

一种海上非作战目标实时清洗方法

林雪原, 李雪腾, 潘新龙, 李敏波, 陈祥光

引用本文:

林雪原, 李雪腾, 潘新龙, 等. 一种海上非作战目标实时清洗方法[J]. 武汉大学学报·信息科学版, 2021, 46(9): 1378–1385.

LIN Xue-yuan, LI Xue-teng, PAN Xin-long, et al. A Real-Time Cleaning Method for Marine Non-Combat Targets[J]. *Geomatics and Information Science of Wuhan University*, 2021, 46(9): 1378–1385.

相似文章推荐 (请使用火狐或IE浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

基于中心点检测和重识别的多行人跟踪算法

A Multi-Pedestrian Tracking Algorithm Based on Center Point Detection and Person Re-identification

武汉大学学报·信息科学版. 2021, 46(9): 1345–1353 <https://doi.org/10.13203/j.whugis20210328>

空天网格化星间通视及路由路径规划算法

Aerospace Grid-Based Algorithm of Inter-satellite Visibility and Route Path Planning for Satellite Constellation

武汉大学学报·信息科学版. 2021, 46(1): 50–57 <https://doi.org/10.13203/j.whugis20200234>

矢量居民地多边形多级图划分聚类方法

A Multi-level Graph Partition Clustering Method of Vector Residential Area Polygon

武汉大学学报·信息科学版. 2021, 46(1): 19–29 <https://doi.org/10.13203/j.whugis20190358>

一种基于改进LCSS的相似轨迹提取方法

A Similar Trajectory Extraction Method Based on Improved LCSS

武汉大学学报·信息科学版. 2020, 45(4): 550–556 <https://doi.org/10.13203/j.whugis20180406>

一种综合利用图像和光谱信息的物体真假模式识别方法

A Method of True and Fake Objects Pattern Recognition Integrating Image Information and Spectral Information

武汉大学学报·信息科学版. 2019, 44(8): 1174–1181 <https://doi.org/10.13203/j.whugis20190139>

DOI:10.13203/j.whugis20190274



文章编号:1671-8860(2021)09-1378-08

一种海上非作战目标实时清洗方法

林雪原¹ 李雪腾^{2,3} 潘新龙³ 李敏波⁴ 陈祥光¹

1 烟台南山学院电气与电子工程系,山东 烟台,265713

2 91827部队,山东 威海,264200

3 海军航空大学信息融合研究所,山东 烟台,264001

4 复旦大学软件学院,上海,200433

摘要:在进行海上作战态势分析时,通常需要剔除目标海域内对态势分析影响较小的非作战目标,只保留参考价值较高、作战相关的目标航迹数据。现有的行为规律挖掘方法大多是基于聚类的思想,作用于非作战目标清洗问题时工作步骤较为复杂、效果较差。结合态势分析需求,基于相似重复记录检测的思想,通过定义多维度记录匹配相似度(multi-dimension record similarity,MDRS),提出了一种海上非作战目标实时清洗方法。通过对多维航迹数据的相似重复检测,实现对非作战目标的实时清洗。在仿真军事场景上进行实验分析,结果表明所提方法能够实时、有效地检测出海上非作战目标。

关键词:态势分析;非作战目标;实时清洗;相似重复记录

中图分类号: P208;TP274.2

文献标志码: A

目前,大多数轨迹数据挖掘技术都是通过聚类算法实现的,但实际应用于干扰因素多、数据量大的海战场环境时,该算法存在以下弊端:(1)聚类算法大多是将空间位置相近的轨迹聚为一类,极少考虑战场态势分析所需的目标位置、速度、航向、属性和类型等多维特征;(2)海战场环境下数据量较大,聚类算法复杂的工作过程大大降低了运行速率;(3)海战场环境下的目标航迹复杂多变,面对不同类型的目标,聚类算法的参数设置很难统一;(4)采用聚类算法时,经常分离出一些不属于任何类的噪声点,而在海战场环境下,这些噪声点很可能是具有单元素的特殊群,若将这些特殊群当作噪声点处理,会对战场态势分析结果产生较大的影响。

本文参考数据预处理技术中的重复数据清洗技术^[1],将每条目标航迹数据看作一条记录^[2]。在处理航迹数据时,可将两条行为规律相似(空间位置相近、速度和航向相似、敌我属性相同、目标类型一致)的航迹看作相似记录,将复杂的目标航迹行为规律挖掘问题^[3-4]转化为较简单的相似重复记录检测问题。并且,在传统的相似重复记录检测^[5-6]中,多源数据通常是包含标识符的

字符串数据,两个字符串是否相等不能仅通过算术得出,往往需要定义一组等价规则,使得重复记录的清洗变得非常复杂。但在海战场环境下,多源数据是经特殊编码后形成的报文数据,即用数值形式表示航迹中相应的属性信息,如目标的敌我属性,分别用“0”“1”“2”表示我方、敌方和中立方,因此,将相似重复记录检测技术应用于海战场数据处理过程,可大大降低计算记录间匹配相似度的复杂程度,有利于进行目标航迹行为规律挖掘。

