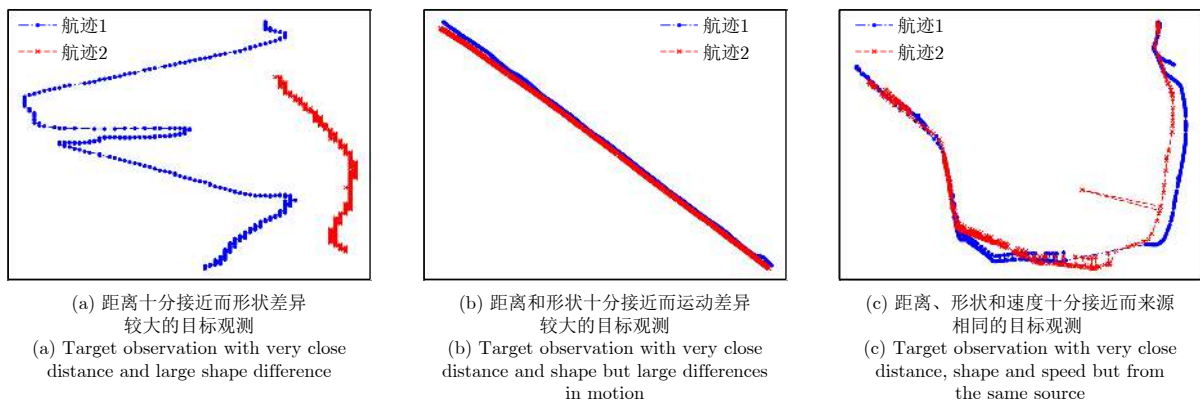


图 11 特征组合ROC曲线

Fig. 11 Feature combination ROC curve

征准确识别的情形,主要原因是两条航迹的距离较近,DTW相似度也高,速度差分布也相近,但两条航迹来自同一观测源,因此判定不是同一目标。图13是距离分布直方图特征关联失败的典型情况,图13(a)和图13(b)是两条航迹源于同一目标,图13(c)为两条航迹源于不同目标。从图13(a)、图13(b)中可以直观地看出,其受噪声干扰较为严重,而图13(c)两条航迹十分接近,通过对它们特征的可视化分析发现,图13(a)和图13(b)的距离、速度整体分布以及DTW相似度呈现较大差异,此时组合特征难以有效地关联图13(a)和图13(b)所示的情况。图13(c)的目标十分接近,且几乎没有移动,因此各项特征



(a) 距离十分接近而形状差异较大的目标观测
(a) Target observation with very close distance and large shape difference
(b) 距离和形状十分接近而运动差异较大的目标观测
(b) Target observation with very close distance and shape but large differences in motion
(c) 距离、形状和速度十分接近而来源相同的目标观测
(c) Target observation with very close distance, shape and speed but from the same source

图 12 单一特征错误关联组合特征正确识别情况

Fig. 12 Correct identification of single feature miscorrelation combination features

十分接近，组合特征在这种情况下难以正确完成关联判断。综上分析以及从表5和图11中可以看出，每增加一组特征，模型的各项指标都有所提升，不仅说明了所提特征的有效性，而且由于不同特征之间具有一定的互补性，多种特征的组合能够更加全面地表征航迹间的关系，进一步提升航迹关联准确性。在噪声干扰较大时航迹间的特征也受到了较大的干扰，此时很难有效地进行关联判定，同时由于

目前航迹点信息有限，可以挖掘的特征有限，因此一些较为特殊的情况下当前组合特征不能准确地完成关联判断。

4.2.5 不同机器学习方法对比实验

表6是利用不同的机器学习方法训练关联模型的各项指标，图14描述了不同机器学习方法的ROC曲线。从表6以及图14可以看出，Bagging各项指标最高，Tree模型各项指标最低。

