电 子 科 技 大 学

UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA

专业学位硕士学位论文

MASTER THESIS FOR PROFESSIONAL DEGREE

论文题目 **基于多源信息融合的 CPS 安全态势要素提取技术研究**

专业学位类别	电子信息
学　　号	202022010520
作者姓名	占勇新
指导教师	段景山　（正）高级实验师
学　　院	信息与通信工程学院

分类号 _____ 密级 _____

UDC [注1] _____

学 位 论 文

基于多源信息融合的 CPS 安全态势要素提取技术研究

（题名和副题名）

占勇新

（作者姓名）

指导教师 段景山 （正）高级实验师

电子科技大学 成 都

（姓名、职称、单位名称）

申请学位级别 硕士 专业学位类别 电子信息

提交论文日期 2023 年 5 月 26 日 论文答辩日期 2023 年 6 月 1 日

学位授予单位和日期 电子科技大学 2023 年 6 月

答辩委员会主席 章小宁

评阅人 潘晔、于富财

注 1：注明《国际十进分类法 UDC》的类号。

Research on Extraction Technology of CPS Security Situation Elements Based on Multi-source Information Fusion

A Master Thesis Submitted to

University of Electronic Science and Technology of China

Discipline	**Electronic Information**
Student ID	**202022010520**
Author	**Yongxin Zhan**
Supervisor	**Senior Laboratory Engineer Jingshan Duan**
School	**School of Information and Communication Engineering**

摘 要

信息物理系统（Cyber-physical system, CPS）是一个综合计算、网络和物理环境的多维复杂系统。相较于传统物理系统，信息物理系统的运行环境从封闭隔离变得更加开放互联，这也导致了 CPS 的安全问题越来越突出。CPS 一旦遭受攻击或者发生异常，不仅会带来财物上的损失，甚至会对国民安全造成威胁。因此保护 CPS 的安全变得至关重要。

态势感知技术是一种重要的安全防护手段，态势要素提取是态势感知中的重要环节。目前关于态势要素提取的研究目标多为网络系统，而 CPS 将网络系统与物理系统相结合，形成了一个智能化系统。只考虑 CPS 环境中的网络安全信息是无法准确地提取 CPS 安全态势要素。所以对于 CPS 安全态势要素提取，需要融合各类安全信息。本文参考网络态势感知模型提出了 CPS 安全态势要素提取模型，并基于多源信息融合技术提取 CPS 安全态势要素。

在 CPS 安全态势要素提取模型中，本文根据 CPS 所面临的安全威胁来源将 CPS 划分为网络子系统、物理子系统和安防子系统。网络子系统涉及计算机网络和通信设施，物理子系统包括传感器和执行器等硬件元件，安防子系统负责监控 CPS 免受异常人员现场入侵。本文针对三个子系统分别提出不同的安全态势要素提取方法，分别提取 CPS 网络安全态势要素、物理安全态势要素和安防安全态势要素。对于 CPS 网络子系统，提出了复合融合模型。在 KDD CUP99 数据集仿真实验中，验证了该复合融合模型相较于全信息源特征级融合可以有效规避个别信息源特征不明显的问题，进一步提高了要素提取效果和精度。对于 CPS 物理子系统，提出了基于时空融合的物理安全态势要素提取方法。在 SWaT 数据集仿真实验中，验证了该方法模型的可行性和有效性。对于 CPS 安防子系统，基于人脸识别技术提取安防安全态势要素。在仿真实验中，可有效识别多姿态人脸信息，并根据人员的合法性提取安防安全态势要素。最后，基于改进的 D-S 证据理论融合三个子系统的态势要素实现 CPS 整体态势要素提取以及 CPS 复合态势要素提取，用于评估 CPS 整体安全态势，也为态势感知之后的态势要素理解、态势预测提供基础。

关键词： 信息物理系统，多源信息融合，安全态势要素提取

ABSTRACT

Cyber-physical system (CPS) is a multi-dimensional complex system that integrates computing, network and physical environment. Compared with traditional physical systems, the operating environment of cyber-physical systems has changed from being closed and isolated to being more open and interconnected, which has also led to more and more prominent security issues of CPS. Once the CPS is attacked or abnormal, it will not only cause property losses, but even pose a threat to national security. Therefore, protecting the security of CPS becomes very important.

Situational awareness technology is an important means of security protection, and situational element extraction is an important part of situational awareness. At present, most of the research targets on situational element extraction are network systems, while CPS combines network systems with physical systems to form an intelligent system. Only considering the network security information in the CPS environment cannot accurately extract the elements of the CPS security situation. Therefore, for the extraction of CPS security situation elements, it is necessary to integrate various security information. This thesis proposes a CPS security situation element extraction model with reference to the network situation awareness model, and extracts CPS security situation elements based on multi-source information fusion technology.

In the CPS security situation element extraction model, this thesis divides CPS into network subsystem, physical subsystem and security subsystem according to the source of security threats faced by CPS. The network subsystem involves computer networks and communication facilities, the physical subsystem includes hardware components such as sensors and actuators, and the security subsystem is responsible for monitoring the CPS from abnormal personnel on-site intrusion. This thesis proposes different situational element extraction methods for the three subsystems, respectively extracting CPS network security situational elements, physical security situational elements and protection security situational elements. For the CPS network subsystem, a compound fusion model is proposed. In the simulation experiment of the KDD CUP99 dataset, it is verified that the composite fusion model can effectively avoid the problem that the characteristics of individual information sources are not obvious compared with the full information source feature level fusion, and further improve the effect and accuracy of

element extraction. For the CPS physical subsystem, a physical security situation element extraction method based on spatio-temporal fusion is proposed. In the simulation experiment of SWaT dataset, the feasibility and effectiveness of the method model are verified. For the CPS security subsystem, elements of the security situation are extracted based on face recognition technology. In the simulation experiment, it can effectively identify multi-pose face information, and extract security situation elements according to the legitimacy of personnel. Finally, based on the improved D-S evidence theory, the situational elements of the three subsystems are integrated to realize the extraction of CPS overall situational elements and the extraction of CPS composite situational elements, which are used to evaluate the overall security situation of CPS, and also provide a basis for situational element understanding and situation prediction after situational awareness.

Keywords: Cyber-physical system, Multi-source information fusion, Security situation element extraction

目 录

第一章 绪论

1.1 研究工作的背景与意义

自信息物理系统（Cyber-physical system, CPS）概念提出以来，引起了各行各业的高度重视。德国《工业 4.0 实施建议》将 CPS 作为工业 4.0 的核心技术, 《中国制造 2025》提出，“基于信息物理系统的智能装备、智能工厂等智能制造正在引领制造方式变革”。在全球大力发展实体经济的背景下，CPS 作为集成计算、通信、控制技术以实现系统稳定、可靠、高效运行的综合系统，已经成为新一轮工业革命中的重要支撑。CPS 已经广泛应用于众多领域，例如智能家居、智能交通、电力系统、医疗系统等，它不仅可以提高生产效率，降低成本，还可以促进产业升级和转型，改善生活质量，具有重要的经济和社会价值。

典型的 CPS 通常由网络系统、物理系统、控制中心构成，CPS 总体结构如图 1-1 所示。网络系统涉及计算机网络和通信设施，主要实现的是为系统提供网络服务，保证实时可靠的通信能力；物理系统包含传感器、执行器等，用于感知和控制物理环境；控制中心根据网络系统传输的传感器信息来制定和发布对物理系统的执行策略[1] 。这三个部分相互协作，共同实现 CPS 的功能。

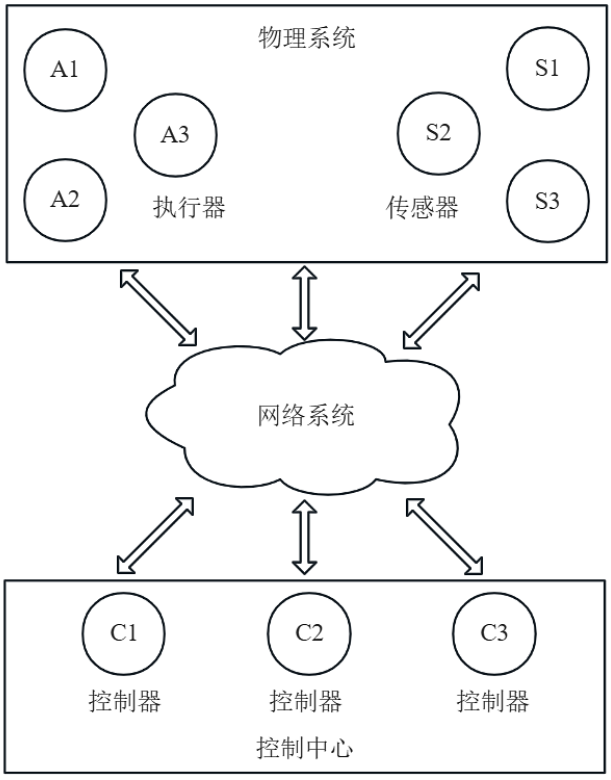


图 1-1 CPS 总体结构

CPS 相较于传统物理系统，其具有复杂、开放、分布式的特性，这就给 CPS 提出了更大的安全挑战，对 CPS 进行安全防护也变得更加困难和重要[2]。近年来，CPS 安全事件层出不穷。2010 年，Stuxnet（震网病毒）通过以 U 盘为媒介成功入侵了伊朗核电站的信息系统和控制系统，并以超速运行离心机的方式破坏了核设施。同时攻击者为了防止异常被及时检测并修复，还伪造了正常的离心机运行数据成功骗过核设施中的采集与监控系统（Supervisory Control and Data Acquisition, SCADA）。2014 年，Havex 木马成功入侵了欧洲工控系统（Industrial control system, ICS）与 SCADA，实现了对电站、电网等关键设施的非法远程操控，造成了难以估量的损失[3]。2017 年，在美国某风电场安全实验中，ARP 病毒成功入侵了风电场的可编程逻辑控制器（Programmable Logic Controller, PLC），实现了对风电发电机的非法远程操控，通过反复制动的方式磨损风力发电机，同时劫持并伪造了风力发电机返回控制中心的传感器信息成功拖延了异常被及时检测并修复的时间[4]。

由以上 CPS 安全事件可知，对 CPS 的攻击行为一般由现场人员入侵或者网络攻击开始，以物理设备损坏结束。CPS 安全和传统信息安全最大的区别在于，攻击者不仅对数字空间的信息系统造成信息泄露、数据损坏等危害，还会对现实世界造成危害，如设备损坏、人员伤亡甚至是危害国家安全。所以对 CPS 进行安全防护是至关重要的。

态势感知技术是一种重要的安全防护手段，其中态势要素提取是态势感知中的重要环节，是对与安全有关的信息进行获取、处理和分析的过程。目前态势要素提取的研究目标多为网络系统，而 CPS 将网络系统与物理系统相结合，形成了一个智能化系统。只考虑 CPS 环境中的网络安全信息无法准确地提取 CPS 安全态势要素。CPS 面临的攻击威胁具有多样性、复杂性、多维性等特点，其安全防护需要采取多层次、多维度的防护策略，以确保 CPS 的安全性和可靠性。CPS 的现场环境通常包含信息系统、物理设备以及现场人员，因此 CPS 所面临的安全威胁可能来源于信息系统、物理设备以及现场人员。在这种情况下，反映 CPS 的安全信息是多层次多角度的。除此之外，按照攻击行为在时间和空间的隐蔽程度划分，CPS 面临的攻击可以分为 4 大类，14 种[5]，如图 1-2 所示。由此可以看出来，CPS 面临的攻击行为是多维度的。综上，对于 CPS 安全态势要素提取，仅考虑单一层次、单一角度、单一维度的安全信息不够全面，需要融合多层次、多角度、多维度的安全信息来提取 CPS 安全态势要素。

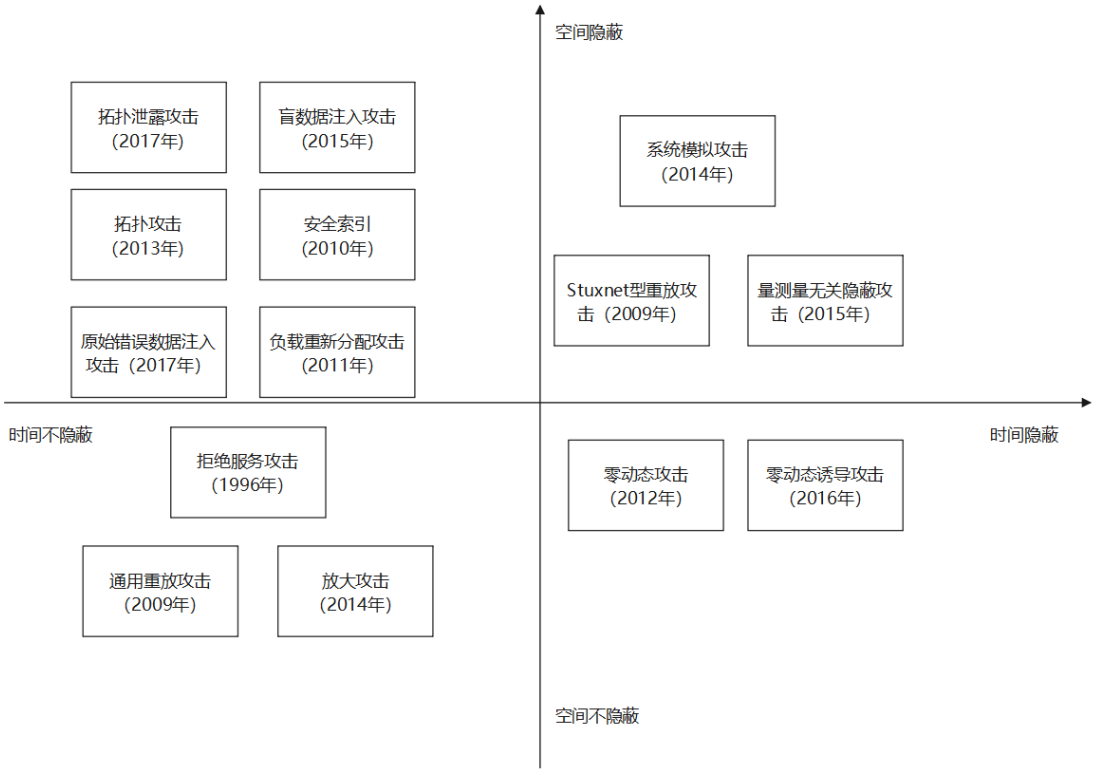


图 1-2 CPS 环境下攻击分类

多源信息融合方法可以尽可能地利用 CPS 环境内所有的安全信息，用来提取 CPS 安全态势要素。目前主流的 CPS 安全防护措施大多单独分析人员、物理系统和网络系统的安全信息。本文希望构建一个 CPS 安全态势要素提取模型，根据 CPS 所面临的安全威胁来源分别提出适合的态势要素提取方法，最终融合多系统，多维度的安全信息来提取 CPS 态势要素。这样可以提高安全态势要素提取的准确性与有效性，为后阶段态势要素理解、态势预测提供基础，对维护 CPS 安全具有重要的意义。

1.2 国内外研究历史与现状

1.2.1 安全态势要素提取研究现状

国外，Endsley 教授在 1988 年的一篇论文[6]中，给出了最初的态势感知的定义，即态势感知是对目标网络空间中的实时信息进行收集、处理、分析和解释，以便对安全威胁进行感知，再应用适当的评估、评价、推理分析机制理解攻击意图，并以此预测未来的安全态势。1995 年，Endsley 教授在之前研究的理论基础上提出了著名的 Endsley 概念模型[7]。该模型将网络安全态势感知分为三个步骤，分别为要素获取、态势评估（或态势理解）以及态势预测。网络安全态势要素提取

是网络安全态势感知的基础，是对与安全有关的信息进行获取、处理和分析的过程。Duan 等人[8]通过对网络安全数据的特征进行分析，选取重要特征作为要素，利用信息增益和随机森林方法进行特征选择和分类，实现了准确且高效的网络安全态势要素提取。Tao 等人[9]基于堆栈式自编码器实现网络安全要素提取，该方法可以保留原始数据的有效特征，还可以降低输入数据的维数、减少存储开销和计算资源。

国内，早期主要是基于知识推理实现态势要素提取。文献[10]提出一种基于特征相异度和 D-S 证据理论的要素提取方法，首先对多源报警信息进行聚类，提取各类报警的特征，计算不同类别之间的相异度，然后使用指数加权 D-S 证据理论对特征进行融合。该方法可以有效提高 CPS 安全态势感知系统的准确度和鲁棒性。文献[11]提出了一种基于本体的要素获取方法，该方法首先对各个组成部分进行分类，然后对每个部分的要素进行提取和描述，构建各部分的本体模型，并在整体上进行融合。该方法的优点在于可以清晰地描述各要素之间的关系和影响，并便于在后续的决策和应对过程中进行理解和解释；除此之外，机器学习方法也被应用于态势要素获取的研究中。文献[12]提出了一种基于堆栈自编码网络的安全态势要素获取方法，通过构建多层自编码器网络，实现了对安全态势要素的深层次提取和分类，同时降低了算法的时间复杂度。文献[13]通过使用深度堆栈编码器网络对原始数据进行特征提取，并结合反向传播算法对网络进行训练，从而解决了反向传播神经网络对过度依赖安全态势信息数据标签的问题。

从以上的研究现状可以看出，对于态势要素提取的研究场景，多伴随在网络态势感知的整体研究中，少部分的研究针对 CPS 场景；对于态势要素提取的研究方向，多侧重于基于空间特征的要素提取，对时间序列的要素提取研究较少。CPS 场景下的安全信息具有多层次多角度的特征，其中不乏属于时间序列的安全信息，例如 CPS 物理子系统中数据多为传感器以固定的时间间隔采集得到，即为时间序列数据。时间序列研究任务可分为时间序列分类、聚类和预测三个主要类别[14]。时间序列分类旨在找到将时间序列映射到类别空间的函数，通过分析区分不同序列的特征来确定序列所属的类别。时间序列聚类则利用一套判别标准将相似的时间序列组织到同一组别中，主要目的是寻找时间序列间的相似特征，实现无监督的序列聚类。时间序列预测则通过分析历史观测结果，寻找时间序列中潜在的趋势特征，建立数学模型来预测未来的发展趋势。这三个类别的任务在要素提取方面有相似性：从时域波形出发，对比全序列和子序列的相似度可以完成时间序列的分类和聚类任务，例如 DTW、Shapelets 和 CNN 等方法；从时间依赖的角度出发，根据序列的内在联系进行预测，例如 FCN、LSTM 和 GRU 等方法；从

时间序列处理的维度出发，例如通过升维、降维、变换等方法，可以提高时间序列要素提取的效率，从而提升算法的整体效率。

1.2.2 多源信息融合研究现状

多源信息融合（Multi-source information fusion）这个概念最早出现于 20 世纪 70 年代的一些文献中，起初是美国中央情报局（CIA）将来自不同情报来源的信息整合起来，以帮助分析员更好地理解全局情况。在此基础上，多源信息融合逐渐发展为一门独立的学科，并被广泛应用于军事、情报、安全、医疗、金融等领域。多源信息融合的基本原则和出发点在于充分利用不同的信息源，以特定标准将不同信息源的信息组合起来，以获得对被测对象的一致性解释或描述。通过这种方式，信息系统能够相对于其所包含的子系统表现出更优秀的性能。

根据信息处理的不同层次可将多源信息融合分为数据级融合、特征级融合和决策级融合。其中，数据是指每个信息源采集的原始数据，特征是指对原始数据加工后的数据，决策是指目标结论。数据级融合是通过将多个传感器提供的各种类型的原始数据直接融合，产生特征或局部决策结果，通常采用统计方法或聚类算法。杜刚[15]对多传感器多系统数据级融合的结构进行了初步探讨，并设计和构成了多系统融合平台。除此之外，他提出的基于逻辑报警控制和多系统融合的方法，可以在一定程度上解决“信息孤岛”问题，并减少应急响应时间。冀少军[16]使用欧氏距离构造距离矩阵，并采用最短距离的聚类算法对其进行数据级融合，最终得到相互关联的传感器组。该方法可以减少主观因素的影响，得到相对客观的融合结果。最终实验结果表明，在煤矿瓦斯预警实验的数据处理过程中具有优越性；特征级融合是提取原始数据的特征信息，为后期决策分析提供支持。近年来，常用的特征级融合算法是神经网络和模糊理论。Turs 等人[17]设计的涡扇发动机信息融合系统可通过特征级融合识别外物损伤事件。该方法将卡尔曼滤波和小波分析提取的轴承加速计信号特征相结合。Jackson[18]将特征级融合应用于军用航空发动机的仿真系统中，通过模糊推理系统隔离各个故障分量。实验表明，该方法在噪声较大的环境下仍然能够高精度地隔离故障分量；决策级融合是将局部决策结果以某种规则进行融合，以获得最终整体的决策结果。常见的决策级融合方法有 D-S 证据理论、Bayes 推理等。温迪[19]基于 D-S 证据理论建立了一个有效的故障决策信息融合框架。实验表明，使用此框架可以融合发动机信息的决策数据，进而提高航空发动机故障诊断的可靠性和安全性。杨亚军[20]则采用 RBF、小波分析和 D-S 证据理论等技术实现了小波神经网络的决策层信息融合。该方法适用于火箭发动机故障诊断分析，实验表明，该成果使得发动机故障诊断的精确度大大提高。

以上关于多源信息融合的研究所融合的信息多为同类信息，CPS 环境的安全信息是多角度多维度的，要想从这些多角度多维度的安全信息中提取安全态势要素，需要针对异类信息融合进行研究。根据融合方法的不同，异类信息融合可以分为基于数学模型的方法和基于机器学习的方法。基于数学模型的方法是将异类信息映射到一个概率空间之后再进行融合。常用的数学模型有模糊集理论、随机有限集合理论等。张崇兴[21]运用模糊粗糙集合建立基本概率分配函数，并使用改进的 D-S 证据理论作为决策融合方法对异类数据进行融合，结果表明能很好地决策出结果。叶宏等人[22]基于随机有限集合理论对异类信息统一表示，从而将异类信息融合问题转化为常见的同类信息融合问题来处理。基于机器学习的方法是对各种信号进行特征提取后利用机器学习技术进行分类或者聚类。宋绪靖[23]分别对文本、语音和视频信息提取特征，并对各类信息的特征配上权重，最后基于神经网络实现了多模态情感识别。唐德权等人[24]在犯罪预测领域，将异类信息进行聚类后获取特征，充分利用特征目标函数计算逻辑损失，从而提高全局特征的正确率和局部特征的精确率。

1.3 本文的研究内容

本文基于多源信息融合技术实现 CPS 安全态势要素提取，即对与安全有关的信息进行获取、处理和分析，提取的结果可以反映系统的安全状况，例如攻击状态信息，脆弱性信息，威胁信息等。为了提升方法模型的适用性并尽可能地提高安全态势要素提取的准确率，主要研究内容包括以下几个方面：

（1）CPS 安全态势要素提取模型：由于 CPS 系统涉及的数据源和环节比较复杂，攻击者可以从多个方面入手，进行各种攻击和破坏。因此，为了保障 CPS 系统的安全稳定运行，需要建立一套完整的安全保障体系，并开展相关研究工作。其中，搭建 CPS 安全态势要素提取模型是其中一个重要的研究内容，提出一个可靠的模型可以事半功倍。该模型可以对来自 CPS 系统中不同数据源的信息进行综合分析和处理，提取出与安全状态相关的关键要素，如攻击、异常等。通过分析这些要素，可以更加全面地了解当前 CPS 系统的安全态势，及时发现潜在的安全风险并采取措施进行预防和应对。

（2）多源信息获取与处理：对于 CPS 系统中涉及到的各种数据源，最好是要建立相应的数据采集和处理系统。具体来说，需要设计和部署传感器网络、网络流量监测设备、监控装置等，以收集来自不同来源的数据。实际上条件不允许，需要选择合适的开源数据集，同时，还需要考虑如何有效处理这些数据，包括数