1 相似重复记录检测原理

在构建数据库时,需要导入大量的多源数据并进行融合^[7-8]。理想情况下,现实世界的每个实体在数据库中都会有唯一一条记录与之对应;而实际情况下,若不同数据源对同一实体的记录存在差异,系统不能有效识别、判定为同一实体,则合并后的数据库就会同时保留这两条记录。若数据库中的两条相似但不完全重复的记录代表同一实体,则称这两条记录为相似记录;若数据库中不存在两条所有属性值都完全相同的记录,则

收稿日期:2020-09-28

项目资助:国家自然科学基金(91538201,61671157);烟台市“双百计划”人才项目(YT201803)。

第一作者:林雪原,博士,教授,主要从事信息融合与组合导航工作。linxy_ytcn@126.com

通讯作者:李雪腾,硕士,助理工程师。574862300@qq.com

称之为重复记录。通常把相似记录和重复记录统称为相似重复记录,即同一实体对应数据库中的多条记录。

相似重复记录检测,就是从多源数据集中找出表达现实世界中同一实体的记录。判断两条记录是否相似重复,首先要比较记录的各对应属性,计算其字段相似度,再根据各属性的权重进行加权平均,得到记录相似度。若两条记录的相似度超过某一阈值,则认为这两条记录是匹配的、相似重复的,将二者合并或删除。

2 多维度记录匹配相似度

2.1 海上目标多维航迹数据

在进行海上态势分析的过程中,目标航迹数据通常是由多维航迹数据点组成的多维序列^[9],每个航迹数据点中包含时间、位置、速度、加速度、航向、敌我属性、类别等多维特征信息^[10]。设目标海域内的全部航迹样本数据集为 T ,表达式为:

$$T = \{T_1, T_2 \dots T_i \dots T_n\} \quad (1)$$

式中, T_i 为航迹样本; $i \in [1, n]$ 为样本编号; n 为样本总数。每条航迹样本 T_i 中包含 m 个按时间顺序排列的多维数据点,表达式为:

$$T_i = \{T_i(1), T_i(2) \dots T_i(j) \dots T_i(m)\} \quad (2)$$

式中, $T_i(j)$ 为第 i 条航迹样本中第 j 个多维航迹数据点; $j \in [1, m]$ 为航迹数据点编号; m 为该航迹样本中的数据点总数。每个航迹数据点 $T_i(j)$ 的表达式为:

$$T_i(j) = \{ \text{time, position, velocity, acceleration, course, attribute, type} \dots \} \quad (3)$$

式中, time、position、velocity、acceleration、course、attribute 和 type 分别表示时间、位置、速度、加速度、航向、敌我属性和类型等多维特征。

2.2 构造过程

本文基于相似重复记录检测的思想,将数据集 T 中的每条航迹样本 T_i 看作一条记录,则构成航迹样本 T_i 的 m 个数据点即为该记录的 m 个子记录,每个数据点 $T_i(j)$ 中包含的时间、位置、速度、航向、敌我属性和类型等特征信息为该子记录的各字段。对于任意两条多维航迹 $T_1, T_2 \in T$, 二者的多维度记录匹配相似度 (multi-dimension record similarity, MDRS) 为 $S_{\text{MDRS}}(T_1, T_2)$, 其构造过程如下。

1) 比较记录的各对应属性。目标航迹数据

中的属性较多,在进行记录匹配时需选择能代表记录特征的属性,这要求操作员对数据本身的含义有深入的了解,能够结合战场需求和专家经验选择合适的属性。

2) 计算字段相似度。字段匹配^[11-12]是记录匹配的基础,是相似重复记录检测的核心问题。字段相似度 S_f 是根据字段信息计算出的表示两字段间相似程度的数值, $0 \leq S_f \leq 1$ 。 S_f 越大,两字段的相似程度越高;若 $S_f = 1$,则两字段为重复字段。针对海上作战环境下的数值型字段,设 T_1, T_2 在同一时刻、针对同一属性 x 的字段值分别为 x_1, x_2 , 则有如下关系:

$$S_f(x_1, x_2) = \left[1 - \frac{|x_1 - x_2|}{\max(x_1, x_2)} \right] \times 100\% \quad (4)$$

3) 属性权重分配和绝对权值优化。属性权重 W 表示了一个属性在决定两条记录相似性中的重要程度^[13],属性越重要,分配到的权重就越大,所有属性权重的和为 1。本文针对海上作战环境,提出绝对权值 V 的概念,用于综合权重的优化。设 T_1, T_2 在同一时刻的位置、速度、航向特征都极为相近,但二者敌我属性不同,则 T_1, T_2 不可能是相似重复记录。因此,只有两条记录的某些属性值完全匹配时,才能进行记录的比较,设该属性的权值 $V = 1$,称为绝对权值,绝对权值对应属性的字段相似度只能为 0 或 1。该方法能够有效提高相似重复记录检测的准确率,减少字段匹配次数,缩短运行时间。在计算记录相似度时,绝对权值对应的属性值与传统属性权重对应的属性值取乘积。

2.3 计算方法

本文采用一种高效的、应用无关的 Pair-wise 比较算法^[14-16]计算 MDRS,其步骤如下:

