



基于动态支持向量提取的复杂装备实时安全评估

焦博韬¹, 郭一楠^{2*}, 何潇³, 蒲佳洋², 巩敦卫⁴

1. 中国石油大学 (华东) 计算机科学与技术学院, 青岛 266580

2. 中国矿业大学 (北京) 机械与电气工程学院, 北京 100083

3. 清华大学自动化系, 北京 100084

4. 青岛科技大学自动化与电子工程学院, 青岛 266061

* 通信作者. E-mail: guoyinan@cumt.edu.cn

收稿日期: 2024-07-15; 修回日期: 2024-09-17; 接受日期: 2024-11-08; 网络出版日期: 2025-01-07

国家重点研发计划 (批准号: 2022YFB4703700)、国家自然科学基金 (批准号: U23A20340, 61973305)、教育部系统控制与信息处理重点实验室基金 (批准号: Scip202203)、深部煤矿开采响应与灾害防控国家重点实验室开放基金 (批准号: SKLMRDPC22KF21) 和中央高校自主创新科研计划 (批准号: 24CX06071A) 资助项目

摘要 在非平稳环境下, 对复杂装备进行实时安全评估有助于及时发现潜在危险. 其中, 在线监测数据的动态变化是最主要的挑战之一. 尽管一些在线安全评估方法可以使用监测数据进行动态的调整, 但它们通常假设监测数据标签是可用的. 然而, 在大部分现实场景中, 仅有部分提前收集的初始数据或历史数据是有标记的, 而实时监测数据则通常是未标记的, 这引发了被称为极端验证延迟的挑战. 为解决该问题, 本文提出了一种基于动态支持向量提取的实时安全评估方法. 首先, 基于支持向量域描述, 在核空间构建包含不同类监测数据的超球面, 捕获数据的几何结构; 其次, 通过对超球的更新, 在未标记的监测数据中传播标签信息, 从而适应监测数据的分布变化; 进一步, 采用支持向量动态加权策略, 消除监测数据类别不平衡对模型性能的影响. 面向蛟龙号深海载人潜水器监测数据的实验结果验证了所提方法的有效性.

关键词 实时安全评估, 非平稳环境, 极端验证延迟, 类别不平衡, 数据流

1 引言

为适应资源勘探、航空航天、汽车制造等领域的特殊需求, 具有繁杂组成部件和结构的复杂装备逐渐涌现^[1]. 组成部件的故障或失效不仅会导致装备运行受限, 而且直接影响生产过程的稳定性和效率, 造成经济损失和人员伤亡^[2]. 为准确地识别复杂装备的运行状态, 通常在关键组成部件上配备多种传感器, 传感器采集的监测信息形成多维数据流, 用于实现对复杂装备的故障识别和隐患排查, 也称为复杂装备的安全评估^[3,4].

目前, 复杂装备的安全评估主要分为离线和在线两类评估方式^[5]. 离线安全评估方法是根据装备的历史运行记录, 结合专家经验, 实现故障建模和识别^[6,7]. 该类方法通常适用于装备设计或维护阶

引用格式: 焦博韬, 郭一楠, 何潇, 等. 基于动态支持向量提取的复杂装备实时安全评估. 中国科学: 信息科学, 2025, 55: 156–171, doi: 10.1360/SSI-2024-0212

Jiao B T, Guo Y N, He X, et al. Real-time safety assessment for complex equipment based on dynamic support vector extraction. Sci Sin Inform, 2025, 55: 156–171, doi: 10.1360/SSI-2024-0212

段,具有低成本、简单、易实施等优点^[8]。然而,非实时监测数据使离线评估方式在安全隐患的感知和响应上存在滞后性。相比而言,在线安全评估方法是根据装备的实时监测数据,构建评估模型,实现装备安全性的及时评价。特别是,复杂装备的运行环境通常具有非平稳特性,采用在线安全评估策略,有利于准确捕捉到装备运行状态的动态变化^[9]。以蛟龙号载人深潜器(deep-sea manned submersible, DSMS)为例^[10],面对不同的探测任务和运行环境,随着设备组件的老化和损耗,监测数据的分布特性也会随之发生时变。若采用离线安全评估模型判断深潜器的运行状态,可能产生不准确的评价,导致设备损害,甚至危及人员安全。

非平稳环境下复杂装备的监测数据分布存在时变性,形成概念漂移^[11]。当漂移发生时,已有的安全评估模型不能准确地反映当前数据分布特性,导致模型偏差,产生错误的装备安全性评价^[12,13]。此外,复杂装备通常处于安全状态,故障或失效等不安全状态的发生频率较低,使监测数据流中的安全状态数据远远多于不安全状态数据,形成类别不平衡^[14]。监测数据的标签通常由工程技术人员标注生成,对于连续采集的监测数据,受到人的易疲劳特性限制,很难实现所有数据的全标注^[15]。仅存在少量标记样本的监测数据流存在标签稀缺问题^[16]。除了以上监测数据流的自身分布特性外,有限的计算和存储资源也是在线安全评估所需面临的挑战^[17]。

为解决上述问题,研究人员分别提出基于在线主动学习和在线半监督学习的安全评估方法^[18]。在线主动学习方法通过有选择地对数据进行标记,以最大化系统性能^[19]。针对非平稳环境中的概念漂移问题,He等^[6]提出了一种动态模型解释引导学习方法,改进了宽度学习系统的增量更新策略,更好地适应非平稳环境中监测数据的变化。此外,设计了一种动态样本标记策略,基于Shapley函数,计算每个样本的平均贡献,并选择具有高贡献的样本进行标记。Ienco等^[20]提出了一种基于块的主动学习框架,在新数据块到来时,首先利用聚类技术,将样本分成多个簇,再从每个簇中选择代表性样本,从而使标记样本更好地覆盖整个特征空间。与之相似,Zgraja等^[21]在聚类算法中引入了加权机制,在具有较高权重的簇中选择更多样本。Liu等^[22]开发了一个基于双查询策略的在线主动学习框架,根据局部密度和不确定性标记代表性样本;同时,使用了一种基于艾宾浩斯定律(Ebbinghaus'law)的遗忘机制,更新滑动窗口,避免噪声数据的影响。针对化工厂设备的状态评估问题,Hu等^[23]提出了一种改进的增量非平衡深度神经网络。采用Best-versus-Second Best准则寻找监测数据中不确定性最大且难以学习的样本,并通过对网络的分层增量更新学习潜在的新故障。Zhang等^[24]提出了一种基于在线学习的配对集成模型,用于标签稀缺的类别不平衡数据流。集成模型由一个静态分类器和一个动态分类器组成,以应对突变式和渐变式的概念漂移。为了选择最具代表性的样本,还提出了一种由不确定性策略和不平衡策略组成的混合标记机制。其中,不确定性策略基于边际不确定性,选择位于决策边界附近的样本进行标记;不平衡策略则更倾向于选择少数类样本。

与在线主动学习不同,在线半监督学习方法假设监测数据中的部分样本已经被标记,并同时从标记数据和未标记数据中学习^[25]。基于流形假设,Shen等^[26]提出了一种双层自组织增量神经网络。其中,第1层负责学习输入数据的概率密度分布,第2层则通过探测低密度区域,有效地分离不同类别的样本。Zhang等^[27]通过聚类算法,从数据流中构建多个微簇,并根据集成分类器的预测标签和簇的内部结构信息,推断每个微簇的类标签。同时,采用标记后的样本对集成分类器加以训练。为解决概念漂移问题,还设计了一个动态加权机制,对集成结构加以动态调整。Soares等^[28]采用具有流形正则化的径向基神经网络(radial basis function network, RBFNN)学习未标记的数据。流形正则化被用来最大化分类器的平滑性,即分类器在测试样本邻域内的变化量总和。通过动态调整RBFNN的节点,所提方法能够很好地适应数据分布变化。在自动驾驶场景中,Zhang等^[29]基于少量标记数据,采用K近邻(K-nearest neighbors, KNN)聚类方法初始化生成多个微簇,并基于未标记数据实时更新微簇。微簇内的样本都具有与微簇相同的类别标签,并用来训练集成分类器,对周围车辆的操纵行为加以分类。

虽然上述在线主动学习和半监督学习方法有效地减少了安全评估模型对标记样本的依赖,但它们仍然需要对部分监测数据进行实时标注,以便更新模型^[30]。然而,在一些极端情况下,安全评估系

统在初始化后,便无法再获得任何监测数据的样本标签.这就导致上述方法在学习过程中无法获得必要的标签信息,从而难以适应监测数据的分布变化,最终可能导致安全评估模型的性能下降,甚至失效^[31].例如,蛟龙号载人深潜器操作员由于要专注于设备操作,缺乏足够的精力对监测数据进行分析,致使在整个任务执行期间都无法提供任何样本的标签信息.这种问题也被称为极端验证延迟,即除部分初始数据外,新获得的监测数据标记将无限延迟^[32].针对该类问题,现有的安全评估方法往往无法满足故障或潜在风险的高精度识别.鉴于此,本文提出一种基于动态支持向量提取的复杂装备状态实时安全评估方法 (real-time safety assessment for complex equipment based on dynamic support vector extraction, DSVE).考虑到特征取值范围可能存在差异,首先提出一种在线数据归一化技术,旨在实现流式数据的特征尺度一致化,并有效地减少噪声数据的影响.基于此,提出一种动态支持向量提取方法,准确地跟踪数据分布的变化,并根据这些支持向量进行标签信息的传播.同时,基于支持向量的动态加权,消除不同标签下监测数据的非平衡性对模型性能的影响.