据预处理、去噪、降维等技术，以减少数据冗余和噪声的影响，提高后续分析的效果。

（3）多场景态势要素提取方法：为应对 CPS 系统的复杂性，需要对 CPS 不同系统提出不同信息融合模型方法。对于 CPS 网络子系统，提出复合融合模型，用来提取网络安全态势要素。CPS 网络子系统涉及计算机网络和通信设施，包含的安全信息来源广泛且异类，有网络流量中提取的安全信息、安全软件检测到的安全信息、系统日志信息等。对于 CPS 物理子系统，考虑到物理安全信息较为相似且具有时间序列特性，提出基于时空融合的物理安全态势要素提取方法。CPS 物理子系统包括传感器和执行器等硬件元件，其安全信息主要包括传感器采集数据，表示执行器是否执行的标志位数据。对于 CPS 整体态势要素提取，提出基于改进的 D-S 证据理论融合方法，通过融合三个子系统的态势要素实现 CPS 整体态势要素提取以及 CPS 复合态势要素提取。提取的 CPS 态势要素可用于评估 CPS 全局安全态势，也为态势感知之后的态势要素理解、态势预测提供基础。

1.4 本文的结构安排

本文共有五个章节，具体的结构为：

（1）绪论：首先介绍了研究工作的背景与意义，然后介绍了国内外关于安全态势要素提取、多源信息融合的研究现状，最后总体上介绍了一下本文的研究内容及结构安排。

（2）相关技术理论：主要介绍了本文中所用到的相关技术理论。包括本文的主要任务，安全态势要素提取相关技术理论。本文基于的方法，多源信息融合相关的技术理论。以及提出的模型方法中用到的机器学习相关技术理论，包括决策树算法、卷积神经网络、长短期记忆网络。

（3）CPS 子系统安全态势要素提取：本章针对 CPS 安全态势要素提取模型中三个子系统分别提出不同的安全态势要素提取方法，分别提取 CPS 网络安全态势要素、物理安全态势要素和安防安全态势要素。对于 CPS 网络子系统，本章在网络安全态势要素提取模型的基础上，提出了复合融合模型，以实现对 CPS 网络子系统多源异类信息的安全态势要素提取。之后选取并分析了合适的开源数据集进行仿真验证；对于 CPS 物理子系统，本章考虑到物理安全信息较为相似且具有时间序列特性，提出了基于时空融合的物理安全态势要素提取方法。具体包括适合物理安全态势要素提取的数据预处理方法以及基于 CNN-LSTM 时空融合的神经网络模型。之后选取并分析了合适的开源数据集进行仿真验证；对于 CPS 安防子系统，本章参考前人方法基于人脸识别技术实现安防子系统安全态势要素提取。

（4）CPS 安全态势要素提取：在 CPS 安全态势要素提取模型的基础上，本章融合 CPS 网络、物理、安防子系统安全态势要素，提取 CPS 整体安全态势要素及复合安全态势要素。首先基于改进的 D-S 证据理论融合实现 CPS 整体安全态势要素提取。随后仿真设计复合攻击场景，使用网络、物理及安防安全态势要素实验结果构造复合安全态势要素提取数据集，最后仿真验证本文的复合态势要素提取方法的正确性。

（5）全文总结与展望：对本文的主要工作内容成果进行了总结，分析了本文研究工作中的不足，并对未来可进一步研究的问题进行了展望。

第二章 相关技术理论

本章介绍和解释与本文相关的理论和技术知识。主要有安全态势要素提取技术相关技术理论、多源信息融合相关的技术理论以及机器学习相关技术理论。此外，本章还讨论了现有技术和方法对于 CPS 环境下安全态势要素提取的局限性和不足之处，为下文的方法和模型选择提供依据。

2.1 安全态势要素提取

安全态势要素提取态势感知中的重要一环，态势感知是指在复杂环境下，通过对各种信息源进行收集、分析和处理，以形成对当前环境状态和未来变化趋势的深入理解和把握，从而支持决策和行动的过程。安全态势感知常用于网络安全领域，用于对网络安全状态进行实时监控、分析和预测，以便及时发现和应对潜在的网络安全威胁和攻击行为。CPS 是由网络、物理、安防设备紧密集成组成的系统，具备系统复杂性、实时性、安全性的特点。故可将网络安全态势感知应用到 CPS 环境中，CPS 系统的本质是将物理系统和网络系统进行深度融合，这种深度融合也带来了更多的安全威胁和风险。安全态势感知可以实时、全面地收集和分析 CPS 系统中的各种数据信息，发现和分析 CPS 系统中的安全威胁和漏洞，及时采取防范和应对措施，保证 CPS 系统的安全稳定运行。

安全态势要素提取是指从各种安全数据源（如传感器、日志、网络流量等）中提取有用的信息和特征，以便更好地理解和诊断系统中的安全状态。本文主要讨论 CPS 安全态势要素提取技术，在 CPS 系统中，安全态势要素提取是保障 CPS 系统安全的重要环节。其主要目的是通过对安全信息进行分析和处理，提取出能够反映 CPS 系统安全状态的关键特征和信息，以便进行安全态势理解与评估。

2.1.1 态势感知模型框架

为了统一态势感知的标准和方法，不少学者提出了态势感知模型框架，影响最深的有 Endsley 模型[7]、JDL 模型[25]。

（1）Endsely 模型：1995 年，Endsley 教授在之前研究的理论基础上提出了著名的 Endsley 概念模型，如图 2-1 所示。该模型提供了一种有效的方法来描述人类在复杂环境中的感知、理解和决策过程。该模型认为在态势感知过程应包含：态势要素提取层，这一层主要是对各种数据源进行处理和分析，从而提取出能够反映系统状态、进程和变化的关键要素；态势理解层，该层的任务是从态势要素中

提取更高层次的信息，并对多个子系统之间的关系进行建模，进一步理解态势的演化过程和态势发展趋势，为最终决策提供基础；态势预测层，该层是指在对当前态势进行了充分提取和理解的基础上，对未来可能出现的态势进行预测和预警，帮助安全管理者及时制定相应的应对策略，预防和减少安全事件的发生，最大程度地保护系统的安全和稳定。

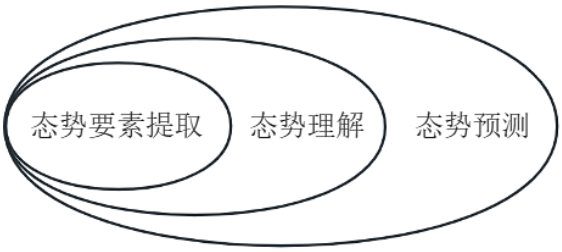


图 2-1 Endsely 模型图

（2）JDL 模型：JDL 模型是信息融合系统中的一种系统化、层次化的信息处理方式，模型如图 2-2 所示。它是由美国国防部的数据融合联合指挥实验室提出，已经成为数据融合领域的执行标准[25]。它的核心思想是将从多个来源采集的数据进行整合处理，提供更全面、准确的信息支持。该模型的实施流程包括数据预处理、数据精炼处理、态势评估和安全状态识别等步骤。在 JDL 模型中，数据预处理是保证数据质量和准确性的基础。数据精炼处理则是对预处理过的数据进行分类、聚类、过滤等精细处理，以提取有用的信息，为后续的态势评估和安全状态识别提供依据。态势评估则是对整合处理后的数据进行分析、判断和预测，以评估可能出现的威胁和风险。安全状态识别则是根据态势评估的结果，识别安全状态并提供预警和决策支持。JDL 模型的应用领域非常广泛，尤其在国防、情报、反恐、网络安全等领域得到了广泛应用。该模型为信息融合技术的发展提供了重要思路和理论基础，对提高信息处理和利用效率，提升决策效力具有重要意义。

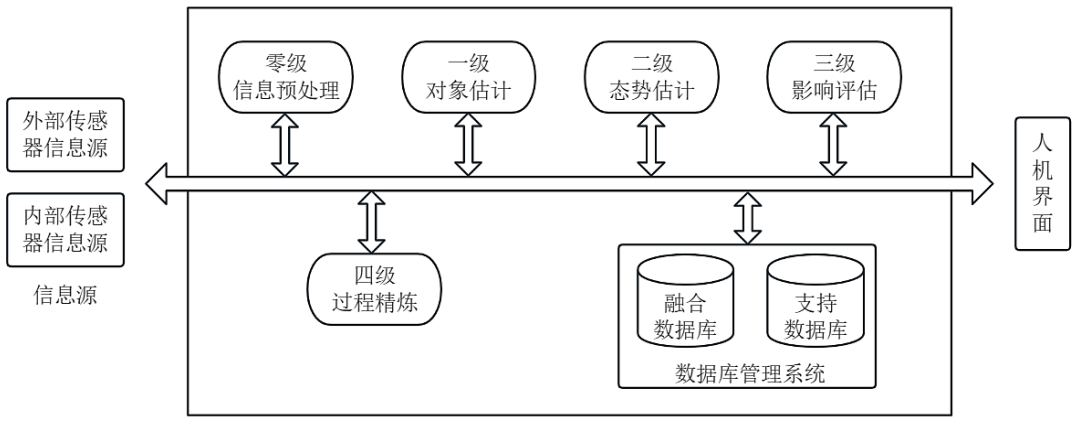


图 2-2 JDL 模型图

Endsley 和 JDL 态势感知模型为安全态势感知模型奠定了基础，后来的学者在此基础上也提出了很多不同的模型，但这些模型的核心都是相似的，可以归纳为三个方面：态势要素提取、态势理解和态势预测。本文主要讨论 CPS 安全态势要素提取技术。

2.1.2 CPS 安全态势要素提取模型

参考网络安全态势感知模型，结合 CPS 系统特点，本文提出 CPS 安全态势要素提取模型，如图 2-3 所示。CPS 所面临的安全威胁有网络攻击、物理攻击、异常人员现场入侵等，本文根据 CPS 所面临的安全威胁来源将 CPS 划分为网络子系统、物理子系统和安防子系统。网络子系统涉及计算机网络和通信设施，物理子系统包括传感器和执行器等硬件元件，安防子系统负责监控 CPS 免受异常人员攻击。这三个子系统密切协同工作，以实现 CPS 的设计目标。本文针对三个子系统分别提出不同的态势要素提取方法，分别提取 CPS 网络安全态势要素、物理安全态势要素和安防安全态势要素。最后融合 CPS 网络、物理、安防子系统安全态势要素，实现 CPS 整体安全态势要素提取和复合安全态势要素提取，用于评估 CPS 全局安全态势，也为态势感知之后的态势要素理解、态势预测提供基础。

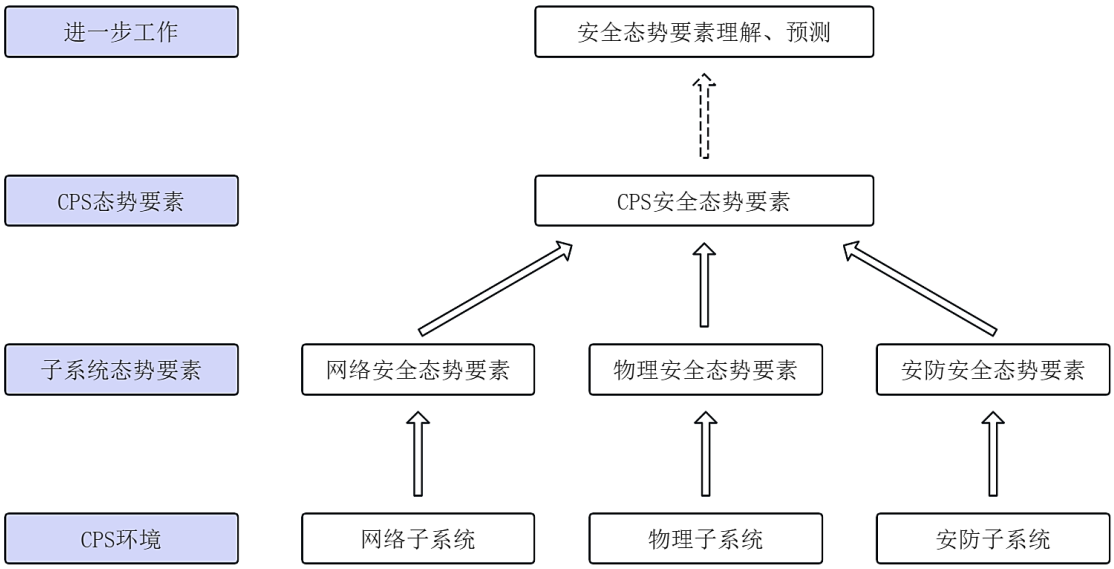


图 2-3 CPS 安全态势要素提取模型

2.1.3 态势要素提取技术

安全态势要素提取技术包括数据预处理、特征提取、特征选择和分类器构建等步骤。其中，数据预处理是指对采集到的数据进行清洗和过滤，保证数据的可靠性和完整性；特征提取是指从原始数据中提取能够反映系统安全状态的关键特

征和信息；特征选择是指对提取到的特征进行筛选和过滤，保留最具有代表性的特征；分类器构建是指利用已有的数据集进行模型训练，构建分类器来对新的数据进行分类和识别。以下是几个常用的态势要素提取技术：

（1）数据挖掘：数据挖掘是指从大量数据中自动地发现隐藏在其中的模式和规律的过程。在态势要素提取中，可以利用数据挖掘技术对历史数据进行分析和挖掘，从中获取关键信息元素，如重要事件、趋势变化等。

（2）机器学习：机器学习是指通过让计算机自动地从数据中学习并改善性能的一类算法。在态势要素提取中，可以利用机器学习算法对大量数据进行分类、聚类、回归等操作，从中提取出关键信息元素。

（3）时间序列分析：时间序列分析是指对随时间变化的数据进行分析和预测的一类统计学方法。在态势要素提取中，可以利用时间序列分析技术对历史数据进行分析和预测，推断未来可能发生的趋势变化等关键信息元素。

（4）多源信息融合：多源信息融合是指将来自不同来源、不同传感器或者不同模态的信息进行集成和处理，提高对目标或场景的认识和理解。在态势要素提取中，可以将来自各种数据源的信息进行融合，对系统的状态进行全方位、多角度的分析，以至于更准确地预测事件发生的趋势并及时应对。

2.2 多源信息融合

多源信息融合是指从不同的信息源中获取的信息，通过一定的技术手段融合起来，形成更完整、更准确、更有用的信息。多源信息融合技术可以克服单一信息源的局限性，提高信息利用效率和可靠性。在不同领域中，如军事、安防、环保、交通、医疗等，都有广泛的应用。本节首先按照多源信息的层次分类介绍了信息融合各层次的融合方法，然后按照融合算法分类介绍了各融合算法的优缺点，最后针对 CPS 安全态势要素提取选择出了最合适的融合理论和融合规则。

2.2.1 多源信息融合层次分类

多源信息融合可以按照不同的层次进行分类。常见的分类方法是按照信息处理的不同层次进行分类，包括以下几个层次：数据级信息融合、特征级信息融合、决策级信息融合。

（1）数据级融合：数据级融合是通过将多个传感器提供的各种类型的原始数据以某种方式集成到一个整体数据集中，并产生特征或局部决策结果，以便进行进一步处理和分析。数据级融合通常采用数学统计方法或聚类算法。数据级融合过程示意图如图 2-4 所示：

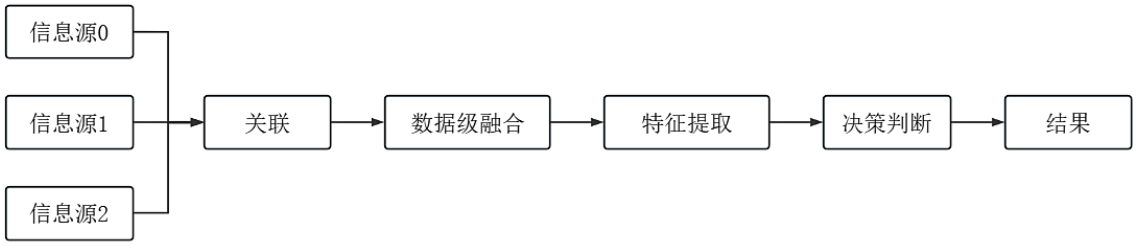


图 2-4 数据级融合过程示意图

（2）特征级融合：特征级融合是将来自不同数据源的特征信息融合在一起，构建一个综合的特征空间，再进行分类或回归等任务。常用的特征级融合算法是神经网络和模糊理论。特征级融合过程示意图如图 2-5 所示：

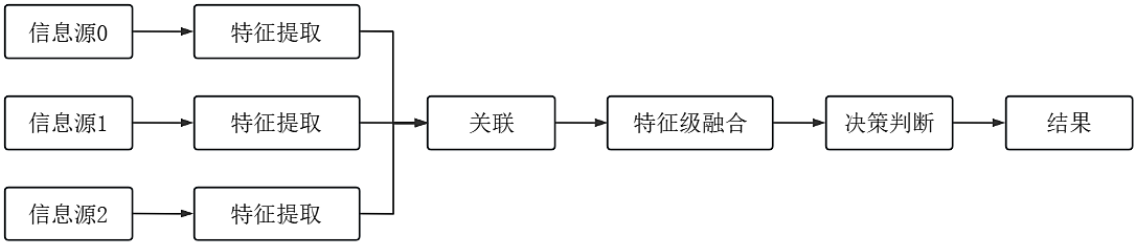


图 2-5 特征级融合过程示意图

（3）决策级融合：决策级融合是将来自多个信息源的不同局部决策结果以某种规则进行融合，以获得最终整体的决策结果。常见的决策级融合方法有 D-S 证据理论、Bayes 推理等。决策级融合过程示意图如图 2-6 所示：

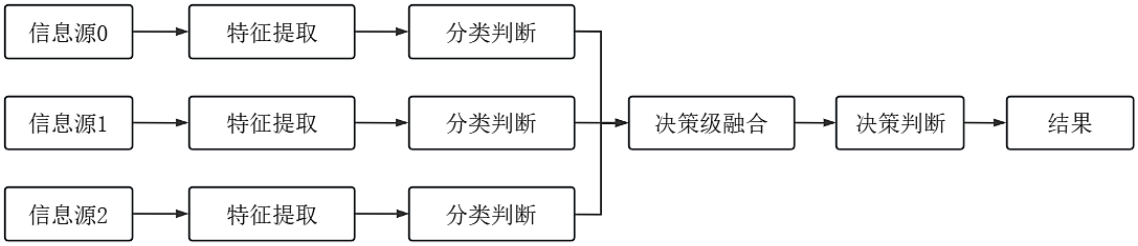


图 2-6 决策级融合过程示意图

2.2.2 多源信息融合方法选择

数据级融合方法主要包括加权平均、卡尔曼滤波算法等。加权平均是将来自不同传感器的数据加权求平均，其中权重可以根据各传感器数据的可靠性、准确性和精度等因素进行调整。卡尔曼滤波算法多用于对实时、动态且冗余的多源数据进行融合以估计动态系统的状态。数据级信息融合方法具有简单易行、计算速

度快的优点，但无法考虑数据之间的相关性和时序性，加上 CPS 系统复杂，数据多样且异类，为简化仿真实验，本文不考虑数据级融合方法。

特征级融合方法主要有以下几种：

（1）特征拼接法：将来自不同数据源的特征按照一定规则拼接在一起，形成一个新的特征向量，再进行分类或回归任务。该方法简单易实现，不需要过多的数学推导，但可能会导致维度灾难的问题，同时在特征选择时也会受到影响。

（2）特征加权法：将来自不同数据源的特征进行加权平均或加权求和，得到一个新的特征向量，再进行分类或回归任务。该方法考虑到不同特征对任务的影响程度，具有较好的鲁棒性和可解释性，但需要在数据集上进行参数调整，特征权重的确定可能会受到一定的主观因素影响。

（3）特征变换法：将来自不同数据源的特征进行线性或非线性变换，得到一个新的特征空间，再进行分类或回归任务。该方法能够更好地考虑特征之间的关联关系，有效地降低特征维度，但需要进行数学建模和计算，且对特征变换的选择和参数调整要求较高。

（4）特征选择法：从来自不同数据源的特征中选择最具有代表性和区分度的特征，构建一个新的特征空间，再进行分类或回归任务。该方法能够有效地降低维度，但需要进行特征选择的算法和模型选择，并可能会忽略一些有用的特征。

特征级融合层次高于数据级融合，对于同类信息融合易于实施。对于 CPS 子系统，安全信息都是同类信息，故特征级融合可作为本文模型中子系统态势要素提取的融合方法，具体方法可根据具体任务选择。

常用决策级融合的方法如下：

（1）数学统计法：数学统计包括投票法、权重法、逻辑运算法等。投票法是选取得票最多的决策作为集成结果。投票法简单易实现，但不考虑不同决策的权重和置信度。权重法是按照权重进行加权平均或加权求和，得到最终的集成结果。权重法考虑了不同决策的权重和置信度，但需要确定权重的分配方式。逻辑运算法是通过逻辑运算对不同的决策进行逻辑组合，得到最终的集成结果。逻辑运算法对不同决策之间的关系进行了考虑，但需要确定逻辑运算符的组合方式。此类方法思路简单，融合方式直接，适用于同类传感器且判决结果统一的情况。

（2）决策树算法：决策树算法将多个数据源的输出结果转化为决策树的形式，根据不同决策树的输出结果，综合得出最终的决策结果。优点是可以很好地处理非线性问题，可解释性较好，但缺点是容易过拟合，且不易处理连续值。

（3）主观 Bayes 估计法：主观 Bayes 的基础是贝叶斯公式，如公式（2-1）所示。主观 Bayes 估计法是一种常用的决策级融合方法。其基本思想是将不同决策

器的决策结果转化为概率分布，然后将这些概率分布进行加权平均，得到最终的融合结果。在主观 Bayes 估计法中，先对每个决策结果设置先验概率 $P(H_i)$，再分别评估各个决策器的决策结果 E 属于分类 H_i 的条件概率 $P(H_i|E)$，最后根据贝叶斯公式由先验概率 $P(H_i)$ 和决策器评估的概率 $P(H_i|E)$ 计算出后验概率 $P(E|H_i)$ 作为最终的判断结果。主观 Bayes 估计法的优点在于能够灵活地处理不同决策的不确定性，同时还可以通过加权平均来平衡各个决策的贡献。然而，主观 Bayes 估计法也存在一些缺点，例如该方法要求假设事件独立且互斥，对先验概率 $P(H_i)$ 的评估需要专家知识等。

$$\begin{cases} P(H_i|E)=\dfrac{P(E|H_i)P(H_i)}{\sum_i P(E|H_i)P(H_i)} \\ \sum_i P(H_i)=1 \end{cases} \tag{2-1}$$

（4）D-S（Dempster - Shafer）证据推理法：D-S 证据推理法是主观 Bayes 估计法的一种推广，用于处理不精确情况下的推理问题。在主观 Bayes 估计法中，必须先给出确切的先验概率才能融合决策判断。然而，D-S 证据推理法则可以在先验概率未知的情况下进行决策级融合。表 2-1 列出了 D-S 证据推理法和主观 Bayes 估计法之间的对比[26]。

表 2-1 主观 Bayes 估计法与 D-S 证据理论的比较

特点	主观 Bayes 估计法	D-S 证据推理法
表示不确定性的方式	概率	信度
必须给出先验概率和条件概率	必须	不必
概率评估的主、客观性	主观或客观	主观
增删规则的方便性	不便	容易

综上所述，决策级信息融合方法可以综合多个信息源的决策结果，提高决策准确性，故决策级融合可作为本文模型中 CPS 安全态势要素提取的融合方法，具体方法可以为基于决策树的学习法和 D-S 证据推理法。

2.2.3 D-S 证据理论

1967 年， Dempster 首先建立了证据理论，后来他的学生 Shafer 进一步完善了 Dempster 提出的证据理论，最终形成了完整的证据理论，即 Dempster-Shafer 证据理论，简称 D-S 证据理论。经过多年的研究和发展，D-S 证据理论已经成为处理不确定性问题的一种有效方法，并在多个领域有着广泛的应用。证据理论对于度量事件的不确定性提供了一种量化方法，即通过对事件掌握的证据和知识进行量

化，得到一组可信度函数，从而实现对不确定性的度量。D-S 证据理论使用确定的组合规则对所有证据进行合成，这个规则保证了最终得到的合成概率依然能够满足基本可靠的性质。D-S 证据理论的基本概念如下[26]：

