

基于多维航迹特征的异常行为检测方法

潘新龙*, 王海鹏*, 何友, 熊伟, 周伟

海军航空工程学院 信息融合研究所, 烟台 264001

摘要: 在信息融合领域, 利用数据挖掘中的异常检测技术, 可以基于目标的多维航迹特征来挖掘目标的异常行为。现有轨迹异常检测方法主要检测目标的位置异常, 没有充分利用目标的属性、类型、位置、速度和航向等多维特征, 在挖掘目标的异常行为时具有局限性。通过定义多因素定向 Hausdorff 距离和构造多维度局部异常因子, 提出了一种基于多维航迹特征的异常行为检测方法, 通过对多维航迹数据的异常检测, 实现对目标异常行为的挖掘。在仿真军事场景和真实的民用场景上进行了实验分析, 所提方法都能有效的检测出目标的异常行为。

关键词: 异常行为; 航迹; 多维特征; 局部异常因子; Hausdorff 距离

中图分类号: V355.1; TP274.2

文献标识码: A

文章编号: 1000-6893(2017)04-320442-10

随着目标探测手段的不断丰富和多传感器信息融合技术^[1]的不断完善, 各种各样的目标被检测、跟踪和识别, 乃至形成连续稳定的目标航迹。海量的目标航迹数据在各种目标情报处理系统、预警监视系统、航行管制系统和视频监控系统中存储和积累, 这些航迹数据中蕴含着大量的信息和知识。作为数据库知识发现中重要步骤的数据挖掘技术^[2-3]依靠跨学科的特点已经在各个领域得到了广泛应用, 轨迹数据挖掘技术^[4]也成为数据挖掘领域的一个热点。在信息融合领域, 利用数据挖掘中的异常检测技术, 可以基于目标的多维航迹特征来挖掘目标的异常行为, 实现对异常情报的智能处理, 这对于态势评估、威胁估计和指挥决策都具有非常重要的意义。

目标航迹数据通常是由多维数据点组成的多维序列, 通过对多维航迹的异常检测, 可以挖掘目标的异常行为。很多学者在异常检测和轨迹异常检测方面进行了大量的研究。Chandola 等对异

常检测技术、分类、应用范围和优缺点进行了全面的概述^[5-6]。现有异常检测方法可以分为基于统计的异常检测^[7]、基于距离的异常检测^[8]、基于密度的异常检测^[9]、基于深度的异常检测^[10]和基于偏离的异常检测^[11]等。轨迹异常检测就是找出偏离其周围轨迹运动规律达到一定程度的轨迹。现有轨迹异常检测方法可以分为基于抽取轨迹全局特征的方法^[12]、基于分类器的方法^[13-14]和基于点集(轨迹片段)相似度检测的方法^[15-16]等。这些方法主要检测目标航迹的位置异常, 没有充分利用目标的属性、类型、位置、速度和航向等多维特征, 在挖掘目标的异常行为时具有局限性。

本文针对现有轨迹异常检测方法的局限性, 提出了一种基于多维航迹特征的异常行为检测方法。利用目标的属性、类型、位置、速度和航向特征, 通过对多维航迹数据的异常检测, 实现对目标异常行为的挖掘。

收稿日期: 2016-05-17; 退修日期: 2016-06-22; 录用日期: 2016-07-18; 网络出版时间: 2016-07-19 17:06

网络出版地址: www.cnki.net/kcms/detail/11.1929.V.20160719.1706.004.html

基金项目: 国家自然科学基金(61531020, 61471383, 91538201); 山东省科技重大专项基金(2015ZDZX01001)

* 通讯作者。E-mail: airadar@126.com; whp5691@163.com

引用格式: 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): 320442. PAN X L, WANG H P, HE Y, et al. Anomalous behavior detection method based on multidimensional trajectory characteristics[J]. Acta Aeronautica et Astronautica Sinica, 2017, 38(4): 320442.

1 多维航迹数据

航迹数据是由目标数据点组成的序列,根据应用场景,可将航迹数据分为预警监视航迹数据,航行管制航迹数据和视频监控航迹数据等,根据目标的类型,可将航迹数据分为飞机航迹数据、船舶航迹数据、车辆航迹数据、行人航迹数据、动物航迹数据和龙卷风航迹数据等。目标航迹数据通常是由多维数据点组成的多维序列,例如,在广播式自动相关监视系统(Automatic Dependent Surveillance Broadcast, ADS-B)^[17-18]中,航迹数据通常包括飞机编号、时间、经度、纬度、高度、速度及航向等多维特征。在预警监视情报处理系统中,航迹数据通常包括目标的批号、属性、类别、数量、型号、机/舷号、时间、经度、维度、高度、速度及航向等多维特征。

航迹可以表示成如下集合的形式:

$$\mathbf{TD} = \{\mathbf{TR}_1, \mathbf{TR}_2, \dots, \mathbf{TR}_i, \dots, \mathbf{TR}_n\} \quad (1)$$

式中: \mathbf{TD} 为航迹集合, $i \in [1, n]$ 为航迹编号, n 为航迹总条数;航迹 \mathbf{TR}_i 为由若干个多维航迹点按照时间顺序组成的一个多维航迹点序列:

$$\mathbf{TR}_i = \{\mathbf{P}_{i1}, \mathbf{P}_{i2}, \dots, \mathbf{P}_{ij}, \dots, \mathbf{P}_{im}\} \quad (2)$$

式中: \mathbf{P}_{ij} 为第*i*条航迹中的第*j*个多维航迹点, $j \in [1, m]$ 为航迹点编号, m 为航迹点总条数,对于不同的航迹 \mathbf{TR}_i ,航迹点总数*m*不一定相同。航迹点 \mathbf{P}_{ij} 为一个具有多维特征的向量,其表达式为

$$\mathbf{P}_{ij} = [\text{label} \quad \text{attribute} \quad \text{category} \quad \dots \quad \text{time} \\ \text{longitude} \quad \text{latiude} \quad \text{altitude} \quad \text{velocity} \quad \text{course}] \quad (3)$$