2 问题描述

记一个监测数据流为 $S = \{d_t\}_{t=1}^T$, 其中, d_t 为第 t 时刻采集到的监测数据. 假设在开始阶段有少量监测数据被标记, 这些数据的标签可以从历史记录或其他与该数据相关的监测系统中获得. 然而, 由于监测数据的标记成本存在限制, 所以监测数据流中的样本通常都是未标记的. 为便于描述, 将初始的少量标记数据合并到监测数据流 S 中. 假设初始的标记样本数量为 L , 则

$$d_t = \begin{cases} (x_t, y_t), & t \leq L, \\ x_t, & t > L, \end{cases} \quad (1)$$

其中, $x_t \in \mathbb{R}^D$ 为第 t 时刻样本的特征向量, y_t 相应的样本状态标签. 本文关注于区分安全与不安全状态, 因此, $y_t \in \{0, 1\}$. 由于非平稳环境中监测数据流可能存在概念漂移, 所以监测数据的分布可能随时间发生变化. 令 $P_t(X, y)$ 表示第 t 时刻样本对应的联合概率分布, 则监测数据流的概念漂移可以定义为

$$\begin{cases} P_t(X, y) \neq P_{t+1}(X, y), & t = t_d, \\ P_t(X, y) = P_{t+1}(X, y), & t \in [t_d + 1, t_d + \tau), \end{cases} \quad (2)$$

其中, t_d 表示漂移发生的时间点. τ 表示一个短的时间间隔, 用于区分新出现的概念与噪声. 漂移可能是由环境变化或设备老化导致的渐变式的数据分布变化, 也可以是新的故障和状态引起的突变式的数据分布变化. 基于数据驱动的在线安全评估目标在于从 $\{d_1, d_2, \dots, d_{t-1}\}$ 中学习一个预测模型, 以准确地预测第 t 时刻复杂装备的运行状态.

3 DSVE 在线安全评估框架

图 1 给出了所提 DSVE 方法的基本框架. 首先, 初始的标记样本被映射到高维核空间, 并用于训练预测器. 在核空间中, 为每个类别的样本创建一个超球, 其中, 位于超球表面的样本称为支持向量. 在原始空间中, 这些支持向量对应于具有不同类的数据簇之间的边界, 从而勾勒出了数据的几何结构. 随后, 采用实时到达的监测数据对超球加以更新. 由于单个样本无法反映数据流的整体分布及其变化, 所以引入一个长度为 W 的滑动窗口, 通过依次存入数据流中到达的样本, 形成数据块序列. 每当一个新的数据块形成, 根据该数据块中落入现有超球内部和邻域的样本, 对超球加以更新; 对于远离所有超球的样本, 会形成新的超球. 当超球数量超过预先设定的上限时, 通过合并和删除超球, 降低内存压力. 随后, 将超球的状态标签分配给对应的支持向量, 并对这些带有标记的支持向量进行采样, 形成新的训练样本集. 根据不同类别样本的规模, 为训练样本分配权重, 并用于更新预测器.

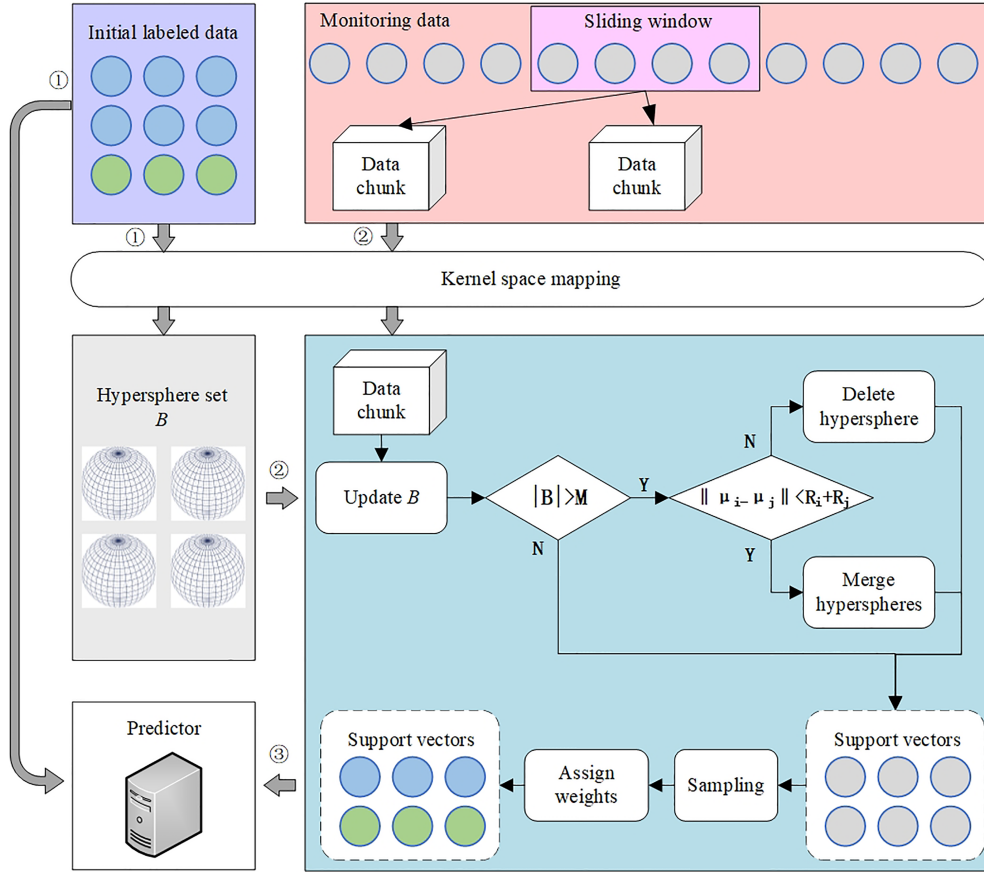


图 1 (网络版彩图) 所提方法框架.

Figure 1 (Color online) Framework of the proposed method.

3.1 在线数据归一化

监测数据中,不同特征的取值范围可能存在差异,可能会对距离度量和预测模型训练产生不良的影响.例如,当两个特征的范围存在较大差异时,样本间的距离计算可能会被具有较大尺度的特征主导,导致模型对数据的敏感性不均衡.为了确保各个特征对模型贡献相对均衡,数据归一化被视为一项重要的预处理技术,用于统一特征的取值范围.

由于无法直接访问所有数据,所以离线的数据归一化方法很难直接应用于监测数据流.因此,提出一种在线自适应归一化方法.对于第 i 个监测数据 d_i ,其归一化后的特征向量表示为

$$\hat{x}_i = (x_i - m)^T N, \quad (3)$$

其中, $m = [m^1, m^2, \dots, m^D] \in \mathbb{R}^D$, $N = \text{diag}(n^1, n^2, \dots, n^D)$,

$$n^i = \frac{1}{M^i - m^i}, \quad (4)$$

M^i 和 m^i 分别为第 i 维特征的动态上限与动态下限.换言之, N 中每一个元素表示为对应特征动态上限与动态下限差的倒数.其中, M^i 和 m^i 根据监测数据的统计特征不断调整.具体来说, M^i 和 m^i 在初始阶段分别为第 i 维特征在初始标记样本中的最大值与最小值.随后,根据统计提取的支持向量样本的最大值与最小值来更新 M^i 和 m^i .由于提取的支持向量反映了监测数据的分布轮廓,过滤了噪声样本,因此可以降低噪声对 M^i 和 m^i 的影响.此外, N 不会改变异常状态数据与其他数据在分布上的差异.为方便描述,本文以下部分将省略归一化向量 N ,使用 x_i 来表示归一化后的样本特征向量.

