

动态系统的实时安全性评估技术

何 潇¹ 刘泽夷¹ 胡嵩乔¹ 刘 畅¹ 周东华^{1,2}

摘 要 动态系统的实时安全性评估 (Real-time safety assessment, RTSA) 在防止潜在安全事故导致重大损失方面发挥着关键作用. 随着系统功能和操作环境复杂性的日益增加, 开发有效的实时安全性评估技术面临着更大的挑战. 鉴于此, 阐述动态系统实时安全性评估的概念定义, 从环境的平稳性及评估模型的构建方式两个维度出发提出一种分类框架, 给出相应的问题描述, 较系统地回顾了动态系统实时安全性评估技术的现有进展, 讨论针对不同实际系统的部署策略, 分析现有技术的发展趋势, 探讨实时安全性评估中亟待解决的问题与未来的发展方向.

关键词 实时安全性评估, 动态系统, 安全事故, 非平稳环境

引用格式 何潇, 刘泽夷, 胡嵩乔, 刘畅, 周东华. 动态系统的实时安全性评估技术. 自动化学报, 2025, 51(2): 249–270

DOI 10.16383/j.aas.c240096 **CSTR** 32138.14.j.aas.c240096

Real-time Safety Assessment Techniques of Dynamic Systems

HE Xiao¹ LIU Ze-Yi¹ HU Song-Qiao¹ LIU Chang¹ ZHOU Dong-Hua^{1,2}

Abstract Real-time safety assessment (RTSA) of dynamic systems plays a critical role in preventing serious losses from potential safety incidents. As the complexity of system functionality and operational environments increases, the development of effective RTSA technologies is faced with greater challenges. In this paper, the conceptual definition of RTSA of dynamic systems is elucidated. A taxonomy framework is introduced based on two dimensions: the stationary properties of the environment and the construction methods of assessment models, along with detailed problem descriptions. Existing safety assessment technologies and discusses deployment strategies for different practical systems are systematically reviewed. We then analyze the developmental trends of current technologies and explore the pressing issues and future development directions in RTSA.

Key words Real-time safety assessment (RTSA), dynamic system, safety incidents, non-stationary environment

Citation He Xiao, Liu Ze-Yi, Hu Song-Qiao, Liu Chang, Zhou Dong-Hua. Real-time safety assessment techniques of dynamic systems. *Acta Automatica Sinica*, 2025, 51(2): 249–270

动态系统是指一类状态随着时间推移而不断变化的模型或系统^[1]. 随着现代科学技术的蓬勃发展, 各类实际系统的复杂程度也随之不断提升. 在工业 4.0 愿景下, 高度复杂的动态系统往往具有部件耦合、相互影响等特点, 对相关运行与维护技术提出更高的要求. 此类系统一旦发生事故, 可能会导致巨大的财产损失和人员伤亡. 因此, 提高动态系统的安全性对于减少重大事故的发生具有重要意义^[2]. 根据美国国防部所制订的第五代安全系统计划标

准 MIL-STD-882E, 安全性可定义为: 免除导致死亡、伤害、职业疾病、设备或财产损失, 或对环境造成损害的条件^[3]. 在我国《GJB900A-2012 装备安全性工作通用要求》中, 安全性这一概念则定义为: 不导致人员伤亡、装备损坏、财产损失或不危及人员健康和环境的能力^[4]. 该要求对系统安全性的标准进行了正式定义, 为我国系统安全性研究工作提供了指引方向. 而依据国际电工委员会 (International Electrotechnical Commission, IEC) 制作的电工百科 (Electropedia) 中针对控制技术领域的定义, 安全性 (Safety) 则描述为: 免除对所考虑的外部功能和物理单元的不可接受的风险^[5]. 而国际自动控制联合会 (International Federation of Automatic Control, IFAC) 故障检测、监督 and 安全性技术委员会所采纳的安全性定义为: 系统不对人员、设备或环境造成危险的能力^[6], 进一步突出在描述动态系统安全性时考虑系统本身与外部单位之间交互关系的重要意义.

图 1 描述了不同学术概念所对应的示意场景,

收稿日期 2024-02-26 录用日期 2024-07-23

Manuscript received February 26, 2024; accepted July 23, 2024
国家重点研发计划 (2022YFB25031103), 国家自然科学基金 (61733009), 华能集团科技研究项目 (HNKJ22-H105) 资助

Supported by National Key Research and Development Program of China (2022YFB25031103), National Natural Science Foundation of China (61733009), and Huaneng Group Science and Technology Research Project (HNKJ22-H105)

本文责任编辑 穆朝絮

Recommended by Associate Editor MU Chao-Xu

1. 清华大学自动化系 北京 100084 2. 山东科技大学电气与自动化工程学院 青岛 266590

1. Department of Automation, Tsinghua University, Beijing 100084 2. College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590

以无人机飞行一段时间后去执行任务的场景为例. 若无人机可如图 1(a) 所示在预期时间内从起飞点成功到达预期落点, 且在运行过程中对周围人、环境及财产造成潜在安全威胁的能力较弱, 则该无人机系统具有高安全性. 相比之下, 如图 1(b) 所示, 若无人机虽能按照预期轨迹在预期时间内从起飞点到达预期落点, 但在运行过程中对周围人、环境及财产造成潜在的安全威胁的能力较大 (如经过人口密集的城市, 无人机在城市上空坠落的影响极大), 则该无人机系统被视为具有低安全性. 表 1 总结了几类相关学术概念的描述. 参考以上标准, 动态系统的安全性在本文中可定义为: 动态系统在运行过程中不对人、财产及环境造成危害的能力.

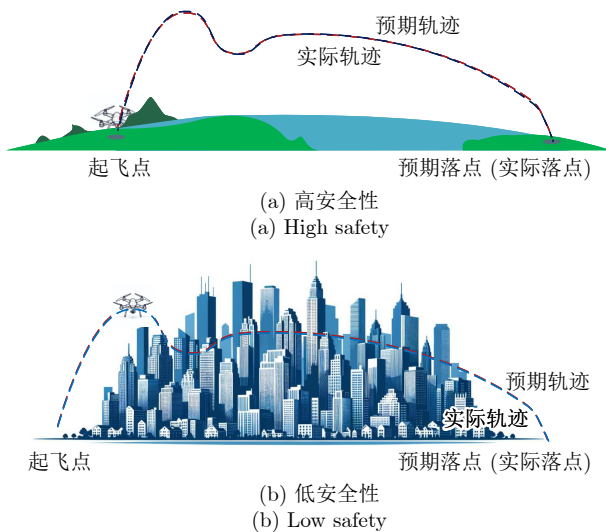


图 1 不同学术概念对应的示意场景:
以无人机飞行过程为例

Fig. 1 Illustrative scenarios corresponding to different academic concepts: A case study of UAV flight processes

自 20 世纪 40 年代提出系统安全的概念以来, 安全性评估相关问题得到国内外学者的广泛关注, 针对不同工业领域中相关问题出现了不同的技术路线^[9-14]. 自 20 世纪 50 年代起, 美国研究人员开始尝试将故障模式及影响分析 (Failure mode and effect analysis, FMEA) 方法应用于航天系统的研发工作中. 随着系统工程技术的不断发展, 该方法被推广至不同的应用领域^[15-16]. 在此基础上, 研究者又提出并发展了确定性安全性评估与以概率论为基础的 Probabilistic safety assessment (PSA) 方法^[17-19]. 从方法的视角看, 采用包含初步危险分析 (Preliminary hazard analysis, PHA)^[20-21]、故障树分析 (Fault tree analysis, FTA)^[22-23]、事件树分析 (Event tree analysis,

ETA)^[24-25]、因果分析图法及运行危险分析 (Hazard and operability analysis, HAZOP)^[26-27] 等, 试图发现系统设计的薄弱环节. 与此同时, 部分研究者尝试将人的行为和习惯所致使的潜在失误纳入考虑范围, 借助人为操作的可靠性分析技术降低安全隐患, 以此形成一套人因安全性 (Human factors in safety, HFS) 方法^[28-29]. 上述研究已在核能、航空航天、化工及医疗等领域得到广泛应用, 并在事故预防及保障维护人员安全等方面做出了举足轻重的贡献^[30-32].

基于传统概率安全性评估框架的安全性评估技术通常无法捕捉系统状态随时间变化的动态性, 需要大量的离线历史数据来估算概率和潜在事件的频率. 因此, 对于当下日趋复杂的复杂系统, 概率安全性评估方法的分析将会变得非常困难. 在过去几十年中, 基于概率框架的安全性评估方法被陆续提出. 持续性概率安全性评估 (Living probabilistic safety assessment, LPSA)^[33-35] 是其中早期的研究工作. 此类方法通过更新来反映系统的变化及当前整体系统相应的安全状态^[36]. 在此基础上, 研究者进一步提出一类基于事故序列前兆 (Accident sequence precursor, ASP) 数据的动态安全性评估 (Dynamic safety assessment, DSA)^[37-38], 可根据一定时间内获得的关于系统组件状态的知识及对危害情景的仿真结果, 更新对恶化系统风险的估计, 并依此捕捉系统风险概况的时间依赖行为¹.

然而, 设备状态反馈的实时性需求的增加、对未知危险情景认知的缺乏、以及事件关联性及级联效应增强等现实问题愈发明显^[44]. 对于具有强耦合、非线性、时变性及高不确定性特点的复杂动态系统而言, 迫切需要研究实时安全性评估技术, 将系统实时状态监测数据及对人、财产与环境的潜在影响考虑进去, 以对系统安全性状态进行及时分析与反馈. 通常而言, 动态系统的实时安全性评估技术在实时场景下利用对系统特征的在线监测, 在危险发生早期能够对安全威胁进行快速响应, 给出系统实时安全性状态的描述, 从而降低事故风险、提高生产效率, 以及保护人员、环境及财产的安全. 经典安全性评估方法与实时安全性评估方法在几个重要维度上的区别与联系如表 2 所示.

本文首先对动态系统的实时安全性评估方法进

¹ 在 Electropedia 中, 风险性定义为危害发生的概率与该危害的严重性之间的结合关系^[39], 该定义更加侧重于描述危害发生的可能性与危害的严重性之间的共同作用. 在系统工程领域, 风险分析通常被建模为危害发生的概率与危害严重性的量化之间的函数关系^[40]. 对于一些非实时的应用场景, 风险性可以被视为与安全性相似的概念^[41-43]. 本文侧重于讨论动态系统的实时安全性评估技术.

表 1 几类相关概念在国家军用系列标准下的描述及主要侧重角度

Table 1 Descriptions and key focuses of several related concepts under national military standards

概念	描述	主要侧重	参考来源
安全性	不导致人员伤亡、装备损坏、财产损失或不危及人员健康和环境的能力	系统状态、环境状态	文献 [4]
可靠性	在特定时间段内, 系统不发生故障且能够按要求执行任务的能力	系统状态	文献 [7]
测试性	及时准确地确定系统状态, 并隔离其内部故障的设计特性	系统状态	文献 [8]

表 2 经典安全性评估方法与实时安全性评估方法的概念辨析

Table 2 Conceptual analysis of classical safety assessment approaches and real-time safety assessment approaches

	经典安全性评估	实时安全性评估
先验知识需求	较多	较少
计算资源需求	较少	较多
主体适用对象	系统级	部件级
主要侧重阶段	方案设计阶段	使用保障阶段
典型应用领域	核能、航空航天	电力、工程结构
现有理论成果	较多	较少
应用价值	较高	较高

行概述, 明确其概念定义、分类框架及问题描述. 随后以环境的平稳性与评估模型构建规则的视角, 对现有实时安全性评估方法进行分类论述. 之后介绍应用部署的情况, 给出以往研究中一些典型实例, 最后指出实时安全性评估研究中亟待解决的问题和未来的发展方向.

1 实时安全性评估技术概述

1.1 概念定义

动态系统的实时安全性评估技术, 是一种在线监测和评估系统安全程度的方法, 为及时的警报与决策提供支持. 在工程实践中, 动态系统在运行过程中的状态随时间不断变化, 所在的环境也可能随时发生改变, 实时安全性评估方法的结果也可能会相应地调整. 动态系统的实时安全性评估过程通常包含以下三步:

1) 设计评估指标. 通过分析系统运行监测数据和影响安全性的关键因素之间的关系, 选取系统运行过程中可表征运行安全性的过程变量, 得出系统实时安全性评估指标集合.

2) 构建评估模型. 基于对动态系统历史运行过程中事故演化机理的认知, 综合考虑任务目标的具体要求, 建立评估指标相关的监测数据与安全性状态之间的映射关系.

3) 优化评估策略. 依据系统运行状态的实时变

化以及决策过程中的反馈信息, 实施对评估模型的持续更新和优化机制, 以确保评估结果的准确性与时效性.

现有研究中对实时故障诊断与实时安全性评估之间的区别与联系讨论较少. 尽管二者都是在系统的在线运行过程中实时得到结果, 二者侧重点和目标仍然有所不同, 其总体流程可总结如图 2 所示. 实时故障诊断技术通常利用系统的输入输出数据, 使用不同方法对故障进行检测、分离与辨识^[45]. 该技术的核心是检测和识别系统中的故障, 侧重于发现可能导致系统性能降级或功能失效的自身问题. 相较之下, 实时安全性评估技术不仅追求系统的内部健康运行, 还着重于系统对外部环境和人员潜在安全威胁的持续了解.

1.2 分类框架

随着理论研究的深入和相关领域的发展, 各种实时安全性评估方法层出不穷. 从评估模型的构建形式出发, 本文将现有动态系统实时安全性评估技术进行分类, 如图 3 所示. 其中, 显式分析是指通过制定规则来直接分析系统的实时安全性状态, 通常可得到明确的系统模型参数, 而隐式分析侧重于识别出测量数据与安全性状态之间难以直接观察到的模式. 前者相对适用于规则清晰、数据结构化、可解释性需求高的场景, 而后者则更适用于复杂程度高、数据非结构化或数据规模较大的场景.

此外, 从环境的平稳性角度出发, 可将实时安全性评估技术分为以下两个方面: 1) 平稳环境 (Stationary environment) 下的实时安全性评估技术; 2) 非平稳环境 (Non-stationary environment) 下的实时安全性评估技术. 其中平稳环境下的实时安全性评估技术假设系统运行过程中处于平稳环境, 该类技术通常适用于对系统的控制和监测具有高度可预测性的场景. 而非平稳环境下的实时安全性评估技术假设系统运行过程处于非平稳环境. 在这种情况下, 针对安全威胁的认知可能出现剧烈变化. 因此, 该类技术通常适用于对系统控制和监测具有不可预测性的情况. 对现有文献进行调研, 不同环境条件下实时安全性评估方法的研究分布情况可总结如表 3 所示.