包含表示第*i*条航迹中的第*j*个多维航迹点的目标编号、属性、类型、时间、经度、维度、高度、速度和航向等特征。对于不同的航迹数据,航迹点 \mathbf{P}_{ij} 包含的多维特征也不一定相同。

2 多因素定向 Hausdorff 距离

在*D*维特征向量空间 $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_D\}$ 中,特征选择是指从原有的*D*维特征空间中删去一些特征描述量,从而得到精简后的特征空间。在这个特征空间中,样本由*n*维的特征向量描述: $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, $n < D$ 。由于 \mathbf{X} 只是 \mathbf{Y} 的一个子集,因此每个分量 \mathbf{x}_i 必然能在原特征集中找

到其对应的描述量 $\mathbf{y}_i = \mathbf{x}_i$ 。在目标航迹数据中,一些特征对于目标行为的异常检测是没有多大贡献的,只需要抽取那些对异常检测最为重要的特征。航迹点 \mathbf{P}_{ij} 是一个包含多维特征的向量,抽取其中的属性、类型、位置、速度和航向特征。属性和类型特征用于构造目标行为标签,位置,速度和航向特征用来构造目标的距离度量。

Hausdorff 距离是描述两组集合之间相似程度的一种度量。在几何计算和图像处理领域非常著名,广泛应用于形状匹配和图形识别方面^[19-20]。给定两条航迹 $\mathbf{TR}_A, \mathbf{TR}_B$,设 \mathbf{TR}_A 到 \mathbf{TR}_B 的定向 Hausdorff 距离为 δ_H ,则

$$\delta_H(\mathbf{TR}_A, \mathbf{TR}_B) = \max_{\mathbf{P}_a \in \mathbf{TR}_A} \left\{ \min_{\mathbf{P}_b \in \mathbf{TR}_B} \{ \text{dist}(\mathbf{P}_a, \mathbf{P}_b) \} \right\} \quad (4)$$

式中: $\text{dist}(\mathbf{P}_a, \mathbf{P}_b)$ 为两点之间只考虑位置特征的欧氏距离; $\delta_H(\mathbf{TR}_A, \mathbf{TR}_B)$ 为 \mathbf{TR}_A 中的点到 \mathbf{TR}_B 中点的最近距离中的最大距离,表示航迹 \mathbf{TR}_A 和 \mathbf{TR}_B 的相似程度。 $\delta_H(\mathbf{TR}_A, \mathbf{TR}_B)$ 并不要求 \mathbf{TR}_B 的每一部分都与 \mathbf{TR}_A 的某一部分相匹配,当 \mathbf{TR}_A 和 \mathbf{TR}_B 不完整时, $\delta_H(\mathbf{TR}_A, \mathbf{TR}_B)$ 也能很好地度量它们之间的相似程度。航迹 \mathbf{TR}_A 和 \mathbf{TR}_B 如图 1 所示。

但是两点之间的欧氏距离只考虑了固定的位置特征,没有考虑速度、航向等动态特征。本文通过构造两点之间的多因素距离 $\text{mfdist}(\mathbf{P}_a, \mathbf{P}_b)$,来定义航迹 \mathbf{TR}_A 到 \mathbf{TR}_B 的多因素定向 Hausdorff 距离 δ_M 。考虑两个矢量点之间的位置特征、速度特征和航向特征,定义两点之间的多因素距离为

$$\text{mfdist}(\mathbf{P}_a, \mathbf{P}_b) = w_d \cdot \text{dist}(\mathbf{P}_a, \mathbf{P}_b) + w_v \cdot \text{dist}(v_{\mathbf{P}_a}, v_{\mathbf{P}_b}) + w_\theta \cdot \text{dist}(\theta_{\mathbf{P}_a}, \theta_{\mathbf{P}_b}) \quad (5)$$

式中: $v_{\mathbf{P}_a}$ 与 $v_{\mathbf{P}_b}$ 为点 \mathbf{P}_a 与点 \mathbf{P}_b 的速度; $\theta_{\mathbf{P}_a}$ 与 $\theta_{\mathbf{P}_b}$ 为点 \mathbf{P}_a 与点 \mathbf{P}_b 的航向; $\text{dist}(v_{\mathbf{P}_a}, v_{\mathbf{P}_b})$ 为点 \mathbf{P}_a 与点 \mathbf{P}_b 之间速度特征的欧氏距离; $\text{dist}(\theta_{\mathbf{P}_a}, \theta_{\mathbf{P}_b})$ 为点

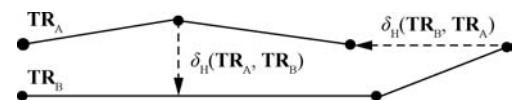


图 1 航迹间的定向 Hausdorff 距离示意图

Fig. 1 Illustration of directed Hausdorff distance between trajectories

\mathbf{P}_a 与点 \mathbf{P}_b 之间航向特征的欧氏距离; w_d 为位置特征的权重因子; w_v 为速度特征的权重因子; w_θ 为航向特征的权重因子, 权重因子的取值取决于多因素距离的应用场景, 满足 $w_d \geq 0, w_v \geq 0, w_\theta \geq 0$ 和 $w_d + w_v + w_\theta = 1$ 。

在一个应用场景中, 如果只需考虑位置异常的情况, 那么 $w_d = 1, w_v = 0, w_\theta = 0$ 即可。如果位置异常、速度异常和航向异常都需要考虑, 就需要合理的选取这 3 个权重因子的取值, 使构造的多因素距离能够均衡的体现出位置、速度和航向特征的影响。由于不同特征的量纲不同, 所以不能平均分配 3 个权重, 可以根据不同特征取值的大小差别程度按比例确定特征权重。在某一个具体的应用场景, 一旦特征权重确定, 在以后的异常检测中, 就可以采用这组权重取值。

基于定义的多因素距离 $mfdist(\mathbf{P}_a, \mathbf{P}_b)$, 可以将定向 Hausdorff 距离的维度进行扩展, 定义多维航迹 \mathbf{TR}_A 和多维航迹 \mathbf{TR}_B 之间的多因素定向 Hausdorff 距离为

$$\delta_M(\mathbf{TR}_A, \mathbf{TR}_B) = \max_{\mathbf{P}_a \in \mathbf{TR}_A} \left\{ \min_{\mathbf{P}_b \in \mathbf{TR}_B} \{mfdist(\mathbf{P}_a, \mathbf{P}_b)\} \right\} \quad (6)$$