（1）识别框架：D-S 证据理论中的识别框架 Θ 是一个有限非空集合，其中包含对某一对象所有可能的命题，使用幂集 2^{Θ} 表示。例如：Θ ={正常，异常}，2^{Θ} ={{正常}，{异常}，{正常，异常}}。

（2）基本概率分配函数（Basic Probability Assignment，BPA）：对于识别框架 Θ 上的所有命题 2^{Θ}，都有基本概率分配函数 m 使得 $m:2^{\Theta}\to[0,1]$，其满足对于空集 $\varnothing$ 有 $m(\varnothing)=0$ 且 $\forall A\in 2^{\Theta}$ 有 $0\le m(A)\le 1$，$\sum m(A)=1$。

（3）焦元：$\forall A\in 2^{\Theta}$，$m(A)$ 表示对命题 A 的信任程度。若 $m(A)>0$，则称命题 A 为一个焦元。

（4）信任函数：识别框架 Θ 上的信度函数表示为 $Bel(A)$，代表对命题 A 的总信任程度。公式如（2-2）所示：

$$Bel(A)=\sum_{B\subset A} m(B),\ A\in 2^{\Theta}, B\in 2^{\Theta} \tag{2-2}$$

（5）似然函数：识别框架 Θ 上的似然函数表示不否定命题 A 的程度。公式如（2-3）所示：

$$Pl(A)=1-Bel(\bar{A})=\sum_{B\cap A\neq\varnothing} m(B),\ A\in 2^{\Theta}, B\in 2^{\Theta} \tag{2-3}$$

（6）信任区间：支持、信任和拒绝命题 A 的判决结果的概率区间被称为 A 的证据区间。A 的信任区间为 $[Bel(A), Pl(A)]$，表示可以信任但是不能肯定的概率区间范围。$Bel(A)$ 表示对 A 命题的支持概率，$|Pl(A)-Bel(A)|$ 表示对某一命题或假设的不清楚程度或者不确定的概率，$1-Pl(A)$ 表示对 A 命题的拒绝概率。

（7）组合规则：D-S 证据理论传统的组合规则是 Dempster 公式，它综合了来自多传感器的基本信度分配，得到一个新的信度分配作为输出。Dempster 组合规则如下：假设 $m_1, m_2, \cdots, m_n$ 为识别框架 $\Theta=\{A_1, A_2, \cdots, A_n\}$ 上不同的基本概率赋值函数（BPA），每一个 BPA 都为识别框架中所有命题分配了基本概率值。则对于 $Z\subseteq 2^{\Theta}$ 且 $Z\neq\varnothing$ 有公式（2-4）。

$$\begin{cases} K=\sum\limits_{\bigcap A_i=\varnothing}\prod\limits_{1\le j\le n} m_j(A_i) \\ m_{组合}(Z)=\dfrac{1}{1-K}\sum\limits_{\bigcap A_i=Z}\prod\limits_{1\le j\le n} m_j(A_i) \\ m_{组合}(\varnothing)=0 \end{cases} \tag{2-4}$$

2.3 机器学习

2.3.1 决策树算法

决策树算法是一种基于树形结构的用于分类和预测的机器学习算法。决策树采用自顶向下的递归方法，通过对数据集进行分裂，构建一棵决策树来预测新样本的分类或回归结果。在决策树中，每个内部节点表示一个属性或特征，每个分支代表该属性或特征的不同取值，而每个叶节点则表示一个类别或回归值。基本思想是以信息熵为度量，构造一棵熵值下降最快的树。

常用信息熵用来衡量随机变量的不确定性，信息熵越小，则包含的信息就越少。假设在数据集 X 中的第 i 类样本所占的比例是 $p_i(i=1,2,...,n)$，那么样本集 X 的信息熵可以用公式（2-5）表示。

$$H(X)=-\sum_{i=1}^{n} p_i \log p_i \tag{2-5}$$

决策树算法的关键在于如何选择最优属性或特征来进行分裂。根据不同的分裂规则，决策树包括 ID3、C4.5、CART 等多种算法。ID3 算法是最早的决策树算法之一，它以信息增益为准则选择属性。信息增益值越大，此属性作为根结点越有效。ID3 算法缺陷是在处理连续属性和缺失值方面存在一定的缺陷。C4.5 算法是 ID3 算法的改进版本，它引入了信息增益率来处理 ID3 算法的缺陷。CART 算法是一种基于二叉决策树的算法，采用 Gini 系数来对特征进行划分，基尼系数象征模型的不纯度，其值越小，不确定的程度越小，特征越明显。假设数据样本包含 n 个类别，第 i 个类别被选中的几率为 p_i ，那么概率分布的基尼系数公式可以用公式（2.6）所示：

$$Gini(p)=\sum_{i=1}^{n} p_i(1-p_i)=1-\sum_{i=1}^{n} p_i^{\,2} \tag{2-6}$$

在实际应用中，决策树算法已经被广泛应用于数据挖掘、机器学习、人工智能等领域。同时，基于决策树的算法还可以通过集成学习的方式来提高分类或回归的准确性和稳定性，例如随机森林和梯度提升树等。

2.3.2 卷积神经网络

卷积神经网络 (Convolutional Neural Network，CNN) 是一种前馈神经网络，特别于处理二维图像和语音识别等任务。与传统神经网络不同的是，CNN 在模型中引入了卷积运算，通过局部连接、权值共享和池化等操作，从原始输入数据中提取特征并实现自动分类。

典型的 CNN 结构如图 2-7 所示： CNN 的基本组成部分包括卷积层、池化层和全连接层。卷积层是 CNN 的核心，通过卷积运算实现对原始数据的特征提取。卷积运算的本质是对一个滤波器在原始数据上进行卷积，提取出局部的特征信息。池化层用于降低卷积层输出的空间维度，同时提高特征的鲁棒性和计算效率。常用的池化方法有最大池化和平均池化等。全连接层将卷积和池化后的特征进行分类。

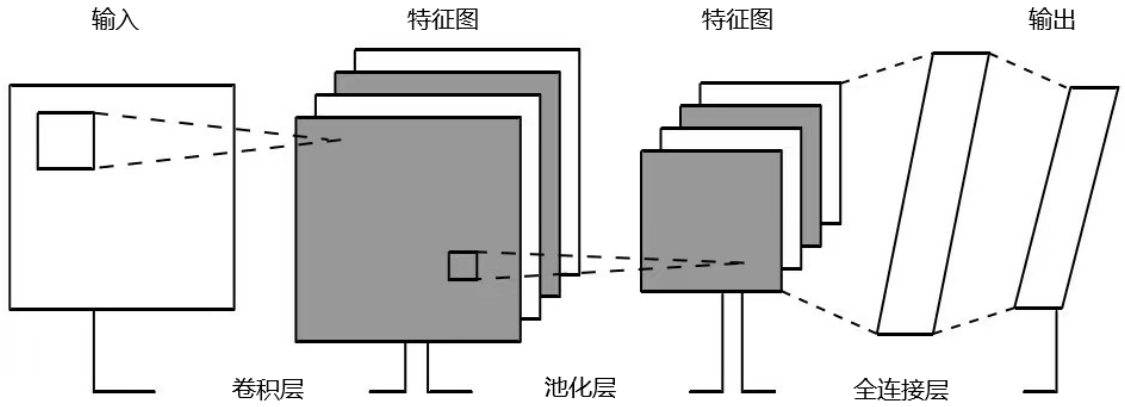


图 2-7 卷积神经网络基本结构

CNN 的优点是具有很强的抽象特征提取能力。由于卷积核是局部连接的，每个卷积核只提取局部特征，通过不同卷积核的组合可以提取出多个不同的特征，如纹理、边缘、形状等。而池化层可以在保留重要信息的同时，降低卷积层输出的空间维度，减少模型参数数量和计算量，提高模型的泛化性能。缺点是需要大量的训练数据和计算资源，同时容易出现过拟合等问题。针对这些问题，研究人员提出了多种改进和优化方法，如数据增强、迁移学习、正则化等。

在实际应用中，CNN 已广泛应用于图像分类、目标检测、人脸识别、自然语言处理等领域。其中，在图像分类任务中，通过对图像的卷积和池化操作，CNN 能够自动学习到图像的局部特征和全局特征，实现对图像的高精度分类。总的来说，CNN 作为一种重要的深度学习算法，已经成为许多视觉和语音任务的基础模型为实现自动化分类和识别提供了有效的工具和方法

2.3.3 长短期记忆网络

长短期记忆网络（Long Short-Term Memory，LSTM）是一种广泛应用于序列数据处理任务的循环神经网络（Recurrent Neural Network，RNN）变种。与传统的 RNN 相比，LSTM 通过引入门控机制来更好地捕捉和记忆序列中的长期依赖关

系。LSTM 具有良好的序列建模能力，被广泛应用于自然语言处理、音频处理、视频处理等领域。

LSTM 的基本思想是引入一个记忆单元（Memory Cell）来存储和管理输入序列的信息，通过一系列的门控机制来控制信息的流动和存储。LSTM 的记忆单元中包含一个状态向量和一个单元向量。状态向量保存了当前时刻的记忆，而单元向量保存了当前时刻的输入数据。如图 2-8 所示，为一个基本的 LSTM 单元结构。

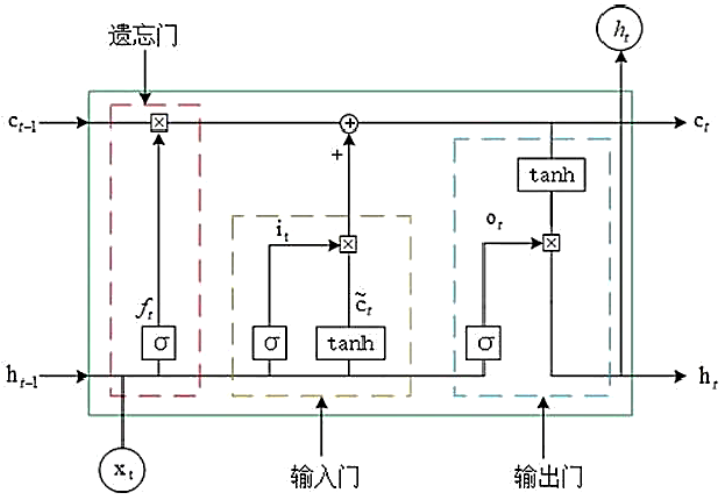


图 2-8 LSTM 基本单元结构

LSTM 的门控机制包括输入门、输出门和遗忘门，它们可以控制记忆单元中的信息流动和存储。输入门（Input Gate）可以控制哪些信息需要更新到记忆单元中。输入门接收前一时刻的状态向量和当前时刻的输入向量，然后输出一个 0 到 1 之间的数值表示需要更新的信息量，1 表示全部更新，0 表示全部忽略。输出门（Output Gate）可以控制从记忆单元中输出哪些信息。输出门接收前一时刻的状态向量和当前时刻的输入向量，然后输出一个 0 到 1 之间的数值表示需要输出的信息量，1 表示全部输出，0 表示全部忽略。遗忘门（Forget Gate）可以控制记忆单元中哪些信息需要保留，哪些信息需要遗忘。遗忘门接收前一时刻的状态向量和当前时刻的输入向量，然后输出一个 0 到 1 之间的数值表示需要保留的信息量，1 表示全部保留，0 表示全部遗忘。LSTM 基本单元的输入门、输出门和遗忘门公式分别如（2-7）、（2-8）和（2-9）所示。

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) \tag{2-7}$$

$$o_t = \sigma\left(W_o x_t + U_o h_{t-1} + b_o\right) \tag{2-8}$$

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{2-9}$$

其中 x_t 表示 t 时刻存储单元的输入，h_{t-1} 表示上一时刻的输出，W 、U 表示输入门、输出门和遗忘门的权值，b 表示偏置值，式中 σ 表示 Sigmod 函数。

已知前一时刻存储单元的状态如公式（2-10）所示。当给定了输入门激活 i_t 、遗忘门激活 f_t 和前一时刻存储单元的状态值 $\tilde{c}_t$ 后，可以计算出 t 时刻存储单元的新状态如公式（2-11）所示。则当前输出单元如公式（2-12）所示。

$$\tilde{c}_t = \tanh\left(W_c x_t + U_c h_{t-1} + b_c\right) \tag{2-10}$$

$$c_t = i_t \odot \tilde{c}_t + f_t \odot c_{t-1} \tag{2-11}$$

$$h_t = o_t \odot \tanh\left(c_t\right) \tag{2-12}$$

公式（2-12）中 $\odot$ 表示向量元素按位相乘。从以上公式可以看出，存储单元 c_t 可以获取上一时刻输出信息的反馈，即神经网络中 t 时刻的输出信息不仅包含了当前时刻的输入信息，还包含了 $t-1$ 时刻的输出信息。

除了以上的门控机制，LSTM 还可以通过堆叠多层 LSTM 来提高建模能力，同时也可以通过注意力机制等方法来进一步增强 LSTM 的性能。总的来说，LSTM 是一种能够捕捉和记忆序列中长期依赖关系的强大模型，它通过引入门控机制来更好地控制信息的流动和存储，从而在序列数据处理任务中取得了很好的效果，适合提取 CPS 环境下具有时序性质数据的特征。

2.4 本章小结

本章首先介绍了安全态势要素提取相关技术理论，并提出了 CPS 安全态势要素提取模型。然后介绍了多源信息融合相关技术理论，重点分析和对比了多源信息融合层次以及相应的方法，并针对本文应用场景选择出了合适的模型方法，作为本文多源信息融合的理论基础。最后简单介绍了本文中使用到的机器学习算法，包括决策树算法、卷积神经网络以及长短记忆网络。

第三章 CPS 子系统安全态势要素提取

在 CPS 安全态势要素提取模型中，本文根据 CPS 所面临的安全威胁来源将 CPS 划分为网络子系统、物理子系统和安防子系统。网络子系统涉及计算机网络和通信设施，物理子系统包括传感器和执行器等硬件元件，安防子系统负责监控 CPS 免受异常人员攻击。这三个子系统密切协同工作，以实现 CPS 的设计目标。本章针对三个子系统分别提出不同的安全态势要素提取方法，分别提取 CPS 网络安全态势要素、物理安全态势要素和安防安全态势要素。对于 CPS 网络子系统，本章使用复合融合模型实现网络子系统安全态势要素提取；对于 CPS 物理子系统，本章使用时空融合模型实现物理子系统安全态势要素提取，并归纳总结了物理安全态势要素提取方法；对于 CPS 安防子系统，本章基于人脸识别技术实现安防子系统安全态势要素提取。

3.1 CPS 网络子系统安全态势要素提取

在 CPS 环境中，物理设备连接到互联网上，让物理设备具有了计算、通信、精确控制、远程协调和自治等功能。联网也给 CPS 带来了新的挑战，网络上的攻击都有可能在 CPS 环境中重现。本节基于复合融合模型实现 CPS 网络子系统安全态势要素提取，并借助 KDD CUP99 数据集，验证基于复合融合的 CPS 网络子系统安全态势要素提取模型的可行性与准确性。

3.1.1 基于复合融合的网络安全态势要素提取模型

CPS 网络信息可以分为流量数据、系统日志、安全软件告警等，分别来自于不同的信息采集源。在 CPS 网络子系统安全态势要素提取中，单一的信息源判决可能会缺失重要的信息，因此多信息源数据的融合判决非常重要。将来自不同信息采集源的数据进行整合和分析，可以获得更全面、准确和可靠的安全态势认知。多源信息融合可以结合多个信息源的优势，弥补各个信息源之间的局限性，提高 CPS 网络子系统安全态势要素提取的效果。

在多源信息融合中，特征级融合和决策级融合是常用的两种融合方法。特征级融合可选择特征拼接法、特征加权法、特征变换法或者特征选择法来实现。特征级融合的优点在于可以利用多特征数据，提取更加全面、准确的特征，从而得到更好的分类或回归效果。同一信息源获取的多特征数据可以互补，减少单一特征数据的误差和不足，提高了分类或回归的准确率和鲁棒性。同时，特征级融合

还可以有效地减少数据冗余，降低特征维度，提高算法的运行效率。决策级融合的优点在于可以利用多信息源的分类或回归结果，结合各自的置信度或权重，得到更加准确的结果。决策级融合可以通过数学统计学方法（如投票法、权重法、逻辑运算法等）、D-S 证据推理法或机器学习方法（如决策树算法、神经网络等）来进行。决策级融合的结果更加稳定可靠，能够有效地减少分类或回归误差。而特征级融合级联决策级融合是将两种方法结合起来，形成的一种复合融合方法。特征级融合级联决策级融合结合了特征级融合和决策级融合的优点，具有更高的分类或回归准确率和鲁棒性。因此，本文基于复合融合方法实现 CPS 网络子系统安全态势要素提取，提出基于复合融合的 CPS 网络子系统安全态势要素提取模型，如图 3-1 所示。

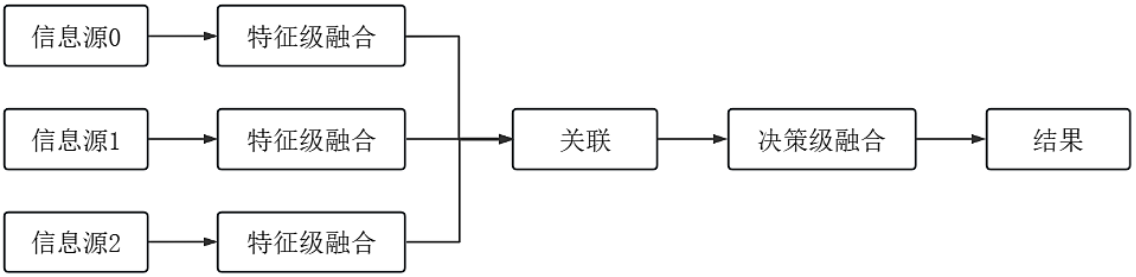


图 3-1 CPS 网络子系统安全态势要素提取模型

在 CPS 网络子系统安全态势要素提取模型中，首先进行单信息源的特征级融合，可采用特征变换的方法将不同特征进行线性或非线性变换，得到一个新的特征空间，得到更加全面、准确的特征，然后对新特征进行分类或回归以得到初步的结果。然后将特征级融合得到的初步结果在时间上进行关联。最后将关联的初步结果作为新的特征集合进行决策级融合，可采用数学统计法、D-S 证据推理法或机器学习方法等，最终实现 CPS 网络子系统安全态势要素提取。

综上所述，特征级融合级联决策级融合的复合融合方法有其独特的优点和适用范围，可根据不同的需求和场景来选择合适的融合方法。在某些情况下，特征级融合和决策级融合可以结合使用，相较于全信息源特征级融合可以有效规避个别信息源特征不明显的问题，以进一步提高融合效果和精度。

3.1.2 实验数据集

3.1.2.1 数据集介绍

在 CPS 网络子系统安全态势要素提取仿真实验中，选取开源数据集 KDD CUP99 作为 CPS 网络子系统安全信息。KDD 是数据挖掘与知识发现（Data Mining

and Knowledge Discovery）的简称，KDD CUP99 数据集是 1999 年举办的 KDD 竞赛官方提供的数据集。KDD CUP99 数据集是入侵检测领域中广泛使用的数据集之一，其中包含大量模拟网络攻击和正常网络流量的数据。它是通过网络仿真器 SWAT（Scenario Weaver for Attack Trace）在模拟的网络环境中模拟攻击产生的。在这个模拟环境中，仿真器 SWAT 对各种攻击类型进行了模拟，包含 DoS、R2L、U2R 和 Probe 四大类共 58 种典型的攻击方式，并记录了攻击过程中的网络数据包、Solaris 审计数据、WinNT 主机审计日志以及和安全有关的备份数据。通过对这些信息进行数据处理和特征提取，得到了 KDD CUP99 数据集中的具体数据。这些数据包含了网络连接的各种特征，例如源地址、目的地址、源端口、目的端口、连接持续时间等等。同时，这些记录还标记了每个连接是否是一次攻击以及攻击类型。因此 KDD CUP99 数据集可以模拟 CPS 环境下多种观测角度的观测数据。

3.1.2.2 数据集分析

KDD CUP99 数据集的一条数据包含 41 个特征，根据不同的观测角度可以分为四类，分别是 TCP 连接的基本特征、TCP 连接的内容特征、基于时间的统计特征以及基于主机的统计特征。第一类特征描述了最基本的网络链接信息，第二类特征是对数据包内容提取的特征；第三类特征是针对网络攻击行为在时间上的关联性提取的特征；第四类特征是针对主机信息提取的特征。各类特征具体内容如表 3-1 所示。

表 3-1 KDD CUP99 特征分类

特征类型	数量	内容
TCP 连接的基本特征	9	协议类型、服务、源到目的主机的字节数等
TCP 连接的内容特征	13	访问次数、登录尝试失败次数、使用 shell 命令次数等
基于时间的统计特征	9	过去两秒内同服务连接出现“SYN”次数等
基于主机的统计特征	10	前 100 个连接中，与当前连接具有相同目标主机连接数等

可以将 KDD CUP99 数据集中的四类特征看作四个信息源，对每类特征分别进行分析。具体信息源划分如表 3-2 所示。

表 3-2 CPS 网络子系统信息源划分

信息源 ID	信息源介绍	对应特征区间
0	TCP 连接基本特征	[1, 9]
1	TCP 连接的内容特征	[10, 22]
2	基于时间的统计特征	[23, 31]
3	基于主机的统计特征	[32, 41]

文献[27]使用决策树算法对 KDD CUP99 数据集中每一种攻击类型进行了特征提取，分析出与每种攻击类别具有强关联性的特征。根据文献[27]的结论，每一种攻击行为在每个信息员中的重要特征数如表 3-3 所示。可以发现每个信息源对不同攻击类型的识别效果有很大的差异，甚至出现了完全无法识别某类攻击的情况。例如信息源 1 对于 Dos 没用重要特征，无法有效识别 Dos 攻击。信息源 2 对于 U2R 没有重要特征，无法有效识别出 U2R 攻击。

表 3-3 每个信息源下对各攻击的重要特征数

攻击类型	信息源 0	信息源 1	信息源 2	信息源 3
Dos	5	0	2	4
Probe	5	1	1	7
R2L	5	4	1	8
U2R	3	3	0	2

3.1.2.3 数据预处理

使用 KDD CUP99 数据集进行网络安全态势要素提取之前，需要对数据集进行预处理。通过数据预处理，可以清洗和转换原始数据，使其更适合机器学习算法。合适的数据预处理可以提高模型的性能、准确性和鲁棒性，从而取得更好的结果。数据预处理过程如图 3-2 所示。

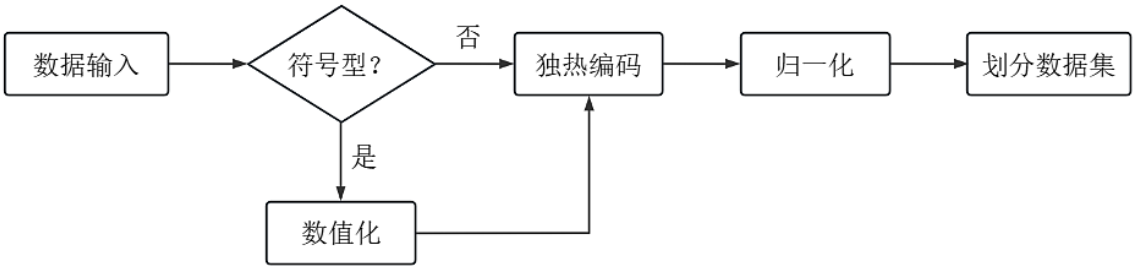


图 3-2 数据预处理过程

（1）符号型数据数值化：KDD99 数据集特征属性中有 3 个特征是字符型，因为计算机在处理数据时只能处理数字，无法直接处理符号型数据，所以首先要将字符型特征数值化。这里通过标签编码函数（Label Encoder Function，LEF）对字符型特征进行处理，即是对字符进行标号。在 KDD99 数据集中，字符型特征索引有[1, 2, 3, 41]，分别对应[protocol_type, service, flag, label]。符号型数据数值化预处理示例如表 3-4 协议类型数值化映射所示。