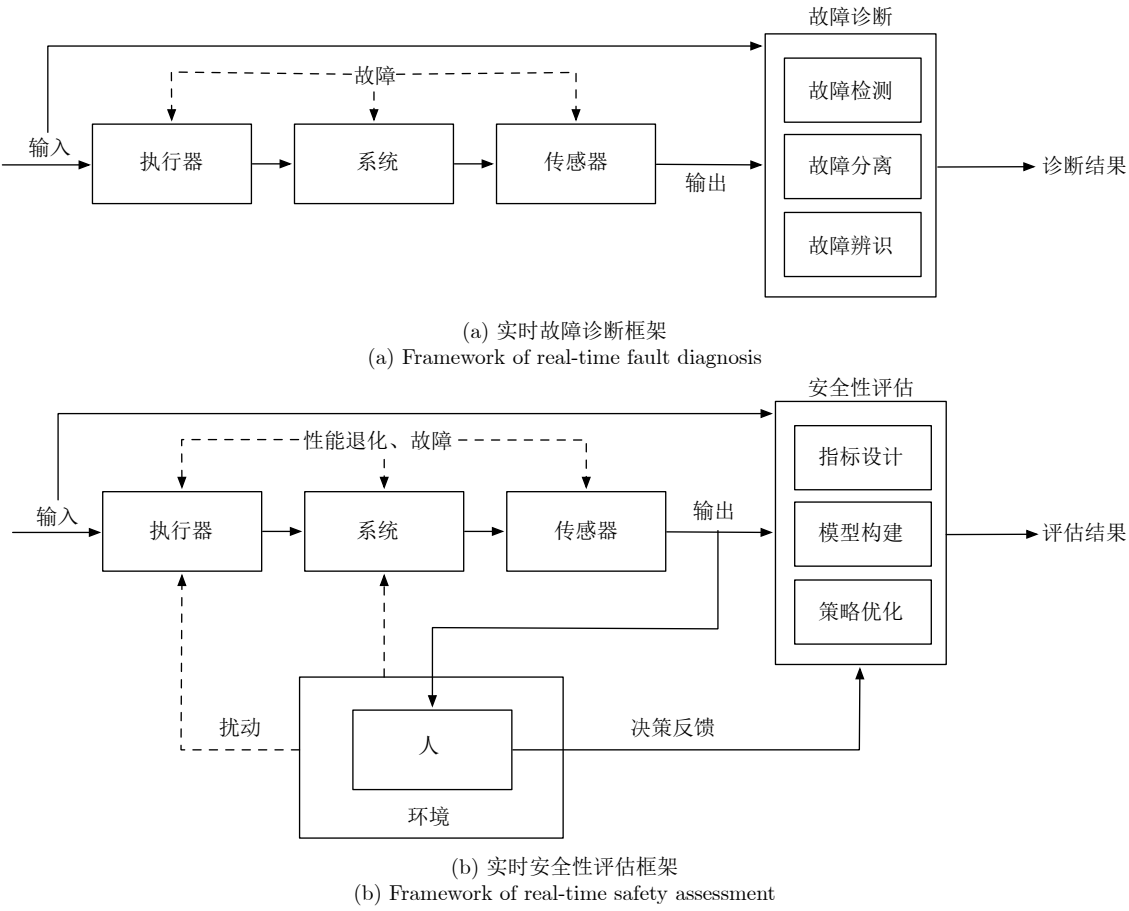


图 2 动态系统实时故障诊断与实时安全性评估示意图
Fig.2 Schematic diagram of real-time fault diagnosis and real-time safety assessment for dynamic systems

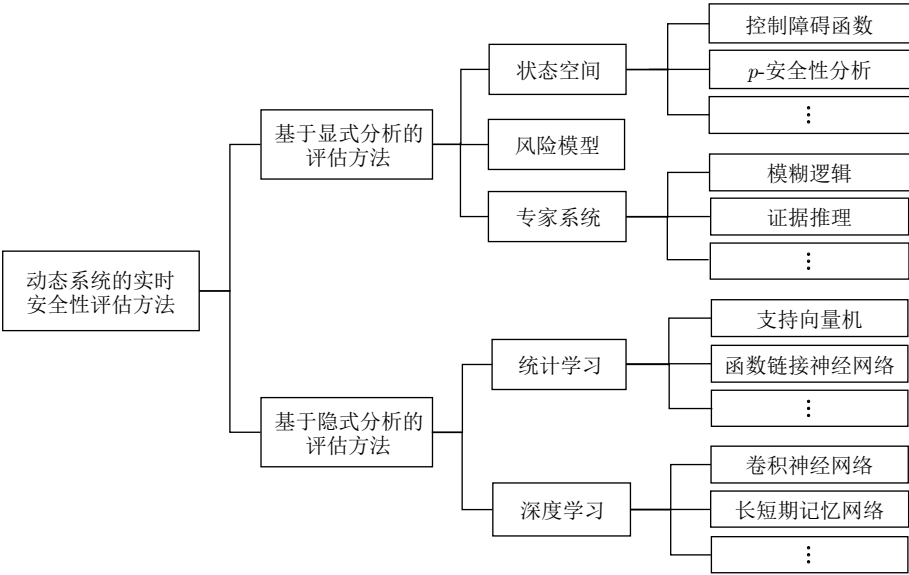


图 3 动态系统的实时安全性评估方法分类示意图
Fig.3 Schematic diagram of the taxonomy of real-time safety assessment approaches for dynamic systems

1.3 问题描述

如前文所述, 系统安全性需要同时考虑系统状 态及环境状态, 令 S 表示系统的安全性状态, 其可描述为

表 3 不同环境条件下实时安全性评估方法的研究情况

Table 3 Research of real-time safety assessment approaches under different environmental conditions

所依据的主要理论体系	平稳环境	非平稳环境
状态空间	极多	极少
风险模型	较少	较多
专家系统	较多	较少
统计学习	较多	较多
深度学习	极多	极少
信号处理	多	少

$$\mathcal{S} = (\mathcal{M}, \mathcal{I}; \mathcal{K}) \quad (1)$$

其中, $\mathcal{M}$ 表示系统测量, $\mathcal{I}$ 表示所构建的安全性指标集, $\mathcal{K}$ 表示完成安全性评估所基于的知识体系, 包含在特定条件下系统对人、环境及财产影响程度的认知, 同时外部环境的影响也可能作用于系统测量 $\mathcal{M}$. 式 (1) 表明安全性评估是以充分考虑专家知识为条件, 关于相关量测及指标的函数.

由于环境的复杂性和多样性, 实时安全性评估技术的建模问题极具挑战性. 参考如图 2(b) 所示的示意图, 实时安全性评估技术可定义如下.

定义 1. 对于一个动态系统, 存在一个安全性指标集 $\mathcal{I} = \{\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_d\}$, 其由部署在系统中的多个传感器收集得到, 通常包含用于监测系统内部的健康状况的系统状态监测传感器, 以及用于监测系统外部的环境条件的环境感知传感器.

对于安全性指标集 $\mathcal{I}$, 假设在离线阶段所收集到的历史测量数据集为 $\mathcal{M}^{\text{off}} \in \mathbf{R}^{n \times d}$, 其可表示为

$$\mathcal{M}^{\text{off}} = \begin{bmatrix} \mathcal{M}_1^{\text{off}} \\ \vdots \\ \mathcal{M}_n^{\text{off}} \end{bmatrix} = \begin{bmatrix} \mathcal{M}_{1,1}^{\text{off}} & \cdots & \mathcal{M}_{1,d}^{\text{off}} \\ \vdots & \ddots & \vdots \\ \mathcal{M}_{n,1}^{\text{off}} & \cdots & \mathcal{M}_{n,d}^{\text{off}} \end{bmatrix} \quad (2)$$

在此背景下, 可根据具体应用场景离散地得到 m 个安全性状态: $\mathcal{S}_k^{\text{off}} \in \{\mathcal{S}_1, \mathcal{S}_2, \dots, \mathcal{S}_m\}$, 或连续地划分为一定范围内的安全性程度: $\mathcal{S}_k^{\text{off}} \in [0, 1]$. 离线阶段所构建的安全性评估模型 Φ 可定义为满足以下条件的映射:

$$\Phi: \mathcal{M}^{\text{off}} \rightarrow \mathcal{S}^{\text{off}} \quad (3)$$

其可通过风险模型或专家系统等显式分析方法得到, 也可结合统计学习或深度学习 (Deep learning, DL) 等隐式分析方法得到. 安全性状态的表示形式通常取决于具体的应用场景, 在一些场景下, 回路中的专家可以根据经验直接实时给出系统的安全性状态, 此时人的因素将直接作用于评估过程; 而在另一些场景下, 可直接依据专家制定的规则来实时

分析系统的安全性状态, 此时人的因素将间接作用于评估过程.

平稳环境所刻画的是数据特性和系统状态表现出相对的稳定性和可预测性. 相比之下, 系统状态在非平稳环境中是动态变化的. 令 $(\mathcal{M}^{\text{on}}, \mathcal{S}^{\text{on}}) = \{(\mathcal{M}^{(t)}, \mathcal{S}^{(t)})\}_{t \in \mathbf{N}}$ 表示在线监测过程中的系统实时测量与实时安全性评估状态, $\theta^{(t)}$ 表示 t 时刻评估模型的实际参数. 评估模型需要依据系统的实时测量对系统的安全性实时地给出评估结果 $\hat{\mathcal{S}}^{(t)}$, 其可表示为

$$\hat{\mathcal{S}}^{(t)} = \Phi(\mathcal{M}^{(t)}; \theta^{(t)}) \quad (4)$$

令 $\mathcal{D}$ 表示描述分布之间差异性的距离度量, $\mathcal{P}(\mathcal{M}^{(t)}, \mathcal{S}^{(t)})$ 表示 t 时刻下系统测量所对应的真实分布, $|\cdot|$ 表示时间序列的长度, 可定义如下:

定义 2. 对于任意时间子序列 T_1 与 T_2 , 若满足 $|T_1|, |T_2| > \tau$ 个单位时间长度且 $\tau \in \mathbf{N}$, 在满足以下假设的情况下, 可称之为平稳环境:

$$\mathcal{D}[\mathcal{P}(\mathcal{M}^{(T_1)}, \mathcal{S}^{(T_1)}), \mathcal{P}(\mathcal{M}^{(T_2)}, \mathcal{S}^{(T_2)})] = 0 \quad (5)$$

定义 3. $\exists$ 时间子序列 T_1 与 T_2 , 若满足 $|T_1|, |T_2| > \tau$ 个单位时间长度且 $\tau \in \mathbf{N}$, 在满足以下假设的情况下, 可称之为非平稳环境:

$$\mathcal{D}[\mathcal{P}(\mathcal{M}^{(T_1)}, \mathcal{S}^{(T_1)}), \mathcal{P}(\mathcal{M}^{(T_2)}, \mathcal{S}^{(T_2)})] \neq 0 \quad (6)$$

定义 2 与定义 3 分别为平稳环境与非平稳环境的数学描述, 在现有文献中, 通常根据系统所处的实际场景对环境的平稳性进行判断, 以合理的方式来构建评估模型. 在一些应用场景中, 若系统内部退化、外部环境干扰等因素所带来的影响相对较小, 则通常可认为该环境具有平稳特征; 若上述因素不可忽视, 则通常可认为该环境具有非平稳特征. 以深海载人潜水器 (Deep-sea manned submersible, DSMS) 的实时安全性评估方法为例, 受到如水流、水压、海洋生物等因素的影响, 其在下潜过程中会遇到不同类型的安全威胁, 实际深海环境为典型的非平稳环境^[46]. 不同安全威胁所对应的安全性等级可总结如图 4 所示. 在一些情况下, 故障等异常事件可能引发安全威胁, 对人、环境及财产造成损害, 但并非所有故障都可能会导致系统具有较低的安全性, 不同异常事件可能会导致不同的安全性等级. 环境带来的安全威胁及人为误操作等异常事件, 也可能对系统安全性造成实时影响. 因此, 合理地设计安全性指标、构建评估模型及优化评估策略, 对于实时得到准确的安全性评估结果具有重要意义.

综上, 动态系统的实时安全性评估方法一般可

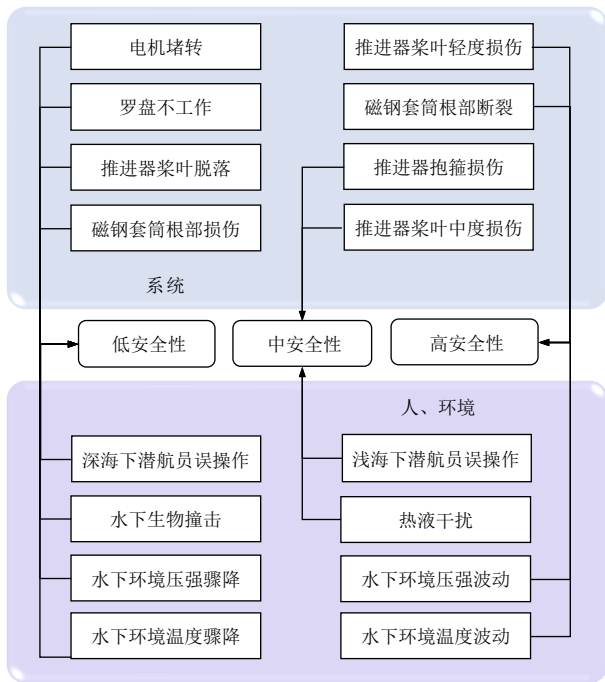


图 4 不同安全威胁对应的安全性等级示意图：
以深潜器为例

Fig.4 Schematic diagram of safety levels corresponding to different safety threats: A case study of a DSMS

描述为：通过利用特定的监测数据和专家知识，寻求合理的映射关系，并将其用于实时估计系统安全性状态的过程。平稳环境下的实时安全性评估技术侧重于在离线阶段对如式 (3) 所示的映射规则进行设计。模型可对已有的历史模式进行学习，其可在监测过程中得到较高的评估准确性。而在非平稳环境下的安全性评估技术更侧重于对适应环境的变化^[47]。通常需要频繁根据最新的数据动态，在在线阶段中调整评估模型的内部参数或结构，以完成对评估逻辑的更新与优化。

2 现有方法与进展

从环境及分析方法特点出发，本节在图 3 所示方法的基础上，对现有动态系统的实时安全性评估方法进行分类总结与讨论，其主要逻辑结构总结如图 5 所示。

2.1 平稳环境下的实时安全性评估技术

平稳环境下动态系统的实时安全性评估技术通常假设评估模型在离线阶段具有足够丰富的认识，且在线阶段相邻时刻的测量所对应的真实分布的变化是有限的。评估模型可基于此建立历史数据与安全性状态的映射关系，使其能够在实时在线监测中迅速识别异常或潜在安全威胁，并作出准确的实时

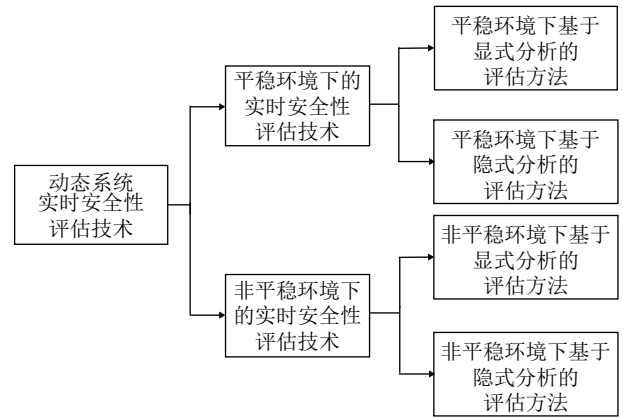


图 5 分类讨论逻辑结构示意图

Fig.5 Schematic diagram of the logical structure for categorized discussion

安全性评估。该类方法可视为忽略人类基于对环境、财产所遭受的安全威胁认知对系统及执行器的干预及调整，同时在一定程度上忽略决策反馈因素对于评估器的影响。

2.1.1 平稳环境下基于显式分析的评估方法

对于动态系统而言，在平稳环境下构建显式的安全性评估模型有助于清晰地展示如何从系统特征、参数或数据中推导出安全性评估结果，使人能够追踪和理解评估的逻辑，提高技术的可解释性，并增强评估结果的可信度。因此，这种方法在学术界受到广泛的讨论和关注。

随着近些年来控制理论技术的不断发展，一类基于状态空间方程来分析系统安全性的方法，在近几十年内引起了相关学者的重视^[48]，在一些文献中也称之为安全性验证 (Safety verification) 问题，尤其是针对故障可能会导致严重安全事故的情况，如人员伤亡、环境破坏或者重大财产损失的安全关键系统 (Safety-critical systems)^[49-51]，此类方法显得尤为重要。

对于有限状态系统的验证，Clarke 提出的模型检查 (Model checking) 技术一经提出便引起学术界的广泛重视^[52]，并促进了用于验证连续系统的类似研究^[53]。考虑到在非线系统上的适用性，Prajna 等^[54-55]参考 Lyapunov 函数的核心思想和稳定性分析理论，提出一种基于障碍函数 (Barrier functions) 的分析方法²，用于限制系统状态变量的变化范围，以确保系统的安全性得以满足。此类方法假设系统在初始时刻是安全的，将运行安全分析问题转化为对障碍函数的约束是否成立的判断问题^[56]。一些研究者将障碍函数推广为控制障碍函数 (Con-

² 在一些文献中也称作障碍验证 (Barrier certificates)。

trol barrier function, CBF), 以此将系统状态上的约束映射到控制输入上的约束^[57-58], 随后发展了指数控制障碍函数与高阶控制障碍函数 (High-order CBF, HOCBF)^[59-60]. 考虑到严格的约束易使得安全性评估结果过于敏感, 近些年来, 一些研究者试图通过寻找障碍函数的较弱约束以提高评估表现^[61-62]. Zhu 等^[63] 开发一类超球面方法, 通过构造一个超球面来覆盖不安全集 (Unsafe set), 使用代数几何方法建立障碍函数. 随后借鉴有限时间稳定性的稳定条件表达式的形式, 一种指数 α 安全准则被提出以实现动态系统运行安全的控制^[64]. 一些研究者也试图利用神经网络技术对高维、非线性系统进行推广^[10, 65-67], 但分析方法仍待进一步发展.