式中: $\delta_M(\mathbf{TR}_A, \mathbf{TR}_B)$ 可以表示多维航迹 \mathbf{TR}_A 和多维航迹 \mathbf{TR}_B 的相似程度, $\delta_M(\mathbf{TR}_A, \mathbf{TR}_B)$ 不要求 \mathbf{TR}_B 的每一部分都与 \mathbf{TR}_A 的某一部分相匹配, 当 \mathbf{TR}_A 和 \mathbf{TR}_B 不完整时, $\delta_M(\mathbf{TR}_A, \mathbf{TR}_B)$ 也能很好地度量它们之间的相似程度。

3 多维度局部异常因子

局部异常因子^[21] (Local Outlier Factor, LOF) 是一个基于密度的异常度量, 在异常检测领域具有非常广泛的应用。LOF 可以很好地解决局部密度不一致的异常检测问题^[22], 比基于距离的异常检测算法更适用于航迹的异常检测。LOF 虽然克服了基于距离的方法的部分缺陷, 但是 LOF 是基于 k 近邻距离提出来的, 距离计算采用传统的欧氏距离, 因此, LOF 只能度量欧氏距离体现的位置异常情况。对于多维的目标航迹数据, 需要度量目标的位置异常、速度异常和航向异常, 此时 LOF 是不能够胜任的。因此, 基于多因素定向 Hausdorff 距离, 构造一个多维度局部异常因子 (Multi Dimensional Local Outlier Factor,

MDLOF) 作为多维航迹的异常度量, MDLOF 可以不仅可以度量目标的位置异常, 还可以同时度量目标的速度异常和航向异常。

对于每条多维航迹 $\mathbf{TR}_i \in \mathbf{TD}$, 分 3 步来构造 MDLOF。

步骤 1 确定每条航迹的近邻航迹。为了确定 \mathbf{TR}_i 的近邻航迹, 先定义它的多因素近邻边界距离为

$$\delta_{M \text{ border}}(\mathbf{TR}_i, k) = \delta_M(\mathbf{TR}_i, \text{NN}(\mathbf{TR}_i, k)) \quad (7)$$

式中: $\text{NN}(\mathbf{TR}_i, k)$ 为多维航迹数据集中与 \mathbf{TR}_i 的多因素定向 Hausdorff 距离第 k 近的航迹。那么, \mathbf{TR}_i 的近邻航迹为

$$\mathcal{N}(\mathbf{TR}_i, k) = \{\mathbf{TR}_j \in \mathbf{TD} \setminus \mathbf{TR}_i \mid \delta_M(\mathbf{TR}_i, \mathbf{TR}_j) \leq \delta_{M \text{ border}}(\mathbf{TR}_i, k)\} \quad (8)$$

步骤 2 估计每条航迹的近邻密度。为了估计 \mathbf{TR}_i 的近邻密度, 先定义航迹 \mathbf{TR}_i 到 \mathbf{TR}_j 的多因素可达距离为

$$\delta_{M \text{ reach}}(\mathbf{TR}_i, \mathbf{TR}_j, k) = \max\{\delta_{M \text{ border}}(\mathbf{TR}_j, k), \delta_M(\mathbf{TR}_j, \mathbf{TR}_i)\} \quad (9)$$

那么, \mathbf{TR}_i 的近邻密度可以定义为

$$\rho(\mathbf{TR}_i, k) = \frac{|\mathcal{N}(\mathbf{TR}_i, k)|}{\sum_{\mathbf{TR}_j \in \mathcal{N}(\mathbf{TR}_i, k)} \delta_{M \text{ reach}}(\mathbf{TR}_i, \mathbf{TR}_j, k)} \quad (10)$$

式中: 使用 $\delta_{M \text{ reach}}$ 代替 δ_M 可以对近邻密度进行平滑, k 的取值越大, 平滑的效果越明显。

步骤 3 计算每条航迹的多维度局部异常因子。通过比较 \mathbf{TR}_i 与 \mathbf{TR}_i 近邻航迹的近邻密度来计算 \mathbf{TR}_i 的多维度局部异常因子:

$$\text{MDLOF}(\mathbf{TR}_i, k) = \frac{\sum_{\mathbf{TR}_j \in \mathcal{N}(\mathbf{TR}_i, k)} \rho(\mathbf{TR}_j, k)}{\rho(\mathbf{TR}_i, k) |\mathcal{N}(\mathbf{TR}_i, k)|} \quad (11)$$

当航迹 \mathbf{TR}_i 的异常程度越大, 多维度局部异常因子的取值越大。

4 异常行为检测方法

本文基于定义的多因素定向 Hausdorff 距离和构造的多维度局部异常因子, 设计了一种基于多维航迹特征的异常行为检测 (MTC-ABD) 方法。MTC-ABD 流程为

步骤 1 输入近邻数 k , 异常阈值 ϵ 和多维航迹数据集 $\mathbf{TD} = \{\mathbf{TR}_1, \mathbf{TR}_2, \dots, \mathbf{TR}_n\}$ 。

步骤 2 利用多维航迹数据中的位置、速度

和航向特征,计算目标航迹间的多因素定向Hausdorff 距离 $\delta_M(\mathbf{TR}_i, \mathbf{TR}_j)$ 。

步骤3 利用近邻数和目标航迹间的多因素定向 Hausdorff 距离,计算每条航迹 \mathbf{TR}_i 的多因素近邻边界距离 $\delta_{M \text{ border}}(\mathbf{TR}_i, k)$ 。

步骤4 利用目标航迹间的多因素定向 Hausdorff 距离和每条航迹 \mathbf{TR}_i 的多因素近邻边界距离 $\delta_{M \text{ border}}(\mathbf{TR}_i, k)$,确定每条航迹 \mathbf{TR}_i 的近邻航迹 $\mathcal{N}(\mathbf{TR}_i, k)$ 。