1) 设每条记录 T_i 中有 $p + q$ 个参与比较的属性字段 $f(1), f(2) \dots f(p + q)$, 其中有 q 个属性对应绝对权值,故各字段的属性权重分别为 $W(1), W(2) \dots W(p), V(1), V(2) \dots V(q)$, 上述权重满足 $\sum_{i=1}^p W(i) = 1$, 且 $V(j) = 1, j = 1, 2 \dots q$ 。

2) 给定两条记录 T_1, T_2 , 其对应的子记录分别为 $T_1(r), T_2(r), r \in [1, m]$, 各字段相似度分别为 $S_f(1), S_f(2) \dots S_f(i) \dots S_f(p + q), i \in [1, p + q]$ 。设 δ_1 为字段相似度阈值,若 $S_f(i) < \delta_1$, 则 T_1, T_2 不是相似重复记录。

3) 根据字段相似度 $S_f(i), i \in [1, p + q]$ 及相应的属性权重,进行加权平均,得到 $T_1(r), T_2(r)$

的相似度 S_r ,计算式为:

$$S_r(T_1(r), T_2(r)) = \frac{\sum_{i=1}^p W(i) \cdot S_f(i)}{\sum_{i=1}^p W(i)} \cdot \left[\prod_{j=1}^q V(j) \cdot S_f(j) \right] \quad (5)$$

4)对所有 m 个子记录的记录相似度求平均,可得 T_1, T_2 的MDRS,计算式为:

$$S_{MDRS}(T_1, T_2) = \sum_{r=1}^m S_r(T_1(r), T_2(r)) / m \quad (6)$$

设 δ_2 为记录相似度阈值,若 $S_{MDRS}(T_1, T_2) > \delta_2$,则认为 T_1, T_2 是相似重复记录。

3 非作战目标实时清洗方法

本文提出了一种海上非作战目标实时清

洗方法(real-time cleaning method for marine non-combat targets, MNCT-RTCM),基本流程如图1所示。

MNCT-RTCM方法的主要步骤包括:(1)对历史数据库中的历史航迹数据进行处理,挖掘历史目标航迹的行为规律;(2)筛选出目标海域内的固定目标(如岛屿、漂浮物)和规律航行的目标(包括规律航行的军用船只和民用船只),对其进行分类存储;(3)将筛选出的历史非作战目标(包括固定目标和规律航行的民用船只)航迹数据点与当前海战场环境下已更新的实时目标航迹数据点进行相似重复检测,若有实时目标航迹与历史非作战目标航迹是相似重复的,则该实时目标就是非作战目标。

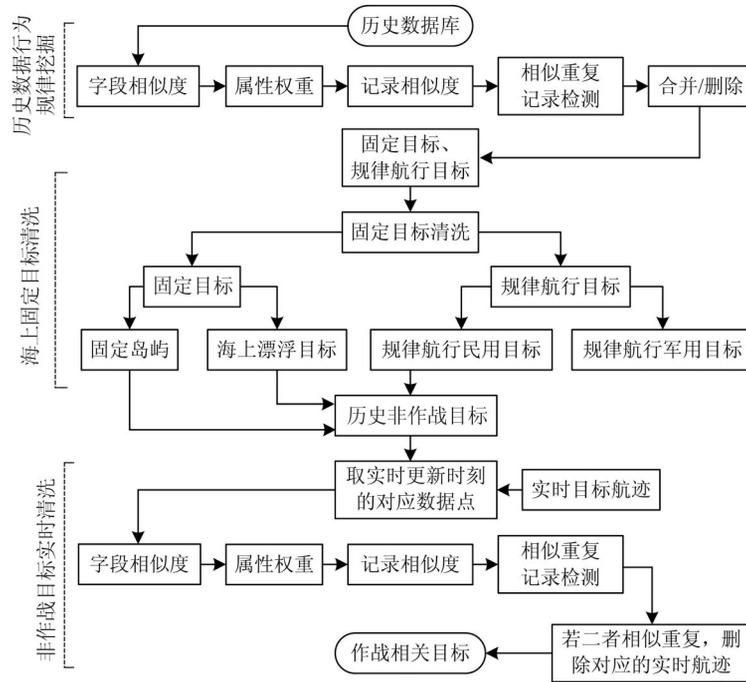


图1 海上非作战目标实时清洗方法流程

Fig.1 Flowchart of Real-Time Cleaning Method for Marine Non-combat Targets

MNCT-RTCM方法的具体流程如下:

1)输入历史数据集。

$$T_p = \{T_{p1}, T_{p2} \dots T_{pi} \dots T_{pp}\} \quad (7)$$

式中, T_p 为过去 N 天内的全部目标航迹集合; p 为历史航迹总数。

2)将 T_p 中的每条记录 T_{pi} 与其他所有记录 $T_{pj}(j \neq i)$ 逐个比较^[17],计算 $S_{MDRS}(T_{pi}, T_{pj})$ 。在海战场环境下,对日期、敌我属性和类型3个字段设定绝对权值,若两条记录的日期相同,则不能进行字段比较;若两条记录的敌我属性和类型不同,则不能进行字段比较。