与此同时, 一类针对随机系统的 p -安全分析 (p -safety) 问题得到广泛讨论, 该类方法旨在从危险事件的视角出发, 通过计算到达-避免概率 (Reach-avoid probability), 对系统的初始状态进行分类, 危险项 (Hazard item) 是导致概率大于 p 的不安全区域的初始状态^[68]. Wisniewski 等^[69] 针对一般类别的马尔科夫过程完善随机避免-到达的概念, 并通过使用扩展生成器和多项式优化来近似搜索最大安全状态集, 以此解决概率安全问题来确定随机系统的安全性^[70]. Girard^[71] 提出一种基于近似双相似抽象的方法, 用于有效计算在可达性规范约束下的安全控制器. 文献^[72] 采用具有时间调度结构的多个 Lyapunov 函数来估计受到停留时间约束的切换系统的输出可达集, 基于输出可达集估计结果, 研究不确定切换系统的安全验证问题. Schürmann 等^[73] 提出一种新的安全控制方法, 将最优控制器和安全控制器结合使用. 在大多数情况下, 系统由最优控制器控制, 而安全控制器则不断监测最优控制器的操作, 在系统变得不安全时介入控制. 安全控制器利用向后可达集计算的新概念, 并开发了一种描述可达集的新方法, 可有效地最大化向后可达集的尺寸, 这些分析方法在控制器设计过程中具有重要价值.

此类方法旨在以控制理论的视角对系统安全性进行建模, 通常将人对系统安全性状态的理解表现在对状态约束或安全区域的划定上. 通常无需预先推演危害事件, 在对系统参数具有足够丰富的认知的条件下, 此类分析方法表现出较强的适用性, 可有效分析由操作失效或误操作导致系统安全状态变化的一类场景. 然而, 障碍函数需要随时间发生变化时, 这类方法的有效性受到了挑战.

另一类主流方法基于专家先验知识来构建规则库, 进而更全面地审视和分析系统的安全性³. 在现

有文献中, 诸如证据推理 (Evidential reasoning, ER) 及模糊规则 (Fuzzy rule, FR) 等融合推理技术广泛用于描述系统测量和安全性状态之间的关系^[74], 其在表示不确定性方面展现了出色的能力. Li 等^[75] 使用条件广义最小方差方法来选择关键特征, 减少了计算复杂性并提高了评估准确性, 随后提出一种用于复杂系统安全性评估的优化模型, 进一步考虑了特征间的相关性和冗余性^[76]. Tang 等^[77] 将传感器的可靠性和扰动因素引入到评估过程. Liu 等^[78] 使用专家群体交互关系对客观融合权重进行修正, 强调专家在评估过程中的影响. Zhou 等^[79] 进一步发展了评估模型可靠性的计算方法, 通过应用一阶局部敏感性方法推导出属性可靠性的敏感性分析, 以定量分析输入可靠性对模型输出和模型可靠性的影响, 在 JRC-7M 航空继电器上取得很好的效果.

这类方法在信息不完整或不精确的情况下通常具有出色的性能, 在系统处于平稳环境时, 可以有效帮助领域内专家理解评估结果的原因. 通常将安全性指标集所对应的传感器视作不同的信息源, 利用系统测量进行显式融合推理, 基于对安全性状态的置信度水平来描述系统的整体安全性. 然而, 专家知识成本昂贵、先验数据较难获取等问题在一定程度上仍较难有效解决. 对于一些具有多个交互组件和相对复杂关系的系统, 该类方法通常难以对复杂模式进行有效观测, 这限制了其在一些具体工程领域的应用价值.

2.1.2 平稳环境下基于隐式分析的评估方法

随着系统复杂性的不断增加, 安全性评估模型的显式构建变得愈发困难. 近几十年来, 针对主要应用领域的系统特性, 逐渐发展出各种不同的基于隐式构建安全性评估规则的方法体系.

一类方法通过引入统计学习模型来分析和理解系统安全性, 其通常基于历史数据对发生过的安全事件进行研究^[80]. 通过上述分析过程, 评估模型能够学习到映射规则, 进而有效地识别和评估潜在的安全威胁. 在过去数十年里, 使用决策树 (Decision tree, DT)^[81-83]、人工神经网络 (Artificial neural network, ANN)^[84] 等技术构建评估模型已经得到广泛研究. Yu 等^[85] 提出一种混合特征选择方案, 用于自动筛选健康指标, 以此提升其安全性评估效果. 文献^[86] 提出一种基于因果关系的特征选择方法, 其利用系统物理学来学习特征间的因果结构并选择特征, 从而实现更快的计算速度、更高的鲁棒性和更好的可解释性. 这类方法需要在引入专家知识描

³ 在一些文献中也称之为系统健康状态 (Health state).

述系统安全性状态的基础上,通过构建隐式模型来捕捉更加复杂的非线性关系,揭示系统测量与安全性状态之间不易察觉的模式和趋势.尽管此类方法进一步降低了对专家经验的依赖性,但也同时对安全性指标的设计方法提出更高的要求,对数据中的噪声、缺失等异常也更加敏感.

进入 21 世纪以来,随着复杂装备的规模性急剧提升,用于分析和描述安全性状态的信息和数据得到显著增长,并且这一趋势仍在持续.在此背景下,随着近年来人工智能技术的蓬勃发展,一些研究人员试图通过提升模型容量的方式来增强学习能力,进而捕捉数据中更复杂和更细微的模式和特征,以此更好地识别安全威胁.集成学习 (Ensemble learning) 框架通常被视作有效的模型结构设计方式^[87-90],以此提高评估模型的泛化性能.通常采用决策树、极限学习机 (Extreme learning machine, ELM)^[91-92]、支持向量机 (Support vector machine, SVM)^[93] 等作为基学习器. Xu 等^[91]采用集成极限学习机,在此基础上设计了决策规则. Sarmadi 等^[94]引入三个层次上的三种马氏距离度量,提出一种应用于非生成式顺序算法的集成学习方法,提高了马氏距离的表现以及损伤检测能力. Dworakowski 等^[95]采用集成 ANNs 来处理单个 ANN 初始权重随机分配的问题,提出一种自主集成设计方法,选择最适合解决特定任务的 ANN 子集.此外,基于无导数贝叶斯优化求解, Liu 等^[96]提出一种基于贝叶斯神经网络的安全约束最优功率流模型,在 IEEE 39-母线系统上说明了其有效性. He 等^[97]使用随机子空间方法离线训练多个小型决策树,并针对缺失数据设计了加权投票机制,以提高评估模型的整体鲁棒性.

随着深度学习在计算机视觉、自然语言处理和语音识别等领域取得卓越表现,其多层次、自动学习特征的能力得到广泛认可^[98-100].不少学者将深度学习引入安全性评估领域中,希望从原始数据中提取关键特征,识别复杂的模式和异常,进而提供更深层次、更全面的解决方案^[101-103].

借助深度学习处理高维度特征的能力和广泛的适应性, Alawad 等^[104]充分融合天气、轨道状况、压载条件、车站人数等数据,利用卷积神经网络 (Convolutional neural network, CNN)^[105]进行物体检测、行为分析,以评估铁路运行过程的安全性. Sun 等^[106]提出一种基于深度学习的特征提取框架,利用深度自编码器将传统状态变量数据转化至低维空间,从而降低安全性评估难度. Sarkar 等^[107]应用深度学习技术来表征复合材料上裂纹形式的损伤,在大量

实际图像数据中验证了方法的鲁棒性和准确性.尽管深度学习展现出卓越的处理高维度复杂数据的能力,其潜在挑战之一在于对数据量的高需求.为了获取丰富多样的训练数据,以捕捉数据中的抽象特征和复杂关系. Azimi 等^[108]采用迁移学习技术将类似形式的数据迁移到结构健康监测中,由此预训练卷积神经网络,进而实现在实际大规模系统内进行损伤识别和定位. Ren 等^[109]在能源系统安全性评估中引入生成对抗网络 (Generative adversarial network, GAN)^[110],利用生成器 (Generator) 产生符合历史数据分布的样本,利用鉴别器 (Discriminator) 区分生成数据和真实历史数据的差异,通过迭代训练达到平衡,用于生成能够准确替代缺失数据的新数据.除此之外,深度学习的可解释性也是限制其应用在安全性评估领域的一大挑战,这一问题也激发了众多学者的研究兴趣. Warnecke 等^[111]提出安全领域内评价解释性方法的指标,如稳定性、效率、鲁棒性,根据这些指标评价现有解释性方法在安全性评估应用中的优劣性.更具体地, Guo 等^[112]提出一种专为安全性应用设计的高保真度解释方法,其核心思想是通过扰动输入、梯度反向传播等简单的可解释模型来近似深度学习决策边界的复杂区域,在恶意软件分类器中得到了较好的应用.针对实际动态系统运行过程中可能出现的未知情景, Liu 等^[113]提出一种基于主动增量学习的在线安全性评估方法,采用具有增量学习能力的分布外检测模型,有效降低了未知情景对评估过程的负面影响.进一步地,文献 [114] 针对实际无标注数据中可能存在未知情景的挑战,提出一种主动标注辅助的半监督安全性评估模型,从而在更广泛的场景下缓解了未知情景所导致的局限性.

尽管近些年来已经出现了大量的研究成果,深度学习在安全性评估领域的研究仍然具有一些客观的局限性.深度学习模型在安全性评估中的表现显著依赖于数据的质量和数量.对于罕见异常状态,数据的缺乏可能导致模型无法准确识别或应对这些事件.此外,模型决策的不透明性可能引发最终用户的理解和信任问题.另外,此类架构通常对计算资源和内存需求较高,这些局限在资源有限的终端设备上尤其突出,在一定程度上限制了其应用范围并增加了部署难度.

2.2 非平稳环境下的实时安全性评估技术

非平稳环境是指在线监测过程中,系统特性随时间发生一定程度变化的情况.以工业系统为例,在设备磨损、环境条件变化、工艺变化、材料老化、

外部干扰等多重因素的影响下, 监测到的数据可能会呈现不规则的波动或漂移, 使得离线阶段所构建的实时安全性评估模型的性能受到挑战^[47]. 因此, 评估模型需要具有极强的适应性, 以确保对系统实时安全性进行准确评估. 该类方法需要充分考虑人类对安全威胁的认知变化, 并需要重点考虑决策反馈因素对于评估器的影响^[11].

2.2.1 非平稳环境下基于显式分析的评估方法

考虑到在非平稳环境下环境条件随着时间不断变化, 离线阶段中的系统模型参数可能不再适用, 因此需要及时更新障碍函数来保持评估的准确性和时效性^[115-116]. 此类方法的研究重点通常是对自适应控制障碍函数 (Adaptive CBF, aCBF) 的设计^[117], 通过对控制障碍函数进行自适应调整, 以反映对系统安全性状态认知的变化. 在文献^[118]中, Lopez 等引入数据驱动的思想, 提出一种具有较少保守性的鲁棒自适应控制障碍函数, 以实现系统的自适应安全性. 通过在 HOCBF 的定义中引入惩罚函数, Xiao 等^[119]提出参数自适应控制障碍函数和弛化自适应控制障碍函数, 以此进一步提高系统的安全性和适应性. 在文献^[120]中, 一种 BarrierNet 架构被进一步提出, 使得控制障碍函数的参数能够自适应环境条件的变化, 并可以从数据中学习. 尽管这类方法有助于使得系统能够适应内部状态和外部环境的变动, 建模复杂性的增加所导致的一系列问题仍然有待进一步研究. 随着障碍函数的动态调整, 对未知参数估计的难度也相应增加. 考虑到对时变动态系统进行状态估计本身所具有的难度, aCBF 更新机制的引入会使得估计过程更加复杂, 需要更高级的算法和计算资源, 同时对数学分析能力提出更高的要求.

近年来, 对融合推理技术进行在线自适应调整的必要性 and 有效性引起了广泛的关注和讨论. 这种自适应调整旨在更好地应对和管理环境带来的复杂性, 通过维护专家知识库或修正置信规则库的方式改变映射规则, 从而动态调整对系统当前所处安全性状态的判据, 以此反映对系统安全性状态的认知变化. Hu 等^[121]提出一种考虑预警功能的自适应在线安全评估方法. 赵福均等^[122]认为由于动态系统在运行过程中的运行状态不断改变, 系统状态的指标也应该随之改变, 以此综合考虑不同阶段的系统安全性特征, 提出一种基于自适应权重求取模型的证据推理方法. 在文献^[123]中, 安全性指标的可靠性也被视作提高评估的准确性的重要因素. 考虑到实际场景通常具有缺乏观测信息、系统复杂度高以及环境干扰等特点, 一种具有属性可靠性的置信规则

库也被提出以考虑属性可靠性和全局不确定性^[124]. Zhao 等^[125]提出基于最大偏差的权重系数计算方法, 为每个安全性指标分配自适应权重. 上述方法的主要思路是通过动态调整对安全性指标重要性估计的方式影响规则库. 然而, 对于一些相对复杂的动态系统, 确定各信息源的可靠性和相应权重变得更加困难, 同时由于系统中可能存在未知的相互作用和非线性关系, 动态调整推理规则的方法可能难以有效预测和应对未知的系统响应.

随着测量技术的发展, 一些研究者也开始尝试将经典动态风险分析方法的分析框架应用于处理动态系统实时安全性评估问题. 然而, 直接应用通常存在一定的局限性^[44]. 传统动态安全性评估方法只使用事故序列前兆数据 (即来自类似系统的事故的计数数据) 来更新风险指标, 该类方法必须等到事故或几乎失误 (前兆) 发生后才能更新风险指标的估计. 且由于统计数据普遍从类似系统中收集, 其结果仅能反映群体特征, 而不能充分考虑目标系统的个体特征.

在此基础上, 一类方法试图通过引入状态监测 (Condition monitoring, CM) 数据来对上述评估过程进行优化⁴, 以此提供有关目标系统个体退化过程的信息^[126-128]. 因此, 多种控制领域的先进技术被用作估计动态系统真实的退化状态, 从而使得评估结果更加真实可靠. 在现有研究中, Zadakbar 等^[129]将 Kalman 滤波器应用于上述过程来对真实的退化状态进行协同估计. 为了应对非线性和非高斯特征, Yu 等^[130]基于状态监测数据开发了一种基于自组织映射的 CMBRA 方法. 而 Wang 等^[131]通过监控过程中的多个关键变量来估计不良事件发生的概率, 并使其随着系统的实时干扰而不断更新. Zeng 等^[132]认为可联合使用状态监测数据与在类似系统运行期间发生事故的后果的统计数据, 因此提供群体信息, 而状态监测数据则可以描述系统特定动态系统的实时特征, 并视作是对 ASP 数据的有效补充.

