

基于 One-Class SVM 的机载塔康测距信息异常检测方法研究

李城梁

(中国电子科技集团公司第二十研究所, 西安 710068)

摘 要: 针对多源导航信息融合系统中导航传感器数据保障的问题, 本文提出了一种基于 One-Class SVM 的机载塔康测距信息异常检测方法。首先, 提取机载塔康测距信息的时域参数构成特征样本空间; 然后, 采用 One-Class SVM 训练出机载塔康测距信息正常状态时的模型, 通过发现非正常状态的样本进行异常检测。利用模拟的机载塔康测距数据进行方法验证, 实验结果表明: 该异常检测方法对机载塔康测距信息中的噪声有一定的鲁棒性, 可以满足实际应用的需要。

关键词: 异常检测; 机载塔康测距; One-Class SVM

中图分类号: TN967

文献标识码: A

文章编号: 1674-7976-(2015)03-282-05

Anomaly Detection of Ranging Information of Airborne TACAN Based on One-Class SVM Classifier

LI Chengliang

Abstract: According to the anomaly detection for ranging information of airborne TACAN in system which fused information of multi-source navigation, this article proposed a detection method based on One-Class SVM. First, the features space is built by extracting time domain features. Then, the model of normal state is built by training the One-Class SVM classifier. Anomaly information is detected by the trained model. The experimental results show that the proposed method has good performance in anomaly detection.

Key words: Anomaly Detection; Airborne TACAN Ranging; One-Class SVM

0 引言

单一的导航系统容易受本身技术体制的限制, 例如: 惯导系统的误差随时间累积, 因而长时间工作难以保持高精度, 卫星导航系统易受干扰, 自主性差等, 难以满足在未来战争环境的要求^[1]。多源导航信息融合技术是解决以上问题的有效途径。战术空中导航系统, 简称塔康 (TACAN, Tactical Air Navigation System) 具有测位、测距及信标识别功

能, 是航空导航体制中主要的无线电导航系统, 是多源导航信息融合系统中重要的导航源之一^[2]。然而, 塔康测距的异常信息会直接影响多源导航信息融合系统的可靠性。因此, 对塔康测距信息进行异常检测, 为导航融合模式提供检测结果, 可以作为调整或重构导航融合模式的依据, 为多源导航信息融合提供数据保障。

传统传感器异常检测的大多数方法需要建立相应的异常状态模型, 并需要提供特定样本所组成的训练数据集来训练, 而且这些基于统计学的模型只是当样本数目趋向于无穷大时的理想极限特性^[3]。现实应用中其观测样本通常是有限的, 难以满足基于传统的统计方法的前提条件, 以致测试误差

较高。由于需要大量训练数据, 计算代价过大, 难以满足实时的在线检测。异常状态可以认为是在特征空间中与正常状态有偏离的现象, 因而异常检测方法本质上就是识别非正常信息的方法。直接建立系统正常工作模型, 并通过检测与正常特征的偏离程度来识别异常, 这样免去了建立故障模型的难题, 从而使问题大大简化。

本文介绍了一种基于 One-Class SVM 机载塔康测距信息异常检测方法。该方法在不使用异常状态类别的条件下, 仅需要正常状态下的时域特征信息来训练 One-Class SVM 以获得检测模型, 有效的提高了在工程应用中的可行性。首先, 第1节中介绍了机载塔康测距信息特征提取的方法。第2节中介绍了采用 One-Class SVM 分类器的理论内容。第3节介绍本文提出的机载塔康测距信息异常检测方法的检测流程。第4节为实验验证。

1 机载塔康测距信息特征提取

1.1 时序列的预处理

多源导航信息融合系统中的惯导数据平稳、连续, 可以作为其他导航传感器的异常检测基准进行特征提取。理想状态下, 机载塔康测距信息的数据更新时刻和惯导数据更新时刻保持同步。则有:

$$d_{ins}^t = \left\| \mathbf{P}_{ins}^t - \mathbf{P}_{stat} \right\|^2 \quad (1)$$