3)将相似的记录合并,只保留一条正确的记录。得到固定目标与规律航行目标数据集 T_C 和不规律航行目标数据集 T_A 。

4)对筛选出的数据集 T_C 进行固定目标清洗,分离出目标海域内的固定目标集 T_G 和规律航行目标集 T_H 。根据目标的属性和类型信息,可进一步将规律航行目标集 T_H 细分为军用目标集 T_J 和民用目标集 T_M ,得到对战场态势分析影响较小的历史非作战目标集 T_Y ,表达式为:

$$T_Y = T_G \cup T_M = \{T_{Y1}, T_{Y2} \dots T_{Yy}\} \quad (8)$$

式中, $T_Y \in T_C$; y 为历史非作战目标航迹总数。

5) 输入海战场环境下已更新的实时数据集。

$$T_X = \{T_{X1}, T_{X2} \cdots T_{Xi} \cdots T_{Xr}\} \quad (9)$$

式中, T_X 为全部实时目标航迹集合; x 为实时航迹总数。每条目标航迹 T_{Xi} 中含有 z ($z \leq m$) 个已更新的多维数据点, 即

$$T_{Xi} = \{T_{Xi}(1), T_{Xi}(2) \cdots T_{Xi}(j) \cdots T_{Xi}(z)\} \quad (10)$$

式中, $T_{Xi}(j)$ 为第 i 条实时航迹样本中的第 j 个多维航迹数据点。

6) 将 T_{Xi} 的数据点 $T_{Xi}(j)$ 与 T_Y 的每条记录 T_{Yk} ($k \in [1, y]$) 中对应时刻的数据点 $T_{Yk}(j)$ 逐个进行比较, 计算 $S_{MDRS}(T_{Xi}, T_{Yk})$ 。

7) 若 T_{Xi} 与 T_{Yk} 为相似重复记录, 则该实时目标航迹记录 T_{Xi} 就是要清洗的非作战目标航迹, 予以剔除。

4 实验验证与分析

4.1 数据集

本文实验参考偷袭珍珠港、中途岛海战等经典战役的作战过程及兵力分布情况, 结合海上复杂环境、现代武器装备性能、作战条令和战术战

法基础等, 模拟军用场景建立了一个历史仿真数据集和一个海上作战实时仿真数据集。

4.1.1 历史仿真数据集

选取某宽阔海域 450 n mile \times 300 n mile 的区域进行战场环境仿真, 设定仿真时间为过去 N 天内的每天 12:00—18:00。数据集包含 $109 \times N$ 条行为规律的目标航迹和 θ 条行为不规律的目标航迹。其中, 行为规律的 109 个目标中, 包括 8 个固定岛群共 69 个固定岛屿、6 个海上漂浮物、我方 3 个军用编队共 12 条目标船只、敌方 3 个军用编队共 11 条目标船只和 11 条民用船只。假设每 10 min 更新一次战场态势信息, 且船只在该海域内的航行高度是相对固定的, 因此, 每条目标航迹数据均由 36 个数据点组成且不考虑目标的高度信息。每条目标航迹数据在传统二维位置信息的基础上, 结合现代海战场实际, 人为添加了日期、初始速度/加速度、敌我属性和类型等多维信息, 计算得出目标的速度、位置和航向特征等 11 个属性值。将全部航迹随机排序, 共同构造出一个 $(109 \times N + \theta) \times 11 \times 36$ 的多维目标航迹历史仿真数据集 T_P , 如图 2 所示。

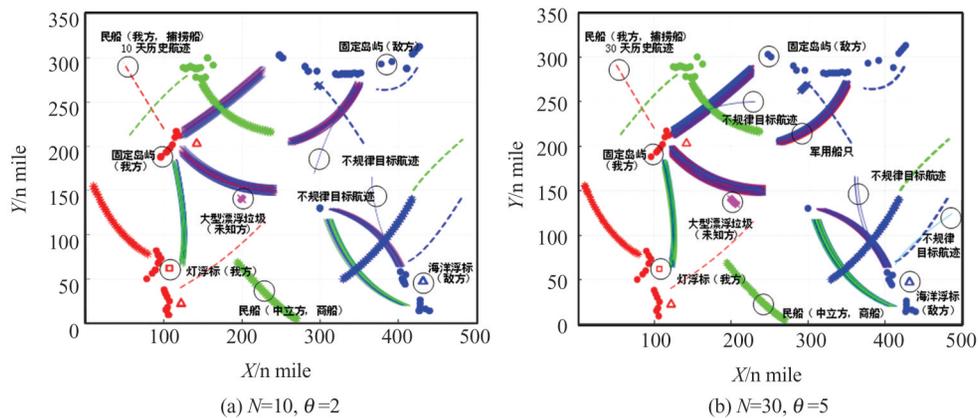


图 2 历史仿真数据集

Fig.2 Historical Simulation Dataset

4.1.2 实时仿真数据集

参照历史仿真数据集的场景设定, 选取某日 12:00—18:00、该宽阔海域 450 n mile \times 300 n mile 的区域进行战场环境仿真。得到的实时仿真数据集 T_X 包括 8 个固定岛群共 69 个固定岛屿、6 个海上漂浮物、我方 5 个作战编队共 113 条目标船只、敌方 3 个作战编队共 43 条目标船只和 10 条民用船只, 共 241 个海上目标的航迹数据。假设每 10 min 更新一次战场态势信息, 构建一个实时的、变化的海上作战过程, 6 h 后形成的全部目标航迹分布如图 3(a) 所示。为了验证非作战目标实时