步骤5 利用目标航迹间的多因素定向 Hausdorff 距离和每条航迹 \mathbf{TR}_i 的多因素近邻边界距离 $\delta_{M \text{ border}}(\mathbf{TR}_i, k)$,计算每条航迹到它的近邻航迹的多因素可达距离 $\delta_{M \text{ reach}}(\mathbf{TR}_i, \mathbf{TR}_j, k)$ 。

步骤6 利用每条航迹到它的近邻航迹的多因素可达距离 $\delta_{M \text{ reach}}(\mathbf{TR}_i, \mathbf{TR}_j, k)$,计算每条航迹的近邻密度 $\rho(\mathbf{TR}_i, k)$ 。

步骤7 利用每条航迹和它的近邻航迹的近邻密度,计算每条航迹的多维度局部异常因子 MDLOF(\mathbf{TR}_i, k)。

步骤8 利用输入的异常阈值对每条航迹进行异常检测判决。

步骤9 输出对每条航迹进行异常检测的异常指示向量 $\mathbf{Anom}_{\text{TD}}^{\epsilon} = [\mathbf{Anom}_{\mathbf{TR}_1}^{\epsilon} \quad \mathbf{Anom}_{\mathbf{TR}_2}^{\epsilon} \quad \dots \quad \mathbf{Anom}_{\mathbf{TR}_n}^{\epsilon}]$ 。

由于本文所提的方法是对静态数据的离线异常检测,所以可以通过实验验证确定近邻数 k 和异常阈值 ϵ ,在某一个具体的应用场景,一旦 k 和 ϵ 的取值确定,在以后的异常检测中,就可以采用这组取值。MTC-ABD 方法的伪代码如算法 1 所示。

算法 1:MTC-ABD

输入: 1. 多维航迹数据集 $\mathbf{TD} = \{\mathbf{TR}_1, \mathbf{TR}_2, \dots, \mathbf{TR}_n\}$ 。
2. 近邻数 k 和异常阈值 ϵ 。

输出: 异常指示向量: $\mathbf{Anom}_{\text{TD}}^{\epsilon}$

for each $\mathbf{TR}_i, \mathbf{TR}_j \in \mathbf{TD} \wedge i \neq j$ do

$$\delta_M(\mathbf{TR}_i, \mathbf{TR}_j) = \max_{P_a \in \mathbf{TR}_i} \left\{ \min_{P_b \in \mathbf{TR}_j} \{ \text{mfdist} (\mathbf{P}_a, \mathbf{P}_b) \} \right\}$$

$$\delta_{M \text{ border}}(\mathbf{TR}_i, k) = \delta_M(\mathbf{TR}_i, \mathbf{NN}(\mathbf{TR}_i, k))$$

```

 $\mathcal{N}(\mathbf{TR}_i, k) = \{ \mathbf{TR}_j \in \mathbf{TD} \setminus \mathbf{TR}_i \mid$ 
 $\delta_M(\mathbf{TR}_i, \mathbf{TR}_j) \leq \delta_{M \text{ border}}(\mathbf{TR}_i, k) \}$ 
end for
for each  $\mathbf{TR}_i$  do
    for  $\mathbf{TR}_j \in \mathcal{N}(\mathbf{TR}_i, k)$  do
         $\delta_{M \text{ reach}}(\mathbf{TR}_i, \mathbf{TR}_j, k) =$ 
         $\max\{\delta_{M \text{ border}}(\mathbf{TR}_j, k), \delta_M(\mathbf{TR}_j, \mathbf{TR}_i)\}$ 
    end for
end for
for each  $\mathbf{TR}_i$  do
     $\rho(\mathbf{TR}_i, k) = \frac{|\mathcal{N}(\mathbf{TR}_i, k)|}{\sum_{\mathbf{TR}_j \in \mathcal{N}(\mathbf{TR}_i, k)} \delta_{M \text{ reach}}(\mathbf{TR}_i, \mathbf{TR}_j, k)}$ 
end for
for each  $\mathbf{TR}_i$  do
    for  $\mathbf{TR}_j \in \mathcal{N}(\mathbf{TR}_i, k)$  do
        MDLOF( $\mathbf{TR}_i, k$ ) =
         $\frac{\sum_{\mathbf{TR}_j \in \mathcal{N}(\mathbf{TR}_i, k)} \rho(\mathbf{TR}_j, k)}{\rho(\mathbf{TR}_i, k) |\mathcal{N}(\mathbf{TR}_i, k)|}$ 
        if  $\text{MDLOF}(\mathbf{TR}_i, k) > \epsilon$  then
             $\mathbf{Anom}_{\mathbf{TR}_i}^{\epsilon} \leftarrow 1$ 
        else
             $\mathbf{Anom}_{\mathbf{TR}_i}^{\epsilon} \leftarrow 0$ 
        end if
    end for
end for

```

被检测为异常的多维航迹可以反映出目标的异常行为。把异常检测后得到的每一条多维航迹作为目标的一个异常行为,通过设置目标属性标签、类型标签和异常航迹序号来组成目标的异常行为标签,从而表示目标的异常行为。例如,在预警监视领域,将目标的属性分为我方,敌方和友方,分别用标签 1、2 和 3 表示,将目标类型分为军用飞机、民用飞机、军舰和民船等,分别用标签 1、2、3 和 4 等来表示。如果对预警监视系统获得的敌方飞机的多维航迹进行异常检测,得到了一个异常指示向量 $\mathbf{Anom}_{\text{TD}}^{\epsilon} = [1 \ 0 \ 0 \ 0 \ 1 \ \dots]$,则检测到的目标异常行为可以表示为 211,212,⋯。

5 实验分析

为了验证本文提出的基于多维特征的异常行

为检测算法的异常检测性能,5.1节在一个模拟军用场景的仿真数据集上进行异常行为检测实验,5.2节在一个民用场景的ADS-B数据集上进行异常行为检测实验,并对结果进行分析。