其中, d_{ins}^t 是在 t 时刻惯导的地心位置信息与塔康台站地心位置信息反算出的斜距值 \mathbf{P}_{ins}^t 是 t 时刻惯导的地心位置信息, \mathbf{P}_{stat} 是塔康台站地心位置信息。 d_{ins}^t 的每时刻的值组成检测基准的观测序 \mathbf{D}_{ins} , d_{tacan}^t 是在 t 时刻机载塔康测量的斜距值, d_{tacan}^t 的每时刻的值组合被检对象的观测序列 \mathbf{D}_{tacan} 。若连续采集 t_0 到 t_n 时刻的 INS 反算出的检测基准值和塔康测量斜距值

$$\mathbf{D}_{ins} = \{d_{ins}^{t_1}, d_{ins}^{t_2}, \dots, d_{ins}^{t_i}, \dots, d_{ins}^{t_n}\} \quad (2)$$

$$\mathbf{D}_{tacan} = \{d_{tacan}^{t_1}, d_{tacan}^{t_2}, \dots, d_{tacan}^{t_i}, \dots, d_{tacan}^{t_n}\} \quad (3)$$

用固定长度为 K 的滑动时间窗, 其中滑动时间窗中, 每新增一个最新的数据点的同时移去一个最旧的数据点。滑动时间窗分别截取 \mathbf{D}_{ins} 和 \mathbf{D}_{tacan} , 则

得到检测基准样本 \mathbf{D}_{ins}^t 和机载塔康测量斜距样本 \mathbf{D}_{tacan}^t 。

$$\mathbf{D}_{ins}^i = \{d_{ins}^{t_i}, \dots, d_{ins}^{t(i+k-1)}\} \quad (4)$$

$$\mathbf{D}_{tacan}^i = \{d_{tacan}^{t_i}, \dots, d_{tacan}^{t(i+k-1)}\} \quad (5)$$

1.2 用于异常检测的特征量提取

对 \mathbf{D}_{ins}^t 和 \mathbf{D}_{tacan}^t 分别进行归一化处理, 然后提取时域特征量^[4], 均值、方差、波形指标、峰值指标等。 \mathbf{D}_{ins}^t 特征提取后构成检测基准样本的特征向量 $\mathbf{X}_{ins}^i = (x_{ins}^{i_1}, \dots, x_{ins}^{i_m})$, \mathbf{D}_{tacan}^t 特征提取后构成被测样本的特征向量 $\mathbf{X}_{tacan}^i = (x_{tacan}^{i_1}, \dots, x_{tacan}^{i_m})$, 其中 m 为特征维数。

由 \mathbf{X}_{ins}^i 和 \mathbf{X}_{tacan}^i 可以得到最终用于异常检测的 m 维特征向量 \mathbf{X}_i 为:

$$\mathbf{X}_i = |\mathbf{X}_{ins}^i - \mathbf{X}_{tacan}^i| \quad (6)$$

2 One-Class SVM 分类算法

Scholkopf 提出的单类支持向量机是支持向量机的一种扩展^[5], 它利用样本集 $\{x_i, i=1, 2, \dots, M\}$ $x_i \in R^n$, 通过非线性核函数映射 K , 在高维特征空间 H 中计算一个包含尽可能多样本的最小超球作为决策边界。这样的 SVM 是一个能描述样本密度分布的二值模型。设

$$f(x) = \begin{cases} +1 & \text{if } x \in S \\ -1 & \text{if } x \in S' \end{cases} \quad (7)$$

其中 S 是样本集合的某个子集。

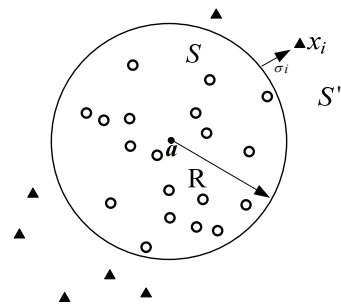


图1 One-Class SVM 分类器原理示意图

如图 1 所示。其中 a 为超球体的圆心, R 为超球体半径, $\sigma_i > 0$ 对应于超球体外面的点。

包含所有训练样本的最小超球体的求解可以转化为以下凸优化问题:

$$\min L(R) = R^2 + C \sum_{i=1}^N \sigma_i \quad (8)$$

$$s.t. (x_i - a)(x_i - a)^T \leq R^2 + \sigma_i \\ \sigma_i \geq 0, i=1, \dots, N$$