清洗的效果, 实时仿真数据集 T_X 中设置了两条与历史非作战目标航迹轨迹不同的特殊航迹, 如图 3(b) 所示。

4.2 实验与分析

选用图 2(b) 所示 $N=30$ 、 $\theta=5$ 时的历史仿真数据集和图 3(a) 所示的海上作战实时仿真数据集, 利用 MNCT-RTCM 方法进行非作战目标的实时清洗。

4.2.1 历史数据处理

确定合适的字段相似度阈值 δ_1 和记录相似度阈值 δ_2 (本实验中取 $\delta_1=0.90$, $\delta_2=0.97$), 基于相

似重复记录检测的思想挖掘历史目标航迹的行为规律,结果如图4所示。其中,图4(a)为筛选出的所有固定目标和规律航行目标的航迹分布图,图4(b)为分离出的不规律航行目标的航迹分布图,此类目

标出现的频率较低,可能为异常航迹目标,应对其重点关注并进行针对性分析。采用循环阈值法对图4(a)数据集进行固定目标清洗,得到历史非作战目标集合,其航迹分布情况如图5所示。

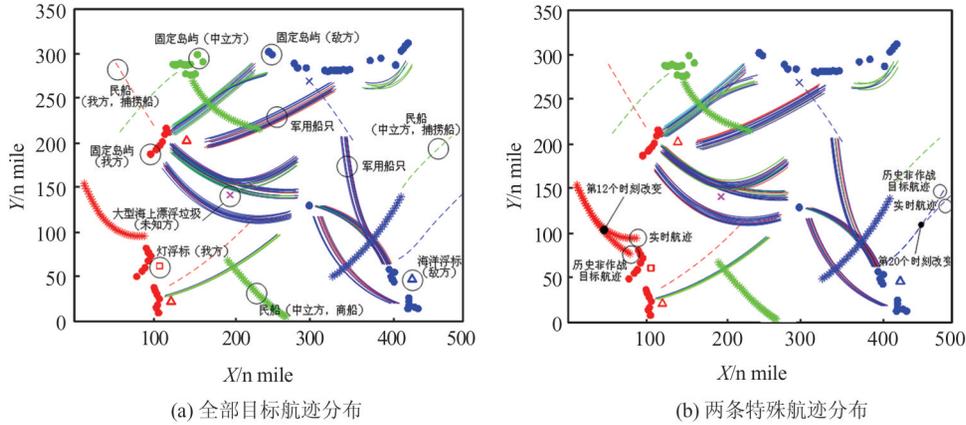


图3 实时仿真数据集

Fig.3 Real-Time Simulation Dataset

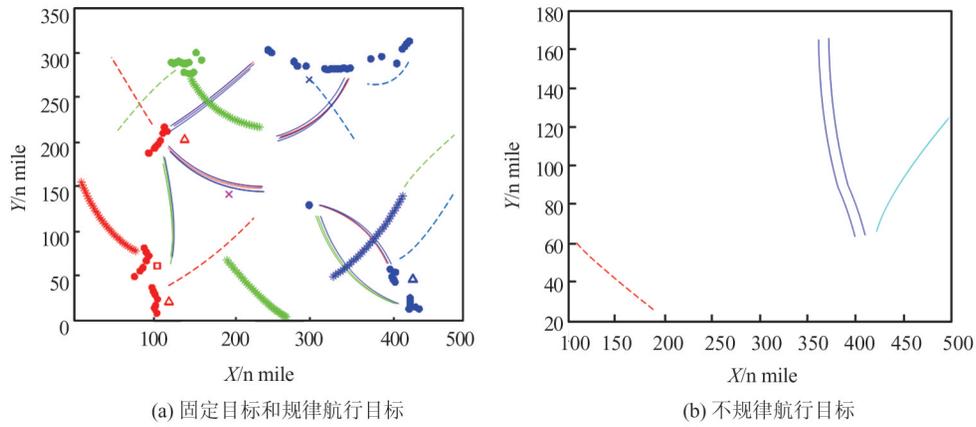


图4 历史目标航迹数据行为规律挖掘结果

Fig.4 Behavior Mining Results of Historical Trajectories

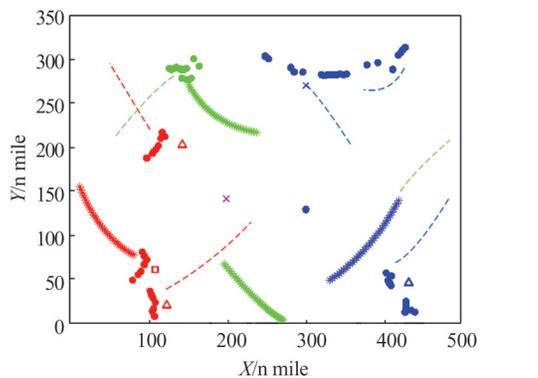


图5 历史非作战目标

Fig.5 Historical Non-combat Targets

4.2.2 非作战目标实时清洗

确定合适的字段相似度阈值 δ_1 和记录相似度阈值 δ_2 (本实验中取 $\delta_1 = 0.85, \delta_2 = 0.87$),将当前海战场环境下已更新的实时目标航迹数据