5.1 军用场景仿真实验分析

在这个实验中,生成一个航迹数据集来模拟敌方军用飞机在一个空域的飞行行为,然后用MTC-ABD算法对其中的异常行为进行检测,并对实验结果进行分析。

5.1.1 数据集

本实验利用Piciarelli公开的目标航迹生成程序^[23],产生包含二维位置特征的目标航迹数据,然后人为添加目标的属性特征、类型特征、速度特征和航向特征,并对其中的部分目标航迹进行修改,构造一个新的目标多维航迹数据集。设定目标的运动方向为从点 (x_i, y_i) 到点 (x_{i+1}, y_{i+1}) ,两点之间的时间间隔相等。如图2所示,目标的速度特征和航向特征可以由式(12)和式(13)计算得出。

$$v_i = \frac{\sqrt{(y_{i+1} - y_i)^2 + (x_{i+1} - x_i)^2}}{t} \quad (12)$$

$$\theta_i = \begin{cases} \arctan \frac{y_{i+1} - y_i}{x_{i+1} - x_i} & x > 0 \\ \arctan \frac{y_{i+1} - y_i}{x_{i+1} - x_i} + \pi & x < 0 \end{cases} \quad (13)$$

这个数据集包括2000条代表目标行为的多维航迹数据,每条航迹数据包括32个数据点。具体构造过程如下:首先,利用目标航迹生成程序产生10个簇的2000条对应目标正常行为的航迹数据和1000条对应目标位置异常的异常航迹数据,目标航迹生成程序中的随机参数设为0.6。将这3000条航迹的属性标签设置为2,类型标签设置为1,速度信息和航向信息由式(12)和式(13)计算得出,这样就得到了3000条目标多

维航迹数据。然后,从每个对应目标正常行为的航迹数据簇中随机取出1条目标航迹,修改这10条航迹数据,使这10条航迹对应的目标速度为正常目标速度的2倍;再从每个对应目标正常行为的航迹数据簇中随机取出1条目标航迹,修改这10条航迹数据,使这10条航迹对应的目标航向与正常目标的航向相反;再从每个对应目标正常行为的航迹数据簇中随机取出1条目标航迹,删除这10条航迹数据;从1000条对应目标位置异常的异常航迹数据中随机取出10条。最后,将10个簇中剩余的1970条对应目标正常行为的航迹数据、10条对应目标位置异常的航迹数据、10条对应目标速度异常的航迹数据和10条对应目标航向异常的航迹数据,共2000条航迹数据随机排列顺序,形成这个实验用的数据集,这个数据集中包括1970条对应目标正常行为的航迹数据和30条对应目标异常行为的航迹数据。图3为这个数据集中的2000条航迹。1970条细航迹对应目标的正常行为模式,30条粗航迹对应目标的异常行为模式,其中10条粗连线画出航迹对应位置异常,10条粗点划线画出的航迹对应速度异常,10条粗点线画出的航迹代表航向异常。这30条异常航迹对应的异常行为可以依次表示为2101,2102,2103,...,2129,2130。

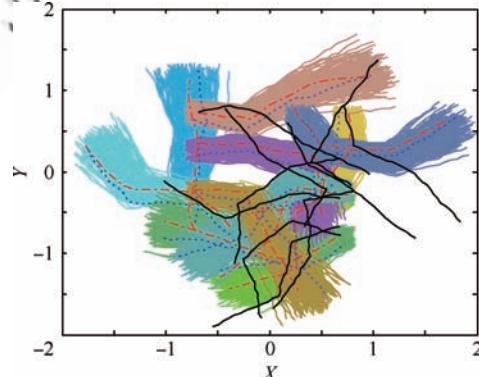


图3 实验数据集中的2000条航迹示意图

Fig. 3 Plot of 2000 trajectories in experiment dataset

5.1.2 结果分析

利用MTC-ABD算法对图3所示的航迹数据集中的多维数据对应的目标行为进行异常检测。

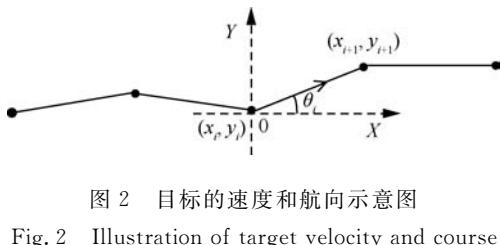


图2 目标的速度和航向示意图

Fig. 2 Illustration of target velocity and course

本实验通过计算异常检测的敏感度(Sensitivity)、准确度(Precision)、虚警率(False alarm rate)和异常检测得分(F_1 -score)来评价MTC-ABD算法的异常检测性能^[24]。敏感度为对应目标异常行为的航迹被检测为异常的比率,准确度为被检测为异常的航迹中真正对应目标异常行为的航迹所占的比率,虚警率为对应目标正常行为的航迹被检测为异常的比率。 F_1 -score由Sensitivity和Precision共同决定:

$$F_1\text{-score} = \frac{2 \times \text{Sensitivity} \times \text{Precision}}{\text{Sensitivity} + \text{Precision}} \quad (14)$$

F_1 -score可以很好地评估异常检测器的分类性能。

本实验场景根据位置、速度和航向特征的取值大小差别程度,按比例确定权重因子 w_d 、 w_v 和 w_θ 的取值,使构造的多因素距离能够均衡地体现出位置、速度和航向特征的影响。

下面通过实验的方法讨论近邻数 k 和异常阈值 ϵ 的取值情况。 k 取计算 k 近邻距离时常用的2、3、4和5。因为MDLOF(\mathbf{TR}_i, k) $\in [0, 1]$,所以 $\epsilon \in [0, 1]$, ϵ 从0到1范围内均匀地取101个点进行实验。在 k 取值固定的情况下,计算 F_1 -score随 ϵ 取值的变化情况,并绘制曲线图。

图4为异常检测得分 F_1 -score随异常阈值 ϵ 取值的变化曲线,从图中可以看出,近邻数 k 取2、3、4、5时,异常检测得分都可以达到较高的水平,而异常阈值 ϵ 的取值对异常检测得分的影响较大。因此,在一个具体的应用场景中,需要在固