随着故障预测与健康管理 (Prognostics and health management, PHM) 领域在近些年的蓬勃发展^[133-134], 一些研究者将 PHM 中的剩余寿命 (Remaining useful life, RUL) 预测及可靠性分析 (Reliability analysis) 等核心技术应用于实时安全性评估过程, 其可对动态系统所遭遇的潜在安全威胁进行辅助估计, 具有重要研究价值. Zhao 等^[135]利用连续时间马尔科夫链来描述退化过程, 并使用 Cox 比例风险模型以描述故障时间的危险率. Dehghani-

⁴ 该类技术在一些文献中又称之为基于状态监测的安全性评估 (Condition monitoring-based safety assessment, CMBSA) 或基于状态监测的风险评估 (Condition monitoring-based risk assessment, CMBRA)^[14].

an 等^[136] 设计了定量方法来定义高压断路器 (High-voltage circuit breakers, HVCB) 陷入每一类衰退/恢复状态的概率, 并利用高压断路器的控制电路条件监测数据对其衰退/恢复状态进行生命周期评估. Kim 等^[137] 的研究着重探讨了随着操作和工厂环境的变化而引起失效概率的变化, 旨在减少在危害时间仿真和频率估计中产生的保守性, 但未考虑监测数据中可能存在的噪声. 考虑到监测数据中的非平稳特性, BahooToroody 等^[138] 将经验模式分解和分层贝叶斯模型相结合, 根据对天然气分配系统的实时监测的数据更新超参数, 进而更新失效概率分布函数. Xing 等^[139] 将条件监测数据和检验数据 (Inspection data) 同时集成到评估过程中, 并说明了其在核能发电系统上的有效性. Ni 等^[140] 结合系统特点, 将事故发生的可能性和严重性表示成反映概率风险的指标, 并利用实际能源管理系统说明了其方法的有效性. Li 等^[141] 通过结合母线/支线和节点/断路器网络模型执行在各种条件下的动态应急分析, 计算实时传输限制, 并提供在线控制建议以减轻运营风险.

尽管此类方法试图将状态监测信息引入到评估过程中, 但绝大多数研究聚焦于将其用来对事故发生的可能性等参数进行调整, 在理论层面仍需利用风险模型的分析框架. 因此可视为是利用事故发生概率与严重性之间的结合关系, 基于系统测量对实时安全性状态进行显式分析⁵.

2.2.2 非平稳环境下基于隐式分析的评估方法

随着系统日益复杂化, 系统由更多异构元素 (硬件、人员、数字) 组成, 这些元素组织成高度互联的结构, 导致难以预测危险行为. 在非平稳环境下, 此类问题所带来的挑战显得尤其重要^[142]. 隐式分析方法能更好地处理系统中不易观察或难以量化的因素, 如人的行为、系统间的复杂相互作用以及难以预测的外部事件, 因此近年来引起了研究人员的广泛关注.

为进一步提升在线阶段评估结果的准确性, 具有增量学习 (Incremental learning, IL) 能力的方法近些年来引起越来越多的关注, 此类方法试图让评估模型在系统运行过程中, 利用所收集到的新数据进行实时更新及优化, 以及时修正评估模型的映射规则. 因此, 该类方法能够适应之前未遇到的情况, 提高对未知安全威胁的识别能力, 相对适用于状态变化频繁或需要对新情况快速响应的系统.

在 20 世纪 80 年代, Sobajic 等^[143] 就已开始探

⁵ 依据第 1.3 节所描述的定义, 使用风险模型进行评估但未使用状态监测数据的方法, 不属于本文所重点讨论的范畴.

索将自适应人工神经网络技术用于解决电力系统关键故障清除时间的预测问题, 为这一领域的发展奠定了基础. 此后, 相关技术得到进一步的发展和完善. Sun 等^[144] 提出一种利用相量测量和决策树的大规模互联电力系统安全性评估方法, 该方案离线构建并定期更新决策树, 以确定关键属性作为安全指标, 进而提升评估结果的可靠性. Diao 等^[145] 提出一种使用同步相量测量和定期更新的决策树的在线电压安全评估方案, 用以评估由电压崩溃 (Voltage collapse) 所带来的安全威胁. He 等^[146] 考虑电力系统在运行期间可能发生的运行条件变化和拓扑变化, 提出一种基于自适应集成学习的鲁棒在线安全性评估方法, 其在离线阶段使用提升算法 (Boosting algorithm) 构建评估模型, 在在线阶段中使用更新后的模型将当前运行条件下的实时系统测量映射到安全性状态上, 并在实际系统上说明了其有效性. Zhang 等^[147] 引入一种基于在线序贯极限学习机的评估方法, 在实际电力系统中进行了验证. 文献^[148] 考虑电力系统参数不确定性、参数变化以及高计算复杂性等问题, 提出一种基于高斯过程 (Gaussian process, GP) 的在线学习方法来评估电力系统的稳定性水平, 并设计了迭代算法来更新评估模型参数. Singh 等^[149] 采用基于粒子群优化 (Particle swarm optimization, PSO) 的加权平均方法来组合基学习器的预测结果, 以此利用互补信息来提高模型的泛化能力, 并在新英格兰 39 节点和 68 节点的测试电力系统中说明了这一策略的有效性. Liu 等^[150] 提出一种基于主动宽度学习系统 (Broad learning system, BLS) 的实时安全性评估方法, 以降低决策反馈所带来的成本, 并进一步考虑了有限计算内存的问题, 在蛟龙号深海载人潜水器的生命支持系统上说明了其方法的有效性^[151]. 在此基础上, 进一步提出基于子模性最大化理论的实时安全性评估方法, 重点考虑了客观存在的极度类别不平衡问题^[152]. He 等^[153] 基于动态模型解释结果, 设计了一种在线主动学习策略, 以此引入主动平稳性检测过程.

综合现有方法所表现出的特点, 可以根据多个维度将不同环境条件下实时安全性评估方法所具有的特点和不足进行总结, 如表 4 所示.

3 应用部署

3.1 典型实例

在当前研究中, 动态系统的实时安全性评估仍处于发展早期, 表 5 中列出了目前实时安全性评估方法有关的一些典型系统对象.

表 4 不同环境条件下实时安全性评估方法的研究特点
Table 4 Features of real-time safety assessment approaches under different environmental conditions

	平稳环境	非平稳环境
是否利用系统输入输出	是	是
系统特性变化	较少	较多
更新能力需求	低	高
外部因素影响程度	低	高
问题困难程度	较低	较高
实际应用范围	较小	较大
评估模型稳定性	较高	较低
决策反馈要求	低	高
现有理论成果	较多	较少

3.2 部署策略

在工程实际中, 不同的动态系统通常展现出各自独特的特征. 这些特征涵盖了系统的响应、行为、性能和特性等方面, 取决于系统的类型、用途和环境条件等多种因素. 因此, 需要强调的是, 为不同类型的动态系统制定标准化的实时安全性评估方法是不切实际的, 这意味着安全性评估方法需要针对具体系统的特性和应用场景进行针对性部署, 以确保准确性和有效性. 本节将以几种典型系统为例, 对实时安全性评估方法的部署策略进行简要讨论.

3.2.1 工程结构系统

工程结构系统通常是指包括建筑物、桥梁、管道等在内的工程和基础设施, 其安全性与人员、财产及环境的安全紧密相关. 通过研究和评估工程结构系统的安全性, 可以及早发现潜在的问题, 预防事故的发生, 具有重要的社会意义和经济意义.

在现有文献中, 一类代表性的方法称为结构化健康监测 (Structural health monitoring, SHM) 技术, 其主要目标是监测和评估各种工程结构系统 (如建筑、桥梁、管道等) 的健康状况和性能, 以确保它们在运行过程中的安全性^[199]. 结构健康监测系统

通常包括各种传感器、数据采集设备和数据分析算法, 通过实时或定期收集有关结构物的信息, 以此及时检测到结构物的异常行为、损伤或劣化^[154, 200–201], 进而分析系统安全性. 自 20 世纪末以来, 针对工程结构系统的健康监测系统开始被陆续建立^[202]. Ye 等^[167] 考虑台风期间桥梁健康监测过程中所表现出的非平稳特征, 采用经验模态分解 (Empirical mode decomposition, EMD) 等方法从原始风数据中提取出时变平均风速, 作为非平稳风模型的趋势项. Liu 等^[181] 提出一种混合模型来处理非平稳环境下风速数据的高波动性、自相关性和随机波动性难以用单一模型预测的问题. Hua 等^[168] 则研究了桥梁在车辆–桥梁相互作用下的随机非平稳特性. 近些年来, 为应对环境作用和结构响应, 研究人员建立了多种环境作用模型及动态性能监测方法^[154, 161, 203], 其通常以振动等监测信号作为系统测量^[204–205], 以进行对安全威胁的早期健康监测. 通常情况下, 系统测量数据具有规模大、复杂性高、维度高等特点, 涵盖了如位移、应力、温度等多个方面的监测信息. 如在文献^[154] 中, Li 等基于长期应变监测数据, 采用断裂力学、损伤累积理论和无损检测等技术, 开发了一种钢箱梁焊接细节的疲劳可靠性评估方法.

在此背景下, 一些学者尝试引入压缩采样 (Compressive sampling)^[206] 等技术来应对上述问题, 取得一定的进展^[155–156]. 然而, 工程结构系统的复杂性使得传统基于显式分析的评估模型较难捕捉数据中的非线性特征, 对局部微损伤敏感性欠佳. 对于一些大型工程结构系统, 一类基于损伤检测 (Damage detection) 的安全性评估方法在近些年得到广泛研究^[162–163, 207]. 随着近年计算机视觉 (Computer vision) 技术的飞速发展, 研究人员已成功将具有高泛化性的深度学习方法应用于工程结构系统表面损伤或变化的分析, 取得了显著的成果^[208]. 而在工业物联网的推动下, 深度学习方法得以更好地与大规模数据集和传感器网络结合, 为工程结构表面的实时监测提供更强大的支持^[157–158]. 例如, Azimi 等^[108] 以

表 5 几类典型动态系统在实时安全性评估框架下的现有进展
Table 5 Current advances in real-time safety assessment frameworks for several typical dynamic systems

系统对象	参考文献			
	平稳环境下显式分析法	平稳环境下隐式分析法	非平稳环境下显式分析法	非平稳环境下隐式分析法
工程结构系统	[154–158]	[74, 84–85, 94–95, 107–108, 159–163]	[126, 129, 164–166]	[167–168]
交通系统	[21, 169]	[170–172]	[173–174]	[150–153, 175]
电力系统	[176]	[80–82, 86, 88, 91–93, 96–97, 177–180]	[135–136, 140]	[143–149, 181]
核能发电系统	[19, 182–183]	[184]	[34, 36, 127, 132, 137, 139, 185]	[186]
化工系统	[187–188]	[189–191]	[44, 130–131]	[192]
航空航天系统	[79, 193–195]	[196]	[37, 128, 197]	[198]

被称为 IASC-ASCE 结构健康监测基准问题的四层四分之一比例钢框架模型为应用对象,采用加速度传感器、电传感器收集应用数据,结合迁移学习技术训练卷积神经网络,并对该应用进行健康监测,平均准确率可达到 90% 以上。

因此,对于大型工程结构系统而言,当下研究主要集中于平稳环境下基于隐式分析的实时安全性评估方法。随着应用场景的发展,相关方法已经开始朝向非平稳环境下基于隐式分析的实时安全性评估方法过渡^[164-167],在非平稳环境下对环境噪声的处理,以及更好地考虑人的决策反馈因素,具有较高的研究价值^[209-210]。

3.2.2 交通系统

另一类在实际场景下得到广泛讨论的应用对象是交通系统,其中替代安全性评估 (Surrogate safety assessment, SSA) 等方法被视为交通系统安全分析的重要组成部分^[211],其通过使用诸如冲突严重性、交通行为模式、车辆轨迹和速度等替代安全性指标来实时评估交通系统的安全性状态⁶,建立系统主动安全防护体系,对系统潜在或即将发生的风险进行感知,并从交通运营服务的稳定性和可靠性保障出发进行针对性风险响应^[212]。通过分析交通流中的潜在安全威胁,实时安全性评估方法可以在事故发生之前识别风险点,从而有助于采取预防措施。其不仅需要充分利用道路传感器、视频监控、GPS 数据等数据源,且依赖于交通模拟软件 and 数据分析平台等先进的计算模型和分析工具^[172, 213]。在该领域中,基于场景 (Scenario) 的方法是一个具有较高应用价值的研究方法^[214],其可通过确定有用的场景来降低测试的工作量。在此背景下,一些数据驱动方法已经被提出来进行场景构建,进而验证车辆安全性^[171-172]。考虑到场景选择的随机性,此类方法具有较高的时间成本。因此一些研究人员试图更侧重于通过设计关键性指标来构建关键场景^[170]。在文献 [174] 中,采用从欧洲收集的、由测试驾驶员在混合驾驶环境中进行的约 25 万公里的驾驶测试数据,利用碰撞时间 (Time-to-collision, TTC) 作为关键度量 (Criticality metrics),并通过比较其与人类驾驶员引起事故的实际距离来分析安全性。Wang 等^[173]根据物理驾驶实例,在 CarMaker 和 ROS 驱动程序中进行仿真模拟,根据车辆感知到的信息做出决策,并使用虚拟自动驾驶实例来感知周围环境,以便其了解对象的正确状态。随后将决策结果输入到触发器中,以评估行为安全性并识别关键场景。

⁶ 自 20 世纪中后期以来,道路安全评估方法 (Road safety assessment) 取得了一定的成功^[215-216],由于其在整体上依赖于对事故数据的分析,而非基于系统测量进行分析,因此不在本文中重点讨论。

综上,考虑到由人类驾驶员的社会交互和与场景的物理交互以及交通系统安全性之间的密切关联,现有交通系统领域的实时安全性评估方法更强调对安全性评估模型的可解释性^[217]。因此,通常采用的方法是基于显式分析的非平稳安全性评估方法。通过利用更先进的测量技术提升系统环境的感知能力,进而更好地分析潜在的安全威胁,具有重要的研究意义。

3.2.3 电力系统

考虑到具有多变性、多元性、规模性、互联性及非线性等特点,电力系统通常被视作是最复杂的动态系统之一。传统的电力系统安全性评估方法依赖于运行连续流和瞬态稳定性模拟^[218],随着现代电力系统变得越来越复杂,此类方法难以在线评估其安全性。在电力系统操作参数和相应的安全性状态之间进行快速和准确映射对于提升评估性能具有迫切性。因此,一类称为动态安全性评估 (DSA) 的技术引起广泛关注^[219]。此类技术在过去几十年间取得了显著的研究进展⁷,旨在通过分析对电力系统瞬态性能至关重要的属性和阈值,以评估电力系统在各种动态条件下的安全性状态^[177-178],其主要分为:瞬态稳定性 (Transient stability, TS)、电压稳定性 (Voltage stability, VS) 和频率稳定性 (Frequency stability)^[220]等细分任务,相关研究通常在 IEEE 标准系统上进行效果验证。

过去几十年内,专家系统等基于显式分析的方法取得一定的成功^[176, 221]。然而随着电力系统的复杂化,理解系统的极限变得越来越困难,这限制了其应用空间。因此,该领域的研究人员在 20 世纪末即开始尝试引入决策树、人工神经网络等统计学习技术作为解决方案,以此学习安全性指标与安全性状态之间的映射关系。例如, Liu 等^[96]为预防动态不安全性,提出基于贝叶斯神经网络的安全约束最优电力流模型,通过无推导贝叶斯优化来进行求解。该研究在 IEEE 39 总线系统中进行了应用分析,结果表明,与传统的点估计神经网络相比,贝叶斯神经网络能够正确可靠地检测电力系统的安全状态,能够避免过拟合问题。在文献 [89] 中,基于部分互信息和皮尔逊相关系数选择关键变量,进而采用迭代随机森林预测系统的暂态稳定裕度 (Transient stability margin, TSM),以此提供电力系统的实时安全性信息。该方法在 IEEE 39 总线系统中进行了应用,并提供了区域 TSM 可视化结果。文献 [222] 提出一种量化拓扑变化对评估模型准确性影响的度量方法,能够在具有高影响拓扑变化的系统中进行有