点与筛选出的历史非作战目标航迹对应时刻的数据点进行相似重复检测,若有实时目标航迹与历史非作战目标航迹是相似重复的,则该实时目标为非作战目标,予以剔除。获取部分时刻更新战场态势信息后,得到非作战目标清洗后的海上作战态势目标航迹分布,结果如图6所示。

从图6中可以看出,针对实时仿真数据集中与历史非作战目标航迹轨迹相同的目标航迹,第2次更新战场态势信息后,MNCT-RTCM方法可以实时检测出仿真数据集中的非作战目标,并对其进行清洗。针对实时仿真数据集中设置的两条与历史非作战目标航迹轨迹不同的特殊航迹:第一条特殊航迹在第12个时刻发生改变,但MNCT-RTCM方法在第14次战场态势信息更新后,才识别出该目标不是非作战目标。因为该目

标航迹开始变化时,其位置、速度和航向等特征变化极小,导致第 13 次战场态势信息更新后,该目标航迹中对应的各字段相似度并没有小于 Pair-wise 比较算法中设定的阈值 δ_1 ,且记录相似度 MDRS 的取值是所有已更新的子记录相似度的均值,故不能立刻判断出目标航迹的异常

变化,做出了错误的判断,认为该目标仍是非作战目标。第二条目标航迹在第 20 个时刻发生变化,MNCT-RTCM 方法在第 21 次战场态势信息更新后,就及时识别出该目标不是非作战目标,实现了对目标航迹异常变化的实时检测。

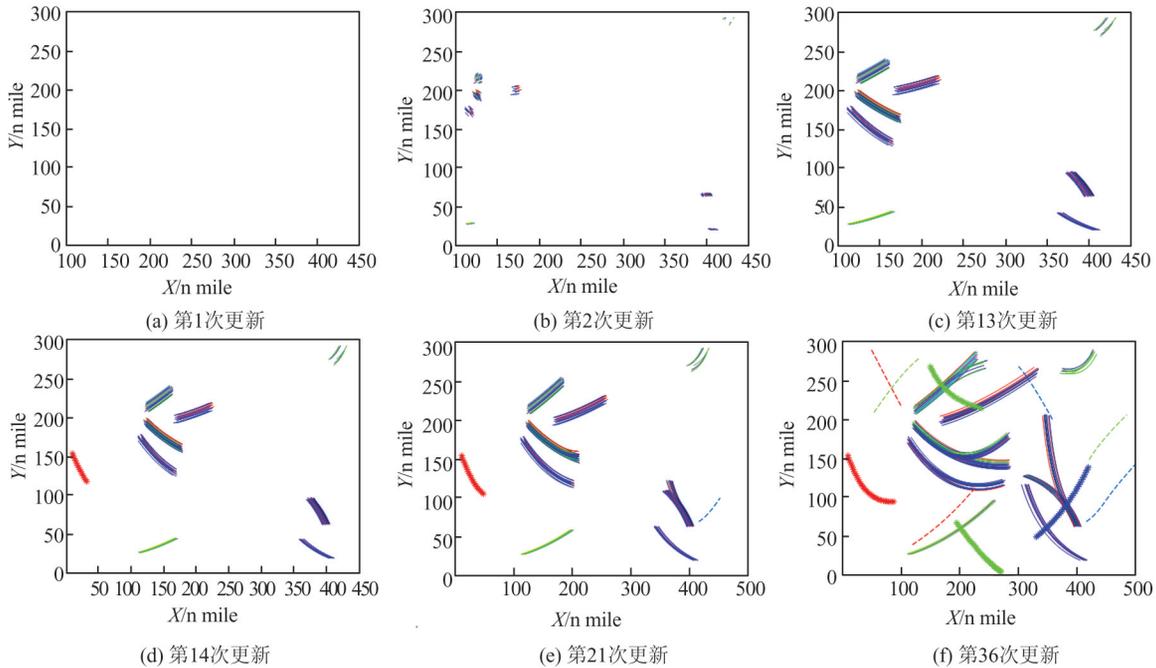


图 6 非作战目标实时清洗结果

Fig.6 Real-Time Cleaning Results of Non-combat Targets

4.2.3 评价与分析

通过计算 MNCT-RTCM 方法的召回率和准确率^[18]对非作战目标实时清洗的质量进行评估。召回率是指被正确识别的非作战目标数量与实际非作战目标数量的百分比,而准确率是指被正确识别的非作战目标数量与识别出的非作战目标数量的百分比,根据上述定义绘制出召回率和准确率随战场态势更新次数变化的曲线图,如图 7 所示。

从图 7 中可以看出,在仿真军用场景中,随着战场态势信息的不断更新,针对实时仿真数据集中与历史非作战目标航迹轨迹相同的目标航迹,MNCT-RTCM 方法可以实时检测出全部非作战目标并进行清洗。针对实时仿真数据集中与历史非作战目标航迹轨迹不同的目标航迹,若目标的各属性特征变化正常,MNCT-RTCM 方法可以实时检测出目标的异常变化,判定该目标不是非作战目标;若目标的各属性特征变化极小,可能会导致 MNCT-RTCM 方法不能立刻检测出航迹的异常变化,仍认为该目标是非作战目标,但在态势更新一段时间后,该方法将识别出该目标