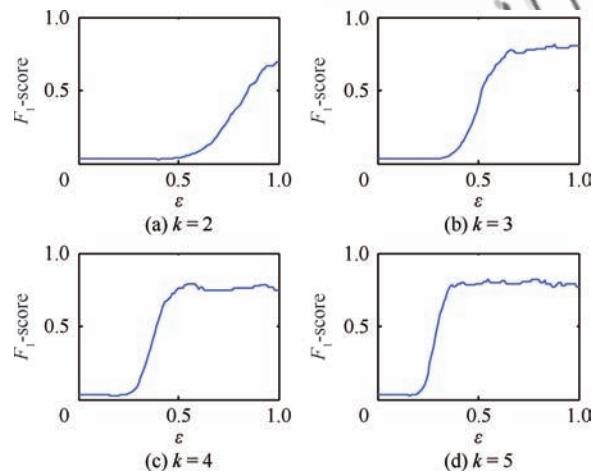


图4 异常检测得分 F_1 -score随 ϵ 的变化情况

Fig. 4 Plot of F_1 -score dependent on the value of ϵ

定 k 值的情况下,选取合适的 ϵ 取值,使异常检测得分达到预期的水平。在本实验中,取 $k=4, \epsilon=0.58$ 分析MTC-ABD算法的性能以及随着数据规模的变化情况。

在逐渐增大数据规模的情况下对该多维航迹数据集进行异常检测,并利用多项式拟合的方法绘制了敏感度、准确度、虚警率和异常检测得分随着数据规模变化的曲线图。

图5(a)为目标异常行为检测的敏感度随数据集规模变化的曲线,从图中可以看出,在数据集规模较小时,随着数据集规模的不断增大,目标异常行为检测的敏感度逐渐增加,当数据集达到一定规模后,敏感度始终保持在很高的水平。这说明在这个仿真的军用场景中,MTC-ABD算法可以很好地将目标的异常行为检测为异常。

图5(b)为目标异常行为检测的准确度随数据集规模变化的曲线,从图中可以看出,随着数据集规模的不断增大,目标异常行为检测的准确度逐渐增加和稍微回落振荡,始终保持在较高的水平。这说明在这个仿真的军用场景中,被MTC-ABD算法检测为异常的行为绝大部分是真正的目标异常行为。

图5(c)为目标异常行为检测的虚警率随数据集规模变化的曲线,从图中可以看出,在数据集规模较小时,随着数据集规模的不断增大,目标异常行为检测的虚警率逐渐减小,当数据集达到一定规模后,虚警率始终保持在非常低的水平。这说明在这个仿真的军用场景中,MTC-ABD算法将目标的正常行为检测为异常行为的概率非常低,这对于很多应用尤其是军事应用来说是非常重要的。

图5(d)为目标异常行为检测得分随数据集规模变化的曲线,从图中可以看出,在数据集规模较小时,随着数据集规模的不断增大,目标异常行为检测得分逐渐增加,当数据集达到一定规模后,异常检测得分始终保持在很高的水平。这说明在这个仿真的军用场景中,MTC-ABD算法具有很好的目标异常行为检测性能,可以非常有效地检测出敌方军用飞机飞行训练的异常行为,这对于掌握敌机的行为改变规律,进行提前预警和防范,具有非常重要的意义。

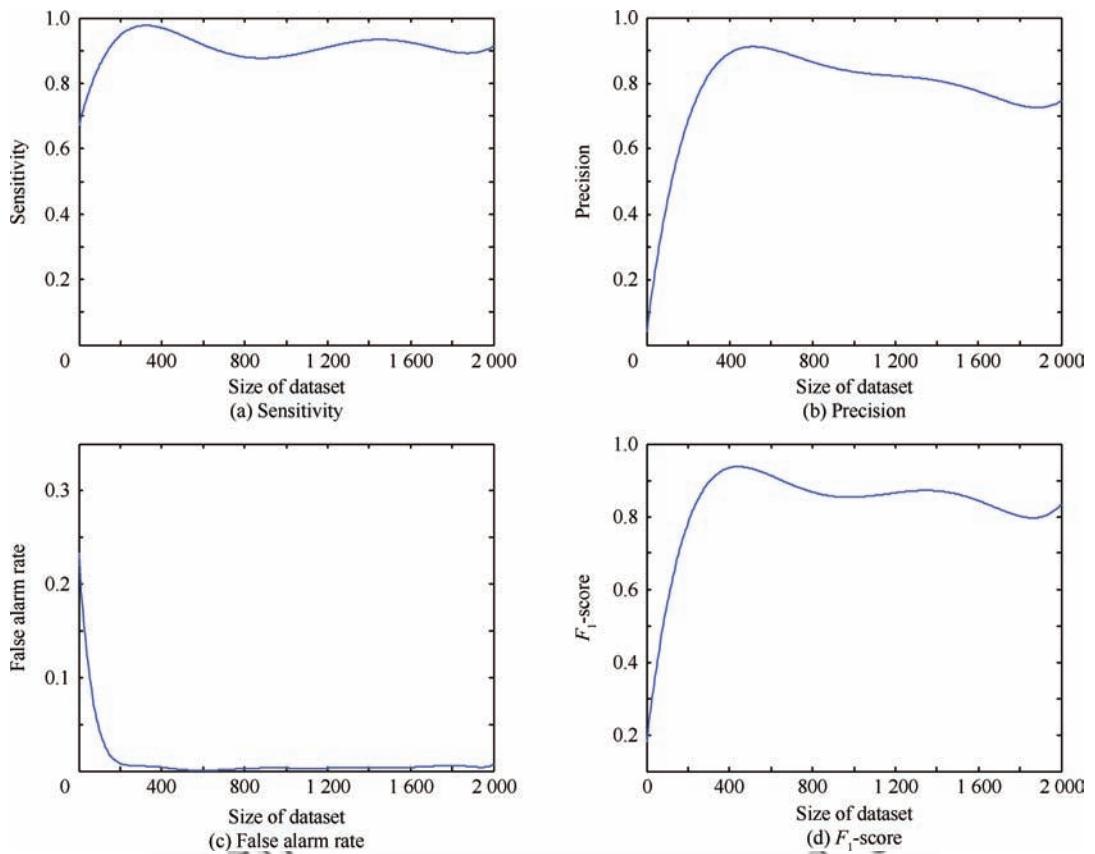


图 5 敏感度、准确度、虚警率和异常检测得分随数据集规模的变化情况

Fig. 5 Plot of sensitivity, precision, false alarm rate and F_1 -score dependent on the size of dataset