表 3-4 协议类型数值化

协议类型	数值型
TCP	0
UDP	1
ICMP	2

（2）离散性数据独热编码：通过对 KDD99 数据集的分析可知，KDD99 数据集中的特征数据分为连续型和离散型两种。其中由于各离散型特征的度量方法不一样，导致不同特征之间的差距较大。为了消除由于属性度量的差异产生的影响，需要对离散型数据进行进一步操作。本文使用 N 位状态寄存器对 N 个状态进行独热编码（One-hot Encoder，OHE），每个状态都有独立的寄存器位，并且在任意时候只有一位有效[28]。在 KDD99 数据集中，离散型特征索引有[1, 2, 3, 6, 11, 13, 14, 20, 21]，分别对应[protocol_type, service, flag, land, logged_in, root_shell, su_attempted, is_hot_login, label]。离散型数据独热编码示例如表 3-5 协议类型独热编码所示。

表 3-5 协议类型独热编码

协议类型	数值型	独热编码
TCP	0	1 0 0
UDP	1	0 1 0
ICMP	2	0 0 1

（3）数据归一化：由于连续型特征具有不同的取值范围，同一特征之间的数值差异较大。为了防止数值差异大对最后的分类结果造成影响，本文采用 Min-Max 处理方法对数据进行归一化，即对数据进行线性变换并将结果映射到[0, 1]之间，进而统一数据的基本度量单位。线性函数转换式如公式（3-1）所示：

$$x_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, 1 \le i \le n \tag{3-1}$$

（4）划分数据集：KDD CUP99 数据的训练集和测试集划分如表 3-6 所示。由表可知，在训练集中 Probe、R2L、U2R 样本数较少，样本分布及其不均衡。例如 Dos 攻击类型样本有 391458 个，而 U2R 攻击类型样本只有 52 个。若使用训练集全部数据进行训练，则会导致识别正确率偏低，所以本文在划分数据集时进行均衡采样。首先将所有训练集数据分为两组，group1 是 Normal，Dos，Probe。gropu2 是 R2L，U2R。之后在 group1 上面使用 RandomUnderSampler 进行欠采样，在 group2 上面使用 SMOTE 进行过采样。最终得到训练集样本比例为[4107, 4107,

4107, 4107, 4107]。对于验证集则使用均衡化的训练集数据按照 test_size = 0.2 比例及样本分类比例划分。测试集采用 KDD CPP99 数据集提供的带标签测试集，其中包含训练集中不存在的攻击。

表 3-6 KDD CUP99 数据集样本分布

数据集	攻击类型					数据总量
	Normal	Dos	U2R	R2L	Probe	
训练集	97278	391458	52	1126	4107	494021
测试集	60593	229853	182	16189	2566	311029

3.1.3 仿真设计及结果分析

为了验证本章多源复合融合模型对 CPS 网络子系统安全态势要素提取的效果和可行性，本节首先对所有信息源的特征进行一次特征级融合，通过经典机器学习算法分析融合效果，目的一是选出适合 KDD CUP99 的机器学习算法，其次是与本章模型方法的效果进行对比。随后按照本章所提模型重新进行实验，在单信息源内进行小范围的特征级融合，之后在决策级融合多信息源特征级融合的结果。最后与全特征融合进行对比，如果本章模型的提取效果优于前者，则可证明本章基于复合融合的 CPS 网络子系统安全态势要素提取模型的效果和可行性。

3.1.3.1 全特征融合仿真设计

KDD CUP99 的一条数据包含 41 种特征，分为四种不同观测角度的特征，分别是各个 TCP 连接的基本功能、TCP 连接的内容特征、基于时间的统计特征、基于主机的统计特征。全特征融合即不区分信息源，利用全特征仅进行一次特征级融合以实现 CPS 网络子系统安全态势要素提取。全特征融合仿真实验的目的一是为本章的复合融合方法选择出适合 KDD CUP99 的机器学习算法，其次是与本章的复合融合模型的效果进行对比，体现出本章模型方法的优势。

利用全特征进行一次特征级融合的方法可以采用数学模型方法，实际应用中特征涉及很多不同的观测维度，难以设计合理的模型进行融合判决，实施困难。也可以通过特征拼接、特征加权或者特征提取的方式结合机器学习等技术进行聚类或者分类的方法。例如文献[29]旨在提取并分析加壳恶意软件运行时的系统调用行为特征，通过对系统调用行为特征进行加权降维，以提高行为特征的有效性。例如文献[30]将异构网络信息的特征进行级联后送入支持向量机中进行分类。后者的优点是可以提高数据的准确性和可靠性，所以本文选择后者进行特征级的融合。具体设计为将 KDD CUP99 数据集各个观测角度的特征按照时间规则拼接在一起，

形成原始特征维度为 41 维的全特征向量。之后通过数据预处理等一系列操作得到训练集、验证集、测试集。最后使用各种经典机器学习算法训练模型，实现分类任务，完成基于全特征融合的 CPS 网络子系统安全态势要素提取仿真实验。

3.1.3.2 全特征融合结果分析

本节实验通过特征拼接的方式结合机器学习进行分类实验，通过模型分类效果进行方法的评价。选择的机器学习算法有 KNN、贝叶斯、MLP、决策树，评价指标有准确率、精准率、召回率、F1-分数。

KNN、贝叶斯、MLP、决策树四种算法实验的混淆矩阵如图 3-3 所示。各算法的准确率分别为 91.84%、90.95%、91.86%、91.97%，所选择的算法对于 KDD CUP99 数据集提取安全态势要素都有较好的表现，基于全特征融合的 CPS 网络子系统安全态势要素提取方法具有可行性。

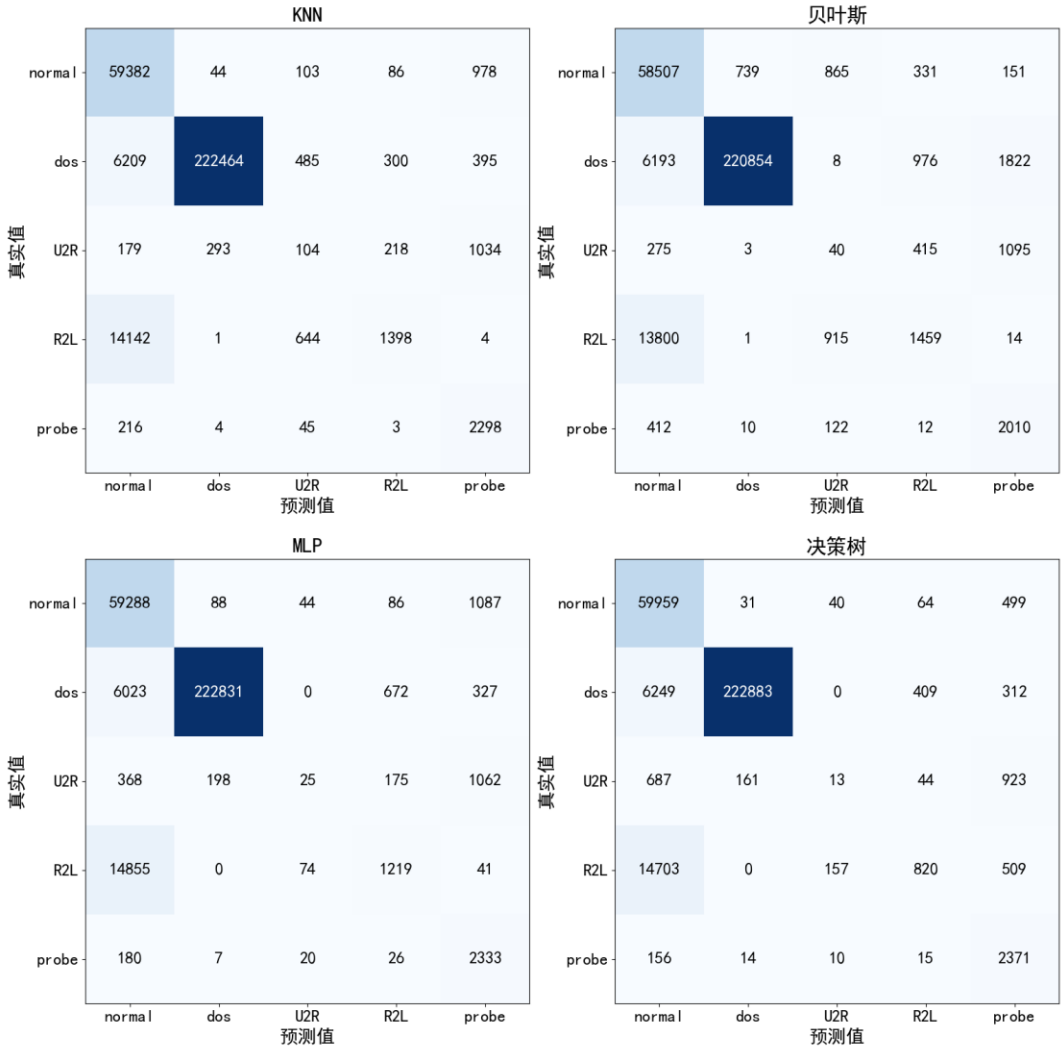


图 3-3 全特征融合实验混淆矩阵

四种方法的精准率性能指标如表 3-7 所示，精准率是指在所有被模型判断为正类的样本中，实际为正类的样本所占的比例。精准率越高，代表模型将负样本误判为正样本的概率越低，可以更好地过滤掉负样本，提高分类器的准确性。由表 3-7 可知，所有算法对于 normal 和 dos 类别的预测精度相对较高，都在 73%以上，而对于其他类别的精度普遍较低，特别是 U2R 类别的精度最低，不足 16%。对于 normal 类别，四个分类算法的精准率相差不大，最高的是 KNN 算法的 74.11%，其次是贝叶斯算法的 73.88%。对于 dos 类别，四个分类算法的精准率都较高，分别达到了 99.85%、99.66%、99.87%、99.91%，说明这些算法在区分 dos 类别时都比较有效。对于 U2R 和 R2L 类别，四个分类算法的精准率相对较低，说明这些算法在区分 U2R 和 R2L 类别时有一定的局限性。对于 probe 类别，KNN 算法和 MLP 算法的精准率相对较低，分别为 48.8%和 48.1%。而贝叶斯算法和决策树算法的精准率则相对较高，分别为 39.47%和 51.39%。这说明 KNN 算法和 MLP 算法在区分 probe 类别时相对较弱，而决策树算法则表现较好。

表 3-7 全特征融合精准率(%)

算法	normal	dos	U2R	R2L	probe
KNN	74.11	99.85	7.53	69.73	48.80
贝叶斯	73.88	99.66	2.05	45.69	39.47
MLP	73.45	99.87	15.34	55.97	48.10
决策树	73.34	99.91	5.91	60.65	51.39

四种方法的召回率性能指标如表 3-8 所示，召回率反映了被正确分类的样本占总样本数的比例，用于描述分类器对于正样本的识别能力。由表 3-8 可知，四种算法在 normal、dos 和 probe 三个类别上的召回率都比较高，都在 90%以上。但在 U2R 和 R2L 两个类别上，四种算法的召回率普遍比较低， R2L 和 U2R 两个类别是网络入侵的类别，较为难以识别，另外因为 KDD CUP99 数据集是一个不平衡的数据集，在训练集中 U2R 只有 52 个样本、R2L 只有 1126 个样本，因此召回率较低。

表 3-8 全特征融合召回率(%)

算法	normal	dos	U2R	R2L	probe
KNN	98.00	96.79	5.69	8.64	89.56
贝叶斯	96.56	96.56	2.19	9.01	78.33
MLP	97.85	96.95	1.37	7.53	90.92
决策树	98.95	96.97	0.71	5.07	92.40

综上所述，决策树算法在各项指标上表现与其他算法相近，没有明显的优劣之分，综合性能较好。对于 KDD CUP99 数据集，决策树是一种非常有效的算法选择。KDD CUP99 数据集具有很高的维度和很大的数据集大小，这意味着我们需要一种计算复杂度较低的算法来快速处理数据。虽然 KNN 算法性能较好，但是对每次判断时需要遍历所有训练集，导致判断过慢，不适合用于对实时性要求较高的安全态势要素提取场景。决策树算法可以高效地对大量数据进行分类和预测，因为它可以利用数据集的结构来快速缩小可能的分类范围。此外，决策树算法易于解释，可以让我们更好地理解数据和模型的运作方式。综上所述，决策树算法作为特征级融合的融合算法具有一定的优势。

3.1.3.3 复合融合仿真设计

在复合融合仿真实验中，区分信息源，分别从不同的观测角度对同一类风险事件进行模型训练。由于缺乏合适的数据集，仿真采用 KDD CUP99 数据集中的四大类特征作为四个信息源。KDD CUP99 中的四类特征实质上来自四种不同类别的信息，所以可以看作是四个信息源，具体信息源划分如表 3-2 所示。

由 3.1.3.2 节可知，决策树算法在 KDD CUP99 数据集中的分类效果要优于其他机器学习算法，所以在复合融合模型的决策级融合阶段选择决策树算法。按照本章的复合融合方法，具体设计如下：首先基于决策树算法进行单信息源的特征级融合以得到初步的结果；然后将特征级融合得到的初步结果在时间上进行关联；最后将关联的初步结果作为新的特征集合进行决策级融合得到最终的结果。最终验证复合融合模型在 CPS 网络子系统环境下的效果和可行性，实现 CPS 网络子系统安全态势要素提取。

本文选择的决策级融合方法包括投票法、加权平均、D-S 融合和学习法。以下对每种方法进行详细说明：

（1）投票法：判决得票最多的类别作为最终的结果，若同时有多个分类结果的票最高，选择预测类别较多的作为最终的结果。

（2）加权平均：投票法的改进，将四个信息源的预测结果乘上基分类器准确率，选择最大值作为融合评估结果。

（3）D-S 融合：D-S 融合是基于证据理论的多源信息融合方法，将各个信息源的不确定性通过信任度的计算，从而得到一个更为准确的决策结果，可以有效地处理信息源之间的不确定性和冲突。

（4）学习法：将各信息源的预测结果再进行一次机器学习，通过学习各信息源的预测结果特征，训练融合模型，本文选择决策树算法。

3.1.3.4 复合融合结果分析

复合融合模型的特征级融合阶段仿真实验显示，各信息源特征级融合准确率分别为 70.09%、15.00%、74.90%、90.77%，混淆矩阵如图 3-4 所示，精准率如表 3-9 所示，召回率如表 3-10 所示。由各项数据可得各信息源的预测效果相对于全特征的特征级融合都要差。主要原因在于单信息源特征数量不足，丢失了部分特征信息，另外训练集数据不平衡且各信息源对不同攻击的识别效果存在差异。结合文献[27]的结论和表 3-3 分析。表 3-3 有 1 处 0 值，表示 Dos 攻击在信息源 1 中没有重要特征，导致信息源 1 对于 Dos 攻击识别召回率只有 0.46%。在 U2R 和 R2L 两个类别上同全特征融合效果一致，识别率较低。总而言之，单信息源特征级融合效果较差，各信息源对各类别的攻击预测效果各有优劣之分，只有将单信息源的预测结果进行决策级融合才能得到更好的效果。

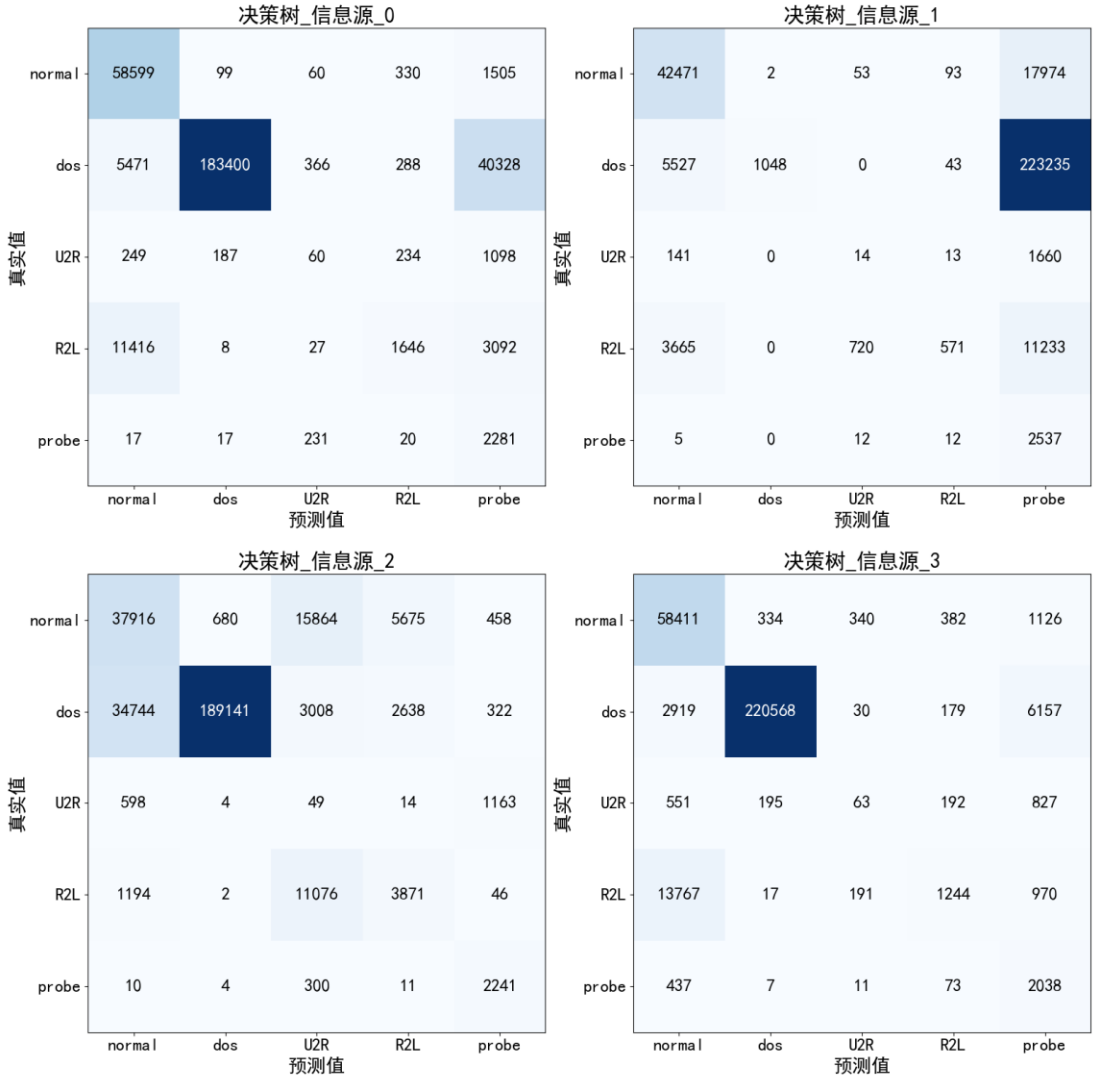


图 3-4 单信息源特征级融合实验混淆矩阵

表 3-9 单信息源特征级融合精准率(%)

信息源	normal	dos	U2R	R2L	probe
信息源 0	77.36	99.83	8.06	65.37	4.72
信息源 1	81.98	99.81	1.75	78.01	0.99
信息源 2	50.92	99.64	0.16	31.71	52.98
信息源 3	76.77	99.75	9.92	60.1	18.33

表 3-10 单信息源特征级融合召回率(%)

信息源	normal	dos	U2R	R2L	probe
信息源 0	96.71	79.79	3.28	10.17	88.89
信息源 1	70.09	0.46	0.77	3.53	98.87
信息源 2	62.57	82.29	2.68	23.91	87.33
信息源 3	96.40	95.96	3.45	7.68	79.42

复合融合模型的决策级融合阶段仿真实验显示，投票法、加权平均法、D-S 融合、学习法各决策级融合方法的准确率分别为 78.99%、92.96%、80.86%、95.33%，全特征融合的决策树算法的准确率为 91.97%。决策级融合算法中加权平均法、学习法都高于全特征融合的决策树算法，尤其是学习法，准确率达到了 95.33%。在准确率方面可以证明本章复合融合模型优于全特征融合模型。

各决策级融合方法的精准率性能指标如表 3-11 所示。可以看到，全特征融合的决策树算法的精准率指标表现相对较差，而投票法、加权平均法、学习法相比之下表现更好，尤其是学习法在 U2R 和 R2L 类别上有较大的提升。D-S 融合算法表现教差，由于对决策树输出的后验概率未进行修正，融合了信息源 1 的错误预测结果，导致最终融合结果较差。而学习法的对各类别攻击的精准率都高于其他决策级融合方法，也高于全特征融合的决策树算法，平均精准率达到了 84%。综上在精准率方面可以证明本章复合融合模型优于全特征融合模型。

表 3-11 决策级融合精准率(%)

方法	normal	dos	U2R	R2L	probe
决策树	73.34	99.91	5.91	60.65	51.39
投票法	76.88	99.71	2.30	85.09	5.19
加权平均法	80.9	97.74	3.60	80.67	55.23
D-S 融合	51.15	99.57	6.73	80.39	66.5
学习法	83.96	99.37	78.18	91.41	68.47

各决策级融合方法的召回率性能指标如表 3-12 所示。不同算法对不同类别的召回率表现存在差异，学习法表现最好，其余三种方法的召回率普遍较低，但是

都比全特征融合的决策树算法效果好。整体上各种方法在 U2R 和 R2L 类别上表现较差，考虑到是使用决策树算法作为基分类器，而决策树容易过拟合，再加上这两个类别的数据样本较少。投票法在分类时只考虑了类别数量，而没有考虑信息源的可信度，这方面加权平均法要优于投票法。D-S 融合效果召回率较差，原因是 D-S 融合过于依赖于证据的一致性，而证据的一致性会受到不同信息源的干扰，导致效果下降。

表 3-12 决策级融合召回率(%)

方法	normal	dos	U2R	R2L	probe
决策树	98.95	96.97	0.71	5.07	92.40
投票法	96.98	79.56	1.59	9.27	97.47
加权平均法	96.19	98.97	1.59	6.37	89.52
D-S 融合	98.21	82.14	2.30	6.91	79.31
学习法	98.58	99.72	15.48	29.56	97.00

总的来说，特征级融合级联决策级融合的复合融合模型在准确率、精准率和召回率上优于单信息源特征级融合，也优于全特征融合模型。由此可以验证，本章模型优于全特征融合模型，在 CPS 网络子系统安全态势要素提取方面具有一定的优势。

3.2 CPS 物理子系统安全态势要素提取

本节基于时空融合对 CPS 物理子系统进行安全态势要素提取。CPS 物理子系统数据多为传感器以固定的时间间隔采集得到，即为时间序列数据。对于时间序列，它有以下四个基本特点：观察值依赖于时间而变化，重视顺序的重要性；每一个时间上的取值有随机性，不可能完全准确地进行预测；前后时间上的取值有一定的相关性；从整体上看，往往有某种趋势或出现周期性现象。正因为以上特点，CPS 物理子系统数据的时间维度特征不可忽略。本节结合 CNN 提取空间特征的优点与 LSTM 提取时间特征的优点，采用在监督模式下在物理数据集上训练的分类器学习过去攻击的特征，并使用它来检测和识别未来的类似攻击。空间特征指的是多传感器在同一时刻采集数据所具有的非时间维度特征，时间特征则是多传感器在一段时间内所采集数据之间的时间关联特征。

3.2.1 基于时空融合的物理安全态势要素提取方法

物理安全态势要素提取可具体为异常检测，攻击检测。异常检测可以检测 CPS 物理系统内部的异常状况，攻击检测则可以提取到 CPS 系统来自外部的安全

威胁，并常伴随异常发生，两者都是系统安全态势要素提取的结果，都可以分析得到当前系统的安全态势。

已经有许多学者提出了异常检测、攻击检测的解决方案。一些论文基于规则的 IDS[31]或确定性有限自动机 (DFA) [32]。在文献[33]中，作者测试了不同的深度学习模型来检测异常主要有 LSTM 和一维卷积神经网络 (1D CNN)。此外，作者建议通过计算当前值与具有给定滞后的过去值之间的差异来添加新特征。作者得出结论，1D-CNN 获得了比 LSTM 更好的结果，同时训练速度更快。在文献[34]中，提出了一种基于 LSTM 神经网络的 IDS。作者提出了一种用于异常检测的累积和 (CUSUM) 方法。 CUSUM 是根据地面实况和模型预测的数据计算的残差计算得出的。文献[35]中提出了一种用于 ICS 场景的新型基于生成对抗网络的异常检测 (GAN-AD) 方法。作者在 GAN-AD 的核心实施了一个 LSTM 网络，以捕获 ICS 中正常条件下传感器和执行器的多元时间序列的分布。除了深度学习技术，也有学者提出了其他机器学习技术。例如，在文献[36]中，作者使用机器学习算法（朴素贝叶斯、PART 和随机森林）开发了一种异常检测 IDS（入侵检测系统），并提取特征以改进其结果。同样，文献[37]的作者提出了一种基于三种机器学习模型（J48 决策树、朴素贝叶斯和 RF）的 IDS，用于检测 SCADA 系统中的分布式拒绝服务 (DDoS) 。在文献[38] 中，作者生成了一个从电力牵引变电站收集的新数据集，并尝试使用半监督和监督机器学习和深度学习模型来检测异常。