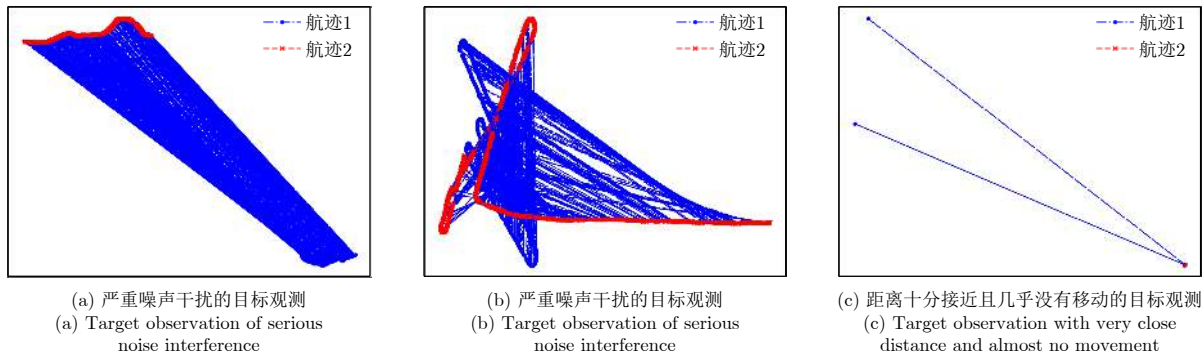


图 13 组合特征关联失败的目标

Fig. 13 The targets of combination features error association

表 6 不同机器学习方法指标对比

Tab. 6 Comparison of different machine learning indicators

学习方法	精度	查准率	查全率	F1	AUC
Tree	0.9863	0.9887	0.9906	0.9896	0.9843
Random forest	0.9873	0.9896	0.9910	0.9903	0.9978
Adaboost	0.9885	0.9915	0.9910	0.9913	0.9965
Bagging	0.9916	0.9925	0.9948	0.9936	0.9979

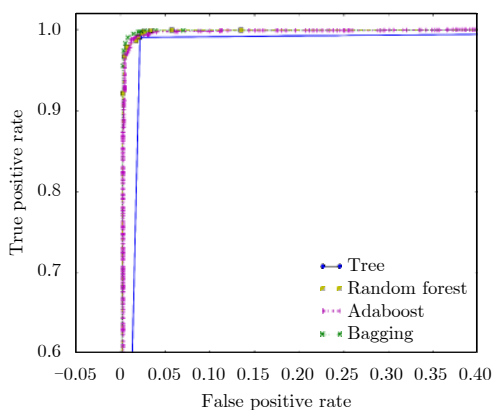


图 14 不同机器学习方法ROC曲线

Fig. 14 ROC curves for different machine learning methods

5 结论

针对传统航迹关联方法存在需要人工设定阈值、参数设置复杂的问题，本文将机器学习的方法引入航迹关联中，针对一些传统方法只考虑航迹单

个点的信息及抗噪声性能较差的问题，本文从航迹的整体出发，提出了一种距离分布直方图特征。与此同时对航迹的特征进行挖掘，提取了航迹间的DTW相似度特征、速度差分布直方图特征以及来源特征，结合机器学习的方法提出了一种基于多特征组合的航迹关联方法，该方法在有限的航迹点信息中挖掘了多种有效的特征信息，获得了较高的关联准确率，并且在实际应用中取得较好的效果，随着样本和航迹点信息不断增多，关联模型的准确率以及泛化性能还将进一步提升。理论分析和实验结果均表明该方法的有效性和合理性。根据算法设计中各环节的假设和约束，本文目前提出的航迹关联算法主要适用于传感器采样频率较高，运动目标速度较慢(船舶、汽车、行人等)的情况，对于高速运动的目标则需要进一步地分析其运动情况，建立更加精细的插值模型，这也是本文后续工作中的一个重要的研究方面。