⁷ 参见引言,其与动态安全性评估技术存在本质区别。

效的安全性评估。

随着同步相量 (Synchrophasor) 等测量技术的出现与发展, 基于相量测量单元 (Phasor measurement unit, PMU) 等测量数据的安全性评估方法也随之取得了突破^[179-180, 223-224]。考虑到电力系统在出现负荷变化、市场变化及政策影响等场景下可能处于非平稳环境, 一类在线动态安全评估 (Online dynamic security assessment, ODSA) 方法在近年来用于对传统动态安全评估技术进行补充^[141, 225-226]。此类方法利用更先进的模型和仿真工具来分析电力系统的动态行为, 考虑当前运行状态或即将发生的运行状态, 试图确保系统在面临负荷变化等异常情况时仍能保持稳定运行。在此背景下, 形成了较多基于隐式分析的非平稳安全性评估方法^[146, 149, 227]。近年来, 一类由相量测量单元故障、通信丢失和相量数据集中器故障等原因所造成的数据缺失问题成为研究热点^[228], 一些研究人员尝试将半监督学习^[229]、主动学习^[230]、迁移学习^[231]等先进的技术应用于解决数据不确定性及高计算复杂度等不同的细分问题。

因此, 目前针对电力系统的实时安全性评估方法在基于隐式分析的非平稳安全性评估方面已经取得显著进展, 这一趋势显著地体现出方法论的创新和应用的深化。尽管如此, 如何在负荷变化或事故发生后, 在电力系统状态监测过程中增强人类决策者的反馈仍是一个值得深入考虑的问题, 其涉及到如何有效整合人工智能和自动化技术的分析结果与人类操作者的直觉和经验, 以应对电力系统中的环境干扰及不确定性因素, 实现更加准确和高效的决策过程。与此同时, 针对电力系统实际问题的优化, 尤其是在提高评估效率、增强系统适应性及处理复杂性方面, 仍然存在着广阔的研究空间。

3.2.4 核能发电系统

核能发电系统具有较高的复杂性, 为评估核能发电系统的安全性, 传统的安全性评估方法通常基于事件树分析框架^[9, 232], 以树形图的形式呈现系统可能发生的某种事故与导致事故发生的各种原因之间的逻辑关系, 找出事故发生的主要原因, 从而进行危险源辨识 (如再循环泵、安全阀、自动减压系统、反应堆)^[139]。然而, 事件树分析本质上是静态的, 即在运行前对风险指标进行评估, 以此分析系统运行过程中组件风险性。在日益发展的核电系统中, 这种方法难以长期准确地评估安全性能。近年来, 研究者试图将状态监测数据引入到风险模型的设计中, 以此来解决传统方法所带来的局限。一种主要的方法是动态事件树分析 (Dynamic ETA) 框架,

利用系统状态监测数据实时更新组件风险指数。在一些核电系统上的实验表明, 动态事件树可以捕获系统特定的特征, 从而提供对目标系统风险的更完整和准确的描述^[132, 139]。例如, Zeng 等^[132]以核电站中的预期瞬态无补偿切断事故 (Anticipated transient without scram accident) 为应用背景, 建立一种用于条件监测数据建模的隐马尔科夫高斯混合模型, 并利用贝叶斯网络将估计的退化状态与动态风险分析数据进行集成。利用集成数据源更新事件树, 并利用事件树评估系统安全性。该研究表明, 将两个数据源集成到动态风险分析中会比使用两个单独数据源中的任何一个提供更准确和稳健的结果。

随着人工智能技术的迅速发展和一些核电系统组件中自主控制系统的发明, 一些研究者开始尝试利用数字孪生 (Digital twin) 等人工智能技术来辅助评估核电系统的安全性^[186], 以维护该控制系统的安全。上述的方法大都聚焦于核能发电系统的内部安全问题, 然而, 一些事故的发生让人们意识到除系统内部安全隐患外, 核恐怖主义 (即系统外部因素) 的威胁也可能导致灾难性后果。于是有工作将层次分析法 (Analytic hierarchy process, AHP) 与人工神经网络相结合以研究核恐怖主义所带来的安全威胁, 输出系统安全范围^[184], 为核能发电系统的实时安全性评估方法提供更全面的视图。

综上, 针对核能发电系统的实时安全性评估方法仍然以显式分析方法为主, 如何应对非平稳环境所带来的问题已经成为新的研究趋势, 利用不断发展的人工智能技术以更深入地挖掘系统潜在的安全隐患仍有较大的研究空间。

3.2.5 化工系统

化工系统通常涉及复杂的化学过程、高压设备和危险品, 存在潜在的安全威胁。在化工系统中, 安全风险因素表 (Safety risk factor table, SRFT) 得到广泛的应用。SRFT 用于确定影响系统整体安全的因素, 如位置、能见度、财产、所有权、是否存在可作为大规模杀伤性武器前驱体的化学物质等^[186-187, 233], 一些方法根据 SRFT 对系统的安全状态进行评级。除此之外, 与其他系统类似, 一些经典的概率统计方法也广泛用于化工系统的实时安全性评估中^[131]。例如, Wang 等^[131]将事故发生的可能性建模为指数分布, 该方法通过监测过程中的多个关键变量来估计不良事件发生的概率, 并且通过考虑变量中的实时扰动使该概率不断更新。该方法在基准的田纳西-伊斯曼过程 (Tennessee-Eastman process, TEP) 中进行了测试, 并验证了其可行性。近年来, 不少研究者也引入多元线性回归、极限学习机、人工神经网络

络、支持向量机和集成技术等机器学习方法来进一步分析化工系统安全性^[190-192]。例如, Amin 等^[192]提出利用人工神经网络识别故障模式, 将贝叶斯网络用于故障检测, 利用事件树建模故障与系统安全性之间的关系。在此基础上, 采用倒置正态损失函数评估一个与时间和故障场景相关的损失, 并采用概率自适应机制来估计每个时间片中的条件概率, 从而实时调整安全性评估模型。该研究也分别在基准的 TE 过程和仿真过程中进行了验证, 展示了其适用性和灵活性。

综上所述, 当下针对化工系统的实时安全性评估主要基于显式分析方法, 但一些基于隐式分析方法的研究正取得越来越多的关注。目前, 非平稳环境下基于隐式分析的实时安全性评估方法相对较少, 通过先进的人工智能技术, 结合监测数据实时挖掘、更新数据潜在模式, 进而更快更准地评估系统安全性, 具有重要的研究价值。

3.2.6 航空航天系统

航空航天系统的事故往往伴随着严重的财产损失, 还直接关系到乘客、机组人员以及地面工作人员的生命安全。因此, 对航空航天系统进行安全性评估是十分重要且必要的^[12]。现有文献中, 针对航空航天系统的安全性评估方法仍然以传统的故障树分析^[195]、模糊逻辑、置信规则库、马尔科夫过程^[128]等传统分析框架为主, 大多属于显式分析框架。例如, 胡昌华等^[193]考虑到环境干扰下火箭监测数据存在不确定性的问题, 提出一种基于距离的监测指标不确定度求取方法, 通过指标不确定度来反映噪声对于监测数据的影响程度。进一步考虑指标的不确定度, 提出一种输入数据与置信规则库之间的匹配度转换方法, 并基于置信规则库融合不同指标的输入信息, 以此得到火箭的安全性等级。在使用监测数据对初始模型中的参数进行调整后, 蓝线能够更加准确地评估火箭的安全性状态, 在专家无法准确判断的部分监测点能够做出准确评估, 有效克服了专家知识存在的不确定性和局部无知性, 达到了监测数据对专家知识的有效补充, 实现了监测数据与专家知识的有效融合。

由于航空航天系统的特殊飞行属性, 其周围环境常常发生变化。其安全性不仅受系统内部情况所制约, 还与碰撞风险、电磁干扰、天气影响等外部环境因素息息相关, 所以只适用于平稳环境下的评估方法具有较大的局限性。随着计算机视觉技术的发展, 一类方法利用航空航天系统飞行过程中收集到的外部数据评估系统的安全性^[198]。Wang 等^[198]利用无线通信技术, 将无人机的飞行数据传输至本地电

脑中, 利用 MATLAB 平台进行数字孪生技术的仿真测试, 采用卷积神经网络来识别无人机是否侵入禁飞区域, 以此判断无人机及其周围的人或物是否存在安全威胁。

总体而言, 目前航空航天系统的实时安全性评估方法主要为显式分析框架。考虑到该系统所处的环境复杂多变, 对于非平稳环境下实时安全性评估方法的需求正在随着计算机视觉等技术的发展而不断提高^[37, 128, 198-199]。

4 总结与展望

动态系统实时安全性评估技术旨在通过高效地利用实时系统测量数据迅速识别潜在的安全威胁, 此类技术使得数据处理和响应过程变得更加及时, 从而在各类实际复杂应用场景下表现出显著的优势。本文首先对动态系统实时安全性评估技术的概念定义进行概述, 从环境的平稳性及评估模型的构建方式的角度给出一类分类框架, 并分别对问题进行具体的形式化描述。随后对现有安全性评估技术进行系统性回顾, 并围绕实际系统的部署策略进行归纳讨论。可以看到, 当前的技术研究进展已经表现出与应用系统高度相关的特点, 这一研究趋势表明评估技术与其应用领域的结合程度在整体上不断提升。各个应用领域整体上从专注于平稳环境下评估技术的研究转向对非平稳环境下评估技术的深入探索。对于系统模型较简单且对可解释性具有较高需求的场景, 研究人员往往更偏好于基于显式分析的评估方法。而对于系统模型较复杂且建模难度高的场景, 基于隐式分析的评估方法获得了更多的关注。

近年来, 随着测量技术及建模能力等方面的蓬勃发展, 对实时安全性评估技术的要求也随之提升。鉴于此, 本文对未来潜在的研究热点进行以下展望:

1) 分布式动态系统的实时安全性评估技术: 随着物联网技术在近年来的发展, 系统复杂性的提升逐步成为不可避免的未来趋势。然而, 终端设备计算资源受限、隐私保护需求高等因素使得安全性评估变得更具挑战性。安全威胁的来源趋于多元化, 其传播特性也呈现出高度的迅速性, 实际终端设备及其配置可能会频繁变化, 也进一步为有效的评估过程带来客观难度。如何同步和分析跨多个节点收集的数据, 从而减少系统停机时间和潜在损失, 具有重要的研究意义, 这要求开发新的算法和协议, 以优化资源使用, 并确保对敏感信息的保护。

2) 高数据复杂性下的实时安全性评估技术: 在实际场景下, 实时收集到的系统测量可能表征出特

征维度高、数据规模大、不确定性强等多重特点, 数据之间可能存在复杂的关联和相互依赖性, 这对安全性指标的设计及评估模型的构建都带来不可忽视的难度. 如何更有效地在高数据复杂性的条件下捕捉关键的特征信息, 并针对相关关系、因果关系、时序关系等特点对模型架构及更新策略进一步优化, 从而降低数据复杂性所带来的消极影响, 值得进一步探索与研究.

3) 具有未知情景的实时安全性评估技术: 随着系统结构的日趋复杂, 动态系统在实际运行过程中通常会不可避免地出现未知情景, 使得现有基于完备认知假设的评估方法容易出现误判现象, 进而降低评估结果准确性. 未知情景所固有的历史监测数据匮乏等特点又使其难以准确辨识, 进一步为评估模型的构建带来了困难与挑战. 因此, 如何在开放的环境下考虑对未知情景的辨识与适应, 对于推动实时安全性评估技术在工程实际上的应用, 具有重要意义.

4) 决策反馈约束下的实时安全性评估技术: 如引言中所述, 系统安全性的判断通常与人的认知水平密切相关, 受到认知差异、场景变化等因素的影响. 当下人、系统与环境之间的交互关系愈发紧密, 使得在评估过程中引入人在回路 (Human-in-the-loop) 特性成为至关重要的因素. 在此背景下, 有效应对决策反馈过程所存在的客观约束, 具有重大现实意义. 如何妥善解决由于标注受限、验证延迟等问题所带来的技术挑战, 并对潜在的误操作行为进行协同分析, 从而促进理解和预测动态系统的复杂行为, 是一个潜在的研究热点.

5) 评估模型动态可解释性的理论研究: 随着这些年来各类基于隐式分析方法的兴起, 安全性评估技术的相关研究取得显著的进展, 但由此也带来对模型可解释性的担忧, 这为大规模模型部署带来难度. 同时非平稳环境的客观特点进一步加深了对评估结果的信任问题. 因此, 如何在确保隐式分析方法捕捉非线性映射关系能力的前提下, 进一步探求针对评估模型的动态模型解释, 从而使得评估过程更加透明和易于理解, 具有重大研究价值.

6) 针对环境非平稳性的主动检测策略设计: 在多重复杂因素的共同作用下, 考虑非平稳环境特点的实时安全性评估技术已经成为当下的研究趋势. 环境对系统行为的影响不断扩大及测量技术的显著发展, 使得监测数据呈现出底层分布变化加剧、变化行为趋于复杂等现象, 评估模型对安全威胁的响应需求也进一步提高. 研究如何主动检测环境的变化行为, 基于预警信息及时对模型更新策略进行优

化, 从而提高安全性评估的准确性和适应性, 值得进一步探索.

References

- 1 Brin M, Stuck G. *Introduction to Dynamical Systems*. Cambridge: Cambridge University Press, 2002.
- 2 Zhou Dong-Hua, Hu Yan-Yan. Fault diagnosis techniques for dynamic systems. *Acta Automatica Sinica*, 2009, **35**(6): 748-758
(周东华, 胡艳艳. 动态系统的故障诊断技术. 自动化学报, 2009, **35**(6): 748-758)
- 3 Smith R E. Quantitative vs. Qualitative ESOH Risk Assessments Using the 882E Risk Matrix, MIL-STD-882E, San Diego, USA, 2012.
- 4 General Equipment Department of the Chinese People's Liberation Army. General Requirements for Materiel Safety Program, GJB 900A-2012, 2012.
(中国人民解放军总装备部. 装备安全性工作通用要求, GJB 900A-2012, 2012.)
- 5 International Electrotechnical Commission. International Electrotechnical Vocabulary Online Database (351-57-05. Safety), 2022.
- 6 Isermann R, Ballé P. Trends in the application of model-based fault detection and diagnosis of technical processes. *Control Engineering Practice*, 1997, **5**(5): 709-719
- 7 Equipment Development Department of People's Republic of China Central Military Commission. General Requirements for Equipment Reliability Work, GJB 450B-2021, 2021.
(中央军委装备发展部. 装备可靠性工作通用要求, GJB 450B-2021, 2021.)
- 8 General Equipment Department of the Chinese People's Liberation Army. General Requirements for Equipment Testing Work, GJB 2547A-2012, 2012.
(中国人民解放军总装备部. 装备测试性工作通用要求, GJB 2547A-2012, 2012.)
- 9 Aldemir T. A survey of dynamic methodologies for probabilistic safety assessment of nuclear power plants. *Annals of Nuclear Energy*, 2013, **52**: 113-124
- 10 Li B Q, Wen S P, Yan Z, Wen G H, Huang T W. A survey on the control lyapunov function and control barrier function for nonlinear-affine control systems. *IEEE/CAA Journal of Automatica Sinica*, 2023, **10**(3): 584-602
- 11 Liu Z Y, Hu S Q, He X. Real-time safety assessment of dynamic systems in non-stationary environments: A review of methods and techniques. In: Proceedings of the CAA Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS). Yibin, China: IEEE, 2023. 1-6
- 12 Chai Yi, Mao Wan-Biao, Ren Hao, Qu Jian-Feng, Yin Hong-Peng, Yang Zhi-Min, et al. Research on operational safety assessment for spacecraft launch system: Progress and challenges. *Acta Automatica Sinica*, 2019, **45**(10): 1829-1845
(柴毅, 毛万标, 任浩, 屈剑锋, 尹宏鹏, 杨志敏, 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**(10): 1829-1845)
- 13 Liu C, He X, Zhou D H, Huang B. Safety assessment for dynamic systems: A survey. *Cybernetics and Intelligence*, DOI: 10.26599/CAI.2024.9390001
- 14 Zio E. The future of risk assessment. *Reliability Engineering and System Safety*, 2018, **177**: 176-190
- 15 Stamatis D H. *Failure Mode and Effect Analysis*. Quality Press, 2003.
- 16 Liu H C, Liu L, Liu N. Risk evaluation approaches in failure