由以上相关研究工作可以发现大部分学者所做的相关工作都是使用机器学习或者深度学习来提取 CPS 物理安全态势要素。由于 CPS 物理数据的性质，物理安全态势要素提取经常被视为时间序列分类问题。因此，许多技术依赖于可以处理时间的算法。在深度学习技术中，卷积神经网络 (CNN) 和长短期记忆 (LSTM) 神经网络，两者都可以应用于需要将输入数据转换为序列的问题，并且都可以处理时间维度的信息。机器学习模型可以有效地提取时间序列特征来处理更复杂的数据，但是在建模时，将时间序列的每个样本简单地看成是高维空间中的一个点，每个点的坐标元素独立于彼此。然而，时间序列中的正常数据通常与相邻数据存在时间相关性，前者的序列可以影响后者的趋势。因此，机器学习模型很难提取时间序列的高级特征。深度学习模型虽然可以在时间序列数据上提取相邻数据的相关特征，但没有充分利用数据点的时间相关性。深度学习模型利用卷积算法等算法捕捉不同特征之间的相关信息，从而提取数据的空间特征。另外，深度学习模型使用诸如 RNN 算法之类的算法来提取数据之间的时间特征。RNN 算法及其变体可以通过记忆先前的数据来提取时间特征。例如作者[39]将时间序列异常检测问题转化为分类问题，并使用滑动窗口和多尺度 LSTM 模型提取时间特征来判断

序列是否异常。一些研究人员使用 LSTM 从历史时间序列数据中预测未来序列，并通过预测序列与真实序列之间的误差来判断序列是否异常[40]。虽然时间深度学习模型可以有效地提取数据的时间特征，但是当序列太长或数据有噪声时，很难提取完整的时间信息。鉴于以上问题，本文提出基于时空融合的深度学习模型，结合了空间深度学习模型和时间深度学习模型，可以充分提取数据的时空信息，更适合本文的 CPS 物理子系统安全态势要素提取。

另外，从以上相关研究工作可以看出物理安全态势要素提取没有标准化的步骤，每项工作都由作者自己决定。总之，大多数相关工作都提出了基于不同机器学习或者深度学习模型的技术。本章结合以上做法优点及网络安全态势要素提取方法，规范了物理安全态势要素提取的方法步骤。总共分为四个步骤，特征处理、数据预处理、模型构建、验证与分析。该方法实施步骤图如图 3-5 所示。尽管这四个步骤对于每个机器学习及深度学习项目都是通用的，但对于 CPS 物理数据有具体的操作。例如对于时序数据的处理，CPS 设备之间的高度依赖性、高相关性特征的优化、新特征提取等。各步骤具体实施如 3.2.3 节所示。

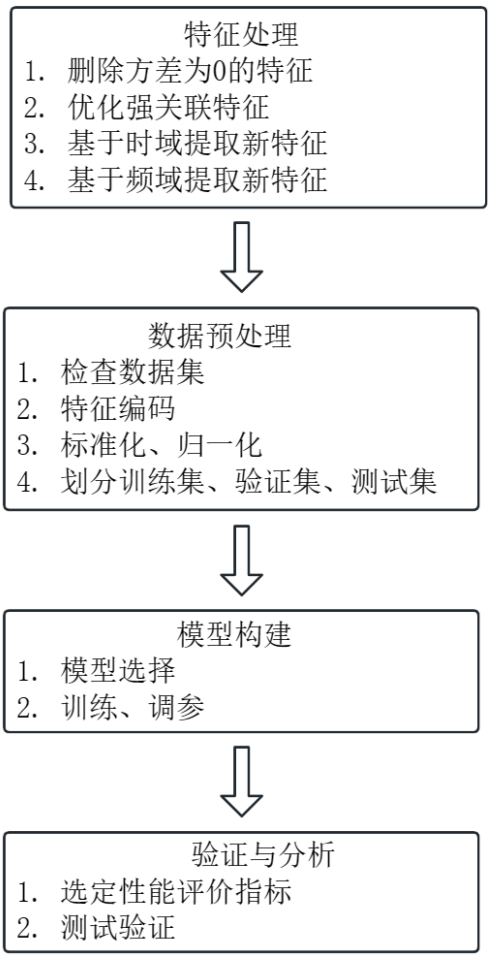


图 3-5 CPS 物理安全态势要素提取方法

3.2.1.1 特征处理

根据 3.2.1 节介绍的物理安全态势要素提取方法，第一步为特征处理，特征处理的目的是去除无用的特征，提取新的特征。物理信息包含连续的传感器数据，也包含离散的阀门开合状态等数据，其中存在整个数据集中未发生变化的特征，存在强相关的特征集合。另外，物理信息作为时序数据，存在明显的规律性，以及特征之间的联系。所以在 CPS 物理安全态势要素提取方法的特征处理应当考虑以下处理方式：

（1）删除方差为 0 的特征：方差为 0 的特征包含的信息量少，对于模型训练影响不大。这种情况在物理信息数据集中比较常见，例如供电标志位，在系统稳定运行期间，供电是一直存在的，在数据集中的体现就是供电标志位一成不变。

（2）优化高度相关的特征：物理信息数据集中特征经常相关。例如，当传感器报告特定水箱已满时，通常控制水箱输入的泵停止运行或控制水箱输出的泵开始运行，这种强相关的特征集合会造成特征冗余，提高数据集的复杂度，提高模型训练时长，所以可以适当优化，选取其中一个。

（3）特征提取：特征提取是利用已有的数学手段从原始特征中提取高阶特征的过程。CPS 物理设备经常执行重复操作，所以相应的数据具有重复的表现，要以特征的形式表示这些重复动作。例如，SWaT 数据集中水箱水位具备周期性。可以使用自相关函数来提取时域的高阶特征，使用 DFT 来提取频域的高阶特征。

（4）主成分分析(Principal Component Analysis，PCA)：PCA 是最重要的降维方法之一，在数据压缩消除冗余和数据噪音消除等领域都有广泛的应用。主要是通过析取主成分显出的最大的个体差异，发现更便于人们能够理解的特征，可以用来削减特征的数目。然而，PCA 返回的主成分是原始特征的线性组合，因此难以解释结果。在许多情况下，可解释性是一个很重要的要求。例如，系统检测出了异常，但是想知道哪个执行器或者传感器导致了这种异常，这就应该避免使用 PCA。所以在系统不需要可解释性时才推荐使用 PCA。

3.2.1.2 数据预处理

根据 3.2.1 节介绍的物理安全态势要素提取方法，第二步为数据预处理，这也是所有机器学习或深度学习所需要的处理。但对于 CPS 环境下物理安全态势要素提取，需要额外考虑以下两种数据预处理的方式：

（1）检查数据集，检查数据是否包含无效或缺失的数据。这些类型的数据在 CPS 预热过程和设备状态转换过程中很常见，例如，当 CPS 物理设备正在初始化时或在被网络攻击之后。所以，在进行数据预处理的时候需要独立绘制每个特征

的数据图，直观地找出无效数据并删除相应的样本。另外，还需要检查数据集是否包含缺失值并修复该数据。处理缺失值有两种常见方式，一种是将缺失值设置为均值或者中值，另一种是删除包含缺失值的样本。第一种方式不适用与 CPS 物理数据的处理，因为 CPS 物理数据集中有许多特征离散的，例如阀门打开或关闭，使用均值或者中值没有意义。所以，最好是采取第二种方式，舍弃包含缺失值的样本。或者是两种方式相结合，在处理缺失值的样本时先分析缺失值是否为连续值，如果为连续值则采用第一种方式；如果为离散值，则删除该样本。这种方式可以更充分的保留数据集包含的样本信息。

（2）在划分数据集的时候，测试集中样本应当包含训练集中不存在的攻击类别，这样可以评估分类器是否学习了过去攻击的特征，使用它是否可以检测和识别未来的类似攻击。这与真实环境中的情况类似，安全检测系统往往只能捕获一部分的攻击类别，对未来未知的攻击缺乏识别能力。所以在 CPS 环境下物理安全态势要素提取仿真实验中，数据集划分应当使用一部分的攻击类别数据进行训练，使用另一部分的攻击类别和未训练过的攻击类别作为测试集。除此之外，在划分数据集的过程中，应当保存数据集的时间顺序，因为 CPS 物理信息往往包含时间维度的特征，不应随机抽样，应当按时间顺序划分数据集。

3.2.1.3 模型构建

本文基于时空融合的深度学习模型实现物理安全态势要素的提取，在实际应用中有不同的实施方法，合适的实施方法对 CPS 物理子系统安全态势要素提取事关重要。本文按照具体的实施过程，将物理安全态势要素提取技术分为两大类：

（1）基于非监督学习的回归模型：非监督学习可以在没有给定样本标签的情况下，从数据中发现规律和模式。回归模型可以通过输入变量来预测输出变量，通过预测值与真实量之间的残差分析发现 CPS 的异常状态。该方法不需要先验知识，可以处理大规模数据。来自 CPS 物理传感器和执行器的信号是时间序列信号，适合做回归预测。其次，采用回归模型并选择合适的阈值可以实现系统的可解释性，可以根据要素提取结果定位异常或者被攻击的传感器或者执行器。缺点是预测效果不稳定，由于非监督学习的回归模型没有目标变量的约束，预测效果不如监督学习模型好，只能提供数据的一些模式和规律。

（2）基于监督学习的分类模型：监督学习的分类模型基于已知标签的数据进行训练，是的该方法具有较高的准确性和稳定性。基于监督学习的分类模型可以适应各种数据类型，包括数值型、符号型、时间序列、文本和图像等数据类型。该方法缺点是需要大量标注好的训练数据，如果数据量不足或标注不准确，模型

的性能可能会受到影响。如果模型的特征选择不恰当，可能会导致过拟合或欠拟合的问题，监督学习的分类模型在处理新的未知数据时可能存在泛化能力差的问题，即训练好的模型对未来未知的异常或者攻击检测效果可能不明显。

3.2.1.4 验证与分析

在验证与分析阶段，需要考虑数据集的性质来定义合适的性能指标。对于物理系统数据往往是一个不平衡的数据集，这意味着与异常类相比，数据集包含大量正常样本，所以准确率不是首选指标。在这种情况下，TP、FP、TN 和 FN 是首选指标。TP、FP、TN 和 FN 具体释义如下：

TP：True Positive，分类器预测结果为正样本，实际也为正样本，即正样本被正确识别的数量。

FP：False Positive，分类器预测结果为正样本，实际为负样本，即误报的负样本数量。

TN：True Negative，分类器预测结果为负样本，实际为负样本，即负样本被正确识别的数量。

FN：False Negative，分类器预测结果为负样本，实际为正样本，即漏报的正样本数量。

为了便于与其他论文结果进行比较，其他常见的指标有准确率、精准率、召回率和 F1-分数指标。这些指标可以使用 TP、FP、TN 和 FN 指标进行计算得到，计算公式如下所示：

Accuracy：准确率，表征的是预测正确的样本比例。不过通常不用这个概念，主要是因为预测正确的负样本没有太大意义。计算公式如（3-2）所示。

$$Accuracy = (TP + TN) / \text{样本总数} \tag{3-2}$$

Precision：精准率，表征的是预测正确的正样本的准确度，精准率等于预测正确的正样本数量/所有预测为正样本数量。Precision 越大说明误检的越少，Precision 越小说明误检的越多。计算公式如（3-3）所示。

$$Precision = \frac{TP}{TP + FP} \tag{3-3}$$

Recall：召回率，表征的是预测正确的正样本的覆盖率，召回率等于预测正确的正样本数量/所有正样本的总和，TP+FN 实际就是 Ground Truth 的数量。Recall 越大说明漏检的越少，Recall 越小说明漏检的越多。计算公式如（3-4）所示。

$$Recall = \frac{TP}{TP + FN} \tag{3-4}$$

F1-score：F1 分数，是统计学中用来衡量二分类模型精确度的一种指标。它同时兼顾了分类模型的精确率和召回率。F1 分数可以看作是模型精确率和召回率的一种调和平均，它的最大值是 1，最小值是 0。计算公式如（3-5）所示。

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3-5}$$

3.2.2 实验数据集

3.2.2.1 数据集介绍

本文选择 SWaT 数据集来进行仿真实验，作为 CPS 环境下的物理数据集。SWaT 数据集是从一个完全运行的缩小水处理厂中采集的，如图 3-6 所示，SWaT 试验台由六个工业过程（标记为 P1 到 P6）组成，它们共同处理水以供分配。具体来说，进程 P1 负责为原水箱提供足够的水来供应其他进程。进程 P2 负责对水进行预处理，以确保其具有可接受的质量水平。如果来自 P1 水箱的原水质量不合格（由传感器 AIT201、AIT202 和 AIT203 测量），将打开泵（P201、P203 和 P205），并应用化学剂量来纠正缺水。一旦水质合格，就会进入进程 P3。在此阶段，水中残留的不需要的物质通过使用精细过滤膜的超滤 (UF) 系统去除。在接下来的进程 P4 中，任何残留的氯都会通过使用紫外线 (UV) 灯脱氯来破坏。下一阶段包括减少水中存在的无机杂质。为了完成这项任务，P4 的水被泵入反渗透 (RO) 进程 P5。最后，最后一个进程 P6 负责使水可用于分配。

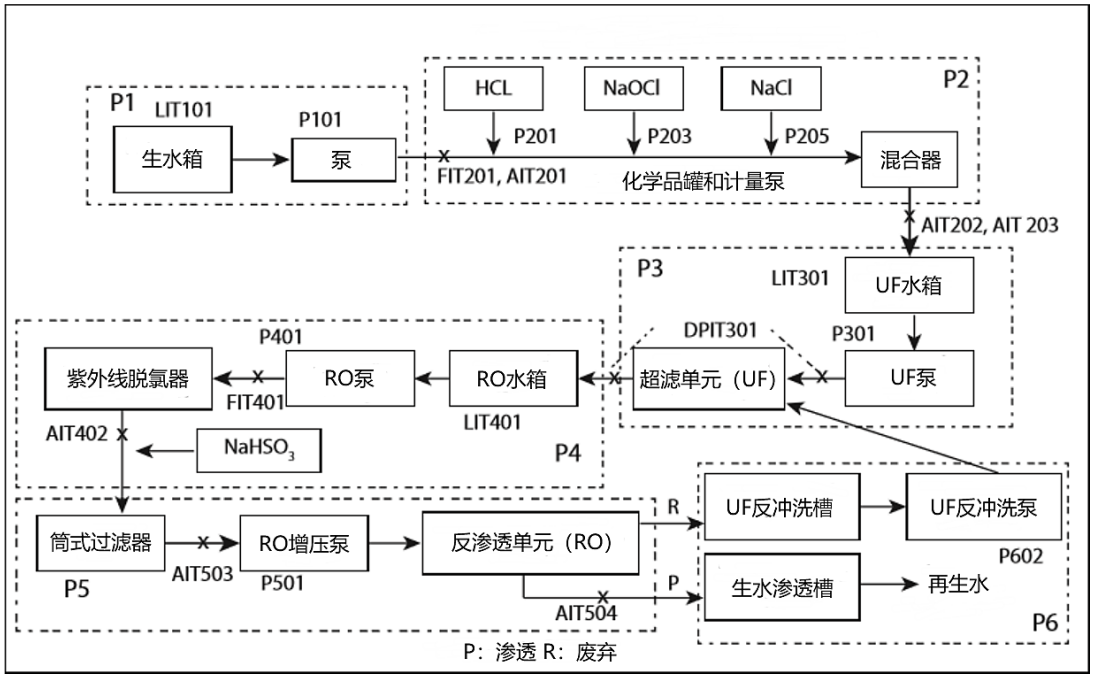


图 3-6 SWaT 测试平台及其流程

3.2.2.2 数据集分析

SWaT 平台在数据收集过程中，共遭受了 36 次模拟攻击，按照攻击过程与攻击目标数量可以将模拟攻击分为以下四类：

（1）单阶段单点：在一个过程中仅针对一个传感器或执行器的网络攻击。

（2）单阶段多点：在一个过程中针对多个传感器或执行器的网络攻击。

（3）多阶段单点：针对多个过程中的一个传感器或执行器的网络攻击。

（4）多阶段多点：针对多个过程中的多个传感器或执行器的网络攻击。

SWaT 数据集每个攻击的类别数目如表 3-13 所示。SCADA 每秒提供一个快照，其中包含 26 个传感器的值和测试台中的 25 个执行器，存储在 CSV 文件中。攻击的持续时间因攻击类型而异。连续进行几次攻击，每次持续 10 分钟，连续攻击之间间隔 10 分钟。一些攻击是通过让系统在后续攻击之前稳定下来执行的。系统稳定的持续时间因攻击而异.一些攻击对系统的动态有更强的影响，并导致系统稳定的时间更长。更简单的攻击，例如那些影响流速的攻击，需要更少的时间来稳定。此外，某些攻击不会立即生效。由于攻击是通过受控过程进行的，因此数据标记非常简单。物理特性的标记与传感器或执行器数据相对应的每个数据项都被单独收集到一个 CSV 文件中。每个 CSV 文件包含服务器名称、传感器名称、该时间点的值、时间戳、可疑、注释和替换。由于属性来自服务器，可疑的、注释的和替换的是多余的，因此被删除。然后将所有剩余数据合并到一个 CSV 文件中。使用攻击日志，根据攻击的开始和结束时间手动标记数据，标签为 Normal / Attack。

表 3-13 每个类别的攻击次数

攻击类型	攻击次数
SSSP	26
SSMP	4
MSSP	3
MSMP	4

3.2.3 仿真设计及结果分析

为了验证基于时空融合模型的 CPS 物理子系统安全态势要素提取的可行性与准确性，本节使用 SWaT 数据集，并按照 3.2.1 节规范的物理安全态势要素提取方法进行仿真实验，具体包括特征处理、数据预处理、具体模型构建、验证与分析四个步骤。

3.2.3.1 特征处理

（1）特征过滤：由 3.2.1 介绍的物理安全态势要素提取方法，在此阶段进行特征的过滤。经过特征的方差分析，删除方差为 0 的特征，分别为['P202', 'P301', 'P401', 'P404', 'P502', 'P601', 'P603']。接下来考虑特征之间的关联，特征之间相关性热图如图 3-7 所示，某些特征对高度相关。例如，相关性最高的特征是 P501 和 UV401，相关性为 99.99%。这是因为当水被脱氯(UV401) 时，它通过泵 P501 传递到 P5。这种类型的相关性在 ICS 中很常见，其中一些传感器或执行器相互依赖。本文中考虑到部分特征是攻击点，所以不过滤这些特征。另外考虑每个特征与结果标签的相关性，特征与结果标签相关性排行如图 3-8 所示。经过分析可得相关性最高的特征是 FIT401，相关性约为 76.33%。此外，如传感器 P403、P602、FIT601、MV303、MV301、P206、P204、AIT202、AIT201、AIT401、AIT504 与标签的相关性不到 10%。考虑到这些特征也是攻击后的表现特征，本文不过滤这些特征。

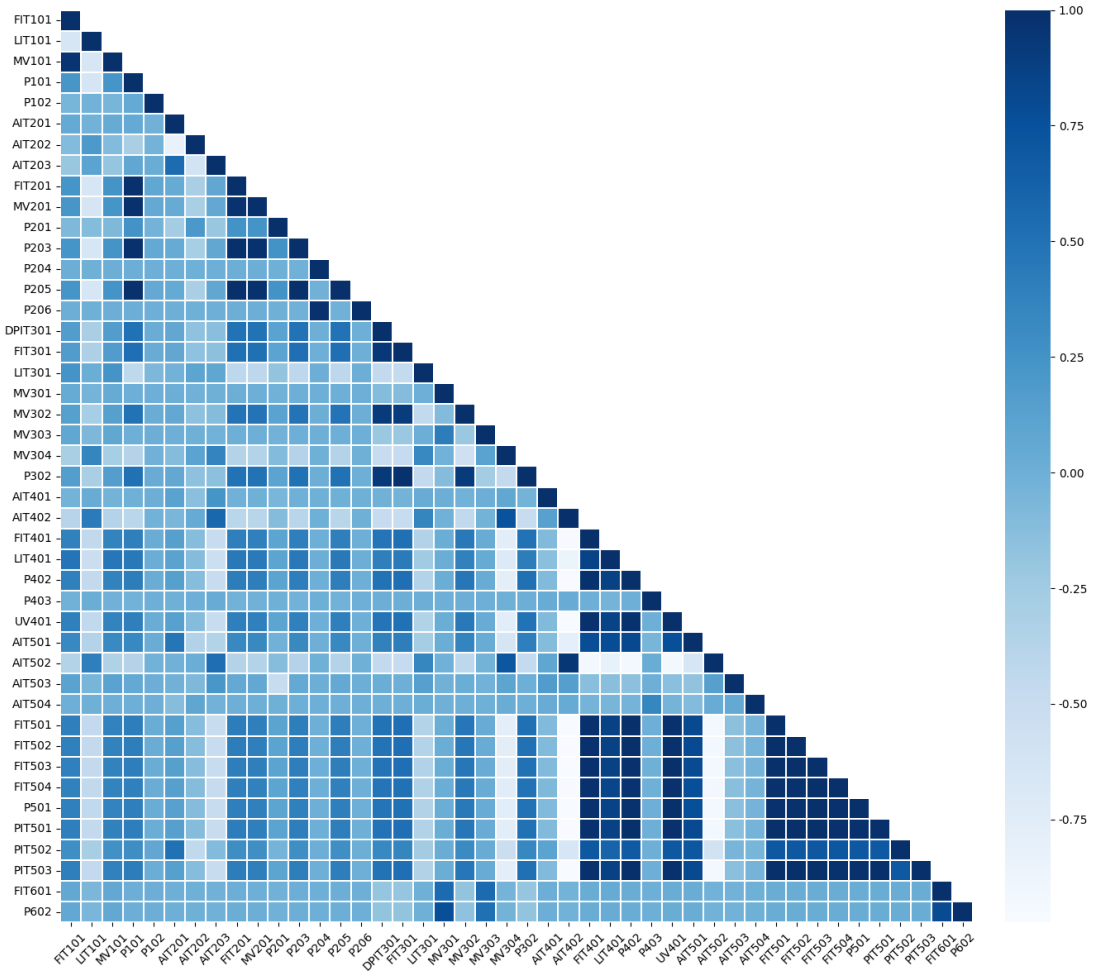


图 3-7 SWaT 数据集特征之间相关性热图（值 1 表示特征高度相关，值-1 表示高度逆相关）

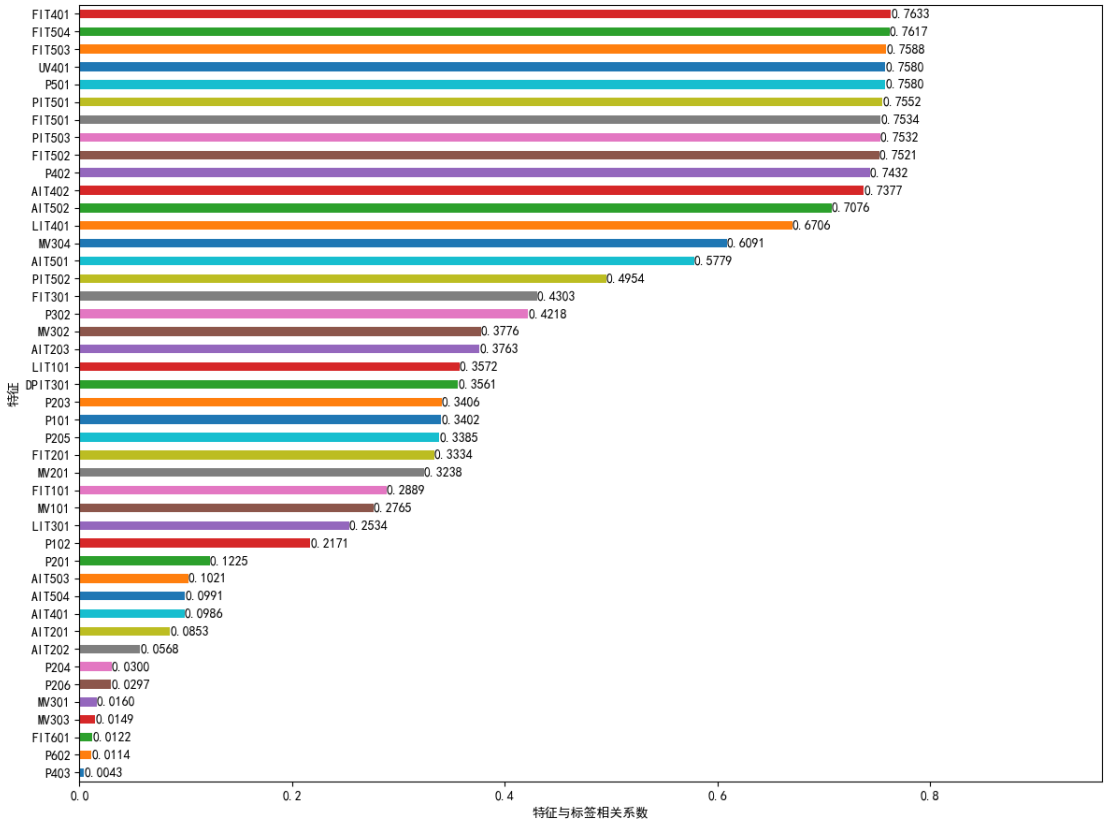


图 3-8 SWaT 数据集特征与结果标签相关性排行图

（2）特征提取：本文使用自相关函数来提取高阶特征，通过自相关特征提取可以提取与传感器和执行器中的信号模式重复相关的特征。自相关定义为信号与自身延迟副本的相关性。本文使用时间步长为W的窗口计算自相关系数，自相关的计算公式如式（3-6）所示，其中x_w表示在w时刻的值，$\hat{x}$是x值在窗口内的平均值。最后将提取的新特征添加到数据集中。

$$autocorr_{x_w,k}=\frac{\sum_{i=w-W+1}^{w-k}\left(x_i-\hat{x}\right)\left(x_{i+k}-\hat{x}\right)}{\sum_{i=w-W+1}^{w}\left(x_i-\hat{x}\right)^2} \tag{3-6}$$