定义 Lagrange 函数:

$$L(R, a, \lambda, \sigma) = R^2 + C \sum_{i=1}^N \sigma_i - \sum_{i=1}^N \alpha_i [R^2 + \sigma_i - (x_i^2 - 2ax_i + a^2)] - \sum_{i=1}^N \gamma_i \sigma_i \quad (9)$$

其中, C 为惩罚系数, $\lambda = \{\alpha_i\}$, 对应的 Lagrange 系数 $\alpha_i \geq 0, \gamma_i \geq 0 (i=1, 2, \dots, N)$ 。将上式子分别对 R 和 α 求偏微分, 并令其等于 0, 得到相关的优化方程:

$$\max L = \sum_{i=1}^N \alpha_i (x_i, x_i) - \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j (x_i, x_j) \quad (10) \\ s.t. \sum_{i=1}^N \alpha_i = 1 \quad 0 \leq \alpha_i \leq C$$

引入高斯径向基核函数 $(x, y) \rightarrow K(x, y)$ 即:

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{\delta^2}\right) \quad (11)$$

得到对应的优化方程:

$$\max L = \sum_{i=1}^N \alpha_i K(x_i, x_i) - \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j K(x_i, x_j) \quad (12) \\ s.t. \sum_{i=1}^N \alpha_i = 1 \quad 0 \leq \alpha_i \leq C$$

根据 KKT 条件, 大部分 α_i 为 0, 只有一小部分 $\alpha_i > 0$, 与那些不为零的 α_i 对应的样本点决定了超球体边界的构成, 为此, 将这些样本点称为支持对象。对于待定的状态数据 z , 它到球心的距离的平方为:

$$f(z) = K(z, z) - 2 \sum_{i=1}^N \alpha_i K(z, x_i) \quad (13)$$

$$\sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j K(x_i, x_j)$$

设任意一个支持对象 x_s , 则球体半径的平方为:

$$R^2 = K(x_s, x_s) - 2 \sum_{i=1}^N \alpha_i K(x_s, x_i) \quad (14) \\ \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j K(x_i, x_j)$$

依据式 (15) 可以判断 z 是否为正类样本:

$$\left. \begin{aligned} f(z) &\leq R^2 & z \text{ 为正类} \\ f(z) &> R^2 & z \text{ 为负类} \end{aligned} \right\} \quad (15)$$

3 基于 One-Class SVM 的机载塔康测距信息异常检测方法

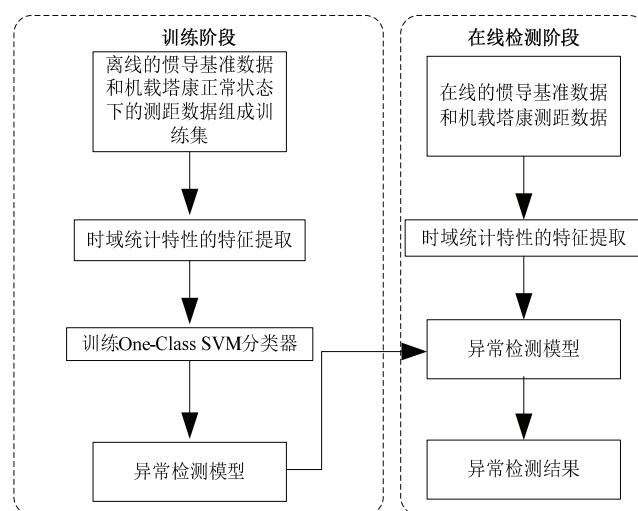


图 2 One-Class SVM 的机载塔康测距信息异常检测方法流程图

如图 2 所示, 基于 One-Class SVM 的机载塔康测距信息异常检测算法分为两个部分: 训练阶段和在线检测阶段。

训练阶段:

(1) 首先, 将离线的惯导基准数据和机载塔康正常状态下的测距数据组成训练集, 按照 1.1 节介绍的方式将这两组是序列进行预处理。

(2) 用 1.2 节介绍的方法构建训练样本的特征空间。

(3) 用(2)中构建好的训练样本训练 One-Class SVM 分类器得到异常检测模型。

在线检测阶段:

(1) 将在线的惯导基准数据和机载塔康测距数据按照 1.1 和 1.2 节中介绍的方法构建样本的特征空间。

(2) 将在线得到的样本输入到已经训练好的异常检测模型中, 最终得到异常检测结果。

4 实验验证

相互匹配的惯导数据和机载塔康正常状态下的测距数据从实验室模拟环境中得到, 惯导位置精度 0.8n mile/h, 速度精度 0.8m/s; 其中, 塔康测距精度 300m, 采样频率均为 25Hz。训练阶段数据是以往飞行正常数据的综合。训练集的数据为模拟 10 次飞行, 每次飞行时间为 20 分钟, 滑动时间窗长度为 10, 每次移动 1 个采样节拍, 共计 29910 个训练样本。在测试样本中, 另外模拟 10 次飞行, 每次飞行时间为 10 分钟, 滑动时间窗长度为 10, 每次移动 1 个采样节拍。为验证噪声对检测方法的影响, 试验中的塔康测距数据, 分别加入均值为 0, 方差为 5, 10, 20 的高斯白噪声, 如图 3~图 5 所示。通过特征选择将均值、方差、波形指标作为检测的特征量。

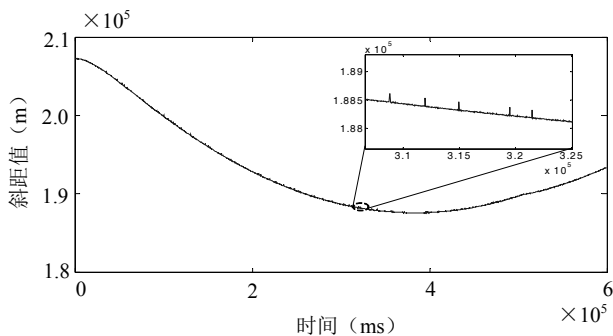


图 3 噪声水平为 $N(0, 5)$ 时的机载塔康数据

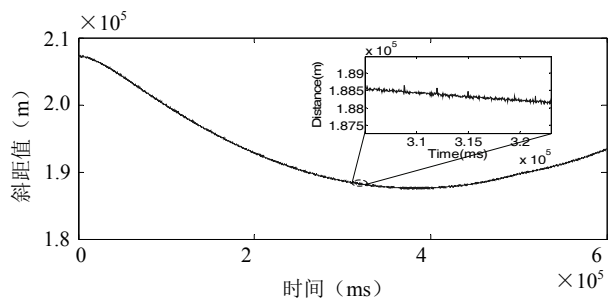


图 4 噪声水平为 $N(0, 10)$ 时的机载塔康数据

异常检测性能通过测试样本的检测概率进行评判, 评判准则为: 根据塔康测距精度, 每组测试样本集中, 在正常的机载塔康测距数据中随机添加 200 个跳变异常样本 (在正常的机载塔康测距数据基础上添加大于 150m 的跳变幅值), 每一次飞行的检测概率记为 p_i , $p_i = n_i / N_i$, n_i 为检测结果为该次飞行检测结果为正常的测试样本个数, N_i 为该次飞行测试样本的总个数, $i=1, 2, \dots, 10$ 。检测概率 $P = \text{average}(p_i)$ 。