航迹的异常变化,判定该目标不是要处理的非作战目标,对其进行保留。

由于需要处理的数据集较大,且记录相似性匹配的复杂度为 $B(B-1)/2$ (N 为数据集中的记录条数),因此 MNCT-RTCM 方法的计算过程较为复杂,需要耗费较长的时间。

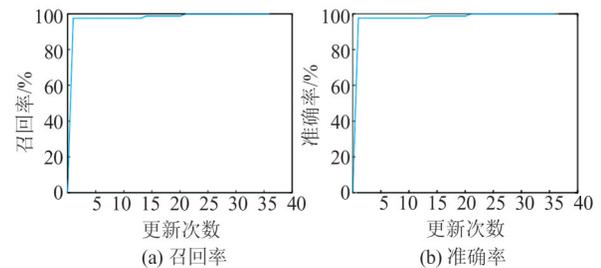


图 7 MNCT-RTCM 方法的召回率与准确率

Fig.7 Recall and Precision of the MNCT-RTCM Method

5 结 语

本文针对海上作战态势生成需求,将数据预处理技术中的重复数据清洗与目标行为规律挖掘相结合,定义了多维度记录匹配相似度,设计

了一种海上非作战目标实时清洗方法,并在仿真军事场景上进行了实验验证,通过计算召回率和准确率对结果进行评估和分析。结果表明, MNCT-RTCM方法能够有效挖掘历史目标航迹的行为规律,得到历史非作战目标集合;随着战场态势的更新,检测出目标海域内的非作战目标,实现海上作战环境下非作战目标的实时清洗,提高海上态势生成的效率和准确性。

参 考 文 献

- [1] Yang Hang. Research on Prediction of Trajectories of Moving Objects Based on Historical Information [D]. Guilin: Guangxi Normal University, 2016 (杨航. 基于历史信息的移动对象轨迹预测研究[D]. 桂林: 广西师范大学, 2016)
- [2] Liu Xianpeng, Zhang Lihua, Wang Tao, et al. TERPM Locating Algorithm Based on Trajectory Tracking[J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(2): 219-226 (刘现鹏, 张立华, 王涛, 等. 基于航迹线追踪的TERPM定位算法[J]. 武汉大学学报·信息科学版, 2018, 43(2): 219-226)
- [3] Pan Xinlong, He You, Wang Haipeng, et al. Mining Regular Behaviors Based on Multidimensional Trajectories [J]. *Expert Systems with Applications*, 2016, 66(12): 106-113
- [4] Pan Xinlong, Wang Haipeng, He You, et al. Online Classification of Frequent Behaviors Based on Multidimensional Trajectories[J]. *IET Radar, Sonar & Navigation*, 2017, 11(7): 1 147-1 154
- [5] Han Jingyu, Xu Lizhen, Dong Yisheng. An Overview of Data Quality Research [J]. *Computer Science*, 2008, 35(2): 1-5 (韩京宇, 徐立臻, 董逸生. 数据质量研究综述[J]. 计算机科学, 2008, 35(2): 1-5)
- [6] Guo Zhimao, Zhou Aoying. Research on Data Quality and Data Cleaning: A Survey [J]. *Journal of Software*, 2002, 13(11): 2 076-2 082 (郭志懋, 周傲英. 数据质量和数据清洗研究综述[J]. 软件学报, 2002, 13(11): 2 076-2 082)
- [7] Liu Yu, Zhan Zhaohui, Zhu Di, et al. Incorporating Multi-Source Big Geo-Data to Sense Spatial Heterogeneity Patterns in an Urban Space [J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(3): 327-335 (刘瑜, 詹朝晖, 朱迪, 等. 集成多源地理大数据感知城市空间分异格局[J]. 武汉大学学报·信息科学版, 2018, 43(3): 327-335)
- [8] Lin Hui, You Lan, Hu Chuanbo, et al. Prospect of Geo-Knowledge Engineering in the Era of Spatio-Temporal Big Data [J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(12): 2 205-2 211 (林珲, 游兰, 胡传博, 等. 时空大数据时代的地理知识工程展望[J]. 武汉大学学报·信息科学版, 2018, 43(12): 2 205-2 211)
- [9] Pan Xinlong, Wang Haipeng, He You, et al. Anomalous Behavior Detection Method Based on Multi-dimensional Trajectory Characteristics [J]. *Acta Aeronautica et Astronautica Sinica*, 2017, 38(4): 249-258 (潘新龙, 王海鹏, 何友, 等. 基于多维航迹特征的异常行为检测方法[J]. 航空学报, 2017, 38(4): 249-258)
- [10] Yang A, Tan X, Baek J, et al. A New ADS-B Authentication Framework Based on Efficient Hierarchical Identity-Based Signature with Batch Verification [J]. *IEEE Transactions on Services Computing*, 2015, 10(2): 165-175
- [11] Bilenko M, Mooney R, Cohen W, et al. Adaptive Name Matching in Information Integration [J]. *IEEE Intelligent Systems*, 2003, 18(5): 16-23
- [12] Monge A E, Elkan C P. The Field Matching Problem: Algorithms and Applications [C]. Proceedings of the 2nd Conference on Knowledge Discovery and Data Mining, Portland, Oregon, 1996
- [13] Yang Qiaqiao. Research on Technologies of Duplicate Record Data Cleaning Under Industrial Big Data [D]. Qingdao: Qingdao University, 2018 (杨巧巧. 工业大数据相似重复记录数据清洗关键技术研究[D]. 青岛: 青岛大学, 2018)
- [14] Wagner R A, Fischer M J. The String-to-String Correction Problem [J]. *Journal of Association for Computing Machine*, 1974, 21(1): 168-173
- [15] Lowrance R, Wagner R A. An Extension of the String-to-String Correction Problem [J]. *Journal of Association for Computing Machine*, 1975, 22(2): 177-183
- [16] Sellers P H. On the Theory and Computation of Evolutionary Distances [J]. *SIAM Journal on Applied Mathematics*, 1974, 26(4): 787-793
- [17] Hernandez M A, Stolfo S J. Real-World Data is Dirty: Data Cleansing and the Merge/Purge Problem [J]. *Data Mining and Knowledge Discovery*, 1998, 2(1): 9-37
- [18] Liu Hua. Research on Data Cleaning in Web Information Integration [D]. Wuhan: Wuhan University of Technology, 2007 (刘华. Web信息集成中数据清洗的研究[D]. 武汉: 武汉理工大学, 2007)