5.2 民用场景真实数据验证

在这个实验中,对一个广播式自动相关监视系统接收到的部分民用航班飞行的航迹数据进行处理,得到一个实验用的多维航迹数据集,然后用MTC-ABD算法对其中的异常行为进行检测。由于该数据集是真实的民用航班飞行航迹数据,事先并不知道其中哪些为异常航迹,所以无法进行如5.1节所示的仿真分析,但是可以将检测出的异常行为表示出来,用于查找民航飞机出现飞行异常的规律,进而查找出现飞行异常的原因,修正飞行程序等。

5.2.1 数据集

该ADS-B数据集包括2015年5月接收到的237条民航飞行航迹,每条航迹包括若干个多维数据点。可以直接读取目标的维度、经度、高度组成的位置特征,速度特征和航向特征,目标属性为

友方,用标签3表示,类型为民用飞机,用标签2表示。

为了计算目标航迹之间的多因素Hausdorff距离,将目标航迹数据点的位置特征由地理坐标中的维度、经度和高度坐标转换到局部直角坐标系中。在转换时,选择北京首都国际机场的地理坐标作为局部直角坐标系的原点,然后将这237条目标航迹的所有数据点都转换到这个局部直角坐标系中。图6为这些民用航班在三维和二维空间中的航迹。

5.2.2 实验结果

利用MTC-ABD算法对这个ADS-B航迹数据集对应的民用航班的飞行行为进行异常检测。采用5.1节所述的方法确定权重因子 w_d 、 w_v 和 w_θ 的取值,在固定 $k=4$ 的情况下,取 $\epsilon=0.4$ 进行异常检测实验。

实验结果得到了6条异常航迹,分别对应

2015年5月份该区域内民用航班飞行的6种异常行为。因为民用航班的目标属性为友方,用标签3表示,类型为民用飞机,用标签2表示,所以这6种异常行为可以表示为3201、3202、3203、3204、3205和3206。图7为异常行为示意图,箭头表示民用航班的飞行方向。

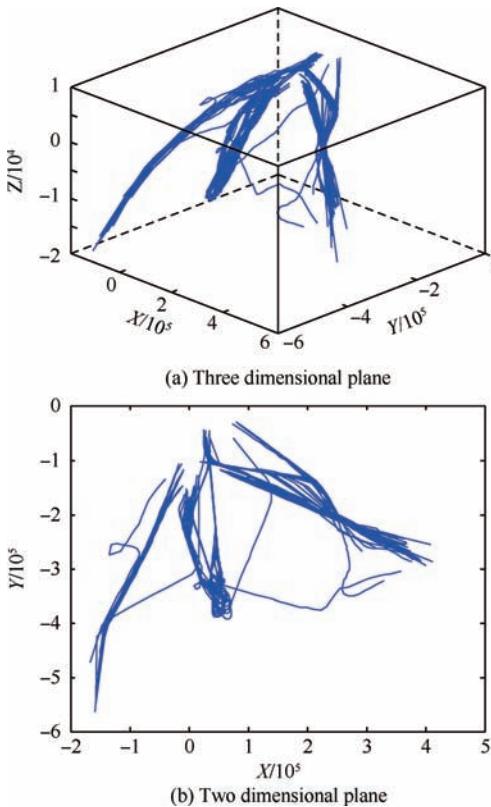


图6 民用航班在三维和二维空间中的航迹

Fig. 6 Plot of civilian flight trajectories in three and two dimensional plane

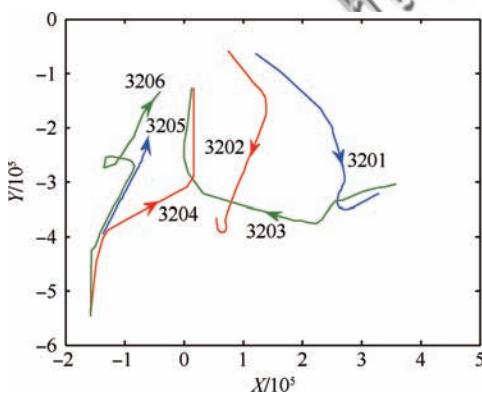


图7 民用航班异常行为示意图

Fig. 7 Plot of anomalous behaviors for civilian flights

由图7可以看出,异常行为3205可能是ADS-B在发射接收过程中造成的目标航迹接收不全。异常行为3201、3202、3203、3204和3206都是该区域内民用航班偏离固定飞行程序的异常行为,发现这些异常行为对于查找民航飞机出现飞行异常的规律,进而查找出出现飞行异常的原因,修正飞行程序等具有非常重要的参考意义。

6 结论

本文利用目标的多维航迹特征,提出了一种基于多维航迹特征的异常行为检测方法,并进行了实验分析验证。

1) 利用目标的位置特征、速度特征和航向特征,定义了多因素距离,并将定向Hausdorff距离进行扩展,定义了多因素定向Hausdorff距离,从而度量目标行为之间的相似程度。

2) 针对局部异常因子不能很好的度量目标行为之间差异程度的问题,本文利用定义的多因素定向Hausdorff距离,构建了多维度异常因子,可以有效的度量目标行为之间的差异程度。

3) 基于定义的多因素定向Hausdorff距离和构造的多维度局部异常因子,设计了一种基于多维特征的异常行为检测(MTC-ABD)算法,并给出了伪代码,用于目标异常行为的检测。

4) 在一个模拟军用场景的仿真数据集上对MTC-ABD算法进行了仿真实验,使用多个指标分析了该算法的异常检测性能。在一个真实的ADS-B民用航班数据集上利用MTC-ABD算法进行了异常行为检测实验,并利用目标的属性特征、类型特征和异常航迹序号设置异常行为标签,对目标的异常行为进行了可视化,得到了非常有意义的结果。

参 考 文 献

- [1] 何友,王国宏,关欣,等.信息融合理论及应用[M].北京:电子工业出版社,2010: 1-15.
- [2] HE Y, WANG G H, GUAN X, et al. Information fusion theory with applications[M]. Beijing: Publishing House of Electronics Industry, 2010: 1-15 (in Chinese).
- [3] CHEN M S, HAN J, YU P S. Data mining: An overview from a database perspective[J]. IEEE Transactions on Knowledge and Data Engineering, 1997, 8(6): 866-883.
- [4] HAN J, KAMBER M. Data mining: Concepts and tech-