- mode and effects analysis: A literature review. *Expert Systems With Applications*, 2013, **40**(2): 828–838
- 17 Zio E. Integrated deterministic and probabilistic safety assessment: Concepts, challenges, research directions. *Nuclear Engineering and Design*, 2014, **280**: 413–419
 - 18 de Vasconcelos V, Soares W A, da Costa A C L, Raso A L. Deterministic and probabilistic safety analyses. *Advances in System Reliability Engineering*. London: Academic Press, 2019. 43–75
 - 19 Holmberg J E, Kahlbom U. Application of human reliability analysis in the deterministic safety analysis for nuclear power plants. *Reliability Engineering and System Safety*, 2020, **194**: Article No. 106371
 - 20 Rausand M. Preliminary Hazard Analysis, Norwegian University of Science and Technology, Norwegian, 2005.
 - 21 Hady-Mabrouk H. Preliminary Hazard Analysis (PHA): New hybrid approach to railway risk analysis. *International Referenced Journal of Engineering and Science*, 2017, **6**(2): 51–58
 - 22 Lee W S, Grosh D L, Tillman F A, Lie C H. Fault tree analysis, methods, and applications—A review. *IEEE Transactions on Reliability*, 1985, **R-34**(3): 194–203
 - 23 Xing L D, Amari S V. Fault tree analysis. *Handbook of Performance Engineering*. London: Springer, 2008. 595–620
 - 24 Andrews J D, Dunnett S J. Event-tree analysis using binary decision diagrams. *IEEE Transactions on Reliability*, 2000, **49**(2): 230–238
 - 25 Ferdous R, Khan F, Sadiq R, Amyotte P, Veitch B. Handling data uncertainties in event tree analysis. *Process Safety and Environmental Protection*, 2009, **87**(5): 283–292
 - 26 Dunjć J, Fthenakis V, Vílchez J A, Arnaldos J. Hazard and operability (HAZOP) analysis. A literature review. *Journal of Hazardous Materials*, 2010, **173**(1–3): 19–32
 - 27 Baybutt P. A critique of the Hazard and Operability (HAZOP) study. *Journal of Loss Prevention in the Process Industries*, 2015, **33**: 52–58
 - 28 Stranks J. *Human Factors and Behavioural Safety*. Oxford: Butterworth-Heinemann, 2007.
 - 29 Booth R T, Lee T R. The role of human factors and safety culture in safety management. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 1995, **209**(5): 393–400
 - 30 Aldemir T, Siu N O, Mosleh A, Cacciabue P C, Göktepe B G. *Reliability and Safety Assessment of Dynamic Process Systems*. Berlin: Springer, 2013. 120
 - 31 Liu Z Y, Xiao F Y. An intuitionistic evidential method for weight determination in FMEA based on belief entropy. *Entropy*, 2019, **21**(2): Article No. 211
 - 32 Zhou Jia-Hong, Xu Kai-Li, Chen Zhi-Yong. On the dynamic assessment of system safety. *Journal of Northeastern University (Natural Science)*, 2008, **29**(3): 416–419
(周家红, 许开立, 陈志勇. 系统动态安全评价研究. 东北大学学报(自然科学版), 2008, **29**(3): 416–419)
 - 33 Holmberg J, Niemelä I. Risk Measures in Living Probabilistic Safety Assessment, VTT-PUB–146, Technical Research Centre of Finland, Finland, 1993.
 - 34 Kančev D, Čepin M, Gjorgiev B. Development and application of a living probabilistic safety assessment tool: Multi-objective multi-dimensional optimization of surveillance requirements in NPPs considering their ageing. *Reliability Engineering and System Safety*, 2014, **131**: 135–147
 - 35 Čepin M. The extended living probabilistic safety assessment. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 2020, **234**(1): 183–192
 - 36 Yang J, Yang M, Wang W L, Li F J. Online application of a risk management system for risk assessment and monitoring at NPPs. *Nuclear Engineering and Design*, 2016, **305**: 200–212
 - 37 Zarei E, Azadeh A, Khakzad N, Aliabadi M M, Mohammadfam I. Dynamic safety assessment of natural gas stations using Bayesian network. *Journal of Hazardous Materials*, 2017, **321**: 830–840
 - 38 Podofilini L, Zio E, Mercurio D, Dang V N. Dynamic safety assessment: Scenario identification via a possibilistic clustering approach. *Reliability Engineering and System Safety*, 2010, **95**(5): 534–549
 - 39 International Electrotechnical Commission. International Electrotechnical Vocabulary Online Database (351-57-03, Risk), 2022.
 - 40 Vališ D. Contribution to reliability and safety assessment of systems. *Safety and Reliability*, 2007, **27**(3): 23–35
 - 41 Siu N. Risk assessment for dynamic systems: An overview. *Reliability Engineering and System Safety*, 1994, **43**(1): 43–73
 - 42 Moradi R, Groth K M. Modernizing risk assessment: A systematic integration of PRA and PHM techniques. *Reliability Engineering and System Safety*, 2020, **204**: 107194.1–107194.11
 - 43 Hollnagel E. *Safety-I and Safety-II: The Past and Future of Safety Management*. CRC Press, 2018.
 - 44 Villa V, Paltrinieri N, Khan F, Cozzani V. Towards dynamic risk analysis: A review of the risk assessment approach and its limitations in the chemical process industry. *Safety Science*, 2016, **89**: 77–93
 - 45 He Xiao, Guo Ya-Qi, Zhang Zhao, Jia Fan-Lin, Zhou Dong-Hua. Active fault diagnosis for dynamic systems. *Acta Automatica Sinica*, 2020, **46**(8): 1557–1570
(何潇, 郭亚琦, 张召, 贾繁林, 周东华. 动态系统的主动故障诊断技术. 自动化学报, 2020, **46**(8): 1557–1570)
 - 46 Hu S Q, Liu Z Y, Li M Y, He X. CADM+: Confusion-based learning framework with drift detection and adaptation for real-time safety assessment. *IEEE Transactions on Neural Networks and Learning Systems*, DOI: 10.1109/TNNLS.2024.3369315
 - 47 Ditzler G, Roveri M, Alippi C, Polikar R. Learning in non-stationary environments: A survey. *IEEE Computational Intelligence Magazine*, 2015, **10**(4): 12–25
 - 48 Ahmadi M, Israel A, Topcu U. Safety assesment based on physically-viable data-driven models. In: Proceedings of the IEEE 56th Annual Conference on Decision and Control (CDC). Melbourne, VIC, Australia: IEEE, 2017. 6409–6414
 - 49 Knight J C. Safety critical systems: Challenges and directions. In: Proceedings of the 24th International Conference on Software Engineering. Orlando, FL, USA: IEEE, 2002. 547–550
 - 50 Rausand M. *Reliability of Safety-critical Systems: Theory and Applications*. Hoboken: Wiley Publishing, 2014.
 - 51 Ames A D, Xu X R, Grizzle J W, Tabuada P. Control barrier function based quadratic programs for safety critical systems. *IEEE Transactions on Automatic Control*, 2017, **62**(8): 3861–3876
 - 52 Clarke E M. Model checking. In: Proceedings of the 17th Conference on Foundations of Software Technology and Theoretical Computer Science. Kharagpur, India: Springer, 1997. 54–56
 - 53 Alur R, Dang T, Ivančić F. Progress on reachability analysis of hybrid systems using predicate abstraction. In: Proceedings of the 6th International Workshop on Hybrid Systems: Computation and Control. Prague, Czech Republic: Springer, 2003. 4–19

- 54 Prajna S, Jadbabaie A, Pappas G J. Stochastic safety verification using barrier certificates. In: Proceedings of the 43rd IEEE Conference on Decision and Control (CDC). Nassau, Bahamas: IEEE, 2004. 929–934
- 55 Prajna S, Rantzer A. On the necessity of barrier certificates. *IFAC Proceedings Volumes*, 2005, **38**(1): 526–531
- 56 Wang G B, Liu J, Sun H Y, Liu J, Ding Z H, Zhang M M. Safety verification of state/time-driven hybrid systems using barrier certificates. In: Proceedings of the 35th Chinese Control Conference (CCC). Chengdu, China: IEEE, 2016. 2483–2489
- 57 Ames A D, Coogan S, Egerstedt M, Notomista G, Sreenath K, Tabuada P. Control barrier functions: Theory and applications. In: Proceedings of the 18th European Control Conference (ECC). Naples, Italy: IEEE, 2019. 3420–3431
- 58 Xiao W, Cassandras C G, Belta C. *Safe Autonomy With Control Barrier Functions: Theory and Applications*. Cham: Springer, 2023.
- 59 Nguyen Q, Sreenath K. Exponential control barrier functions for enforcing high relative-degree safety-critical constraints. In: Proceedings of the American Control Conference (ACC). Boston, MA, USA: IEEE, 2016. 322–328
- 60 Xiao W, Belta C. Control barrier functions for systems with high relative degree. In: Proceedings of the IEEE 58th Conference on Decision and Control (CDC). Nice, France: IEEE, 2019. 474–479
- 61 Romdlony M Z, Jayawardhana B. Stabilization with guaranteed safety using control Lyapunov-barrier function. *Automatica*, 2016, **66**: 39–47
- 62 Xu X R, Tabuada P, Grizzle J W, Ames A D. Robustness of control barrier functions for safety critical control. *IFAC-PapersOnLine*, 2015, **48**(27): 54–61
- 63 Zhu Z R, Chai Y, Yang Z M. A novel kind of sufficient conditions for safety judgement based on control barrier function. *Science China Information Sciences*, 2021, **64**(9): Article No. 199205
- 64 Zhu Z R, Chai Y, Yang Z M, Huang C H. Exponential-alpha safety criteria of a class of dynamic systems with barrier functions. *IEEE/CAA Journal of Automatica Sinica*, 2022, **9**(11): 1939–1951
- 65 Liu S M, Liu C L, Dolan J. Safe control under input limits with neural control barrier functions. In: Proceedings of the 6th Conference on Robot Learning. Auckland, New Zealand: PMLR, 2023. 1970–1980
- 66 Zhang Z Y, Zhao Q C, Sun K L. A learning-based method for computing control barrier functions of nonlinear systems with control constraints. *IEEE Robotics and Automation Letters*, 2023, **8**(7): 4259–4266
- 67 Liu S H, Liu L J, Yu Z. Safe reinforcement learning for affine nonlinear systems with state constraints and input saturation using control barrier functions. *Neurocomputing*, 2023, **518**: 562–576
- 68 Bujorianu M L, Wisniewski R, Boulougouris E. p -safety and stability. *IFAC-PapersOnLine*, 2021, **54**(9): 665–670
- 69 Wisniewski R, Bujorianu L M. Safety of stochastic systems: An analytic and computational approach. *Automatica*, 2021, **133**: Article No. 109839
- 70 Wisniewski R, Bujorianu M L, Sloth C. p -safe analysis of stochastic hybrid processes. *IEEE Transactions on Automatic Control*, 2020, **65**(12): 5220–5235
- 71 Girard A. Controller synthesis for safety and reachability via approximate bisimulation. *Automatica*, 2012, **48**(5): 947–953
- 72 Xiang W M, Tran H D, Johnson T T. Output reachable set estimation for switched linear systems and its application in safety verification. *IEEE Transactions on Automatic Control*, 2017, **62**(10): 5380–5387
- 73 Schürmann B, Klischat M, Kochdumper N, Althoff M. Formal safety net control using backward reachability analysis. *IEEE Transactions on Automatic Control*, 2022, **67**(11): 5698–5713
- 74 Zhang Yan, Zhou Wei, Cong Pei-Jiang. Fuzzy inference-based deformation prediction model for dam safety monitoring. *Hydropower Automation and Dam Monitoring*, 2009, **33**(2): 51–54 (张燕, 周伟, 丛培江. 基于模糊规则推理的大坝安全监测变形预测模型. 水电自动化与大坝监测, 2009, **33**(2): 51–54)
- 75 Li G L, Zhou Z J, Hu C H, Chang L L, Zhou Z G, Zhao F J. A new safety assessment model for complex system based on the conditional generalized minimum variance and the belief rule base. *Safety Science*, 2017, **93**: 108–120
- 76 Li G L, Zhou Z J, Hu C H, Chang L L, Zhang H T, Yu C Q. An optimal safety assessment model for complex systems considering correlation and redundancy. *International Journal of Approximate Reasoning*, 2019, **104**: 38–56
- 77 Tang S W, Zhou Z J, Hu C H, Zhao F J, Cao Y. A new evidential reasoning rule-based safety assessment method with sensor reliability for complex systems. *IEEE Transactions on Cybernetics*, 2022, **52**(5): 4027–4038
- 78 Liu Z Y, Deng Y, Zhang Y, Ding Z J, He X. Safety assessment of dynamic systems: An evidential group interaction-based fusion design. *IEEE Transactions on Instrumentation and Measurement*, 2021, **70**: Article No. 3523014
- 79 Zhou Z J, Feng Z C, Hu C H, Hu G Y, He W, Han X X. Aeronautical relay health state assessment model based on belief rule base with attribute reliability. *Knowledge-Based Systems*, 2020, **197**: Article No. 105869
- 80 Tomin N V, Kurbatsky V G, Sidorov D N, Zhukov A V. Machine learning techniques for power system security assessment. *IFAC-PapersOnLine*, 2016, **49**(27): 445–450
- 81 Wehenkel L, Pavella M. Decision tree approach to power systems security assessment. *International Journal of Electrical Power and Energy Systems*, 1993, **15**(1): 13–36
- 82 Krishnan V, McCalley J D, Henry S, Issad S. Efficient database generation for decision tree based power system security assessment. *IEEE Transactions on Power Systems*, 2011, **26**(4): 2319–2327
- 83 Hatziaargyriou N D, Contaxis G C, Sideris N C. A decision tree method for on-line steady state security assessment. *IEEE Transactions on Power Systems*, 1994, **9**(2): 1052–1061
- 84 Nazarko P, Ziemiański L. Application of artificial neural networks in the damage identification of structural elements. *Computer Assisted Methods in Engineering and Science*, 2017, **18**(3): 175–189
- 85 Yu J B. A hybrid feature selection scheme and self-organizing map model for machine health assessment. *Applied Soft Computing*, 2011, **11**(5): 4041–4054
- 86 Bellizio F, Cremer J L, Sun M Y, Strbac G. A causality based feature selection approach for data-driven dynamic security assessment. *Electric Power Systems Research*, 2021, **201**: Article No. 107537
- 87 Liu C X, Tang F, Bak C L. An accurate online dynamic security assessment scheme based on random forest. *Energies*, 2018, **11**(7): Article No. 1914
- 88 Liu S K, Liu L H, Yang N, Mao D, Zhang L, Cheng J Z, et al. A data-driven approach for online dynamic security assessment with spatial-temporal dynamic visualization using random bits forest. *International Journal of Electrical Power and*