A Real-Time Cleaning Method for Marine Non-Combat Targets

LIN Xueyuan¹ LI Xueting^{2,3} PAN Xinlong³ LI Minbo⁴ CHEN Xiangguang¹

1 Department of Electrical and Electronic Engineering, Yantai Nanshan University, Yantai 265713, China

2 91827 Troops, Weihai 264200, China

3 Institute of Information Fusion, Naval Aviation University, Yantai 264001, China

4 Software School, Fudan University, Shanghai 200433, China

Abstract: Objectives: It is usually necessary to eliminate non-combat targets for conducting marine combat situation analysis. Because non-combat targets have little impacts on situational analysis and the retained combat-correlated targets have high reference value. Most of the existing behavior mining methods, which are based on the idea of clustering, are complicated and ineffective in cleaning of non-combat targets. **Methods:** Therefore, this paper defines multi-dimension record similarity (MDRS) and proposes a real-time cleaning method for marine non-combat targets (MNCT-RTCM). The proposed method realizes the real-time cleaning of non-combat targets by similar duplicate record detection of multi-dimensional track data. **Results:** The experiments are carried out on simulated military scenarios, and the results are evaluated and analyzed by calculating the recall rate and the precision rate. **Conclusions:** The results show that the MNCT-RTCM method can effectively detect non-combat targets and achieve real-time cleaning of non-combat targets in the marine combat environment.

Key words: situation analysis; non-combat targets; real-time cleaning; similar duplicate record

First author: LIN Xueyuan, PhD, professor, specializes in information fusion and integrated navigation. E-mail: linxy_ytcn@126.com

Corresponding author: LI Xueting, master, assistant engineer. E-mail: 574862300@qq.com

Foundation support: The National Natural Science Foundation of China (91538201, 61671157); Double Hundred Plan Talent Project of Yantai City(YT201803).

引文格式: LIN Xueyuan, LI Xueting, PAN Xinlong, et al. A Real-Time Cleaning Method for Marine Non-Combat Targets[J]. Geomatics and Information Science of Wuhan University, 2021, 46(9):1378-1385. DOI:10.13203/j.whugis20190274(林雪原,李雪腾,潘新龙,等.一种海上非作战目标实时清洗方法[J].武汉大学学报·信息科学版,2021,46(9):1378-1385. DOI:10.13203/j.whugis20190274)

(上接第 1369 页)

ment, which is used for generating traversability labels in the optimization method. Based on the terrain information and labels, we train a U-Net-like network for predicting labels according to the given multi-dimensional information, and the network performs well on test dataset with the accuracy of 93.8% on average. **Conclusions:** The proposed data-driven method for traversability analysis is effective in virtual environment. **Key words:** extraterrestrial terrain; multi-source map merging; traversability analysis; procedural content generation

First author: ZHANG Zhiyu, postgraduate, specializes in computer aided design and computer graphics. E-mail: 541600517@qq.com

Corresponding author: TANG Min, PhD, professor. E-mail: tang_m@zju.edu.cn

Foundation support: The National Key Research and Development Program of China (2018AAA0102703); the National Natural Science Foundation of China (61972341, 61832016, 51775496, 61732015).

引文格式: ZHANG Zhiyu, ZHU Chang'an, TANG Min, et al. A Data-driven Method for Traversability Analysis and Dataset Generation on Extraterrestrial Terrain[J]. Geomatics and Information Science of Wuhan University, 2021, 46(9): 1274-1282. DOI: 10.13203/j.whugis20210308(张智宇,朱昶安,唐敏,等.数据驱动的地外星表通行性分析及数据集生成方法[J].武汉大学学报·信息科学版,2021,46(9):1274-1282. DOI:10.13203/j.whugis20210308)