3.2.3.2 数据预处理

对于 SWaT 数据集的数据预处理过程如图 3-9 所示。

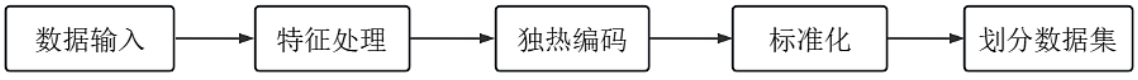


图 3-9 数据预处理过程

（1）检查数据集：首先检查数据集是否包含无效数据，首先独立绘制每个特征的数据图，分析发现在攻击结束后，SWaT 数据集中的样本被过早地标记为 Normal。按照该数据集的标注方法，只有在攻击时间段内的数据才会被标注为 Attack，其实 CPS 物理系统在被攻击后需要一些时间才能再次稳定并达到其正常状态。出于这个原因，本文修改了部分攻击活动之后一段时间内的样本标签，这一段时间内系统还是处于异常状态的。其次是处理缺失值样本，对于连续性缺失值样本采取窗口内平均值的做法；对于离散性缺失值样本做删除处理。

（2）特征编码：对于标签列 Normal / Attack，数据类型为字符串，本文将字符串类型数据数值化，对应关系为 Normal：0，Attack：1。接下来使用独热编码对离散特征进行特征编码，经过分析需独热编码列有['MV101', 'P101', 'P102', 'MV201', 'P203', 'P204', 'P205', 'P206', 'MV301', 'MV302', 'MV303', 'MV304', 'P302', 'P402', 'P403', 'UV401', 'P501', 'P602']。

（3）数据标准化：由于连续性数值特征的取值范围不同，每条数据中有的数值会相差很大，为了防止数值大的属性会对最后的分类结果造成影响，需要统一数据的基本度量单位。本文中采用的是标准化处理方法，对于数据集中的每个连续特征，减去平均值，并将所得值除以标准差，将结果映射到[0, 1]之间。标准化后的每个特征都具有零均值和单位方差，标准化如公式（3-7）所示：

$$z=\frac{x-\mu}{\sigma} \tag{3-7}$$

（4）划分数据集：对于时间序列而言，本文使用滑动时间窗方法，对数据集进行构造，将时间序列数据转化为监督问题数据，初始滑动窗口大小为 40，每次滑动 1 个时间步。之后才划分数据集，才可以输入到所构造的神经网络之中。构造方法如式（3-8）所示，其中 $x_{i,j}$ 表示 i 时刻特征 j 的取值，y_i 表示 i 时刻的标签结果。k 表示滑动窗口大小。

$$\begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} & y_1 \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,n} & y_m \end{bmatrix} \Rightarrow \begin{bmatrix} \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{k,1} & x_{k,2} & \cdots & x_{k,n} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} x_{m-k+1,1} & x_{m-k+1,2} & \cdots & x_{m-k+1,n} \\ x_{m-k+2,1} & x_{m-k+2,2} & \cdots & x_{m-k+2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m,2} & \cdots & x_{m,n} \end{bmatrix} \end{bmatrix} + \begin{bmatrix} y_k \\ \vdots \\ y_m \end{bmatrix} \tag{3-8}$$

按照本文提到的 CPS 物理安全态势要素提取方法，在划分数据集的时候，测试集中样本应当包含训练集中不存在的攻击类别，且在划分数据集的过程中，应当保存数据集的时间顺序，按时间顺序划分数据集。遵循以上规则，将训练集、验证集和测试集按照 6:2:2 的比例划分，分布如表 3-14 所示：

表 3-14 SWaT 数据集划分布

数据集	样本标签		数据总量
	Normal	Attack	
训练集	237316	32612	269928
验证集	79051	10926	89977
测试集	78893	11083	89976

3.2.3.3 模型构建

按照 3.2.1.3 介绍，物理安全态势要素提取技术分为两大类：基于非监督学习的回归模型、基于监督学习的分类模型。由于本文选取数据集特征多维度，多攻击阶段，若选取基于非监督学习的回归模型会难以指定合适的阈值提取安全态势要素。所以本文选择方法二：基于监督学习的分类模型，该方法具有较高的准确性和稳定性。

图 3-10 展示了本章设计的 CPS 环境中物理安全态势要素提取网络模型示意图。由图可以看出，态势要素提取模型主要由 2 个部分组成，分别是以卷积操作为核心的神经网络结构以及长短时间记忆网络。首先，将预处理的后的数据作为整体模型的输入。之后，经过 CNN-LSTM 网络模型处理。最后经过归一化指数函数（softmax）得到最终的分类结果。

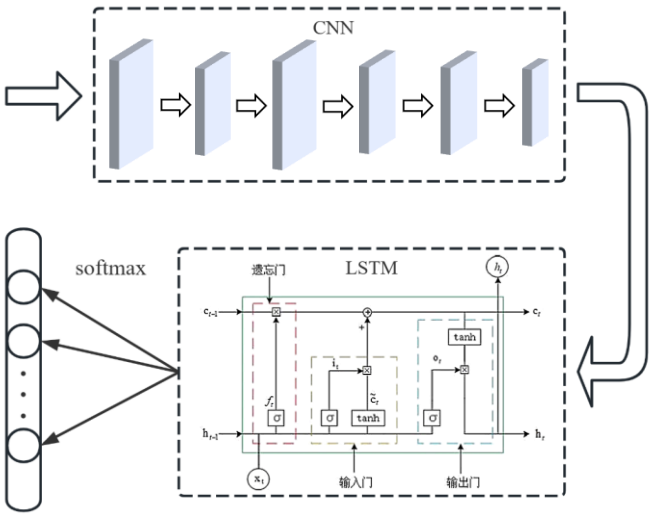


图 3-10 CNN-LSTM 网络模型示意图

对于实验中的每个模型，本文使用 Adam 优化器和相同的训练参数。使用网格搜索方法调参，通过尝试定义的超参数的所有可能值来执行详尽搜索。对于超参数搜索策略的每次迭代，都会使用所选的超参数值训练模型。为了检查超参数的优化效果，选择部分训练集数据作为验证数据集。

在这个任务中，还需要监测训练误差，以确定是否存在欠拟合或过拟合。一方面，当训练和验证误差很高时，就面临欠拟合问题，这意味着 DL 算法无法很好地对数据建模。另一方面，如果训练误差远低于验证误差，就面临过度拟合问题，这意味着模型正在记忆而不是泛化。解决过度拟合的典型方法是正则化网络以限制学习，收集更多数据以及使用交叉验证方法。然而要解决欠拟合，需要增加模型复杂度或增加训练时间。

3.2.3.4 验证与分析

为了训练和评估本文设计的方法模型，本文使用配备 80GB 内存、14 VCPU Intel(R) Xeon(R) Gold 6330 CPU @ 2.00GHz 和 24GB 显存的 RTX 3090 显卡的云服务器作为工作台。

本文按照 3.2.1 节介绍的 CPS 物理安全态势要素提取方法，进行特征处理、数据预处理、模型构建，最后进行验证与分析。处理后的 SWaT 数据集中训练集、验证集和测试集的比例为 6:2:2，CNN-LSTM 网络模型初始窗口大小为 40。鉴于 SWaT 数据集的不平衡性，并且为了便于与其他论文结果进行比较，本文使用精准率、召回率和 F1-分数指标。

实验结果如表 3-15 所示，本文 CNN-LSTM 网络模型的实验结果分别达到了 0.976 的精准率、0.855 的召回率和 0.911 的 F1-分数指标。对于消融实验结果，CNN 与 LSTM 模型的精准率、召回率、F1-分数相差无几，但是都稍差于 CNN-LSTM 模型。由此验证了本文的时空融合模型结合了 CNN 提取空间特征的优点与 LSTM 提取时间特征的优点。

除了消融实验结果对比，本文也与使用 SWaT 数据集的其他论文的不同处理方法的结果进行对比。实验结果表明，本文的方法模型在 F1-分数方面达到了最好的结果，文献[41]的 GAN 模型给出了 0.954 的召回率，该结果优于本文模型，文献[41]的 DNN 模型以 0.982 的精准率超过本文 0.980 的精准率。总的来说，使用本文规范总结的 CPS 物理安全态势要素提取方法的 CNN-LSTM 模型在综合性能上表现优异，精准率与召回率也高于其他方法模型的平均水平。由此验证了本文基于时空融合的 CPS 物理安全态势要素提取方法的可行性和有效性。

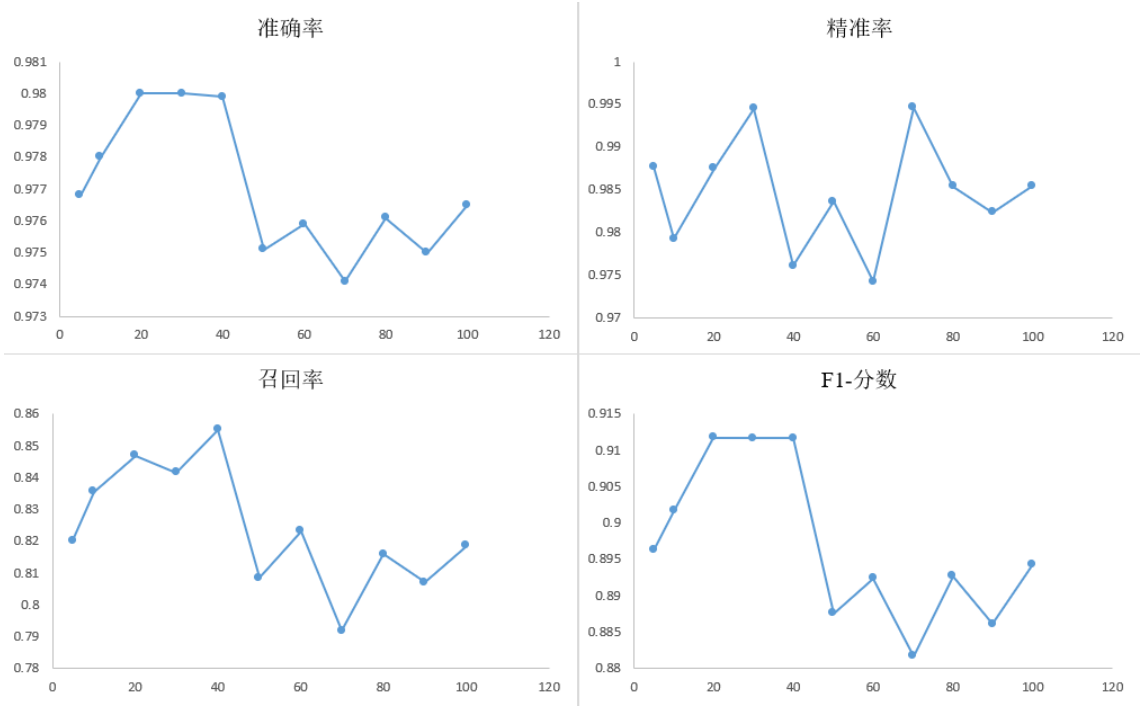
表 3-15 SWaT 数据集不同处理方法结果对比

方法模型	精准率	召回率	F1-分数
1D CNN[33]	0.969	0.791	0.871
DNN[41]	0.982	0.678	0.802
OCSVM[41]	0.925	0.699	0.796
GAN[41]	0.700	0.954	0.810
MLP[42]	0.967	0.696	0.812
RNN[42]	0.936	0.692	0.796
AE[43]	0.924	0.827	0.873
DIF[44]	0.935	0.835	0.882
CNN	0.976	0.810	0.885
LSTM	0.975	0.802	0.880
CNN-LSTM	0.976	0.855	0.911

本文方法模型优于其他方法模型的原因有：对于文献[33]，只使用了 CNN 进行空间特征提取，没有考虑到 CPS 物理数据的时间特征；对于文献[42]作者未对特征进行提取; 对于文献[44]，该作者未对特征进行过滤。本文方法为基于监督学习的分类模型，以上部分模型为基于非监督学习的回归模型，所有在性能评价上略有优势。但是基于监督学习的分类模型需要已知攻击的样本数据，这在实际应用中具有一定的局限性。在某些场景上不如基于非监督学习的回归模型，例如，新部署的 CPS 环境中，新环境中并没有已知的攻击样本，只包含大量的正常样本，基于监督学习的分类模型就没有合适的样本数据，而基于非监督学习的回归模型依然能够工作，这是本文方法的一个局限性。

为了证明滑动窗口大小对 CNN-LSTM 模型性能的影响，本文设置了各种大小的滑动窗口 (k = 5、10、20、30、40、50、60、70、80、90、100)，得到了不同时间步长下的性能指标，包括精准率、召回率和 F1 得分，如图 3-11 所示。随着时间步长的增加，精准率无明显趋势，召回率、F1 得分呈现先上升后下降的趋势。这是因为较长的时间序列可以提供更多的上下文信息，有助于提高分类准确率，但同时也可能导致模型更关注过去的信息而忽略当前的信息，从而降低召回率。

从整体上看，模型在不同时间步长下的性能表现较为稳定，F1 得分在 0.88~0.91 之间波动。其中，在时间步长为 20、30、40 时，F1-分数较高，可以认为这些时间步长能够较好地平衡精确度和召回率。模型在时间步长为小于 20 时表现相对较差，可能是因为这些时间步长过短，难以捕捉到足够的序列信息。模型在时间步长大于 50 时召回率较低，这可能是因为该时间步长过长，使得模型更注重历史信息，而忽略了当前的信息。



3-11 CNN-LSTM 模型在不同滑动窗口下的性能对比

综上，选择合适的时间步长对于模型的性能至关重要。应该根据具体数据集和任务需求选择合适的时间步长，以达到最佳的要素提取效果。

3.3 CPS 安防子系统安全态势要素提取

CPS 安防子系统安全态势要素提取是指从各种监控设备中获取数据，并对这些数据进行处理和分析，从中提取出关键的态势要素，以实现对 CPS 的全面监控和实时预警。CPS 安防系统可获取的态势要素包括：

（1）异常人员检测：通过分析人员身份合法性，检测出异常行为，如越界、停留等，以实现对工作人员权限管理。

（2）人员和车辆的轨迹：通过分析监控视频，提取出人员和车辆的位置和移动轨迹，以便对其进行实时追踪和分析。

（3）安全事件检测：通过分析监控视频，检测出各种安全事件，如火灾、泄漏等，以实现对系统物理设备运行状态的视频监控和实时预警。

考虑到实验条件，本文在 CPS 安防子系统中提取人员信息作为态势要素。

3.3.1 仿真实验设计

在 CPS 环境下，人员异常情况可能会对系统的稳定性和安全性产生负面影响，人员异常信息包括恶意攻击、非法人员入侵、合法人员跨区域访问。合法人员的

恶意访问无法避免，非法人员入侵、合法人员跨区域访问均可通过人脸识别检测。文献[45]研究了 CPS 环境下多姿态人脸识别算法，文献[46]在此基础上完成人员监测场景建模。本文在前人的研究基础上设计 CPS 安防子系统态势要素提取仿真实验。

本文使用文献[46]中使用到的数据源，为若干段涵盖 4 名测试人员在 CPS 环境下的监控视频以及 VGGFace2 数据集。数据预处理阶段，提取 CPS 环境下的监控视频中的多姿态人脸数据作为合法人员测试数据，并选择 4 张正脸图片作为样本库，从 VGGFace2 数据集中随机选择部分多姿态人脸数据作为非法人员测试数据。由于选取的样本库较小，不适合直接训练神经网络模型，本文使用 FaceNet 开源库提供的 20180402-114759 预训练模型进行迁移学习获取人脸识别模型。由于现实中非法人员样本无法预训练，所以不适合对人脸进行分类识别。该模型将检测到的人脸提取出特征向量，将特征向量与样本库中所有人脸的特征向量计算欧式距离，选取最小的作为距离 d，之后采用阈值法判断人员的合法性。

3.3.2 结果分析

本节按照 3.3.1 节仿真实验设计进行仿真，从 CPS 环境下的监控视频中提取 1000 张多姿态人脸数据作为正样本，并选择 4 张正脸图片作为样本库，从 VGGFace2 数据集中随机选择 1000 张多姿态人脸数据作为负样本。经过人脸识别模型得到正样本、负样本与样本库中正脸图片的最小距离数组。正样本距离数组均值 0.676935，最小值 0.288563，最大值 1.335279。负样本距离数组均值 1.404757，最小值 0.881776，最大值 1.664579。根据 ROC 曲线分析可设置阈值为 1.100000，特征距离小于 1.100000 为正常人员，特征距离大于 1.100000 则为异常人员，最终使得准确率取得最大值 0.997，此时人脸识别混淆矩阵如图 3-12 所示。实验表明，该模型能准确的识别 CPS 环境下的多姿态异常人员，识别结果可作为 CPS 安防子系统的安全态势要素。

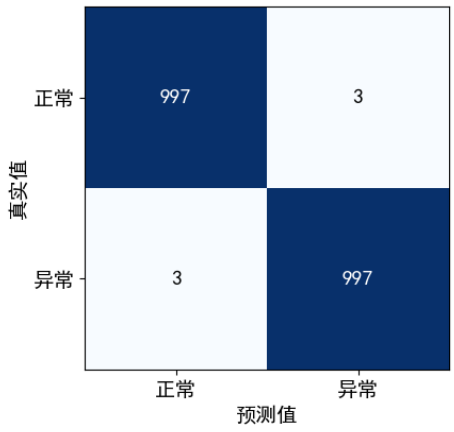


图 3-12 人脸识别混淆矩阵

3.4 本章小结

本章对 CPS 安全态势要素提取模型中三个子系统的安全态势要素提取问题进行研究，分别提出了适合的安全态势要素提取方法，并完成了仿真实验。

对于 CPS 网络子系统，本章基于复合融合模型对 CPS 网络子系统进行安全态势要素提取，并进行了仿真验证，得出了本章模型优于全特征融合模型的结果，验证了本章模型在 CPS 网络子系统安全态势要素提取效果和可行性。在复合融合仿真设计中使用基于决策树算法的特征级融合方法级联决策级融合方法，最终效果优于 D-S 融合算法及其他基于数学统计的方法。本章提出的模型方法适合非时间序列数据，仅从空间维度提取态势要素，适用于 CPS 网络子系统安全态势要素提取场景。

对于 CPS 物理子系统，本章提出了基于时空融合的物理安全态势要素提取方法。该方法重点强调了 CPS 物理数据的空间维度特征和时间维度特征同样重要。本章使用 CNN-LSTM 网络模型，在监督模式下训练的分类器来提取物理安全态势要素。该模型结合了 CNN 提取空间特征的优点和 LSTM 提取时间特征的优点，可以融合提取空间特征和时间特征。在验证与结果分析中，本章的方法模型达到了 0.976 的精准率、0.855 召回率和 0.911 的 F1-分数指标，均高于参考文献中方法模型的平均值，验证了该方法模型的可行性和有效性。

对于 CPS 安防子系统，本章基于人脸识别技术实现安防子系统安全态势要素提取，使用 FaceNet 开源库提供的 20180402-114759 预训练模型进行迁移学习获取人脸识别模型，提取人员信息作为安全态势要素。仿真实验表明，该模型能准确的识别 CPS 环境下的多姿态异常人员，识别结果实现了 99.7%的准确率。

第四章 CPS 安全态势要素提取

CPS 安全态势要素提取根据时间跨度分为整体安全态势要素提取和复合安全态势要素提取。CPS 整体安全态势要素提取指的是根据同一时刻三个子系统的安全态势要素进行 CPS 整体安全态势要素的提取。复合态势要素是指在某一场景或环境中，多个要素相互作用，形成的综合效应，定位为涉及多系统的复合攻击行为或者异常。在 CPS 环境下，安全态势要素包括网络子系统安全态势要素、物理子系统安全态势要素、安防子系统安全态势要素。本章融合 CPS 网络、物理、安防子系统态势要素，提取 CPS 整体安全态势要素及复合安全态势要素。第三章分别提出基于复合融合的 CPS 网络子系统安全态势要素提取模型；基于时空融合的物理安全态势要素提取方法实现 CPS 物理子系统安全态势要素提取；基于人脸识别技术实现安防子系统安全态势要素提取。本章首先基于改进的 D-S 证据理论融合实现 CPS 整体安全态势要素提取，然后设计复合攻击场景，仿真验证 CPS 复合安全态势要素提取方法的正确性。

4.1 CPS 整体安全态势要素提取

CPS 整体安全态势要素提取阶段需要综合考虑网络、物理和安防子系统的安全态势要素信息，以便获取 CPS 全面、准确的安全态势要素。前文已就各子系统分别提取了安全态势要素，具体而言，网络子系统通过复合融合模型方法融合网络流量、系统日志等信息提取网络安全态势要素；物理子系统通过基于 CNN-LSTM 深度学习网络模型融合传感器、执行器等时间序列数据提取物理安全态势要素；安防子系统通过人脸识别技术识别异常人员人脸信息提取安防子系统安全态势要素。本阶段通过 D-S 证据理论融合方法融合各子系统安全态势要素提取 CPS 整体安全态势要素，使之能够更加全面、准确地反映 CPS 安全状态，进而指导决策和应对措施，有助于提高 CPS 安全性和可靠性。

由 2.2.3 节介绍的 D-S 证据理论相关内容可知，使用 D-S 证据理论进行多源信息融合主要分为三步：首先确定识别框架，然后通过基本概率分配函数给焦元分配基本概率值，最后根据组合规则对基本概率值进行融合。流程图如图 4-1 所示：