- Energy Systems*, 2021, **124**: Article No. 106316
- 89 Liu S K, Liu L H, Fan Y P, Zhang L, Huang Y H, Zhang T, et al. An integrated scheme for online dynamic security assessment based on partial mutual information and iterated random forest. *IEEE Transactions on Smart Grid*, 2020, **11**(4): 3606–3619
 - 90 He M, Zhang J S, Vittal V. A data mining framework for online dynamic security assessment: Decision trees, boosting, and complexity analysis. In: Proceedings of the IEEE PES Innovative Smart Grid Technologies (ISGT). Washington, DC, USA: IEEE, 2012. 1–8
 - 91 Xu Y, Dong Z Y, Zhao J H, Zhang P, Wong K P. A reliable intelligent system for real-time dynamic security assessment of power systems. *IEEE Transactions on Power Systems*, 2012, **27**(3): 1253–1263
 - 92 Liu R D, Verbič G, Xu Y. A new reliability-driven intelligent system for power system dynamic security assessment. In: Proceedings of the Australasian Universities Power Engineering Conference (AUPEC). Melbourne, VIC, Australia: IEEE, 2017. 1–6
 - 93 Rizwan-ul-Hassan, Li C G, Liu Y T. Online dynamic security assessment of wind integrated power system using SDAE with SVM ensemble boosting learner. *International Journal of Electrical Power and Energy Systems*, 2021, **125**: Article No. 106429
 - 94 Sarmadi H, Entezami A, Razavi B S, Yuen K V. Ensemble learning-based structural health monitoring by Mahalanobis distance metrics. *Structural Control and Health Monitoring*, 2021, **28**(2): Article No. e2663
 - 95 Dworakowski Z, Stepinski T, Dragan K, Jablonski A, Barszcz T. Ensemble ANN classifier for structural health monitoring. In: Proceedings of the 15th International Conference on Artificial Intelligence and Soft Computing. Zakopane, Poland: Springer, 2016. 81–90
 - 96 Liu T J, Liu Y B, Liu J Y, Wang L F, Xu L X, Qiu G, et al. A Bayesian learning based scheme for online dynamic security assessment and preventive control. *IEEE Transactions on Power Systems*, 2020, **35**(5): 4088–4099
 - 97 He M, Vittal V, Zhang J S. Online dynamic security assessment with missing PMU measurements: A data mining approach. *IEEE Transactions on Power Systems*, 2013, **28**(2): 1969–1977
 - 98 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 2015, **521**(7553): 436–444
 - 99 Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience*, 2018, **2018**(1): Article No. 7068349
 - 100 Minaee S, Boykov Y, Porikli F, Plaza A, Kehtarnavaz N, Terzopoulos D. Image segmentation using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, **44**(7): 3523–3542
 - 101 Ayodeji A, Amidu M A, Olatubosun S A, Addad Y, Ahmed H. Deep learning for safety assessment of nuclear power reactors: Reliability, explainability, and research opportunities. *Progress in Nuclear Energy*, 2022, **151**: Article No. 104339
 - 102 Ye X W, Jin T, Yun C B. A review on deep learning-based structural health monitoring of civil infrastructures. *Smart Structures and Systems*, 2019, **24**(5): 567–585
 - 103 Liu C, Zhang Y, He X. Expert-augmented data-driven safety level assessment scheme with incremental learning. In: Proceedings of the CAA Symposium on Fault Detection, Supervision, and Safety for Technical Processes (SAFEPROCESS). Chengdu, China: IEEE, 2021. 1–6
 - 104 Alawad H, Kaewunruen S, An M. A deep learning approach towards railway safety risk assessment. *IEEE Access*, 2020, **8**: 102811–102832
 - 105 Li Z W, Liu F, Yang W J, Peng S H, Zhou J. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, **33**(12): 6999–7019
 - 106 Sun M Y, Konstantelos I, Strbac G. A deep learning-based feature extraction framework for system security assessment. *IEEE Transactions on Smart Grid*, 2019, **10**(5): 5007–5020
 - 107 Sarkar S, Reddy K K, Giering M, Gurvich M R. Deep learning for structural health monitoring: A damage characterization application. *Annual Conference of the PHM Society*, 2016, **8**(1): 1–7
 - 108 Azimi M, Pekcan G. Structural health monitoring using extremely compressed data through deep learning. *Computer-Aided Civil and Infrastructure Engineering*, 2020, **35**(6): 597–614
 - 109 Ren C, Xu Y. A fully data-driven method based on generative adversarial networks for power system dynamic security assessment with missing data. *IEEE Transactions on Power Systems*, 2019, **34**(6): 5044–5052
 - 110 Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath A A. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 2018, **35**(1): 53–65
 - 111 Warnecke A, Arp D, Wressnegger C, Rieck K. Evaluating explanation methods for deep learning in security. In: Proceedings of the IEEE European Symposium on Security and Privacy (EuroS&P). Genoa, Italy: IEEE, 2020. 158–174
 - 112 Guo W B, Mu D L, Xu J, Su P R, Wang G, Xing X Y. LEMNA: Explaining deep learning based security applications. In: Proceedings of the ACM SIGSAC Conference on Computer and Communications Security. Toronto, Canada: ACM, 2018. 367–379
 - 113 Liu C, Zhang Y, Ding Z J, He X. Active incremental learning for health state assessment of dynamic systems with unknown scenarios. *IEEE Transactions on Industrial Informatics*, 2023, **19**(2): 1863–1873
 - 114 Liu C, He X, Li M Y, Zhang Y, Ding Z J. Active labeling aided semi-supervised safety assessment with task-related unknown scenarios. *IEEE Transactions on Reliability*, 2024, **73**(4): 1792–1804
 - 115 Xiao W, Cassandras C G, Belta C. Adaptive control barrier functions. *Safe Autonomy With Control Barrier Functions: Theory and Applications*. Cham: Springer, 2023. 73–94
 - 116 Dhiman V, Khojasteh M J, Franceschetti M, Atanasov N. Control barriers in Bayesian learning of system dynamics. *IEEE Transactions on Automatic Control*, 2023, **68**(1): 214–229
 - 117 Taylor A J, Ames A D. Adaptive safety with control barrier functions. In: Proceedings of the American Control Conference (ACC). Denver, CO, USA: IEEE, 2020. 1399–1405
 - 118 Lopez B T, Slotine J J E, How J P. Robust adaptive control barrier functions: An adaptive and data-driven approach to safety. *IEEE Control Systems Letters*, 2021, **5**(3): 1031–1036
 - 119 Xiao W, Belta C, Cassandras C G. Adaptive control barrier functions. *IEEE Transactions on Automatic Control*, 2022, **67**(5): 2267–2281
 - 120 Xiao W, Wang T H, Hasani R, Chahine M, Amini A, Li X, et al. BarrierNet: Differentiable control barrier functions for learning of safe robot control. *IEEE Transactions on Robotics*, 2023, **39**(3): 2289–2307

- 121 Hu J Q, Zhang L B, Liang W. An adaptive online safety assessment method for mechanical system with pre-warning function. *Safety Science*, 2012, **50**(3): 385–399
- 122 Zhao Fu-Jun, Zhou Zhi-Jie, Hu Chang-Hua, Chang Lei-Lei, Wang Li. Online safety assessment method based on evidential reasoning for dynamic systems. *Acta Automatica Sinica*, 2017, **43**(11): 1950–1961
(赵福均, 周志杰, 胡昌华, 常雷雷, 王力. 基于证据推理的动态系统安全性在线评估方法. *自动化学报*, 2017, **43**(11): 1950–1961)
- 123 Zhao F J, Zhou Z J, Hu C H, Chang L L, Zhou Z G, Li G L. A new evidential reasoning-based method for online safety assessment of complex systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018, **48**(6): 954–966
- 124 Feng Z C, He W, Zhou Z J, Ban X J, Hu C H, Han X X. A new safety assessment method based on belief rule base with attribute reliability. *IEEE/CAA Journal of Automatica Sinica*, 2021, **8**(11): 1774–1785
- 125 Zhao F J, Zhou Z J, Hu C H, Cao Y, Han X X, Feng Z C. A new safety assessment method based on evidential reasoning rule with a prewarning function. *IEEE Access*, 2018, **6**: 31862–31871
- 126 Wenzel H. Monitoring based risk assessment and asset management of civil infrastructures. In: *Proceedings of the Structural Health Monitoring*. 2019.
- 127 Adumene S, Islam R, Amin T, Nitonye S, Yazdi M, Johnson K T. Advances in nuclear power system design and fault-based condition monitoring towards safety of nuclear-powered ships. *Ocean Engineering*, 2022, **251**: Article No. 111156
- 128 Compare M, Martini F, Mattafirri S, Carlevaro F, Zio E. Semi-Markov model for the oxidation degradation mechanism in gas turbine nozzles. *IEEE Transactions on Reliability*, 2016, **65**(2): 574–581
- 129 Zadakbar O, Imtiaz S, Khan F. Dynamic risk assessment and fault detection using a multivariate technique. *Process Safety Progress*, 2013, **32**(4): 365–375
- 130 Yu H Y, Khan F, Garaniya V, Ahmad A. Self-organizing map based fault diagnosis technique for non-Gaussian processes. *Industrial and Engineering Chemistry Research*, 2014, **53**(21): 8831–8843
- 131 Wang H Z, Khan F, Ahmed S, Imtiaz S. Dynamic quantitative operational risk assessment of chemical processes. *Chemical Engineering Science*, 2016, **142**: 62–78
- 132 Zeng Z G, Zio E. Dynamic risk assessment based on statistical failure data and condition-monitoring degradation data. *IEEE Transactions on Reliability*, 2018, **67**(2): 609–622
- 133 Zio E. Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice. *Reliability Engineering and System Safety*, 2022, **218**: Article No. 108119
- 134 Hu Y, Miao X W, Si Y, Pan E S, Zio E. Prognostics and health management: A review from the perspectives of design, development and decision. *Reliability Engineering and System Safety*, 2022, **217**: Article No. 108063
- 135 Zhao S, Makis V, Chen S W, Li Y. Health assessment method for electronic components subject to condition monitoring and hard failure. *IEEE Transactions on Instrumentation and Measurement*, 2019, **68**(1): 138–150
- 136 Dehghanian P, Guan Y F, Kezunovic M. Real-time life-cycle assessment of high-voltage circuit breakers for maintenance using online condition monitoring data. *IEEE Transactions on Industry Applications*, 2019, **55**(2): 1135–1146
- 137 Kim H, Lee S H, Park J S, Kim H, Chang Y S, Heo G. Reliability data update using condition monitoring and prognostics in probabilistic safety assessment. *Nuclear Engineering and Technology*, 2015, **47**(2): 204–211
- 138 BahooToroody A, Abaei M M, BahooToroody F, de Carlo F, Abbassi R, Khalaj S. A condition monitoring based signal filtering approach for dynamic time dependent safety assessment of natural gas distribution process. *Process Safety and Environmental Protection*, 2019, **123**: 335–343
- 139 Xing J D, Zeng Z G, Zio E. A framework for dynamic risk assessment with condition monitoring data and inspection data. *Reliability Engineering and System Safety*, 2019, **191**: Article No. 106552
- 140 Ni M, McCalley J D, Vittal V, Tayyib T. Online risk-based security assessment. *IEEE Transactions on Power Systems*, 2003, **18**(1): 258–265
- 141 Li H F, Diao R S, Zhang X H, Lin X, Lu X, Shi D, et al. An integrated online dynamic security assessment system for improved situational awareness and economic operation. *IEEE Access*, 2019, **7**: 162571–162582
- 142 Tchernykh A, Babenko M, Chervyakov N, Miranda-López V, Avetisyan A, Drozdov A Y, et al. Scalable data storage design for nonstationary IoT environment with adaptive security and reliability. *IEEE Internet of Things Journal*, 2020, **7**(10): 10171–10188
- 143 Sobajic D J, Pao Y H. Artificial neural-net based dynamic security assessment for electric power systems. *IEEE Transactions on Power Systems*, 1989, **4**(1): 220–228
- 144 Sun K, Likhate S, Vittal V, Kolluri V S, Mandal S. An online dynamic security assessment scheme using phasor measurements and decision trees. *IEEE Transactions on Power Systems*, 2007, **22**(4): 1935–1943
- 145 Diao R S, Sun K, Vittal V, O’Keefe R J, Richardson M R, Bhatt N, et al. Decision tree-based online voltage security assessment using PMU measurements. *IEEE Transactions on Power Systems*, 2009, **24**(2): 832–839
- 146 He M, Zhang J S, Vittal V. Robust online dynamic security assessment using adaptive ensemble decision-tree learning. *IEEE Transactions on Power Systems*, 2013, **28**(4): 4089–4098
- 147 Zhang R, Xu Y. Data-driven dynamic security assessment and control of power systems: An online sequential learning method. *Journal of Energy Engineering*, 2019, **145**(5): Article No. 04019019
- 148 Zhai C, Nguyen H D, Zong X F. Dynamic security assessment of small-signal stability for power grids using windowed online Gaussian process. *IEEE Transactions on Automation Science and Engineering*, 2023, **20**(2): 1170–1179
- 149 Singh M, Chauhan S. A hybrid-extreme learning machine based ensemble method for online dynamic security assessment of power systems. *Electric Power Systems Research*, 2023, **214**: Article No. 108923
- 150 Liu Z Y, Zhang Y, Ding Z J, He X. An online active broad learning approach for real-time safety assessment of dynamic systems in nonstationary environments. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, **34**(10): 6714–6724
- 151 Liu Z Y, He X. Real-time safety assessment for dynamic systems with limited memory and annotations. *IEEE Transactions on Intelligent Transportation Systems*, 2023, **24**(9): 10076–10086
- 152 Liu Z Y, He X. Dynamic submodular-based learning strategy in imbalanced drifting streams for real-time safety assessment in nonstationary environments. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, **35**(3): 3038–3051
- 153 He X, Liu Z Y. Dynamic model interpretation-guided online