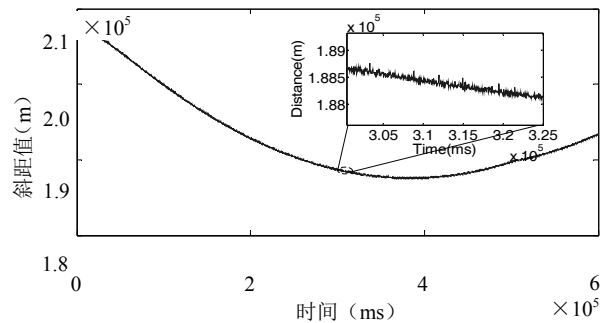


图 5 噪声水平为 $N(0, 20)$ 时的机载塔康数据

三种高斯白噪声环境下的异常检测概率结果如图 6 所示。

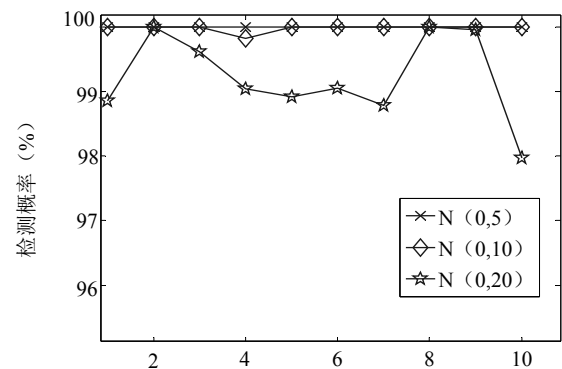


图 6 三种白噪声环境下的异常检测概率结果

从图 6 中可以看出, $N(0, 5)$ 情况下的检测概率为 100%, 在 $N(0, 10)$ 情况下为 99.99%, $N(0, 20)$ 在第 10 次试验中的检测概率最低, 为 97.97%, 但 10 次试验平均后的结果仍大于 99%。

实验结果表明, 噪声水平在一定范围内变化时对异常检测性能有微弱的影响, 但是可以满足该机载塔康测距信息异常检测方法能够满足实际应用。

(下转第 309 页)

目标特征信号控制技术,1995: 24-36.

[4] 王昭群, 朱长江, 陈作如等. 毫米波无源干扰技术进展及途径分析[J]. 南京理工大学学报, 1998: 22-25.

[5] 张德文. 反雷达的电子隐身技术[J]. 航天电子对抗, 1996: 30-39.

[6] 曲长文. 反雷达隐身和雷达反隐身技术[J]. 电子工程,

1997: 6-11.

[7] 黄培康, 殷红成, 许小剑. 雷达目标特性[M]. 北京: 电子工业出版社, 2005: 265-269.

[8] 丁建江. 防空雷达目标识别技术[M]. 北京: 国防工业出版社, 2008.

(上接第 285 页)

5 总结

本文提出在多源导航信息融合系统的应用中, 用 One-Class SVM 实现对机载塔康测距信息进行异常检测的方法, 它借助于惯导基准数和机载塔康测距数据构建时域特征量, 通过 One-Class SVM 获取机载塔康测距信息在正常状态下的模型, 从而实现在线异常检测。仿真实验表明了此方法的可行性和有效性。

从实验结果中还可以看到, 噪声水平对异常检测存在影响。如何进一步降低强噪声背景环境对异常检测性能的影响是在实际应用中需要解决的问题, 也是今后研究的重点。

参考文献:

- [1] 何友等. 多传感器信息融合及应用[M]. 北京: 电子工业出版社, 2000: 129 -138.
- [2] 张忠兴, 李晓明, 张景伟等. 无线电导航理论与系统[M]. 西安: 陕西科学技术出版社, 1998: 1-8.
- [3] Funt B V, Finlayson G D. Color Constant Color Indexing [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995, 17(5): 522-529.
- [4] 钟秉林, 黄仁. 机械故障诊断学[M]. 北京: 机械工业出版社, 1997.
- [5] Scholkopf B, Platt J C, Shawe-Taylor J, Smola A J, Williamson R C. Estimating the support of a high-dimension distribution[R]. Technical Report MSR-TR-99-87, Microsoft Research, 1999.