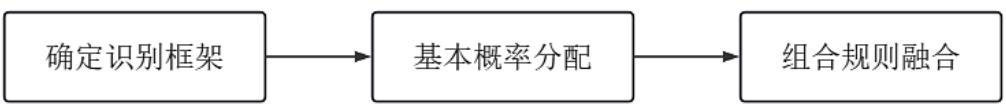


图 4-1 D-S 证据理论融合步骤

4.1.1 确定识别框架

通过 D-S 证据理论融合各子系统安全态势要素提取 CPS 整体安全态势要素，首先要确定 D-S 证据理论的识别框架Θ，即 CPS 整体安全态势要素结果。参考子系统提取的安全态势要素结果，将 D-S 证据理论的识别框架设置为Θ={N，A}，2^{Θ}={{N}，{A}，{N，A}}。对于 CPS 整体态势要素提取，焦元{N}表示 CPS 系统正常，处于低风险；焦元{A}表示 CPS 系统异常，处于高风险；焦元{N，A}表示系统不确定、处于预警状态。

4.1.2 基本概率分配函数构建

建立合适的基本概率分配函数是融合的关键之一，基本概率分配函数的主要作用是将可用的证据或信息分配到假设或命题上，并给出它们在各个假设或命题上的置信度或可信度。D-S 证据理论的融合过程就是根据基本概率分配函数计算多个证据的组合，并给出最终的融合结果的过程。因此，基本概率分配函数是 D-S 证据理论中非常重要的一环，它为信息融合提供了一种灵活而有效的方法，可以在处理不确定性和矛盾性时提供更好的结果。

在本文提出的 CPS 安全态势要素提取模型中，通过学习法提取的网络子系统态势要素可使用决策树输出的各分类概率值作为 D-S 融合中网络子系统证据的基本概率值，通过深度学习网络提取的物理子系统态势要素可使用网络模型输出的二分类值作为 D-S 融合中物理子系统证据的基本概率值。由于本文模型中安防态势要素提取是通过人脸识别技术提取人脸特征向量，再计算与样本库中人脸的距离的最小值，最后根据阈值法确定人员的合法性。所以对于安防子系统态势要素缺乏基本概率值，故需要提出适用于安防子系统的基本概率分配函数。

目前建立基本概率分配函数的方式有以下几种：基于专家系统，专家通过对领域知识的理解和经验，来对假设的可信度进行估计，然后基于经验和直觉来分配基本概率值。基于贝叶斯理论，基于贝叶斯理论的方法可以将先验概率和观测数据结合起来，以获得后验概率分布，然后使用这些分布来分配基本概率值。基于模糊集理论，模糊集理论提供了一种处理不确定性的方法，可以用于建立基本概率分配函数。其中基于专家系统的方法容易受到个人经验和主观意识的影响，容易出现主观性和误判的情况。同时，如果专家的经验和知识不全面或不准确，可能导致基本概率分配函数的准确性下降。基于贝叶斯理论的方法在实际应用中，需要大量的计算和处理，算法复杂度高，时间和资源消耗较大。而基于模糊集理论的方法能够有效处理不确定性问题，减少数据的歧义性和模糊性，对于实际应用中存在的一些模糊信息具有很好的适用性。

由于客观事物的不确定性和复杂性，构建的基本概率分配函数应具有随机性和模糊性。为了实现 D-S 融合中安防子系统证据的基本概率值更好的客观性与有效性，本文基于高斯隶属函数的模糊集理论来构建概率分配函数。

4.1.2.1 模糊集理论介绍

模糊集理论，也称模糊集合论，是由美国数学家 LA Zadeh 于 1965 年在数学上创立的一种描述模糊现象的方法[47]。这种方法把待考察的对象及反映它的模糊概念作为一定的模糊集合，建立适当的隶属函数，通过模糊集合的有关运算和变换，对模糊对象进行分析。该理论是为了解决传统集合论中难以处理模糊性和不确定性问题而提出的。模糊集理论将传统的二元关系扩展到了一个元素与模糊概念之间的关系上，从而能够更好地描述模糊现象。

模糊集理论的形式化定义是一个二元组$(A,\ \mu_A)$，假设存在一个元素集合 A，集合 A 上的一个映射 μ_A 满足公式（4-1）：

$$\mu_A : x \to \mu_A(x),\ x \in A,\ \mu_A(x) \in [0,\ 1] \tag{4-1}$$

则称 μ_A 是从 A 的隶属函数，隶属函数将元素 x 映射为实数 $\mu_A(x)$，表示元素 x 在模糊集中的隶属度，隶属度的范围为单位区间[0, 1]。隶属度取值的多少表示元素 x 属于集合 A 的可能性的大小。$\mu_A(x)$ 取值越大，元素 x 隶属于 A 的可能性越大；$\mu_A(x)$ 取值越小，元素 x 隶属于 A 的可能性越小。

常见的隶属度函数包括三角形隶属度函数、梯形隶属度函数、高斯隶属度函数、S 形隶属度函数和锯齿形隶属度函数等。隶属度函数的一般形式如公式（4-2）所示，其中：$0 \le L(x) \le 1$，$0 \le R(x) \le 1$：

$$\mu_A(x) = \begin{cases} L(x),\ l \le x \le m \\ R(x),\ \mathrm{m} \le x \le \mathrm{r} \end{cases} \tag{4-2}$$

4.1.2.2 基于模糊集理论建立基本概率分配函数

本文基于高斯隶属函数的模糊集理论来构建概率分配函数，高斯隶属函数公式如（4-3）所示：

$$f(x,\sigma,c) = e^{-\frac{(x-c)^2}{2\sigma^2}} \tag{4-3}$$

其中 c 相当于正态分布中的均值，σ 相当于正态分布中的标准差。

对于基于特征距离判别的多分类问题，涉及多个分类结果 $C_i(i=1,2,...,m)$，每个分类结果的元素集合为 $A_i(i=1,2,...,m)$，不同分类结果对应的判别区间可以表示为 $[x_i(L),\ x_i(R)](i=1,2,...,m)$。其中 $x_i(L)$ 为该分类判别区间的最小值，$x_i(R)$

为该分类判别区间的最大值，且 $\forall x_i \in [x_i(L),\ x_i(R)](x_i \in A_i,\ i=1,2,...,m)$ 。对于每个分类结果都定义了对应的高斯隶属函数，每个高斯隶属函数具有两个参数，分别是均值 c，标准差 σ，各参数计算方式如下。

（1） c 的计算：对于分类结果 $C_i(i=1,2,...,m)$ 的隶属函数的均值可以使用该分类结果的元素集合 $A_i(i=1,2,...,m)$ 的均值表示，如公式（4-4）所示：

$$c=\frac{\sum_{x_{ij}\in A_i} x_{ij}}{n_i}(i=1,2,...,m,\ j=1,2,...,n_i) \tag{4-4}$$

（2） σ 的计算：对于分类结果 $C_i(i=1,2,...,m)$ 的隶属函数的标准差计算较为复杂，本文考虑计算的 σ 对于两个相邻分类结果对应的高斯隶属函数在判别区间的阈值处的隶属度相等且都为 0.5。即对于分类结果 C_i 的判别区间为 $[x_i(L),\ x_i(R)]$，分类结果 C_{i+1} 的判别区间为 $[x_{i+1}(L),\ x_{i+1}(R)]$，有 $x_i(R)=x_{i+1}(L)=d$，d 即为两个判别区间的阈值，使得以上任一分类结果的高斯隶属函数在 d 处的隶属度相等且都为 0.5。如公式（4-5）所示：

$$f_i(d,\sigma_i,c_i)=f_{i+1}(d,\sigma_{i+1},c_{i+1})=\frac{1}{2} \tag{4-5}$$

即：

$$e^{-\frac{(d-c)^2}{2\sigma^2}}=\frac{1}{2} \tag{4-6}$$

得：

$$\sigma=\sqrt{\frac{(d-c)^2}{2\ln 2}} \tag{4-7}$$

由于本文关于 c 的计算方法不能保证元素集合 $A_i(i=1,2,...,m)$ 的均值为分类结果对应的判别区间的中值，故对于多分类问题不能保证所有相邻判别区间阈值处的基本概率值相等且都为 0.5，存在一定局限性。一个可行的解决方法为修改 c 的计算方法，使之为分类结果对应的判别区间的中值。由于本文场景为二分类场景且人脸特征距离在判别区间内非均匀分布，故不做修改，使用元素集合 $A_i(i=1,2,...,m)$ 的均值作为 c 更加合理有效。

基于公式（4-3）、公式（4-4）、公式（4-7）计算得到的高斯隶属函数，在分类结果的元素集合 $A_i(i=1,2,...,m)$ 的均值处计算的隶属度最高，在两个相邻分类结果对应的判别区间的阈值处的隶属度相等且都为 0.5。对分类结果进行基本概率分配的情况，隶属度值不适合直接作为基本概率值，会导致各分类的基本概率值

和不为 1。对于此问题，常用的解决方法是通过 x_i 所属判别区间的高斯隶属函数计算得到隶属度 p_i 作为该区间的基本概率值 p_i'，其余区间的基本概率值都为 0，并将多余的概率值分配到 Θ 不确定度上，如公式（4-8）所示，假设各分类结果 $C_i(i=1,2,...,m)$ 对应的高斯隶属函数计算得到的隶属度为 $p_i(i=1,2,...,m)$，则：

$$p_\Theta' = 1 - \sum p_i (i=1,2,...,m) \tag{4-8}$$

本文在基本概率分配阶段不考虑将部分基本概率值分配到 Θ 不确定度上，仅考虑对分类结果进行概率分配，故本文采取的解决方法是不限制判别区间，对于任意 x 值，通过所有判别区间对应的高斯隶属函数计算得到 x 属于各判别区间的隶属度，之后对各分类的隶属度值进行归一化，再将归一化后的值作为基本概率值，如公式（4-9）所示。

$$p_i' = \frac{p_i}{\sum p_i} (i=1,2,...,m) \tag{4-9}$$

4.1.2.3 基本概率分配函数验证

在本小节对 3.3 节提取的安防子系统安全态势要素分配基本概率，一是为了证明本文构建的基本概率分配函数的合理性与正确性，二是为下一步融合提取 CPS 整体态势要素做准备。

在 CPS 整体安全态势要素提取中，安防子系统作为一个信息源，安防子系统提取的态势要素作为 D-S 融合中的证据。由 3.3 节可知，经过人脸识别模型得到正样本、负样本与样本库中正脸图片的最小距离数组中，正样本距离数组均值 0.676935，最小值 0.288563，最大值 1.335279，负样本距离数组均值 1.404757，最小值 0.881776，最大值 1.664579。结合 4.1.1 节确定的识别框架，对安防系统态势要素进行基本概率分配。为保证公式（4-9）的正确性，选取两集合元素均值的中值作为阈值，即 $d=1.040846$。对于分类结果 N 对应的判别区间为[0, 1.040846]，对于分类结果 A 对应的判别区间为[1.040846, $+\infty$]。对于属于 N 的元素集合均值为 0.676935，结合公式（4-4）可得 $c_N=0.676935$，结合公式（5-7）可得 $\sigma_N=0.309078$，结合公式（4-3）可得判别区间[0, 1.040846]所属的高斯隶属函数如公式（4-10）所示：

$$f_N(x) = e^{-\frac{(x-0.676935)^2}{2\times 0.309078^2}} \tag{4-10}$$

同理可得，$c_A=1.404757$，$\sigma_A=0.309078$，判别区间[1.040846, $+\infty$]所属的高斯隶属函数如公式（4-11）所示：

$$f_A(x)=e^{-\frac{(x-1.404757)^2}{2\times 0.309078^2}} \tag{4-11}$$

综上所述，分类结果 N、分类结果 A 的高斯隶属函数如图 4-2 所示：

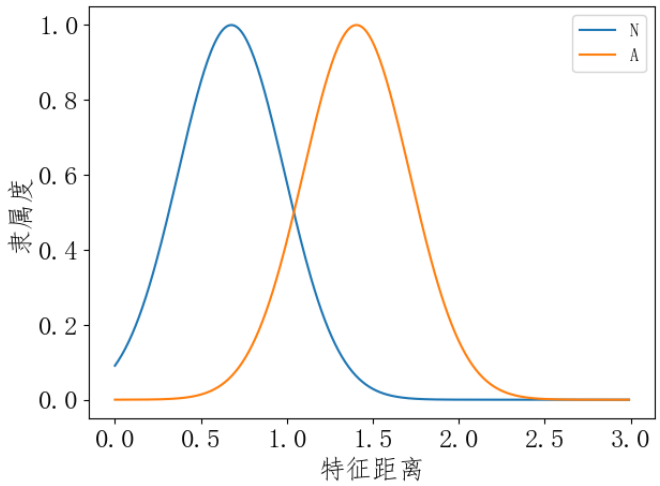


图 4-2 安防态势要素高斯隶属函数

结合公式（4-9）可得安防态势要素基本概率分配函数如图 4-3 所示。当人脸特征距离处于判别区间[0, 1.040846]内，属于正常人员的基本概率处于1～0.5区间内，且下降速度先慢后快。当人脸特征距离处于判别区间[1.040846, +∞]内，属于正常人员的基本概率处于0.5～0区间内，且下降速度先快后慢。

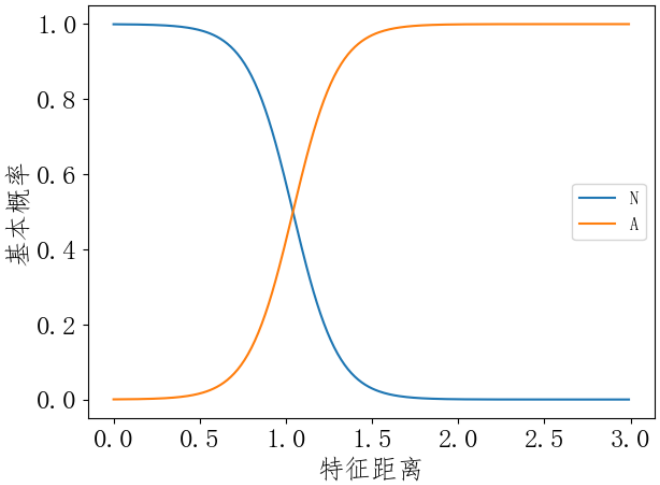


图 4-3 安防态势要素基本概率分配函数

将基本概率大于 0.5 的样本认定为分类正确，则分类结果如表 4-1 所示。其中阈值法为 3.3 节根据 ROC 曲线分析设置阈值的分类结果。为了证明本文概率分配函数的优越性，与文献[46]根据期望偏差定义的基本概率分配函数的方法进行对比。

其做法是通过时间轴滑窗的方式，获取前一段时间该参数历史数据的均值和方差，用来评估参数偏离的概率，再根据观测值与正常情况下均值的偏差量$|x-E(X)|$获取对参数的基本概率分配函数。综合图 4-3 与表 4-1 可证明本文构建的基本概率分配函数对安防态势要素分配基本概率的合理性与正确性

表 4-1 安防态势要素识别结果

方法	TP	FP	TN	FN
阈值法	997	3	997	3
期望偏差法[46]	960	40	999	1
本文方法	981	19	999	1

4.1.3 改进的 D-S 证据理论

D-S 证据理论原始的组合规则是 Dempster 公式，它综合了来自多传感器的基本信度分配，得到一个新的信度分配作为输出。Dempster 组合规则因为冲突因子 K 的引入带来了许多问题，例如一般冲突问题、一票否决问题等。一般冲突问题即当冲突因子 K 值很大时，Dempster 组合规则对证据融合判决的最终结果可能与直觉相悖。且证据间的冲突程度越大，融合结果的不确定性越高，因此 Dempster 组合规则对这些证据的融合失去了意义。一票否决问题即如果证据中某一个焦元的基本概率值为零时，即该证据对该命题完全不支持，那么即使其他证据支持该命题，融合结果对该命题的支持度还是为零。因此一票否决问题表示只要存在证据对某一命题的焦元为零，则会使其他证据的对该命题的支持失去作用。

由于传统的 D-S 证据理论存在不足，所以改进 D-S 证据理论很有必要。为解决上述问题，许多学者都提出了改进 D-S 证据理论的方法。改进方法主要有两个方向，一是修正证据源，二是修改组合规则。修正证据源的方法有 Murphy 方法[48]等。Murphy 认为既然证据存在冲突，那就对证据进行某种修改来降低冲突从而可以使用组合规则来得到相对正确的结果。Murphy 方法主要是在证据组合之前将所有证据进行平均，然后对修正后的证据源使用 D-S 证据理论，这一方法可以有效避免高冲突证据情况。修改 D-S 证据理论对冲突证据组合规则的有 Yager 方法[49]等。Yager 认为原始的 D-S 证据组合规则存在缺陷从而导致出现违背常理的结果，所以在合成冲突证据时，将冲突的那部分概率赋给 Θ 或额外命题 X。该方法在某种程度上可以解决以上问题，且能够提高不确定事件的融合概率，扩展融合结果。

为了在提取 CPS 整体态势要素过程中解决子系统要素高冲突、子系统误报、态势要素结果单一等问题，本文在以上方法的基础上，提出基于 BJS (Belief Jensen–Shannon)散度修正证据源并将部分概率分配给 Θ 的改进 D-S 证据理论方法。

BJS 散度通过结合 JS 散度和信念熵，同时获得证据的可靠性和相对重要性，因此可以表示信息的不确定性，具有灵活性和有效性的优点[50]。BJS 散度计算公式如（4-12）所示。

$$BJS\left(m_1,m_2\right)=\frac{1}{2}\left[\sum_i m_1\left(A_i\right)\log\left(\frac{2m_1\left(A_i\right)}{m_1\left(A_i\right)+m_2\left(A_i\right)}\right)+\sum_i m_2\left(A_i\right)\log\left(\frac{2m_2\left(A_i\right)}{m_1\left(A_i\right)+m_2\left(A_i\right)}\right)\right] \tag{4-12}$$

假设在识别框架$\Theta=\{A,\ B\}$下有两个信息源，其基本概率分配函数分别为m_1和m_2，两条证据如（4-13）所示，则随着参数α在[0, 1]内变化，BJS 散度度量的变化如图 4-4 所示。

$$\begin{aligned}&\mathrm{m}_1: m_1(A)=\alpha,\ m_1(B)=1-\alpha\\&\mathrm{m}_2: m_2(A)=0.999999,\ m_2(B)=0.000001\end{aligned} \tag{4-13}$$

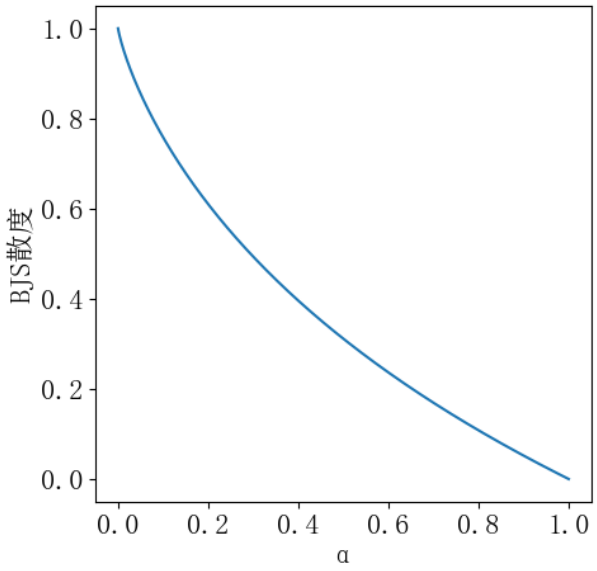


图 4-4 具有变化参数α的 BJS 散度示例

由图 4-4 可知，随着α趋于 1，m_1和m_2之间的 BJS 散度趋于 0，表示在这种情况下m_1和m_2具有相似性，冲突因子较小。在α接近于 0 的情况下，m_1和m_2之间的 BJS 散度度量趋于 1，表示m_1和m_2具有冲突性，冲突因子较大。因此本例验证了 BJS 衡量证据冲突的合理性。

本文提出的基于 BJS (Belief Jensen–Shannon)散度修正证据源并将部分概率分配给Θ的改进 D-S 证据理论方法主要分为五步。假设识别框架为Θ，存在k个焦元，l个证据源，每个证据源对焦元j的基本概率分配为$m_{ij}(i=1,2,...l,\ j=1,2,...k)$。

第一步先计算初始证据源对所有焦元的基本概率均值，如公式（4-14）所示：

$$\mathbf{m}_{\mathbf{avg}} = \frac{\sum \mathbf{m}_{\mathbf{i}}}{l}(i = 1, 2, ...l) \tag{4-14}$$

其中$\mathbf{m}$表示证据基本概率向量，如公式（4-15）所示：

$$\mathbf{m}_{\mathbf{i}} = \begin{bmatrix} m_{i1} \\ m_{i2} \\ \vdots \\ m_{ik} \end{bmatrix}(i = 1, 2, ..., l) \tag{4-15}$$

第二步根据公式（4-12）计算各初始证据源基本概率与证据基本概率均值的 BJS 散度，得到 BJS 散度向量如（4-16）所示：

$$\mathbf{BJS} = \begin{bmatrix} BJS_1 \\ BJS_2 \\ \vdots \\ BJS_l \end{bmatrix}(i = 1, 2, ..., l) \tag{4-16}$$

第三步将初始证据源与证据均值的 BJS 散度作为权重，对初始证据源进行加权处理，以此降低初始证据源冲突因子，如公式（4-17）所示：

$$\mathbf{m}_{\mathbf{i}} = (1 - BJS_i) \times \begin{bmatrix} m_{i1} \\ m_{i2} \\ \vdots \\ m_{ik} \end{bmatrix}(i = 1, 2, ..., l) \tag{4-17}$$

第四步将多余的概率值分配给 Θ，扩展融合结果，如公式（4-18）所示：

$$m_{\Theta} = BJS_i = 1 - \sum m_{ij}(i = 1, 2, ...l,\ j = 1, 2, ...k) \tag{4-18}$$

最后通过 Dempster 组合规则处理修正后的证据源得到最终的融合结果。

4.1.4 仿真设计及结果分析

本文使用 D-S 证据理论融合方法融合各子系统安全态势要素实现 CPS 整体安全态势要素提取，以便获取 CPS 全面、准确的安全态势要素。以上工作已经给出了 CPS 整体安全态势要素提取的识别框架、基本概率分配以及改进的 D-S 证据理论，本节对 CPS 整体安全态势要素提取进行仿真设计及结果分析。CPS 整体安全态势要素提取仿真实验具体实现流程如图 4-5 所示：