- active learning scheme for real-time safety assessment. *IEEE Transactions on Cybernetics*, 2024, **54**(5): 2734–2745
- 154 Li A Q, Ding Y L, Wang H, Guo T. Analysis and assessment of bridge health monitoring mass data—Progress in research/development of “Structural Health Monitoring”. *Science China Technological Sciences*, 2012, **55**(8): 2212–2224
 - 155 Bao Y Q, Beck J L, Li H. Compressive sampling for accelerometer signals in structural health monitoring. *Structural Health Monitoring*, 2011, **10**(3): 235–246
 - 156 Bao Y Q, Tang Z Y, Li H. Compressive-sensing data reconstruction for structural health monitoring: A machine-learning approach. *Structural Health Monitoring*, 2020, **19**(1): 293–304
 - 157 Harshitha C, Alapati M, Chikkakrishna N K. Damage detection of structural members using internet of things (IoT) paradigm. *Materials Today: Proceedings*, 2021, **43**: 2337–2341
 - 158 Abdelgawad A, Yelamarthi K. Internet of things (IoT) platform for structure health monitoring. *Wireless Communications and Mobile Computing*, 2017, **2017**(1): Article No. 6560797
 - 159 Goulet J A, Michel C, der Kiureghian A. Data-driven post-earthquake rapid structural safety assessment. *Earthquake Engineering and Structural Dynamics*, 2015, **44**(4): 549–562
 - 160 Catelani M, Ciani L, Galar D, Patrizi G. Optimizing maintenance policies for a yaw system using reliability-centered maintenance and data-driven condition monitoring. *IEEE Transactions on Instrumentation and Measurement*, 2020, **69**(9): 6241–6249
 - 161 Nyman J, Rosengren P, Kool P, Karoumi R, Leander J, Petursson H. Smart condition monitoring of a steel bascule railway bridge. *Life-Cycle of Structures and Infrastructure Systems*. London: CRC Press, 2023. 229–236
 - 162 Bandara R P, Chan T H T, Thambiratnam D P. Structural damage detection method using frequency response functions. *Structural Health Monitoring*, 2014, **13**(4): 418–429
 - 163 Avci O, Abdeljaber O, Kiranyaz S, Inman D. Structural damage detection in real time: Implementation of 1D convolutional neural networks for SHM applications. *Structural Health Monitoring and Damage Detection, Volume 7*. Cham: Springer, 2017. 49–54
 - 164 Entezami A, Shariatmadar H. Structural health monitoring by a new hybrid feature extraction and dynamic time warping methods under ambient vibration and non-stationary signals. *Measurement*, 2019, **134**: 548–568
 - 165 Avendano-Valencia L D, Spiridonakos M D, Fassois S D. In-operation identification of a wind turbine structure via non-stationary parametric models. In: Proceedings of the 8th International Workshop on Structural Health Monitoring. Stanford, CA, USA: Stanford University, 2011. Article No. 2611
 - 166 Xu C, Ni Y Q, Wang Y W. A novel Bayesian blind source separation approach for extracting non-stationary and discontinuous components from structural health monitoring data. *Engineering Structures*, 2022, **269**: Article No. 114837
 - 167 Ye X W, Xi P S, Su Y H. Analysis of non-stationary wind characteristics at an arch bridge using structural health monitoring data. *Journal of Civil Structural Health Monitoring*, 2017, **7**(4): 573–587
 - 168 Hua X, Xiao F, Chen G S, Zatar W, Hulsey L. Stochastic non-stationary characteristics of vehicle-induced bridge vibrations. *Journal of Low Frequency Noise, Vibration and Active Control*, 2023, **42**(2): 759–770
 - 169 Klischat M, Althoff M. Generating critical test scenarios for automated vehicles with evolutionary algorithms. In: Proceedings of the IEEE Intelligent Vehicles Symposium (IV). Paris, France: IEEE, 2019. 2352–2358
 - 170 Feng S, Sun H W, Yan X T, Zhu H J, Zou Z X, Shen S Y, et al. Dense reinforcement learning for safety validation of autonomous vehicles. *Nature*, 2023, **615**(7953): 620–627
 - 171 Krajewski R, Moers T, Nerger D, Eckstein L. Data-driven maneuver modeling using generative adversarial networks and variational autoencoders for safety validation of highly automated vehicles. In: Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITSC). Maui, HI, USA: IEEE, 2018. 2383–2390
 - 172 Jenkins I R, Gee L O, Knauss A, Yin H, Schroeder J. Accident scenario generation with recurrent neural networks. In: Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITSC). Maui, HI, USA: IEEE, 2018. 3340–3345
 - 173 Wang C, Storms K, Winner H. Online safety assessment of automated vehicles using silent testing. *IEEE Transactions on Intelligent Transportation Systems*, 2022, **23**(8): 13069–13083
 - 174 Åsljung D, Nilsson J, Fredriksson J. Using extreme value theory for vehicle level safety validation and implications for autonomous vehicles. *IEEE Transactions on Intelligent Vehicles*, 2017, **2**(4): 288–297
 - 175 Liu Z Y, He X. Contrastive preference-guided active learning approach based on ranking correlation for real-time safety assessment. *IEEE Transactions on Automation Science and Engineering*, DOI: 10.1109/TASE.2024.3401470
 - 176 Fouad A A, Vekataraman S, Davis J A. An expert system for security trend analysis of a stability-limited power system. *IEEE Transactions on Power Systems*, 1991, **6**(3): 1077–1084
 - 177 Wehenkel L, van Cutsem T, Ribbens-Pavella M. An artificial intelligence framework for online transient stability assessment of power systems. *IEEE Transactions on Power Systems*, 1989, **4**(2): 789–800
 - 178 Rovnyak S, Kretsinger S, Thorp J, Brown D. Decision trees for real-time transient stability prediction. *IEEE Transactions on Power Systems*, 1994, **9**(3): 1417–1426
 - 179 Kamwa I, Grondin R, Loud L. Time-varying contingency screening for dynamic security assessment using intelligent-systems techniques. *IEEE Transactions on Power Systems*, 2001, **16**(3): 526–536
 - 180 Diao R S, Vittal V, Logic N. Design of a real-time security assessment tool for situational awareness enhancement in modern power systems. *IEEE Transactions on Power Systems*, 2010, **25**(2): 957–965
 - 181 Liu D, Niu D X, Wang H, Fan L L. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. *Renewable Energy*, 2014, **62**: 592–597
 - 182 Sui Y, Ding R, Wang H Q. A novel approach for occupational health and safety and environment risk assessment for nuclear power plant construction project. *Journal of Cleaner Production*, 2020, **258**: Article No. 120945
 - 183 Shin J, Son H, Heo G. Cyber security risk evaluation of a nuclear I&C using BN and ET. *Nuclear Engineering and Technology*, 2017, **49**(3): 517–524
 - 184 Jang K B, Baek C H, Woo T H. Assessment for nuclear security using analytic hierarchy process (AHP) incorporated with neural networking method in nuclear power plants (NPPs). *Kerntechnik*, 2022, **87**(5): 607–614
 - 185 Cohn B, Noel T, Cardoni J, Haskin T, Osborn D, Aldemir T. Integrated safety and security analysis of nuclear power plants using dynamic event trees. *Nuclear Science and Engineering*, 2023, **197**(sup1): S45–S56

- 186 Yockey P, Erickson A, Spirito C. Cyber threat assessment of machine learning driven autonomous control systems of nuclear power plants. *Progress in Nuclear Energy*, 2023, **166**: Article No. 104960
- 187 Bajpai S, Sachdeva A, Gupta J P. Security risk assessment: Applying the concepts of fuzzy logic. *Journal of Hazardous Materials*, 2010, **173**(1-3): 258-264
- 188 Zhou J F, Reniers G, Zhang L B. A weighted fuzzy Petri-net based approach for security risk assessment in the chemical industry. *Chemical Engineering Science*, 2017, **174**: 136-145
- 189 Peng T, Li C, Zhou X B. Application of machine learning to laboratory safety management assessment. *Safety Science*, 2019, **120**: 263-267
- 190 Gao Y C, Zhang J C, Cui S X, Wu Y Q, Huang M L, Zhuang S L. Machine learning-based QSAR for safety evaluation of environmental chemicals. *QSAR in Safety Evaluation and Risk Assessment*. Academic Press, 2024. 89-99
- 191 Wang Z H, Wen H Q, Su Y, Shen W F, Ren J Z, Ma Y J, et al. Insights into ensemble learning-based data-driven model for safety-related property of chemical substances. *Chemical Engineering Science*, 2022, **248**: Article No. 117219
- 192 Amin T, Khan F. Dynamic process safety assessment using adaptive Bayesian network with loss function. *Industrial and Engineering Chemistry Research*, 2022, **61**(45): 16799-16814
- 193 Hu Chang-Hua, Feng Zhi-Chao, Zhou Zhi-Jie, Hu Guan-Yu, He Wei, Cao You. A safety assessment method for a liquid launch rocket based on the belief rule base with environmental disturbance. *Scientia Sinica Informationis*, 2020, **50**(10): 1559-1573
(胡昌华, 冯志超, 周志杰, 胡冠宇, 贺维, 曹友. 考虑环境干扰的液体运载火箭结构安全性评估方法. 中国科学: 信息科学, 2020, **50**(10): 1559-1573)
- 194 Li Q Y, Wu Q G, Tu H Y, Zhang J P, Zou X, Huang S. Ground risk assessment for unmanned aircraft focusing on multiple risk sources in urban environments. *Processes*, 2023, **11**(2): Article No. 542
- 195 Tabassum A, Sabatini R, Gardi A. Probabilistic safety assessment for UAS separation assurance and collision avoidance systems. *Aerospace*, 2019, **6**(2): Article No. 19
- 196 Jiao R H, Peng K X, Zhang K, Ma L, Pi Y T. A novel scheme for remaining useful life prediction and safety assessment based on hybrid method. In: Proceedings of the CAA Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS). Xiamen, China: IEEE, 2019. 395-400
- 197 Dheedan A A. On-line safety monitor based on a safety assessment model and hierarchical deployment of a multi-agent system. *International Journal on Advances in Internet Technology*, 2012, **5**(3-4): 95-113
- 198 Wang W X, Li X M, Xie L F, Lv H B, Lv Z H. Unmanned aircraft system airspace structure and safety measures based on spatial digital twins. *IEEE Transactions on Intelligent Transportation Systems*, 2022, **23**(3): 2809-2818
- 199 Farrar C R, Worden K. An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2007, **365**(1851): 303-315
- 200 Gordan M, Sabbagh-Yazdi S R, Ismail Z, Ghaedi K, Carroll P, McCrum D, et al. State-of-the-art review on advancements of data mining in structural health monitoring. *Measurement*, 2022, **193**: Article No. 110939
- 201 Cury A, Ribeiro D, Ubertini F, Todd M D. *Structural Health Monitoring Based on Data Science Techniques*. Cham: Springer, 2022.
- 202 Ko J M, Ni Y Q. Technology developments in structural health monitoring of large-scale bridges. *Engineering Structures*, 2005, **27**(12): 1715-1725
- 203 He Z G, Li W T, Salehi H, Zhang H, Zhou H Y, Jiao P C. Integrated structural health monitoring in bridge engineering. *Automation in Construction*, 2022, **136**: Article No. 104168
- 204 Niyirora R, Ji W, Masengesho E, Munyaneza J, Niyonyungu F, Nyirandayisabye R. Intelligent damage diagnosis in bridges using vibration-based monitoring approaches and machine learning: A systematic review. *Results in Engineering*, 2022, **16**: Article No. 100761
- 205 Kaartinen E, Dumphly K, Sadhu A. LiDAR-based structural health monitoring: Applications in civil infrastructure systems. *Sensors*, 2022, **22**(12): Article No. 4610
- 206 Donoho D L. Compressed sensing. *IEEE Transactions on Information Theory*, 2006, **52**(4): 1289-1306
- 207 Stepinski T, Uhl T, Staszewski W. *Advanced Structural Damage Detection: From Theory to Engineering Applications*. West Sussex: John Wiley & Sons, 2013.
- 208 Feng D M, Feng M Q. Computer vision for SHM of civil infrastructure: From dynamic response measurement to damage detection—A review. *Engineering Structures*, 2018, **156**: 105-117
- 209 Worden K, Baldacchino T, Rowson J, Cross E J. Some recent developments in SHM based on nonstationary time series analysis. *Proceedings of the IEEE*, 2016, **104**(8): 1589-1603
- 210 Worden K, Iakovidis I, Cross E J. New results for the ADF statistic in nonstationary signal analysis with a view towards structural health monitoring. *Mechanical Systems and Signal Processing*, 2021, **146**: Article No. 106979
- 211 Tarko A P. Surrogate measures of safety. *Safe Mobility: Challenges, Methodology and Solutions*. Leeds: Emerald Publishing Limited, 2018. 383-405
- 212 Ning Bin. A number of scientific and technical problems in intelligent transportation. *Scientia Sinica Informationis*, 2018, **48**(9): 1264-1269
(宁滨. 智能交通中的若干科学和技术问题. 中国科学: 信息科学, 2018, **48**(9): 1264-1269)
- 213 Arun A, Haque M, Bhaskar A, Washington S, Sayed T. A systematic mapping review of surrogate safety assessment using traffic conflict techniques. *Accident Analysis and Prevention*, 2021, **153**: Article No. 106016
- 214 Riedmaier S, Ponn T, Ludwig D, Schick B, Diermeyer F. Survey on scenario-based safety assessment of automated vehicles. *IEEE Access*, 2020, **8**: 87456-87477
- 215 Rassafi A A, Ganji S S, Pourkhani H. Road safety assessment under uncertainty using a multi attribute decision analysis based on Dempster-Shafer theory. *KSCE Journal of Civil Engineering*, 2018, **22**(8): 3137-3152
- 216 de Leur P, Sayed T. Development of a road safety risk index. *Transportation Research Record: Journal of the Transportation Research Board*, 2002, **1784**(1): 33-42
- 217 Wang W S, Wang L T, Zhang C Y, Liu C L, Sun L J. Social interactions for autonomous driving: A review and perspectives. *Foundations and Trends® in Robotics*, 2022, **10**(3-4): 198-376
- 218 Morison K, Wang L, Kundur P. Power system security assessment. *IEEE Power and Energy Magazine*, 2004, **2**(5): 30-39
- 219 Grigsby L L. Dynamic security assessment. *Power System Stability and Control*. Boca Raton: CRC Press, 2007. 421-430
- 220 Alimi O A, Ouahada K, Abu-Mahfouz A M. A review of machine learning approaches to power system security and stability. *IEEE Access*, 2020, **8**: 113512-113531

- 221 Fouad A A, Vittal V. *Power System Transient Stability Analysis Using the Transient Energy Function Method*. Englewood: Pearson, 1992.
- 222 Bellizio F, Cremer J L, Strbac G. Machine-learned security assessment for changing system topologies. *International Journal of Electrical Power and Energy Systems*, 2022, **134**: Article No. 107380
- 223 Li Q Q, Xu Y, Ren C, Zhao J H. A hybrid data-driven method for online power system dynamic security assessment with incomplete PMU measurements. In: *Proceedings of the IEEE Power and Energy Society General Meeting (PESGM)*. Atlanta, GA, USA: IEEE, 2019. 1–5
- 224 Makarov Y V, Du P W, Lu S, Nguyen T B, Guo X X, Burns J W, et al. PMU-based wide-area security assessment: Concept, method, and implementation. *IEEE Transactions on Smart Grid*, 2012, **3**(3): 1325–1332
- 225 Jardim J L. Online dynamic security assessment. *Real-Time Stability in Power Systems: Techniques for Early Detection of the Risk of Blackout*. Cham: Springer, 2014. 159–197
- 226 Vaahedi E, Mansour Y, Tse E K. A general purpose method for on-line dynamic security assessment. *IEEE Transactions on Power Systems*, 1998, **13**(1): 243–249
- 227 Zhang Y, Xie L. Online dynamic security assessment of microgrid interconnections in smart distribution systems. *IEEE Transactions on Power Systems*, 2015, **30**(6): 3246–3254
- 228 Zhang Y C, Xu Y, Bu S Q, Dong Z, Zhang R. Online power system dynamic security assessment with incomplete PMU measurements: A robust white-box model. *IET Generation, Transmission and Distribution*, 2019, **13**(5): 662–668
- 229 Liu R D, Verbič G, Ma J. A new dynamic security assessment framework based on semi-supervised learning and data editing. *Electric Power Systems Research*, 2019, **172**: 221–229
- 230 Zhang Y Q, Zhao Q, Tan B D, Yang J. A power system transient stability assessment method based on active learning. *The Journal of Engineering*, 2021, **2021**(11): 715–723
- 231 Ren C, Xu Y. Transfer learning-based power system online dynamic security assessment: Using one model to assess many unlearned faults. *IEEE Transactions on Power Systems*, 2020, **35**(1): 821–824
- 232 Čepin M. Event tree analysis. *Assessment of Power System Reliability: Methods and Applications*. London: Springer, 2011. 89–99
- 233 Bajpai S, Gupta J P. Site security for chemical process industries. *Journal of Loss Prevention in the Process Industries*, 2005, **18**(4–6): 301–309



何 潇 清华大学自动化系长聘教授. 2010 年获得清华大学博士学位. 主要研究方向为动态系统的故障诊断与容错控制. 本文通信作者.

E-mail: hexiao@tsinghua.edu.cn

(HE Xiao Tenured professor in the Department of Automation, Tsinghua University. He received his Ph.D. degree from

Tsinghua University in 2010. His research interest covers fault diagnosis and fault-tolerant control for dynamic systems. Corresponding author of this paper.)



刘泽夷 清华大学自动化系博士研究生. 主要研究方向为动态系统的故障诊断与安全性评估.

E-mail: liuzy21@mails.tsinghua.edu.cn

(LIU Ze-Yi Ph.D. candidate in the Department of Automation, Tsinghua University. His research in-

terest covers fault diagnosis and safety assessment for dynamic systems.)



胡嵩乔 清华大学自动化系博士研究生. 主要研究方向为动态系统的故障诊断与安全性评估.

E-mail: hsq23@mails.tsinghua.edu.cn

(HU Song-Qiao Ph.D. candidate in the Department of Automation, Tsinghua University. His research

interest covers fault diagnosis and safety assessment for dynamic systems.)



刘 畅 清华大学自动化系博士研究生. 主要研究方向为动态系统的故障诊断与安全性评估.

E-mail: liuc19@mails.tsinghua.edu.cn

(LIU Chang Ph.D. candidate in the Department of Automation, Tsinghua University. His research

interest covers fault diagnosis and safety assessment for dynamic systems.)



周东华 山东科技大学和清华大学教授. 主要研究方向为动态系统的故障诊断与容错控制, 故障预测与最优维护技术.

E-mail: zdh@mail.tsinghua.edu.cn

(ZHOU Dong-Hua Professor at Shandong University of Science and

Technology and Tsinghua University. His research interest covers fault diagnosis and fault-tolerant control, fault prediction, and optimal maintenance.)