- niques[M]. San Francisco: Morgan Kaufmann, 2001: 1-18.
- [4] ZHENG Y. Trajectory data mining: An overview[J]. ACM Transactions on Intelligent Systems and Technology, 2015, 6(3): 1-41.
- [5] CHANDOLA V, BANERJEE A, KUMAR V. Anomaly detection: A survey[J]. ACM Computing Surveys, 2009, 41(3): 1-58.
- [6] CHANDOLA V, BANERJEE A, KUMAR V. Anomaly detection for discrete sequences: A survey [J]. IEEE Transactions on Knowledge and Data Engineering, 2012, 24(99): 1-16.
- [7] BARNETT V, LEWIS T. Outliers in statistical data[M]. New York: Wiley, 1994: 24-26.
- [8] KNORR E M, NG R T, TUCAKOV V. Distance-based outliers: Algorithms and applications [J]. International Journal on Very Large Data Bases, 2000, 8(3): 237-253.
- [9] JIN W, TUNG A K H, HAN J, et al. Ranking outliers using symmetric neighborhood relationship[J]. Lecture Notes in Computer Science, 2006, 3918: 577-593.
- [10] STRUYF A, ROUSSEEUW P J. High-dimensional computation of the deepest location[J]. Computational Statistics and Data Analysis, 2000, 34: 415-426.
- [11] SARAWAGI S, AGRAWAL R, MEGIDDO N. Discovery driven exploration of OLAP data cubes [M]. Berlin: Springer, 1998: 67-69.
- [12] KNORR E M, NG R T. Algorithms for mining distance-based outliers in large datasets[C]//Proceedings of 24th International Conference on Very Large Data Bases. New York: Morgan Kaufmann, 1998: 392-403.
- [13] LI X, HAN J, KIM S. Motion-Alert: Automatic anomaly detection in massive moving objects[C]//IEEE International Conference on Intelligence and Security Informatics. Piscataway, NJ: IEEE Press, 2006: 166-177.
- [14] LI X, HAN J, KIM S, et al. ROAM: Rule-based and motif-based anomaly detection in massive moving object data sets[C]//Proceedings of 7th SIAM International Conference on Data Mining. Philadelphia: SIAM Press, 2007: 273-284.
- [15] LEE J G, HAN J, LI X. Trajectory outlier detection: A partition-and-detect framework[C]//Proceedings of 24th International Conference on Data Engineering. Washington, D. C.: IEEE Computer Society, 2008: 140-149.
- [16] 刘良旭, 乔少杰, 刘宾, 等. 基于R-Tree的高效异常轨迹检测算法[J]. 软件学报, 2009, 20(9): 2426-2435.
- [17] LIU L X, QIAO S J, LIU B, et al. Efficient trajectory outlier detection algorithm based on R-Tree[J]. Journal of Software, 2009, 20(9): 2426-2435 (in Chinese).
- [18] 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, 1939, PP(99): 1.
- [19] 沈笑云, 唐鹏, 张思远, 等. ADS-B统计数据的位置导航不确定类别质量分析[J]. 航空学报, 2015, 36(9): 3128-3136.
- [20] SHEN X Y, TANG P, ZHANG S Y, et al. Quality analysis of navigation uncertain category for position based on ADS-B statistical data[J]. Acta Aeronautica et Astronautica Sinica, 2015, 36(9): 3128-3136 (in Chinese).
- [21] HUTTENLOCHER D P, KLANDERMAN G A, RUCKLIDGE W A. Comparing images using the Hausdorff distance[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1993, 15(9): 850-863.
- [22] GAO Y, LEUNG M K H. Line segment Hausdorff distance on face matching[J]. Pattern Recognition, 2002, 35(2): 361-371.
- [23] BREUNIG M M, KRIESEL H P, NG R T, et al. LOF: Identifying density-based local outliers[C]//Proceedings of ACM SIGMOD International Conference on Management of Data. 2000, 29(2): 93-104.
- [24] JANSSENS J, POSTMA E. One-class classification with LOF and LOCI: An empirical comparison[C]//Proceedings of the 18th Belgian-Dutch Conference on Machine Learning, 2009: 1021-1040.
- [25] PICIARELLI C, MICHELONI C, FORESTI G. Trajectory-based anomalous event detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2008, 18(11): 1544-1554.
- [26] LAXHAMMAR R, FALKMAN G. Online learning and sequential anomaly detection in trajectories [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(6): 1158-1173.

(责任编辑: 苏磊)

Anomalous behavior detection method based on multidimensional trajectory characteristics

PAN Xinlong^{*}, WANG Haipeng^{*}, HE You, XIONG Wei, ZHOU Wei

Institute of Information Fusion, Naval Aeronautical and Astronautical University, Yantai 264001, China

Abstract: In the information fusion domain, anomalous behaviors could be mined based on multidimensional trajectory characteristics by using the anomalous detection technique in data mining. Previous trajectory anomaly detection algorithms mainly detect the position anomalies, without making full use of the attribute, category, position, velocity, and course characteristics. In order to overcome this limitation, we define the multi-factor Hausdorff distance, construct the multidimensional local outlier factor, and propose a method for detecting anomalous behaviors based on multidimensional trajectory characteristics. The method can mine anomalous behaviors based on detecting multidimensional trajectories. We conducted experiments on simulated military scenario and real civilian scenario, the proposed method can effectively detect the anomalous behavior of the target.

Keywords: anomalous behavior; trajectory; multidimensional characteristics; local outlier factor; Hausdorff distance

Received: 2016-05-17; **Revised:** 2016-06-22; **Accepted:** 2016-07-18; **Published online:** 2016-07-19 17:06

URL: www.cnki.net/kcms/detail/11.1929.V.20160719.1706.004.html

Foundation items: National Natural Science Foundation of China (61531020, 61471383, 91538201); Major Science and Technology Projects in Shandong Province (2015ZDZX01001)

* **Corresponding author.** E-mail: airadar@126.com; whp5691@163.com