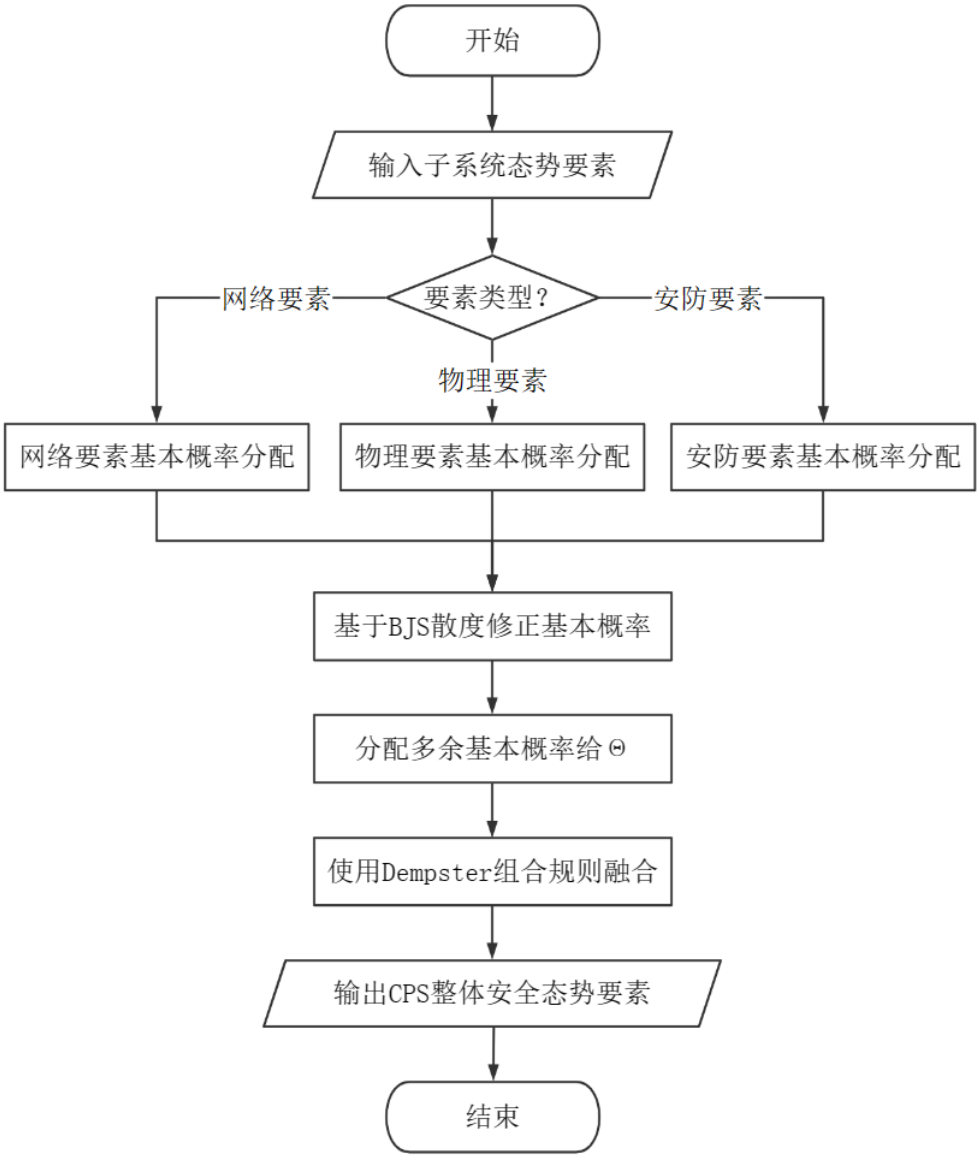


图 4-5 CPS 整体态势要素提取流程图

首先，给网络子系统、物理子系统、安防子系统获取的安全态势要素分配基本概率。对于网络子系统安全态势要素，使用决策树输出的各分类概率值作为网络子系统证据源的基本概率值，并将五分类转化为二分类，以适配 CPS 整体态势要素提取的识别框架。对于物理子系统安全态势要素，使用 CNN-LSTM 深度学习网络模型输出值作为物理子系统证据源的基本概率值。对于安防子系统安全态势要素，使用 4.1.1 节提出的基于高斯隶属函数的模糊集理论构建的概率分配函数对人脸特征距离分配基本概率。

然后从网络子系统证据源、物理子系统证据源、安防子系统证据源中各随机抽取 80 条证据全组合，可得到 51200 个用于提取 CPS 整体态势要素的实验样本。

最后根据 4.1.2 节提出的基于 BJS 散度修正证据源并将部分概率分配给Θ的改进 D-S 证据理论方法提取 CPS 整体安全态势要素。总共进行十次实验取平均值，最终实验结果如表 4-2 所示。为了证明改进的 D-S 证据理论的优越性，本文还进行了原始的 D-S 证据理论实验。实验结果如表 4-3 所示。

表 4-2 改进的 D-S 证据理论提取 CPS 整体安全态势要素结果

子系统异常数	低风险	高风险	预警
0	108178	0	0
1	209726	9603	29919
2	24827	98858	24800
3	0	6085	0

表 4-3 原始的 D-S 证据理论提取 CPS 整体安全态势要素结果

子系统异常数	低风险	高风险	预警
0	109140	0	0
1	232601	16314	0
2	25838	121710	0
3	0	6395	0

改进的 D-S 证据理论提取 CPS 整体安全态势要素结果显示，对于子系统要素异常个数为 0 的情况，所有 CPS 整体态势要素结果都为正常，即低风险。对于子系统要素异常个数为 1 的情况，大部分 CPS 整体态势要素结果为正常，即低风险，少部分为异常，即高风险，少部分提取为不确定，即预警。由此可知，改进的 D-S 证据理论可有效解决子系统误报问题、并扩展态势要素融合提取结果。对于子系统要素异常个数为 2 的情况，大部分 CPS 整体态势要素结果为异常，即高风险，少部分为正常，即低风险，少部分提取为不确定，即预警。对于子系统要素异常个数为 3 的情况，所有 CPS 整体态势要素结果都为异常，即高风险。原始的 D-S 证据理论提取 CPS 整体安全态势要素结果显示，由于原始的 D-S 证据理论未分配基本概率给Θ，所以 CPS 整体态势要素结果无预警结果。对于子系统异常数为 0、1 的情况，与本文方法结果一致。对于子系统要素异常个数为 1 的情况，相较于本文方法，具有更高的概率融合为异常，即高风险。对于子系统要素异常个数为 2 的情况，相较于本文方法，也具有更高的概率融合为异常，即高风险。由此可见本文方法可以缓解子系统要素高冲突，子系统误报、扩展态势要素融合结果的优点。综上，本文证明了改进的 D-S 证据理论对于 CPS 整体安全态势要素提取的合理性与正确性以及相较于原始的 D-S 证据理论的优越性。

4.2 CPS 复合安全态势要素提取

CPS 复合安全态势要素提取定位为能正确识别涉及多系统的复合攻击行为或者异常，为态势要素理解、态势要素预测提供基础。本节与 4.1 节的差异在于前者融合的子系统态势要素提取的结果在时间上具有异步性，后者是融合同一时刻各子系统安全态势要素提取的结果。由于缺乏真实的复合攻击行为数据，本文基于震网病毒进行复合攻击场景设计。最后基于 4.1.3 节改进的 D-S 证据理论进行 CPS 复合安全态势要素提。

4.2.1 复合攻击场景设计

复合攻击是指通过利用多种攻击方式来实现对目标系统的攻击。由于实验条件的限制，缺乏真实的复合攻击行为数据，无法在实际环境中进行复合攻击实验。本文基于震网病毒攻击流程设计本文的复合攻击场景。震网病毒是世界上首个专门针对工业控制系统编写的破坏性病毒，能够利用对 windows 系统和西门子 SIMATIC WinCC 系统的 7 个漏洞进行攻击。特别是针对西门子公司的 SIMATIC WinCC 监控与数据采集 (SCADA) 系统进行攻击。震网病毒最初由工作人员通过 U 盘进行传播，随后病毒修改 PLC 控制软件代码，最终引起物理设备的异常。本文参考震网病毒设计复合攻击场景为以操作人员的非法操作开始，继而引发网络异常，最终由物理设备异常结束。在要素提取层体现为在一定时间间隔内，首先安防子系统识别到攻击或者异常，随后网络子系统识别到攻击或者异常，最后物理子系统识别到攻击或者异常。

考虑到现实情况下，随着时间的推移，各子系统识别到攻击的风险概率会不断下降直到接近零，因此本文引入了时间衰减函数。关于时间衰减函数的选择，正态分布累积概率密度函数以其先平稳，后迅速下降，再平稳的波形特点，适合作为本文引入的时间衰减函数，时间衰减函数如图 4-6 所示，其中窗口大小为 10。

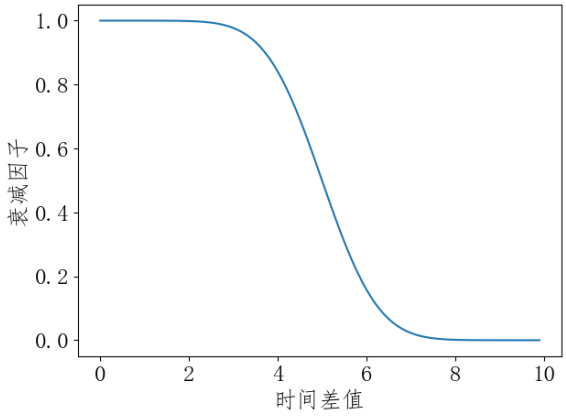


图 4-6 时间衰减函数

基于时间衰减函数 f 来计算 t_0 时刻子系统的攻击风险概率 p_0 在当前时刻 t 的剩余攻击风险概率 p，计算公式如（4-19）所示，最终基于剩余攻击风险概率进行融合，以达到识别复合攻击的目的。

$$p = f(t - t_0) \times p_0 \tag{4-19}$$

4.2.2 仿真设计及结果分析

由于缺乏真实场景的数据集，所以利用公开数据库，分别从网络、物理、人员维度仿真得到子系统态势要素。分别利用 3.1 节 KDD CUP99 数据集的仿真结果作为复合态势要素中的网络安全态势要素，3.2 节 SWaT 数据集的仿真结果作为复合态势要素中的物理安全态势要素，3.3 节的仿真结果作为复合态势要素中的安防安全态势要素。

首先需要对各类态势要素进行预处理。对于网络安全态势要素，由五分类处理成二分类，仅考虑网络正常与异常两种情况，随机抽取 2000 个样本作为网络安全态势要素。对于物理安全态势要素，随机抽取 2000 个样本作为物理安全态势要素。对于安防安全态势要素，由于正常人员与异常人员仿真实验是分开做的，故将实验结果随机打乱合并，共 2000 个样本作为安防安全态势要素。最后将三类安全态势要素对齐，作为复合安全态势要素提取仿真的实验数据集；然后根据 4.2.1 节设计的复合攻击场景对以上数据集进行仿真实验，基于 4.1.3 节改进的 D-S 证据理论进行 CPS 复合安全态势要素提；最后通过对比同一时刻的 CPS 整体态势要素提取结果与加入 CPS 复合安全态势要素提取结果的差异，来验证能否发现数据集中的复合攻击，最终说明本文 CPS 复合安全态势要素提取方法的可行性与有效性。

使用原始 D-S 证据理论融合方法进行 10 次实验的平均结果如表 4-4 所示，使用改进 D-S 证据理论融合方法 10 次实验的平均结果如表 4-5 所示。

表 4-4 原始的 D-S 证据理论提取 CPS 复合安全态势要素结果

态势要素类型	低风险	高风险	预警
整体	1453	538	0
整体+复合	1335	656	0

表 4-5 改进的 D-S 证据理论提取 CPS 复合安全态势要素结果

态势要素类型	低风险	高风险	预警
整体	1328	453	210
整体+复合	1216	573	202

由表 4-4，表 4-5 分析可得，在原始的 D-S 证据理论融合方法下，使用本节方法提取的高风险安全态势要素较整体安全态势要素高出 21.93%；在改进的 D-S 证据理论融合方法下，这一指标高出 26.49%。实验结果表明，在两种 D-S 融合方法下，都不同程度地识别到 4.2.1 节设计的复合攻击场景，最终验证了本文 CPS 复合安全态势要素提取方法在该场景下的可行性与有效性。

4.3 本章小结

本章主要介绍了 CPS 整体安全态势要素提取和复合态势要素提取的方法和实验结果。首先，提出了改进的 D-S 证据理论，用于 CPS 整体安全态势要素提取。然后根据前文仿真实验提取的网络子系统安全态势要素、物理子系统安全态势要素与安防子系统安全态势要素进行 CPS 整体安全态势要素的提取仿真实验。其次，在 CPS 复合态势要素提取方面，本文基于震网病毒攻击流程设计复合攻击场景，用来验证复合态势要素提取方法的准确性。实验结果表明，本文的方法能够准确地实现 CPS 整体安全态势要素提取和复合态势要素提取，为态势要素理解、态势要素预测提供基础。

第五章 总结与展望

5.1 全文总结

本文基于多源信息融合技术实现 CPS 安全态势要素提取，根据 CPS 所面临的安全威胁来源将 CPS 划分为网络子系统、物理子系统和安防子系统，并针对三个子系统分别提出不同的态势要素提取方法，分别提取 CPS 网络安全态势要素、物理安态势要素和安防安全态势要素。最后融合 CPS 网络、物理、安防子系统安全态势要素，实现 CPS 整体安全态势要素提取和复合安全态势要素提取，用于评估 CPS 全局安全态势，也为态势感知之后的态势要素理解、态势预测提供基础。本文的主要工作体现在以下几个方面：

（1）对于 CPS 网络子系统，提出了复合融合模型，用来提取网络安全态势要素。CPS 网络子系统涉及计算机网络和通信设施，包含的安全信息来源广泛且异类，可以是网络流量中提取的安全信息，可以是安全软件检测到的安全信息，可以是系统日志信息。本文将每一个安全信息来源看作是一个网络子系统信息源，在每一个信息源内使用特征级融合提取单信息源安全态势要素信息，之后再级联决策级融合提取多信息源的安全态势要素。在 KDD CUP99 数据集仿真实验中，验证了该复合融合模型相较于全信息源特征级融合可以有效规避个别信息源特征不明显的问题，进一步提高了融合效果和精度。

（2）对于 CPS 物理子系统，提出了基于时空融合的物理安全态势要素提取方法。CPS 物理子系统包括传感器和执行器等硬件元件，其安全信息主要包括传感器采集的数据，表示执行器是否执行的标志位。考虑到物理安全信息较为相似且具有时间序列特性，提出了基于时空融合的物理安全态势要素提取方法。具体包括适合物理安全态势要素提取的数据预处理方法以及基于 CNN-LSTM 时空融合的神经网络模型。在 SWaT 数据集仿真实验中，验证了该方法模型的可行性和有效性。

（3）对于 CPS 安防子系统，基于人脸识别技术提取安防安全态势要素。本文使用 FaceNet 开源库提供的 20180402-114759 预训练模型进行迁移学习获取人脸识别模型。在仿真实验中，可有效识别多姿态人脸信息，根据人员的合法性提取安防安全态势要素。

（4）在文章的最后，进行了 CPS 整体安全态势要素提取和复合态势要素提取。首先，基于高斯隶属函数的模糊集理论构建概率分配函数，让 D-S 融合中安防子系统证据的基本概率值具有更好的客观性。然后，通过基于 BJS 散度修正证据源

并将部分概率分配给Θ的改进 D-S 证据理论方法融合 CPS 网络、物理、安防子系统安全态势要素，实现了 CPS 整体安全态势要素提取，解决了在提取 CPS 整体态势要素过程中子系统要素高冲突、子系统误报、态势要素结果单一的问题。最后，根据设计的复合攻击场景，基于改进的 D-S 证据理论提取复合安全态势要素，验证了本文的复合态势要素提取方法的可行性与有效性。

5.2 后续工作展望

虽然本文提出并验证了 CPS 安全态势要素提取的方法模型，但还有很多不足，可以从以下几个方面改善当前工作并展开后续研究：

（1）由于场景限制，未能在真实的 CPS 环境中进行实验验证。同时由于 CPS 数据的多样性与复杂性，也未能找到一个涵盖网络、物理、人员多种维度信息的真实场景数据集，所以本文只能在单独的数据集上进行仿真实验。后续研究可以在条件允许的情况下，使用真实 CPS 环境中采集的数据进行仿真实验，甚至实地部署。

（2）在安防子系统中仅考虑了人员异常信息作为安防子系统的安全态势要素。CPS 安防系统可获取的态势要素包括异常人员检测、人员和车辆的轨迹、安全事件检测等。进一步研究可以提取更多更详细的安全态势要素。

（3）在复合安全态势要素提取仿真实验中，缺乏真实的复合攻击数据，只能通过已有数据集构造复合安全态势要素提取仿真实验数据集。其次为进行复合态势要素提取的数据集维度较少，可设计的复合攻击场景较为简单。后续研究可以扩展态势要素数据集维度，设计更多更复杂的复合攻击场景。

参考文献

[1] 李必信, 周颖. 信息物理融合系统导论[M]. 科学出版社, 2014.

[2] 周济. 智能制造是“中国制造 2025”主攻方向[J]. 企业观察家, 2019(11): 54-55.

[3] Cherdantseva Y, Burnap P, Blyth A , et al. A review of cyber security risk assessment methods for SCADA systems[C]. Elsevier Advanced Technology Publications, 2016, 56.

[4] Staggs J, Ferlemann D, Shenoi S. Wind farm security: attack surface, targets, scenarios and mitigation[J]. International Journal of Critical Infrastructure Protection, 2017, 17: 3-14.

[5] 刘烃, 田决, 王稼舟等. 信息物理融合系统综合安全威胁与防御研究[J]. 自动化学报, 2019, 45(01): 5-24.

[6] Endsley M R. Design and evaluation for situation awareness enhancement[C]. Proceedings of the Human Factors Society Annual Meeting. Sage CA: Los Angeles, CA: Sage Publications, 1988, 32(2): 97-101.

[7] Endsley M R. Measurement of situation awareness in dynamic systems[J]. Human Factors, 1995, 37(1): 65-84.

[8] Duan Y, Li X, Yang X, et al. Network security situation factor extraction based on random forest of information gain[C]. Proceedings of the 4th International Conference on Big Data and Computing. 2019: 194-197.

[9] Tao X, Kong K, Zhao F, et al. An efficient method for network security situation assessment[J]. International Journal of Distributed Sensor Networks, 2020, 16(11): 1550147720971517.

[10] 赖积保, 王慧强, 郑逢斌, 等. 基于 DSimC 和 EWDS 的网络安全态势要素提取方法[J]. 计算机科学, 2010, 37(11): 64-69.

[11] 司成, 张红旗, 汪永伟, 等. 基于本体的网络安全态势要素知识库模型研究[J]. 计算机科学, 2015, 42(5): 173-177.

[12] 朱江, 明月, 王森. 基于深度自编码网络的安全态势要素获取机制[J]. 计算机应用, 2017, 37(3): 771-776.

[13] 寇广, 王硕, 张达. 基于深度堆栈编码器和反向传播算法的网络安全态势要素识别[J]. 电子与信息学报, 2019, 41(09): 2187-2193.

[14] 任守纲, 张景旭, 顾兴健等. 时间序列特征提取方法研究综述[J]. 小型微型计算机系统, 2021, 42(02): 271-278.

[15] 杜刚. 基于多系统数据级融合的煤矿监测监控逻辑报警分析[J]. 山西煤炭, 2017, 37(5): 13-21.

[16] 冀少军. 多传感器数据融合技术在煤矿瓦斯预警中的应用研究[D]. 邯郸: 河北工程大学, 2016.

[17] Turso J A, Litt J S. A foreign object damage event detector data fusion system for turbofan engines[J]. Journal of Aerospace Computing, Information, and Communication, 2005, 2(7): 291-308.

[18] Rouben N O. The application of fuzzy logic to the construction of the ranking function of information retrieval[J]. Computer Modelling and New Technologies, 2006, 10(1): 20-27.

[19] 温迪. 基于 D－S 证据理论的航空发动机气路故障信息融合与 FMECA 分析[D]. 成都: 电子科技大学, 2014.

[20] 杨亚军, 王福明. 基于 D－S 的小波神经网络信息融合方法[J]. 弹箭与制导学报, 2006, 26(4): 266-268.

[21] 张崇兴. 基于预测和 D-S 证据理论的多传感器数据融合研究[D]. 电子科技大学, 2021.

[22] 叶宏, 曹学军, 李军, 等. 基于 FISST 的多源异类信息配准算法[J]. 太赫兹科学与电子信息学报, 2014, 12(06): 865-869.

[23] 宋绪靖. 基于文本、语音和视频的多模态情感识别的研究[D]. 山东大学, 2019.

[24] 唐德权, 史伟奇, 张波云. 基于多模态信息特征融合的犯罪预测算法研究[J]. 计算机应用与软件, 2018, 35(07): 221-225+262.

[25] Steinberg A N, Bowman C L, White F E. Revisions to the JDL data fusion model[J]. Proceedings of SPIE - The International Society for Optical Engineering, 1999, 3719(01):1-17.

[26] 孙力帆. 多传感器信息融合理论技术及应用[M]. 中国原子能出版社, 2019.

[27] Staudemeyer R C, Omlin C W. Extracting salient features for network intrusion detection using machine learning methods[J]. South African computer journal, 2014, 52(1): 82-96.

[28] 梁杰, 陈嘉豪, 张雪芹, 等. 基于独热编码和卷积神经网络的异常检测[J]. 清华大学学报: 自然科学版, 2019, 059(007):523-529.

[29] 陈岑, 李暖暖, 蔡军飞, 等. 基于动态行为特征加权聚类的加壳恶意软件未知变种检测方法[J]. 重庆大学学报, 2023, 46(03): 129-136.

[30] 李辉, 管晓宏, 昝鑫等. 基于支持向量机的网络入侵检测[J]. 计算机研究与发展, 2003(06): 799-807.

[31] Vávra J, Hromada M. Comparison of the intrusion detection system rules in relation with the SCADA systems[C]. Software Engineering Perspectives and Application in Intelligent Systems: Proceedings of the 5th Computer Science On-line Conference 2016 (CSOC2016), Vol 2 5. Springer International Publishing, 2016: 159-169.

[32] Kleinmann A, Wool A. A statechart-based anomaly detection model for multi-threaded SCADA systems[C]. Critical Information Infrastructures Security: 10th International Conference, CRITIS 2015, Berlin, Germany, October 5-7, 2015, Revised Selected Papers. Cham: Springer International Publishing, 2016: 132-144.

[33] Kravchik M, Shabtai A. Detecting cyber attacks in industrial control systems using convolutional neural networks[C]. Proceedings of the 2018 workshop on cyber-physical systems security and privacy. 2018: 72-83.

[34] Zizzo G, Hankin C, Maffeis S, et al. Intrusion Detection for Industrial Control Systems: Evaluation Analysis and Adversarial Attacks[J]. arXiv Preprint arXiv: 1911.04278, 2019.

[35] Li D, Chen D, Jin B, et al. MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks[C]. Artificial Neural Networks and Machine Learning–ICANN 2019: Text and Time Series: 28th International Conference on Artificial Neural Networks, Munich, Germany, September 17–19, 2019, Proceedings, Part IV. Cham: Springer International Publishing, 2019: 703-716.

[36] Khan A A Z. Misuse intrusion detection using machine learning for gas pipeline scada networks[C]. Proceedings of the international conference on security and management (SAM). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2019: 84-90.

[37] Alhaidari F A, AL-Dahasi E M. New approach to determine DDoS attack patterns on SCADA system using machine learning[C]. 2019 International conference on computer and information sciences (ICCIS). IEEE, 2019: 1-6.

[38] Gómez Á L P, Maimó L F, Celdrán A H, et al. On the generation of anomaly detection datasets in industrial control systems[J]. IEEE Access, 2019, 7: 177460-177473.

[39] Cheng M, Xu Q, Jianming L V, et al. MS-LSTM: A multi-scale LSTM model for BGP anomaly detection[C]. 2016 IEEE 24th International Conference on Network Protocols (ICNP). IEEE, 2016: 1-6.

[40] Wang Y, Du X, Lu Z, et al. Improved lstm-based time-series anomaly detection in rail transit operation environments[J]. IEEE Transactions on Industrial Informatics, 2022, 18(12): 9027-9036.

[41] Inoue J, Yamagata Y, Chen Y, et al. Anomaly detection for a water treatment system using unsupervised machine learning[C]. 2017 IEEE international conference on data mining workshops (ICDMW). IEEE, 2017: 1058-1065.

[42] Shalyga D, Filonov P, Lavrentyev A. Anomaly detection for water treatment system based on neural network with automatic architecture optimization[J]. arXiv Preprint arXiv:1807.07282, 2018.

[43] Kravchik M, Shabtai A. Efficient cyber attack detection in industrial control systems using lightweight neural networks and pca[J]. IEEE Transactions on Dependable and Secure Computing, 2021, 19(4): 2179-2197.

[44] Elnour M, Meskin N, Khan K, et al. A dual-isolation-forests-based attack detection framework for industrial control systems[J]. IEEE Access, 2020, 8: 36639-36651.

[45] 陈律. CPS 中目标识别关键技术应用研究[D]. 电子科技大学, 2020.

[46] 王昱人. CPS 环境下异类信息融合技术应用研究[D]. 电子科技大学, 2021.

[47] Zadeh L A. Fuzzy sets[J]. Information and Control, 1965, 8(3): 338-353.

[48] Lefevre E, Colot O, Vannoorenberghe P. Belief function combination and conflict management[J]. Information Fusion, 2002, 3(2): 149-162.

[49] Yager R R. On the Dempster-Shafer framework and new combination rules[J]. Information Sciences, 1987, 41(2): 93-137.

[50] Xiao F. A new divergence measure for belief functions in D–S evidence theory for multisensor data fusion[J]. Information Sciences, 2020, 514: 462-483.