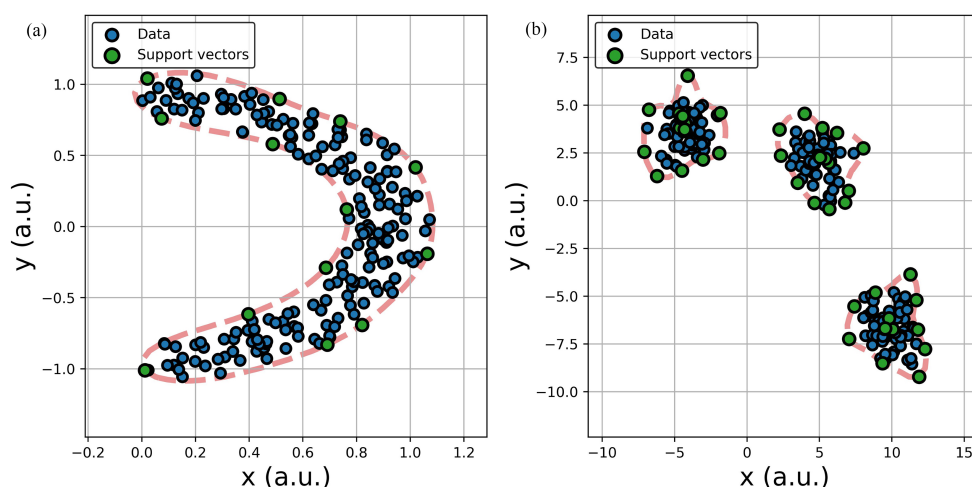


图2 (网络版彩图) 二维数据集上基于支持向量构建的簇边界. (a) 单个簇; (b) 多个簇.

Figure 2 (Color online) Cluster boundaries based on support vectors on two-dimensional datasets. (a) Single cluster; (b) multiple clusters.

3.2 动态支持向量提取

对于存在极端验证延迟的监测数据流, 准确地传播标签信息并适应变化的数据分布是确保安全评估模型性能的关键^[33]. 通常, 标签信息的传播取决于数据的内在结构. 通过分析样本之间的相似程度和聚合关系^[34], 可以将标记样本的标签传播到相似的未标记样本^[35,36]. 支持向量域描述 (support vector domain description, SVDD)^[37] 是一种经典的数据结构描述方法. 它可以通过少量的支持向量对任意形状的数据结构加以描述, 还可以借助松弛变量有效地检测和消除噪声数据. 图2展示了SVDD在两组二维数据上的分类结果. 可以看出, SVDD不仅能够有效勾勒出不规则数据的边界, 还能准确识别数据之间的聚合关系. 因此, 本文利用SVDD提供的数据结构, 实现标签信息传递, 可以有效地识别新的异常状态, 并避免噪声数据对状态评估性能的干扰.

为了准确获得支持向量, SVDD利用核函数将监测数据从输入空间映射到高维空间. 在高维空间中, SVDD找到一个最小的超球体, 将大多数数据包围在其内部. 此时, 处于超球体边界上的数据就被认定为支持向量. 本文提出的DSVE方法中, 通过动态维护由多个超球体组成的集合 B , 实现对监测数据内部结构的准确追踪, 并及时地适应数据分布的变化. B 中的每一个元素代表一个高维空间的超球体, 结构如下:

$$b_n = \{\mu_n, L_n, SV_n, \tau_n\}, \quad (5)$$

其中, μ_n 为超球体的中心, L_n 为超球体对应的状态标签, SV_n 是支持向量集, τ_n 是从当前时刻到超球最近一次更新的时间间隔, 其初始值设置为0.

3.2.1 初始化

在DSVE方法中, 首先为每一类的数据构建一个初始超球, 用于发现不同类别样本的簇边界. 对于第 i 类, 构建其在核空间的超球体 b_n 可以表示为以下约束优化问题:

$$\begin{aligned} \min_{\mu_i, R_i, \xi_{i,j}} \quad & R_i^2 + C \sum_{j=1}^{n_i} \xi_{i,j} \\ \text{s.t.} \quad & \|\phi(x_{i,j}) - \mu_i\|^2 \leq R_i^2 + \xi_{i,j}, \\ & \xi_{i,j} > 0, \quad \forall i = 1, 2, \dots, n_c, \end{aligned} \quad (6)$$

其中, $x_{i,j}$ 为初始获得的标记样本中第 i 类的第 j 个样本, $\xi_{i,j}$ 表示其所对应的松弛因子, n_i 为第 i 类样本的数量, C 一个惩罚参数, 用于权衡超球体体积和误分率. ϕ 表示高斯 (Gauss) 核函数. 尽管原则上可以采用任意形式的核函数, 但是根据文献 [38] 的分析, 高斯核函数可以提供更紧密的轮廓表示.

通过引入拉格朗日 (Lagrange) 乘子 $\alpha_{i,j} \geq 0$, $\beta_{i,j} \geq 0$, 式 (6) 所示的约束优化问题可以转化为

$$\min L(R_i, \mu_i, \xi_{i,j}, \alpha_{i,j}, \beta_{i,j}) = R_i^2 + C \sum_{j=1}^{n_i} \xi_{i,j} - \sum_{j=1}^{n_i} \alpha_{i,j} (R_i^2 + \xi_{i,j} - \|\phi(x_{i,j}) - \mu_i\|^2) - \sum_{j=1}^{n_i} \beta_{i,j} \xi_{i,j}. \quad (7)$$

分别求解 L 相对于 R_i , μ_i 和 $\xi_{i,j}$ 的偏导 $\frac{\partial L}{\partial R_i}$, $\frac{\partial L}{\partial \mu_i}$, $\frac{\partial L}{\partial \xi_{i,j}}$, 并令其为 0, 可以得到

$$\sum_{j=1}^{n_i} \alpha_{i,j} = 1, \quad \mu_i = \sum_{j=1}^{n_i} \alpha_{i,j} \phi(x_{i,j}), \quad \alpha_{i,j} = C - \beta_{i,j}. \quad (8)$$

此外, 基于 KKT 条件 (Karush-Kuhn-Tucker condition), 有

$$\begin{aligned} \beta_{i,j} \xi_{i,j} &= 0, \\ \alpha_{i,j} (R_i^2 + \xi_{i,j} - \|\phi(x_{i,j}) - \mu_i\|^2) &= 0. \end{aligned} \quad (9)$$

通过消除变量 R_i , μ_i , $\xi_{i,j}$ 和 β_i , 式 (7) 所示的约束优化问题可以转化为以下对偶问题:

$$\begin{aligned} \max_{\alpha_{i,j}} & \sum_{j=1}^{n_i} \alpha_{i,j} K(x_{i,j}, x_{i,j}) - \sum_{j,k=1}^{n_i} \alpha_{i,j} \alpha_{i,k} K(x_{i,j}, x_{i,k}) \\ \text{s.t.} & \sum_{j=1}^{n_i} \alpha_{i,j} = 1, \quad 0 \leq \alpha_{i,j} \leq C, \quad \forall j = 1, 2, \dots, n_i, \end{aligned} \quad (10)$$

式中, $K(x_{i,j}, x_{i,k})$ 表示 $\phi(x_{i,j})$ 与 $\phi(x_{i,k})$ 的点积. 在高斯核空间中,

$$K(x_{i,j}, x_{i,k}) = e^{-q\|x_{i,j} - x_{i,k}\|^2}. \quad (11)$$

通过求解式 (10) 所示的优化问题, 可以得到每个样本对应的拉格朗日乘子 $\alpha_{i,j}$, 其中, 满足 $0 < \alpha_{i,j} < C$ 对应的样本位于超球体表面, 构成了支持向量集 SV_n . 这些支持向量共同描述了第 i 类数据分布的边界. 因此, 对应的超球状态标签 $L_n = i$. 基于 $\alpha_{i,j}$ 和 SV_n , 超球的球心通过下式计算获得:

$$\mu_n = \sum_{j=1}^{n_i} \alpha_{i,j} \phi(x_{i,j}). \quad (12)$$

3.2.2 支持向量更新

为及时地适应数据分布的变化, 集合 B 中的超球需要根据新收集的样本进行更新, 从而获得当前数据分布所对应的支持向量. 监测数据的变化形式通常可以划分为两类, 其一是已有数据簇的移动与变形, 其二为新数据簇的出现.

在第 1 种情况下, 属于该簇的新样本通常会落在对应的超球内部或其邻域范围内. 可以通过这些新样本, 更新超球结构及其对应的支持向量. 令 x_i 表示滑动窗口形成的新数据块中的第 i 个样本, 其到超球 b_n 的球心距离记为

$$\begin{aligned} d(x_i, \mu_n) &= \|\phi(x_i) - \mu_n\| \\ &= \left(1 - 2 \sum_{j=1}^{|SV_n|} \alpha_{n,j} K(x_i, x_{n,j}) + \sum_{j,k=1}^{|SV_n|} \alpha_{n,j} \alpha_{n,k} K(x_{n,j}, x_{n,k}) \right)^{\frac{1}{2}}. \end{aligned} \quad (13)$$

如果 $d(x_i, \mu_n)$ 满足下式, 则认为 x_i 位于超球 b_n 的内部或与 b_n 接近. 因此, 将 x_i 分配给超球 b_n .

$$d(x_i, \mu_n) = \min \{d(x_i, \mu_j) \mid d(x_i, \mu_j) \leq \gamma R_j\}, \quad (14)$$

$$R_j = \max \{d(x, \mu_j) \mid x \in \text{SV}_j\}, \quad (15)$$

其中, R_j 为超球 b_j 的半径, $\gamma > 1$ 为半径松弛因子, 用于发现与超球接近的样本. 记 J_n 为分配给超球 b_n 的样本集合, 其他远离现有超球的样本则被分配到集合 O 中.

$$O = \{x_i \mid d(x_i, \mu_j) > \gamma R_j, \forall j = 1, 2, \dots, |B|\}. \quad (16)$$

如果 J_n 中的样本数量大于 2, 本文所提 DSVE 方法便采用式 (9) 更新超球 b_n . 随后, 根据式 (14) 进一步判断集合 O 中是否有样本可以分配给超球 b_n . 如果有符合的样本, 将这些样本从集合 O 中移除, 并添加到 J_n 当中. 重复此过程, 直到集合 O 为空集, 或不再发生变化. 此外, 对于任何更新过的超球, 其活跃度权重都会被重置为 0, 否则, 按 $\tau_i \leftarrow \tau_i + 1$ 进行更新.