参 考 文 献

- [1] 杨威, 陈杰, 李春升. 面向目标特性精细提取的SAR数据融合成像处理方法[J]. 雷达学报, 2015, 4(1): 29–37. doi: [10.12000/JR15017](https://doi.org/10.12000/JR15017).
YANG Wei, CHEN Jie, and LI Chun-sheng. SAR data fusion imaging method oriented to target feature extraction[J]. *Journal of Radars*, 2015, 4(1): 29–37. doi: [10.12000/JR15017](https://doi.org/10.12000/JR15017).
- [2] 李伟, 王兴亮, 邹鲲, 等. 基于数据融合和陷波滤波的MIMO雷达抗欺骗干扰算法[J]. 雷达学报, 2012, 1(3): 246–252. doi: [10.3724/SP.J.1300.2012.20060](https://doi.org/10.3724/SP.J.1300.2012.20060).
LI Wei, WANG Xing-liang, ZOU Kun, *et al.* Anti deceptive jamming for MIMO radar based on data fusion and notch filtering[J]. *Journal of Radars*, 2012, 1(3): 246–252. doi: [10.3724/SP.J.1300.2012.20060](https://doi.org/10.3724/SP.J.1300.2012.20060).
- [3] LIGGINS M E, HALL D L, and LLINAS J. Handbook of Multisensor Data Fusion: Theory and Practice[M]. 2nd ed., Boca Raton, FL: CRC Press, 2009.
- [4] 何友, 王国宏, 陆大经, 等. 多传感器信息融合及应用[M]. 第2版, 北京: 电子工业出版社, 2007.
HE You, WANG Guo-hong, LU Da-jin, *et al.* Multisensor Information Fusion with Applications[M]. 2nd ed., Beijing: Publishing House of Electronics Industry, 2007.
- [5] CHANG C and YOUENS L C. Measurement correlation for multiple sensor tracking in a dense target environment[J]. *IEEE Transactions on Automatic Control*, 1982, 27(6): 1250–1252. doi: [10.1109/TAC.1982.1103107](https://doi.org/10.1109/TAC.1982.1103107).
- [6] Bar-Shalom Y. On the track-to-track correlation problem[J]. *IEEE Transactions on Automatic Control*, 1981, 26(2): 571–572. doi: [10.1109/TAC.1981.1102635](https://doi.org/10.1109/TAC.1981.1102635).
- [7] AZIZ A M, TUMMALA M, and CRISTI R. Fuzzy Logic Data Correlation Approach in Multisensor-Multitarget Tracking Systems[M]. Elsevier North-Holland, Inc, 1999. <https://www.sciencedirect.com/science/article/pii/S0165168499000080>.
- [8] SMITH III J F. Fuzzy logic multisensor association algorithm[C]. Proceedings of SPIE Signal Processing, Sensor Fusion, and Target Recognition VI, Orlando, FL, USA, 1997: 76–87. doi: [10.1117/12.280788](https://doi.org/10.1117/12.280788).
- [9] YANG Y T, LIANG Y, YANG Y B, *et al.* Asynchronous track-to-track association algorithm based on dynamic time warping distance[C]. Proceedings of the 34th Control Conference, Hangzhou, China, 2015: 4772–4777.
- [10] 董凯, 王海鹏, 刘瑜. 基于拓扑统计距离的航迹抗差关联算法[J]. 电子与信息学报, 2015, 37(1): 50–55. doi: [10.11999/JEIT140244](https://doi.org/10.11999/JEIT140244).
DONG Kai, WANG Hai-peng, and LIU Yu. Anti-bias track association algorithm based on topology statistical distance[J]. *Journal of Electronics & Information Technology*, 2015, 37(1): 50–55. doi: [10.11999/JEIT140244](https://doi.org/10.11999/JEIT140244).
- [11] STONE L D, TRAN T M, and WILLIAMS M L. Improvement in track-to-track association from using an adaptive threshold[C]. Proceedings of the 12th International Conference on Information Fusion, Seattle, WA, USA, 2009: 1627–1633.
- [12] AGRAWAL R, FALOUTSOS C, and SWAMI A N. Efficient similarity search in sequence databases[C]. Proceedings of the 4th International Conference on Foundations of Data Organization and Algorithms, Berlin, Heidelberg, Germany, 1993: 69–84.
- [13] BARCZEWSKA K and DROZD A. Comparison of methods for hand gesture recognition based on dynamic time warping algorithm[C]. Proceedings of 2013 Federated Conference on Computer Science and Information Systems, Poland, 2013: 207–210.
- [14] BANKÓ Z and ABONYI J. Correlation based dynamic time warping of multivariate time series[J]. *Expert Systems with Applications*, 2012, 39(17): 12814–12823. doi: [10.1016/j.eswa.2012.05.012](https://doi.org/10.1016/j.eswa.2012.05.012).
- [15] CASACUBERTA F, VIDAL E, and RULOT H. On the metric properties of dynamic time warping[J]. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1987, 35(11): 1631–1633. doi: [10.1109/TASSP.1987.1165065](https://doi.org/10.1109/TASSP.1987.1165065).
- [16] HUANG B and KINSNER W. ECG frame classification using dynamic time warping[C]. Proceedings of 2002 Canadian Conference on Electrical and Computer Engineering, Winnipeg, Manitoba, Canada, 2002: 1105–1110.
- [17] JABLONSKI B. Quaternion dynamic time warping[J]. *IEEE Transactions on Signal Processing*, 2012, 60(3): 1174–1183. doi: [10.1109/TSP.2011.2177832](https://doi.org/10.1109/TSP.2011.2177832).
- [18] FU C X, ZHANG P L, JIANG J, *et al.* A Bayesian approach for sleep and wake classification based on dynamic time warping method[J]. *Multimedia Tools and Applications*, 2017, 76(17): 17765–17784. doi: [10.1007/s11042-015-3053-z](https://doi.org/10.1007/s11042-015-3053-z).
- [19] BAUMANN M, OZDOGAN M, RICHARDSON A D, *et al.* Phenology from landsat when data is scarce: Using MODIS and dynamic time-warping to combine multi-year landsat imagery to derive annual phenology curves[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2017, 54: 72–83. doi: [10.1016/j.jag.2016.09.005](https://doi.org/10.1016/j.jag.2016.09.005).
- [20] VARATHARAJAN R, MANOGARAN G, PRIYAN M K, *et al.* Wearable sensor devices for early detection of Alzheimer disease using dynamic time warping algorithm[J]. *Cluster Computing*, 2017(6): 1–10. doi: [10.1007/s10586-017-0977-2](https://doi.org/10.1007/s10586-017-0977-2).
- [21] ZHOU N C, WANG J, and WANG Q G. A novel estimation method of metering errors of electric energy based on membership cloud and dynamic time warping[J].