当 B 中的超球不再变化时, 如果集合 O 不为空集, 则说明现有超球体集已经无法描述整个数据块. 因此, 基于集合 O 中的样本创建一个新的超球, 并将其状态标签设置为最近邻超球体所对应的状态标签. 通过上述策略, 所提出的 DSVE 方法可以有效地跟踪现有簇的变化, 同时发现潜在的新簇. 此外, 更新后的支持向量可以作为当前概念下数据结构的高度概括, 而无需保留所有数据.

由于可用内存和计算资源的限制, 将集合 B 中超球数量上限设为 M . 一旦 $|B| > M$, 则合并彼此相近的超球. 如果任意两个超球 b_i 和 b_j 满足式 (17), 则认为它们描述的数据存在重叠.

$$\|\mu_i - \mu_j\| < R_i + R_j. \quad (17)$$

即两个超球的球心距小于它们的半径和. 其中, 球心距可以计算为

$$\|\mu_i - \mu_j\| = \|\mu_i\|^2 + \|\mu_j\|^2 - 2 \sum_{m \in \text{SV}_i, n \in \text{SV}_j} \alpha_{i,m} \alpha_{j,n} K(x_{i,m}, x_{j,n}). \quad (18)$$

如果这两个超球具有相同的状态标签, 则合并它们的支持向量集 SV_i 和 SV_j , 并基于式 (10) 形成一个新超球. 相应的活跃度权重被设定为两个超球的活跃度权重的较小值. 当不存在可以合并的超球时, 则删除集合 B 中最不活跃的超球, 即

$$B \leftarrow B \setminus \left\{ b_i \mid i = \arg \max_j \tau_j \right\}. \quad (19)$$

3.3 预测器更新

为了在不断变化的环境中保持良好的性能, 所选择的预测器应具备适应数据分布变化的能力. 霍夫丁自适应树 (Hoeffding adaptive tree, HAT) 是由 Bifet 等^[39] 提出的一种增量分类算法. 该算法能够基于对样本的单次学习, 在线构建分类模型, 并通过自适应的分支调整, 在非稳定环境中保持对新概念的准确追踪. 因此, 选择它作为 DSVE 的预测器.

初始化阶段, 全部的初始的标记样本都用于训练预测器. 采用代价敏感法为不同类样本赋予不同的权重, 从而消除类别不平衡的影响. 令 $P(c)$ 表示第 c 类中样本的比例, 则该类样本的权重为

$$w_c = 1 - P(c). \quad (20)$$

当集合 B 中的超球在新数据块上完成更新后, 便从每个超球的支持向量集 SV_i 中采用无放回的方式采样 $\lceil (1/\tau_i) |\text{SV}_i| \rceil$ 个样本, 其中, $\lceil \cdot \rceil$ 表示向上取整函数. 这些数据采用超球体的状态标签 L_i 作为其标签, 同时根据式 (20) 分配权重, 用于更新预测器. 可见, 新更新的超球可以提供更多的训练数据, 使预测器可以更加关注新概念, 从而更快地适应变化.

3.4 时间复杂度分析

本文所提 DSVE 方法主要包含 3 个核心步骤,分别是在线数据归一化、动态支持向量提取与预测器更新.对于长度为 W 的滑动窗口,在线数据归一化的时间复杂度为 $\mathcal{O}(2WD)$.动态支持向量提取过程中,更新超球集的复杂度为 $\mathcal{O}(W^2DK)$,其中, K 表示参与更新的超球数量.预测器更新的时间复杂度主要取决于所选用的在线分类算法,对于本文使用的霍夫丁自适应树,其时间复杂度为 $\mathcal{O}(WD \log W)$.综上,保留最高次项,DSVE 的时间复杂度为 $\mathcal{O}(W^2DK)$.

4 实验结果与对比分析

为了验证本文所提方法的有效性,面向蛟龙号载人深海潜水器数据集,开展了以下 3 组实验.首先,将本文所提方法 DSVE 分别与两种经典的在线方法及两种离线方法加以对比实验,分析其对不同状态的在线评估准确性;其次,分析了 DSVE 方法中 3 个关键参数对在线评估性能的影响程度;最后,通过消融实验,验证了 DSVE 方法中核心策略的有效性.所有实验均采用 scikit-multiflow 库^[40]实现,在 Inter(R) Core(TM) i7-8700 处理器、16 GB 内存、Nvidia 2070 计算条件下进行.每个实验独立运行 15 次.

4.1 实验数据集

本文实验采用的 DSMS 数据集由山东省青岛市国家深海中心提供,来源于 2017 年 3 月 19 日蛟龙号载人深海潜水器第 151 次探测任务.所有数据以多元时间序列形式呈现,相应的监测时长约为 9 h.

所有数据都包含 14 个特征,其中,8 个来源于蛟龙号载人深海潜水器的内部生命支持系统,包括:舱内氧气浓度、舱内二氧化碳浓度、舱内压力、舱内湿度、舱内温度、应急液压源压力,以及来自另一组传感器的氧气和二氧化碳浓度.其余 6 个特征是蛟龙号载人深海潜水器的外部监测数据,及其液压系统监测数据.表 1 给出了所有特征的取值范围.

在该 DSMS 数据集中,总计有 64186 个样本,其中,56949 个是安全状态样本,剩余的 7237 个为不安全状态样本.为了直观地观察监测数据流是否存在概念漂移,采用主成分分析方法提取了监测数据流的关键属性.图 3 展示了降维后监测数据流在 5 个不同时间段的数据分布,可以看出数据分布随时间发生显著的变化.本文所提方法的核心任务在于如何从上述监测数据流中实时、准确地识别蛟龙号载人深海潜水器的当前状态,即安全或不安全.

4.2 算法评价方法与指标

实验中,假设监测数据流中前 L 个样本的标签已知,并用于对模型进行初始化,而后续样本的标签未知.采用事前评估方法对所提模型的性能进行评估.在每个时间步下,模型首先对样本标签加以预测,然后将模型的预测结果与真实标签进行比较.这意味着模型总是在未曾见过的样本上进行测试.该方法的优点在于测试时不需要保留任何测试集,并且可以最大限度地利用数据.

由于 DSMS 数据集存在样本标签的非平衡性,所以采用几何平均值 (geometric mean, G-mean) 作为算法性能的评价指标.同时,采用 false positive rate (FPR) 和 false negative rate (FNR) 进一步度量所提方法对于不安全状态的误报率和漏报率.上述性能评价指标的具体计算如下所示:

$$\text{G-mean} = \sqrt{\frac{\text{TP}}{\text{TP} + \text{FN}} \times \frac{\text{TN}}{\text{TN} \times \text{FP}}}, \quad (21)$$

$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}}, \quad (22)$$

$$\text{FNR} = \frac{\text{FN}}{\text{TP} + \text{FN}}, \quad (23)$$

表 1 蛟龙号 DSMS 数据特征.
Table 1 Features of JiaoLong DSMS dataset.

	Feature	Range of feature values
Life support system	Oxygen concentration	[0.0, 21.3]
	Carbon dioxide concentration	[0.0, 0.1]
	Pressure	[0.0, 104.2]
	Humidity	[0.0, 69.9]
	Temperature	[0.0, 34.4]
	Emergency hydraulic source pressure	[0.0, 1.0]
	Oxygen concentration backup	[−0.2, 0.7]
	Carbon dioxide concentration backup	[−0.3, 0.3]
Hydraulic system	Fuel tank pressure	[0.0, 22.7]
	Fuel tank temperature 1	[−141.1, 0.0]
	Fuel tank temperature 2	[0.0, 60.7]
External environment	Salinity	[0.0, 5.8]
	Temperature	[0.0, 28.6]
	Pressure	[0.0, 1.0]

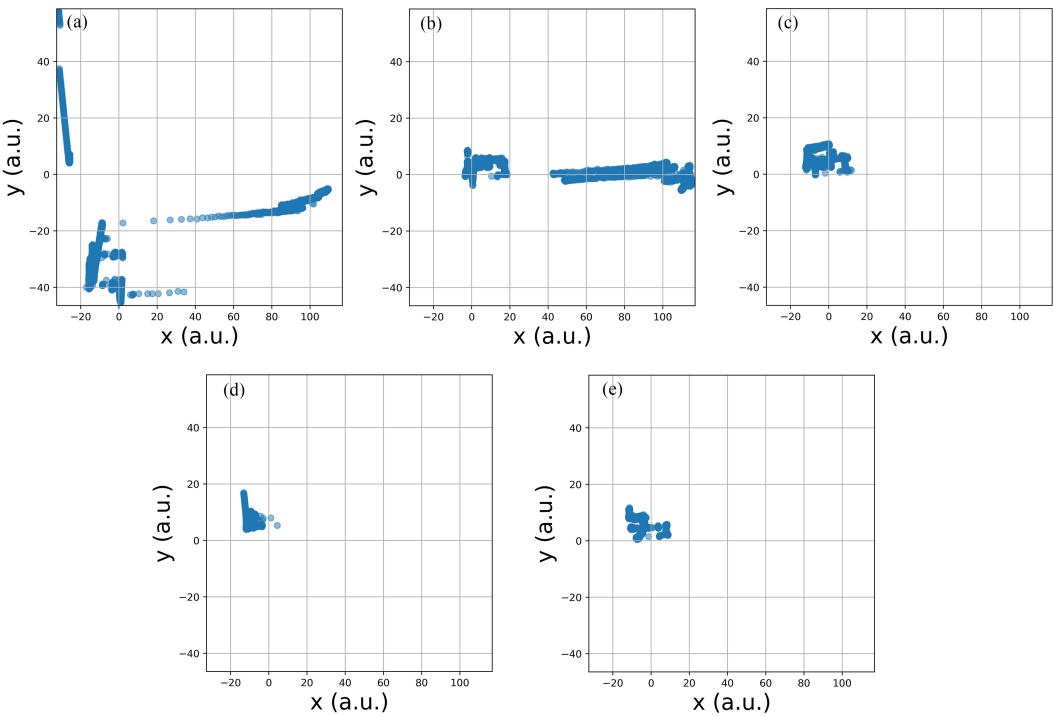


图 3 (网络版彩图) 蛟龙号 DSMS 数据流在不同时间段的数据分布. (a) 数据间隔 [1, 10000]; (b) 数据间隔 [10001, 20000]; (c) 数据间隔 [20001, 30000]; (d) 数据间隔 [30001, 40000]; (e) 数据间隔 [40001, 50000].

Figure 3 (Color online) Data distribution of JiaoLong DSMS dataset in different time periods. (a) Data interval: [1, 10000]; (b) data interval: [10001, 20000]; (c) data interval: [20001, 30000]; (d) data interval: [30001, 40000]; (e) data interval: [40001, 50000].

其中, TP 表示正确预测到的不安全状态的样本个数, TN 表示正确预测到的安全状态的样本个数, FN 表示不安全状态预测为安全状态的样本个数, FP 表示安全状态预测为不安全状态的样本个数.

表 2 DSVE 与对比方法在不同 L 下的性能. 最优结果用粗体表示.Table 2 Performance of DSVE and comparison methods under different L . The best results are in bold.

		DSVE	ARF	LB	HAT	SVM	XGBoost
$L = 500$	G-mean (%)	96.51±2.71	87.56±0.78	77.79±1.16	78.19±0.00	89.13±0.00	80.99±1.61
	FPR (%)	3.53±0.08	2.36±0.16	2.78±0.18	2.52±0.00	2.52±0.00	11.70±0.18
	FNR (%)	4.41±0.19	21.47±0.20	37.76±0.44	37.29±0.00	18.50±0.00	25.71±0.38
	Time (s)	379.18±25.48	32.71±5.03	20.98±4.11	17.61±2.39	28.98±6.41	37.71±6.38
$L = 1000$	G-mean (%)	95.86±1.40	86.70±2.92	78.14±1.82	80.00±0.00	88.79±0.00	45.14±1.08
	FPR (%)	3.99±0.11	5.88±0.20	7.27±0.20	14.43±0.00	2.20±0.00	73.89±0.44
	FNR (%)	7.19±0.12	20.13±0.65	34.16±0.38	25.20±0.00	19.38±0.00	21.96±0.27
	Time (s)	367.25±29.19	39.25±4.91	20.22±3.10	19.88±3.42	39.88±7.80	44.39±8.58
$L = 1500$	G-mean (%)	96.72±1.44	92.83±1.22	93.95±1.92	86.15±0.00	64.81±0.00	85.10±1.57
	FPR (%)	2.01±0.00	1.52±0.44	0.25±0.08	1.64±0.00	1.43±0.00	14.09±0.43
	FNR (%)	6.46±0.41	12.49±0.18	11.51±0.18	24.54±0.00	57.39±0.00	15.70±0.18
	Time (s)	391.77±38.24	30.40±3.99	29.10±3.88	17.44±3.55	49.80±8.52	58.99±7.45

4.3 对比方法及参数设置

面向存在极端验证延迟的复杂装备安全评估问题, 本文选取了两种典型的离线学习方法, 包括支持向量机 (support vector machine, SVM)^[41] 和极限梯度提升 (extreme gradient boosting, XGBoost)^[42], 以及两种经典的在线学习方法, 即自适应随机森林 (adaptive random forest, ARF)^[43] 和改进的在线袋装方法 (leveraging bagging, LB)^[44] 用于与本文所提 DSVE 方法加以性能对比. 此外, 选择霍夫丁自适应树 (HAT) 作为基线方法, 以验证所提动态支持向量提取的有效性. 上述对比算法是数据流分析的典型方法, 但并不是针对极端验证延迟的安全评估问题提出的.

实验中, 为平衡模型适应变化的效率与稳定性, 将 DSVE 中的滑动窗口 W 设为 500. 超球体惩罚参数 C 设定为 0.9, 以有效地区分噪声样本. 设置半径松弛因子 γ 为 1.4, 用于区分新簇, 并有效追踪当前簇的分布变化. 考虑到内存资源的限制, 将最大超球体数量 M 设为 12. DSVE 中的预测器采用 Scikit-multiflow 库中的默认参数. 对比方法中, SVM 和 XGBoost 分别采用了 Scikit-learn 和 XGBoost 库中的实现, 而两种在线方法和霍夫丁自适应决策树的源码则来自于 Scikit-multiflow 库. 所有对比方法均采用相应库中的默认参数值.

4.4 算法性能对比实验

由于 4 种对比方法并不是面向极端验证延迟下的安全评估问题提出的, 所以采用初始的 L 个标记样本对上述对比方法加以训练, 并采用事前评估方法对它们的性能进行测试. 为充分对比分析算法性能, 实验中采用了 3 种不同的初始标记样本规模, 记为 $L = \{500, 1000, 1500\}$. 表 2 展示了所有对比方法的性能表现. 由 G-mean 指标可以看出, DSVE 在整体预测性能上优于所有对比方法. 与基线的霍夫丁自适应决策树算法相比, DSVE 在不同 L 下, 分类准确率分别取得了 18.32%, 15.86% 和 10.57% 的提升, 这证明了, 所提出的动态支持向量提取策略可以在具有极端验证延迟的安全评估问题中有效地传递标签信息, 并及时准确地追踪数据分布的变化.

从误报率和漏报率来看, DSVE 在所有情况下都维持了最低的漏报率, 但具有相对较高的误报率. 分析表明, 这可能是由于 DSVE 方法中不安全状态的超球将正常状态下的样本纳入其邻域所导致的, 这种情况可以通过缩小半径松弛因子 γ 来解决. 但是, 过小的半径松弛因子会导致邻域内的样本无法有效地被捕获, 从而降低对监测数据流中概念漂移的适应效率. 然而, 相比于安全状态被误报, 对不安全状态的漏报可能会导致更加严重的后果, 甚至设备损害和人员伤亡. 因此, 在实际工业系统中, 通常

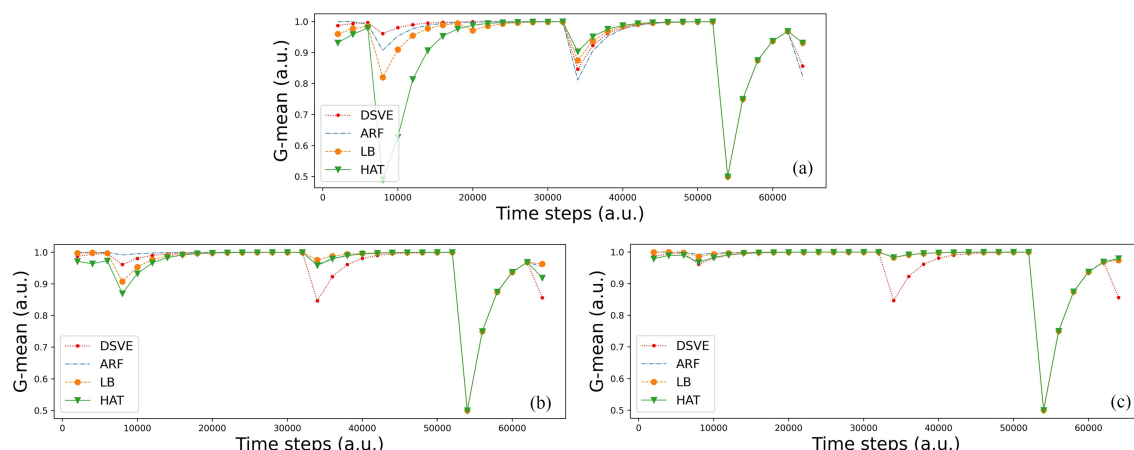


图 4 (网络版彩图) 不同标记比例下各方法的 G-mean 曲线. (a) 标记比例为 1%; (b) 标记比例为 5%; (c) 标记比例为 10%.