- IEEE Transactions on Smart Grid*, 2017, 8(3): 1318–1329. doi: [10.1109/TSG.2016.2619375](https://doi.org/10.1109/TSG.2016.2619375).
- [22] WAN Y, CHEN X L, and SHI Y. Adaptive cost dynamic time warping distance in time series analysis for classification[J]. *Journal of Computational and Applied Mathematics*, 2017, 319: 514–520. doi: [10.1016/j.cam.2017.01.004](https://doi.org/10.1016/j.cam.2017.01.004).
- [23] SINGER R A and KANYUCK A J. Computer Control of Multiple Site Track Correlation[M]. Tarrytown, NY, USA: Pergamon Press, Inc., 1971.
- [24] DITZLER W R. A demonstration of multisensor tracking[C]. Proceedings of the 1987 Tri-Service Data Fusion Symposium, JHU/APL, Laurel, Md., 1987: 303–311.
- [25] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
ZHOU Zhi-hua. Machine Learning[M]. Beijing: Tsinghua University Press, 2016.
- [26] 周晓安. VTS系统中多传感器目标航迹关联算法研究[D]. [硕士学位论文], 南京信息工程大学, 2014.
ZHOU Xiao-an. The research of track association for multi-sensor target in VTS system[D]. [Master dissertation], Nanjing University of Information Science & Technology, 2014.
- [27] 周晓安, 高贵明. 自适应阈值航迹关联算法与实现[J]. 现代防御技术, 2014, 42(4): 193–198. doi: [10.3969/j.issn.1009-086x.2014.04.031](https://doi.org/10.3969/j.issn.1009-086x.2014.04.031).
ZHOU Xiao-an and GAO Gui-ming. Adaptive threshold algorithm and implementation for track-to-track association[J]. *Modern Defense Technology*, 2014, 42(4): 193–198. doi: [10.3969/j.issn.1009-086x.2014.04.031](https://doi.org/10.3969/j.issn.1009-086x.2014.04.031).
- [28] HAN Jia-wei and KAMBER M. 数据挖掘: 概念与技术[M]. 范明, 孟小峰, 译. 北京: 机械工业出版社, 2007.
HAN Jia-wei and KAMBER M. Data Mining: Concepts and Techniques[M]. FAN Ming, MENG Xiao-feng, trans. Beijing: China Machine Press, 2007.

作者简介



徐亚圣(1993–), 男, 湖南耒阳人。2015年在吉林大学获得工学学士学位, 现为中国科学院大学, 中国科学院电子学研究所硕士研究生。主要研究方向为航迹关联、机器学习。

E-mail: xuyasheng93@163.com



丁赤飏(1969–), 男, 陕西西安人, 研究员, 博士生导师。1997年在北京航空航天大学获得博士学位, 现任中国科学院电子学研究所副所长, 微波成像技术国家重点实验室主任, 国家863计划专家。主要研究方向为雷达信号处理。

E-mail: cbding@mail.ie.ac.cn



任文娟(1982–), 女, 河南焦作人, 副研究员, 博士。2011年在中国科学院电子学研究所获得博士学位, 现为中国科学院电子学研究所中国科学院空间信息处理与应用系统技术重点实验室副研究员, 主要研究方向为多源遥感信息融合

处理与应用技术。

E-mail: wjren@mail.ie.ac.cn



许光銮(1978–), 男, 浙江天台人, 研究员, 博士生导师。2005年在中国科学院电子学研究所获得博士学位, 现为中国科学院电子学研究所研究员, 中国科学院空间信息处理与应用系统技术重点实验室主任, 主要研究方向为地理空间信

息挖掘与应用技术。

E-mail: gluanxu@mail.ie.ac.cn