Figure 4 (Color online) G-mean curves of each method under different labeling proportions. (a) Labeling proportion 1%; (b) labeling proportion 5%; (c) labeling proportion 10%.

会优先考虑具有低漏报率的方法. 此外, 随着初始可用标记样本数量的增加, 对比方法并没有表现出显著的性能提升, 甚至 XGBoost 与 SVM 的性能会随着 L 增大而呈现恶化趋势. 分析表明, 出现上述现象的主要原因在于 4 种对比方法无法充分利用在线监测数据对模型结构加以实时调整. 这就导致无论采用多少标记样本加以模型训练, 都无法在概念漂移发生后对数据分布做出准确的评估. 相比而言, DSVE 方法可以利用实时监测数据对分类模型加以在线更新. 更多的标记样本有助于 DSVE 方法更好地传播标签信息, 从而提高整体分类性能和对概念漂移的适应能力. 进一步, 从表 2 可以看出, 相比于其他方法, DSVE 方法具有相对较高的时间成本. 其运行时间主要消耗在核空间超球的在线更新过程. 但是, 实时监测数据的采集周期大约为 9 h, 远远高于 DSVE 方法的运行时间. 因此, 所提方法可以在新样本到来之前, 完成对历史监测数据的处理, 满足时效性要求.

为进一步比较本文所提方法与 3 种在线学习方法的分类性能, 随机标记数据流中的部分样本用于更新在线模型. 需要注意的是, 这些额外的标记仅用于 3 种在线学习方法, 但是所提出的 DSVE 仍然仅从未标记的样本中学习. 初始标记样本数量固定选取为 $L = 1000$, 样本标记比例分别设置为 1%, 5% 和 10%. 为对比分析不同方法在学习过程中的分类性能变化, 图 4 显示了在不同的样本标记比例下, DSVE 与 3 种在线学习方法所获得的 G-mean 性能曲线. 表 3 给出了 G-mean 和其他性能指标的数值结果. 由表 3 可见, 在 1% 和 5% 的样本标记比例下, 尽管对比方法获得了额外的样本标签信息, 但是 DSVE 方法仍然取得了最佳的 G-mean 值. 然而, 在 10% 的样本标记比例下, 对比方法呈现出了更好的预测精度, 但是额外地消耗了接近 6000 个样本的标记开销. 尽管如此, DSVE 方法依然表现出了与 ARF 和 LB 相似的安全评估性能. 综上, 所提出的动态支持向量提取策略可以高效地传播标签信息, 从而有效地降低标记数据的成本.

4.5 参数敏感性分析

本文所提出的 DSVE 方法包含 3 个关键参数, 分别为滑动窗口长度 W 、超球数量上限 M 和半径松弛因子 γ . 设置初始标记样本数量 $L = 1000$, 本小节分析了不同参数值对所提方法分类性能的影响.

由图 5(a) 可见, DSVE 方法所获得的 G-mean 值会先随着 W 的增加呈现出上升趋势, 随后趋于稳定. 显然, 较大的 W , 使滑动窗口内包含更多的样本, 从而有利于模型构建更加准确的数据分类边界. 图 5(b) 表明增大超球数量上限, 可以一定程度地改善算法性能. 但是, 与 γ 相比而言, 超球数量上限对算法性能的影响并不显著. 当采用较小的 M 值时, 通过移除最不活跃的超球, 可以在有效地控制

表 3 DSVE 与在线学习方法在不同标记比例下的性能比较. 最优结果用粗体表示.

Table 3 Performance comparison of DSVE and online learning methods under different labeling proportions. The best results are in bold.

		DSVE	ARF	LB	HAT
Labeling proportion 1%	G-mean (%)	96.33±2.05	92.93±1.41	91.43±0.85	93.89±3.01
	FPR (%)	3.26±0.24	0.60±0.18	0.68±0.14	2.31±0.42
	FNR (%)	5.99±1.88	11.82±0.11	19.60±0.24	12.35±0.10
	Time (s)	334.11±22.32	49.11±4.47	27.76±5.44	20.89±3.39
Labeling proportion 5%	G-mean (%)	96.33±2.05	94.64±1.54	93.33±1.79	95.91±2.31
	FPR (%)	3.26±0.24	0.18±0.04	0.38±0.07	0.43±0.12
	FNR (%)	5.99±1.88	6.66±0.08	7.50±0.28	10.76±0.31
	Time (s)	334.11±22.32	55.41±7.80	29.47±5.74	23.33±5.85
Labeling proportion 10%	G-mean (%)	96.33±2.05	99.84±0.39	99.82±2.81	99.68±6.53
	FPR (%)	3.26±0.24	0.06±0.11	0.05±0.01	0.08±0.01
	FNR (%)	5.99±1.88	0.46±0.08	0.54±0.12	1.34±0.26
	Time (s)	334.11±22.32	69.79±10.36	38.66±6.91	25.58±8.04

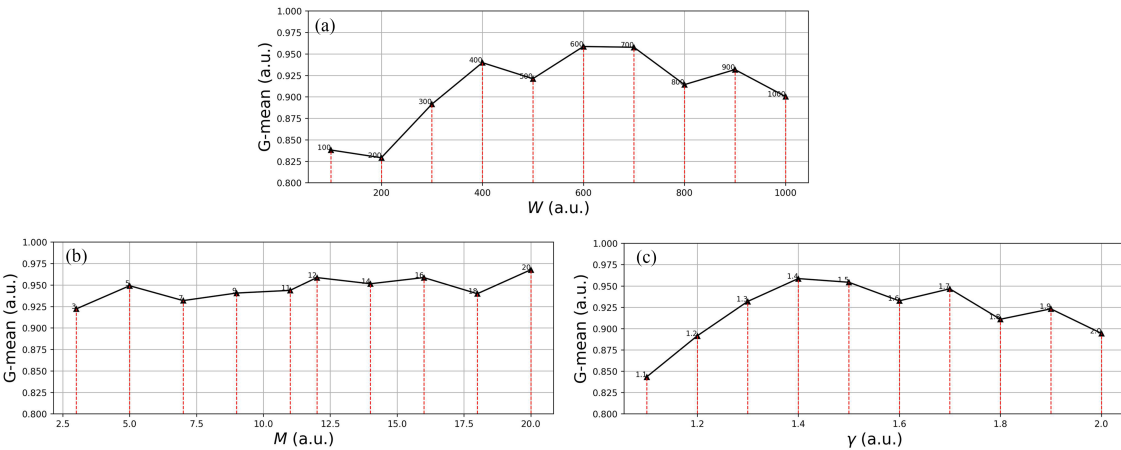


图 5 (网络版彩图) 不同参数的敏感性分析. (a) 滑动窗口长度; (b) 超球的最大数目; (c) 半径松弛系数.

Figure 5 (Color online) Sensitivity analysis of different parameters. (a) Length of sliding window; (b) maximum number of hypersphere; (c) radius relaxation factor.

超球规模的同时, 最大程度地保留有价值的信息. 从图 5(c) 可以看出, 随着 γ 的增加, 模型的预测性能首先得到提升, 但当 γ 超过 1.5 h, 预测性能开始下降. 可见, 较小的 γ 不利于超球发现其邻域内的样本, 从而无法有效地跟踪数据分布的变化. 然而, 过大的 γ 值也可能导致超球错误地将其他簇的样本误判为由变化后的新数据, 从而影响模型的准确性.

4.6 消融实验

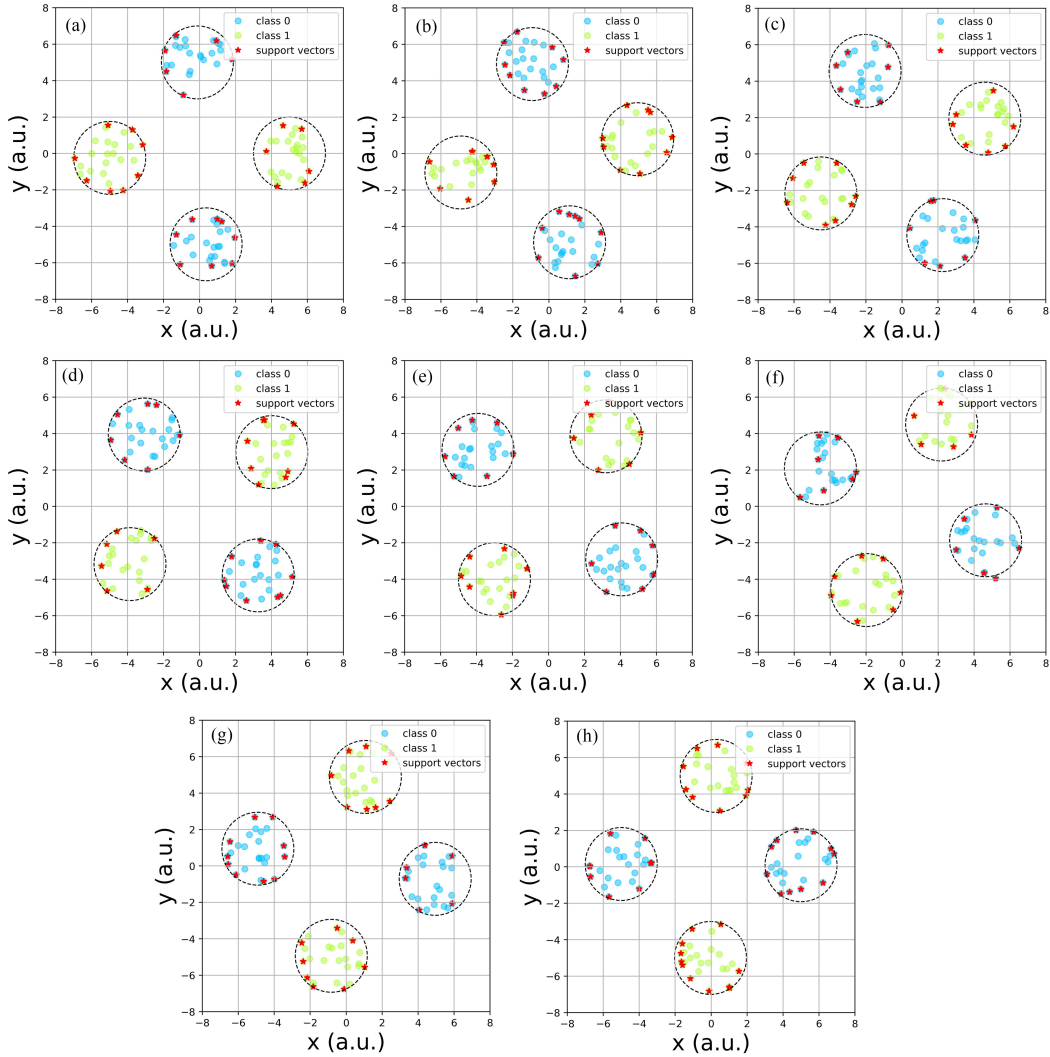
为了验证 DSVE 方法中引入在线数据归一化方法、动态支持向量更新机制、支持向量加权策略与重采样策略, 以及霍夫丁自适应树的有效性, 在消融实验中采用了 DSVE 的 5 种变体:

- (1) DSVE_{NoNOR}: 不采用在线数据归一化方法;
- (2) DSVE_{NoDSV}: 移除了动态支持向量更新机制;
- (3) DSVE_{NoDW}: 将初始样本及所有支持向量的权重固定为 1;
- (4) DSVE_{NoRE}: 移除了支持向量重采样策略, 采用超球的所有支持向量加以模型训练;

表 4 DSVE 及其变体的性能对比. 最优结果用粗体表示.

Table 4 Classification performance comparison of DSVE and its variants. The best results are in bold.

	DSVE	DSVE _{NoNOR}	DSVE _{NoDSV}	DSVE _{NoDW}	DSVE _{NoRE}	DSVE _{HT}
G-mean (%)	95.21±2.41	90.29±4.33	77.90±2.66	89.31±6.79	94.02±0.00	92.54±3.07
FPR (%)	2.37±0.31	4.71±1.01	1.22±0.11	1.09±0.20	1.59±0.00	3.77±0.69
FNR (%)	6.88±0.28	7.15±1.02	19.45±2.81	9.88±1.96	9.44±0.00	10.81±1.46

图 6 (网络版彩图) DSVE 在不同时间步下提取的支持向量. (a) $t = 100$; (b) $t = 300$; (c) $t = 600$; (d) $t = 900$; (e) $t = 1200$; (f) $t = 1500$; (g) $t = 1800$; (h) $t = 2000$.Figure 6 (Color online) Support vectors extracted by DSVE at different time steps. (a) $t = 100$; (b) $t = 300$; (c) $t = 600$; (d) $t = 900$; (e) $t = 1200$; (f) $t = 1500$; (g) $t = 1800$; (h) $t = 2000$.

(5) DSVE_{HT}: 将预测器由霍夫丁自适应树替换为霍夫丁树.

根据表 4 给出的 $L = 1000$ 下的实验结果可知, 缺乏数据归一化的 DSVE_{NoNOR} 整体预测精度低于 DSVE, 同时具有更高的误报率. 由于缺少对新概念的适应能力, 移除动态支持向量更新机制的 DSVE_{NoDSV} 具有显著恶化的预测精度, 并且漏报率明显提高. 统一的样本权重导致了 DSVE_{NoDW} 较低的 G-mean, 而移除支持向量重采样策略会导致 DSVE_{NoRE} 的漏报率增加. 由于霍夫丁树缺少适

应数据分布变化的能力, DSVE 在所有指标上都优于 $DSVE_{HT}$. 综上, DSVE 中的在线数据归一化方法, 动态支持向量提取机制, 支持向量加权策略及重采样策略能够有效地提高模型的预测精度, 同时降低误报率与漏报率.

为进一步直观地分析动态支持向量更新策略的合理性, 在一个合成二维数据流上可视化了支持向量的变化过程. 初始化 4 个以 $[0, 5]$, $[5, 0]$, $[-5, 0]$ 和 $[0, -5]$ 为圆心, 半径为 2 的圆. 为了模拟概念漂移, 这些圆的中心沿着弧线以恒定速度移动. 在这些圆内均匀生成 2000 个样本, 包含两个类别, 分别用绿色和蓝色表示. 将初始标记样本数量和 DSVE 的滑动窗口尺寸设为 100, 图 6 描述了 DSVE 方法所提取支持向量的连续变化图. 可以看出, 随着数据分布的变化, SSALD 中的支持向量能够持续跟踪当前的数据分布, 并准确的描述不同类别样本当前的数据结构, 为标签信息的传播奠定了基础.

5 结束语

针对存在极端验证延迟的复杂装备安全评估问题, 提出了一种基于动态支持向量提取的实时安全评估方法, 从而有效降低了在线安全评估系统对样本标签的依赖性. 在所提方法中, 考虑到监测数据流的特征可能具有不同取值范围, 采用在线数据归一化方法, 对流式数据特征尺度加以统一. 随后, 设计了一种动态支持向量提取方法, 用于准确地跟踪数据分布的变化, 更好地传播标签信息. 最后, 通过对支持向量的重采样和加权, 消除不同状态下数据比例的差异对模型性能的影响. 在蛟龙号载人深潜器数据集上进行的实验表明, 提出的方法相较于现有的在线与离线方法, 具有更高的预测精度, 并且在安全评估过程中能够显著降低漏报率.

虽然本文所提方法有效地降低了实时安全评估方法对样本标签的依赖, 并且可以在非平稳环境中保持良好的预测精度, 但是, 对问题相关的领域知识利用不够充分. 已有研究表明, 有效利用领域知识不仅可以进一步降低数据的标记成本, 同时可以提高模型的预测性能. 鉴于此, 设计整合领域专业知识与数据驱动模型的方法将是我们未来的工作.

参考文献

- 1 Liu Z, Deng Y, Zhang Y, et al. Safety assessment of dynamic systems: an evidential group interaction-based fusion design. *IEEE Trans Instrum Meas*, 2021, 70: 1–14
- 2 Chai Y, Mao W B, Ren H, et al. Research on operational safety assessment for spacecraft launch system: progress and challenges. *Act Autom Sin*, 2019, 45: 1829–1845 [柴毅, 毛万标, 任浩, 等. 航天发射系统运行安全性评估研究进展与挑战. *自动化学报*, 2019, 45: 1829–1845]
- 3 Keller W, Modarres M. A historical overview of probabilistic risk assessment development and its use in the nuclear power industry: a tribute to the late Professor Norman Carl Rasmussen. *Reliab Eng Syst Saf*, 2005, 89: 271–285
- 4 Aldemir T. A survey of dynamic methodologies for probabilistic safety assessment of nuclear power plants. *Ann Nucl Energy*, 2013, 52: 113–124
- 5 Prajna S, Jadbabaie A, Pappas G J. Stochastic safety verification using barrier certificates. In: *Proceedings of the 43th IEEE Conference on Decision and Control (CDC)*, 2004. 929–934
- 6 He X, Liu Z. Dynamic model interpretation-guided online active learning scheme for real-time safety assessment. *IEEE Trans Cybern*, 2024, 54: 2734–2745
- 7 Zhu Z R, Chai Y, Yang Z M, et al. Safety criteria based on barrier function under the framework of boundedness for some dynamic systems. *Sci China Inf Sci*, 2022, 65: 122203
- 8 Wisniewski R, Bujorianu L M. Safety of stochastic systems: an analytic and computational approach. *Automatica*, 2021, 133: 109839
- 9 Liu Z, He X. Dynamic submodular-based learning strategy in imbalanced drifting streams for real-time safety assessment in nonstationary environments. *IEEE Trans Neural Netw Learn Syst*, 2024, 35: 3038–3051
- 10 Cui W. Development of the Jiaolong deep manned submersible. *Mar Technol Soc J*, 2013, 47: 37–54
- 11 Liu Z, He X. Real-time safety assessment for dynamic systems with limited memory and annotations. *IEEE Trans Intell Transp Syst*, 2023, 24: 10076–10086

- 12 Jiao B, Guo Y, Gong D, et al. Dynamic ensemble selection for imbalanced data streams with concept drift. *IEEE Trans Neural Netw Learn Syst*, 2024, 35: 1278–1291
- 13 Liu C, He X, Zhou D, et al. Safety assessment for dynamic systems: a survey. *Cybern Intell*, 2023. doi: 10.26599/CAI.2024.9390001
- 14 Malialis K, Panayiotou C G, Polycarpou M M. Online learning with adaptive rebalancing in nonstationary environments. *IEEE Trans Neural Netw Learn Syst*, 2020, 32: 4445–4459
- 15 Lu J, Liu A, Dong F, et al. Learning under concept drift: a review. *IEEE Trans Knowl Data Eng*, 2019, 31: 2346–2363
- 16 O'Reilly C, Gluhak A, Imran M A, et al. Anomaly detection in wireless sensor networks in a non-stationary environment. *IEEE Commun Surv Tut*, 2014, 16: 1413–1432
- 17 Krempel G, Žliobaite I, Brzeziński D, et al. Open challenges for data stream mining research. *SIGKDD Explor Newsl*, 2014, 16: 1–10
- 18 Jiao B, Guo Y, Yang C, et al. Incremental weighted ensemble for data streams with concept drift. *IEEE Trans Artif Intell*, 2022, 5: 92–103
- 19 Fahy C, Yang S, Gongora M. Scarcity of labels in non-stationary data streams: a survey. *ACM Comput Surv*, 2022, 55: 1–39
- 20 Ienco D, Bifet A, Žliobaite I, et al. Clustering based active learning for evolving data streams. In: *Proceedings of the 16th International Conference on Discovery Science*, 2013. 79–93
- 21 Zgraja J, Gama J, Wozniak M. Active learning by clustering for drifted data stream classification. In: *Proceedings of the 18th European Conference on Machine Learning and Knowledge Discovery in Databases Workshops: DMLE 2018 and IoTStream*, 2018. 10–14
- 22 Liu S, Xue S, Wu J, et al. Online active learning for drifting data streams. *IEEE Trans Neural Netw Learn Syst*, 2021, 34: 186–200
- 23 Hu Z, Jiang P. An imbalance modified deep neural network with dynamical incremental learning for chemical fault diagnosis. *IEEE Trans Ind Electron*, 2018, 66: 540–550
- 24 Zhang H, Liu W, Shan J, et al. Online active learning paired ensemble for concept drift and class imbalance. *IEEE Access*, 2018, 6: 73815–73828
- 25 Shen Y Y, Zhang Y M, Zhang X Y, et al. Online semi-supervised learning with learning vector quantization. *Neurocomputing*, 2020, 399: 467–478
- 26 Shen F, Yu H, Sakurai K, et al. An incremental online semi-supervised active learning algorithm based on self-organizing incremental neural network. *Neural Comput Applic*, 2011, 20: 1061–1074
- 27 Zhang P, Zhu X, Tan J, et al. Classifier and cluster ensembles for mining concept drifting data streams. In: *Proceedings of the 10th IEEE International Conference on Data Mining*, 2010. 1175–1180
- 28 Soares R G F, Minku L L. OSNN: an online semisupervised neural network for nonstationary data streams. *IEEE Trans Neural Netw Learn Syst*, 2021, 34: 6029–6041
- 29 Zhang H, Fu R. An ensemble learning-online semi-supervised approach for vehicle behavior recognition. *IEEE Trans Intell Transp Syst*, 2021, 23: 10610–10626
- 30 Wen Y M, Liu S, Miao Y Q, et al. Survey on semi-supervised classification of data streams with concept drifts. *J Softw*, 2021, 33: 1287–1314 [文益民, 刘帅, 缪裕青, 等. 概念漂移数据流半监督分类综述. *软件学报*, 2021, 33: 1287–1314]
- 31 Li Y H, Wang T T, Wang S G, et al. Online active learning method for imbalanced data stream. *Act Autom Sin*, 2024, 50: 1389–1401 [李艳红, 王甜甜, 王素格, 等. 非平衡概念漂移数据流主动学习方法. *自动化学报*, 2024, 50: 1389–1401]
- 32 Costa Júnior J D, Faria E R, Silva J A, et al. Novelty detection for multi-label stream classification under extreme verification latency. *Appl Soft Comput*, 2023, 141: 110265
- 33 Idrees M M, Stahl F, Badii A. Adaptive learning with extreme verification latency in non-stationary environments. *IEEE Access*, 2022, 10: 127345
- 34 Guo Y, Zheng Z, Pu J, et al. Robust online active learning with cluster-based local drift detection for unbalanced imperfect data. *Appl Soft Comput*, 2024, 165: 112051
- 35 Zhang C, Jin C Q, Zhou A Y. Clustering algorithm over uncertain data streams. *J Softw*, 2010, 21: 2173–2182 [张晨, 金澈清, 周傲英. 一种不确定数据流聚类算法. *软件学报*, 2010, 21: 2173–2182]
- 36 Chu R H, Wang H J, Yang Y, et al. Clustering ensemble based on density peaks. *Act Autom Sin*, 2016, 42: 1401–1412 [褚睿鸿, 王红军, 杨燕, 等. 基于密度峰值的聚类集成. *自动化学报*, 2016, 42: 1401–1412]
- 37 Tax D M J, Duin R P W. Support vector domain description. *Pattern Recogn Lett*, 1999, 20: 1191–1199
- 38 Ben-Hur A, Horn D, Siegelmann H T, et al. Support vector clustering. *J Mach Learn Res*, 2001, 2: 125–137
- 39 Bifet A, Gavaldà R. Adaptive learning from evolving data streams. In: *Proceedings of the 8th International Symposium*

- on Intelligent Data Analysis, 2009. 249–260
- 40 Montiel J, Read J, Bifet A, et al. Scikit-multiflow: a multi-output streaming framework. *J Mach Learn Res*, 2018, 19: 2915–2914
 - 41 Zhang X G. Introduction to statistical learning theory and support vector machines. *Act Autom Sin*, 2000, 26: 32–42 [张学工. 关于统计学习理论与支持向量机. *自动化学报*, 2000, 26: 32–42]
 - 42 Sagi O, Rokach L. Approximating XGBoost with an interpretable decision tree. *Inf Sci*, 2021, 572: 522–542
 - 43 Gomes H M, Bifet A, Read J, et al. Adaptive random forests for evolving data stream classification. *Mach Learn*, 2017, 106: 1469–1495
 - 44 Bifet A, Holmes G, Pfahringer B. Leveraging bagging for evolving data streams. In: *Proceedings of the 18th Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2010. 135–150

Real-time safety assessment for complex equipment based on dynamic support vector extraction

Botao JIAO¹, Yinan GUO^{2*}, Xiao HE³, Jiayang PU² & Dunwei GONG⁴

1. *College of Computer Science and Technology, China University of Petroleum (East China), Qingdao 266580, China*

2. *School of Mechanical and Electrical Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China*

3. *Department of Automation, Tsinghua University, Beijing 100084, China*

4. *College of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266061, China*

* Corresponding author. E-mail: guoyinan@cumt.edu.cn

Abstract In non-stationary environments, real-time safety assessment of complex equipment holds immense significance as it enables the timely detection of potential hazards. The dynamic changes in online monitoring data emerge as the most prominent obstacle. While some incremental learning-based methods can dynamically adjust based on monitoring data, they typically operate under the assumption that labels for all monitoring data are available. However, in most real-world scenarios, only part of the initial data collected in advance or historical data are labeled, leaving real-time monitoring data predominantly unlabeled, which poses a difficult challenge known as the extreme verification latency problem. To address this issue, a real-time safety assessment method based on dynamic support vector extraction is proposed. First, several hyperspheres with different class monitoring data are constructed in the kernel space using the support vector domain description, with the purpose of capturing the geometric structure of monitoring data. Second, the label information is propagated within the unlabeled monitoring data by updating the hypersphere, enabling adaptation to distribution changes. Additionally, a dynamic support vector weighting strategy is employed to mitigate the impact of class imbalance on model performance. Finally, experiments on the JiaoLong deep-sea manned submersible dataset verified the effectiveness of the proposed method.

Keywords real-time safety assessment, non-stationary environment, extreme verification latency, class imbalance, data stream