**2024年全国大学生信息安全竞赛**

**需求分析说明书**

**作品名称：**  Fed Privacy Shield

**电子邮箱：** 2948953101@qq.com

**提交日期：** 2024年6月5号

YOUR **COMPANY NAME**

**目录**

**[第一章 作品概述](#_Toc14518)** [1](#_Toc14518)

[1.1作品背景及研发意义 1](#_Toc9687)

[1.2相关工作 3](#_Toc2810)

[1.2.1联邦学习隐私保护现状 4](#_Toc2736)

[1.2.2联邦学习助力隐私保护 4](#_Toc8224)

[1.2.3基于边缘计算的联邦学习隐私保护研究 6](#_Toc31820)

[1.3作品简介 6](#_Toc17342)

[1.4创新性综述 7](#_Toc1331)

[1.4.1平台方便高效： 7](#_Toc13927)

[1.4.2 平台智能共享 8](#_Toc19273)

[1.4.3平台安全可信 8](#_Toc14230)

[1.5 应用前景 9](#_Toc17535)

[1.6 本章小结 10](#_Toc28737)

**[第二章 需求分析与确定](#_Toc23992)** [11](#_Toc23992)

[2.1需求概述 11](#_Toc19089)

[2.2功能需求 11](#_Toc3577)

[2.2.1 训练模型用户Web端需求 11](#_Toc26839)

[2.2.2 平台管理人员Web端需求 13](#_Toc24624)

[2.3 非功能需求 15](#_Toc15546)

[2.3.2 易用性 15](#_Toc4111)

[2.3.3 健壮性 16](#_Toc2343)

[2.3.4 可扩展性 16](#_Toc31433)

[2.3.5 可移植性 16](#_Toc11417)

[2.4 应用场景 16](#_Toc27318)

[2.4.1 医疗任务场景 16](#_Toc16038)

[2.4.2 车联网 17](#_Toc15662)

[2.4.3 金融服务 19](#_Toc8956)

[2.5 本章小结 19](#_Toc7190)

**[参考文献](#_Toc1322)** [21](#_Toc1322)

第一章 作品概述

本节主要介绍本作品的社会背景以及研发意义，隐私保护现状，相关工作，作品创新性及项目的应用前景。

## 1.1作品背景及研发意义

2021年，全球范围内总共泄露了11亿条个人记录，2022年，这个数字达到了15亿条，2023年全球**数据泄漏**规模创下历史新高，比2022年全年高出20%[1]。

《2023年数据泄露风险年度报告》对2023年数据泄露风险概况、黑产数据交易市场等进行具体分析，数据显示：2023年，全网监测并分析验证有效的数据泄露事件超过19500起，涉及金融、物流、航旅、电商、汽车等20余个行业；金融行业超过物流行业，成为2023年公民个人信息泄露事件数量最多的行业；2023年航旅行业公民个人信息泄露事件数量也出现**大幅增长**，在公民个人信息泄露的行业分布中首次排名前三；图1-1显示了数据泄露行业的主要分类，图1-2展示了2023年数据泄露数量Top5行业。展现了目前隐私数据保护的严峻性和必要性。

在大数据环境下，数据隐私安全至关重要。传统的集中式学习将各方数据收集并进行建模（机器学习、深度学习），但底层未经审核、不透明的数据收集和集合协议，很可能造成严重的**数据安全威胁**和隐私风险。由于网络存储能力巨大，用户在互联网的全部活动都会留下记录，致使攻击者可以通过技术手段对这些数据进行追踪和还原。在如此背景下，隐私信息泄露事件就变得异常普遍。

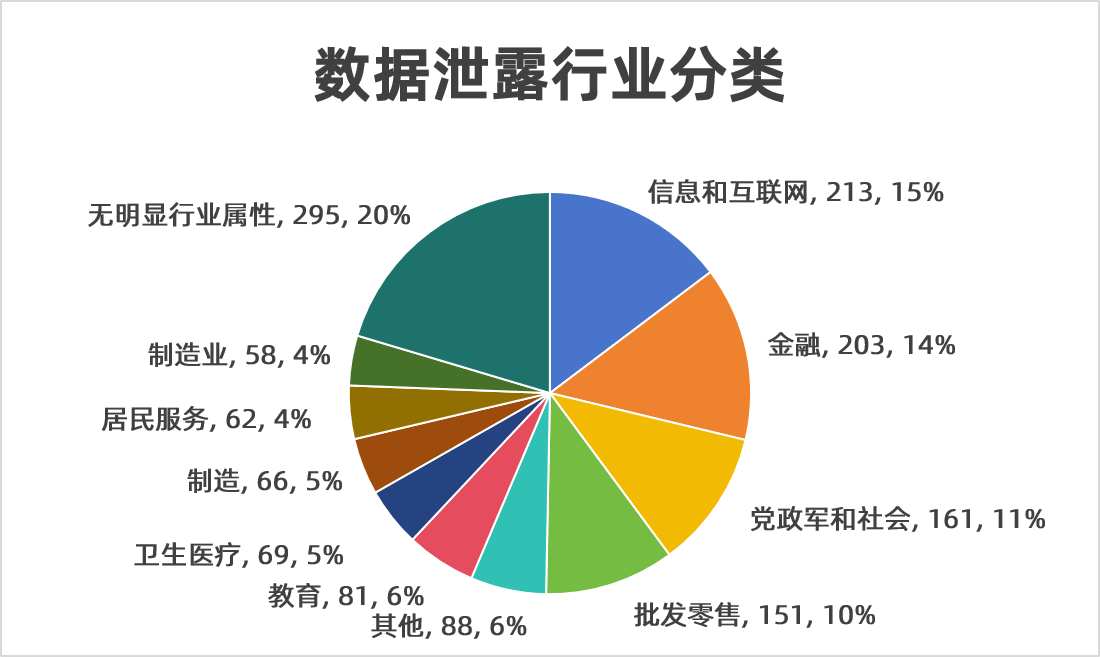


图1-1 数据泄露行业分类

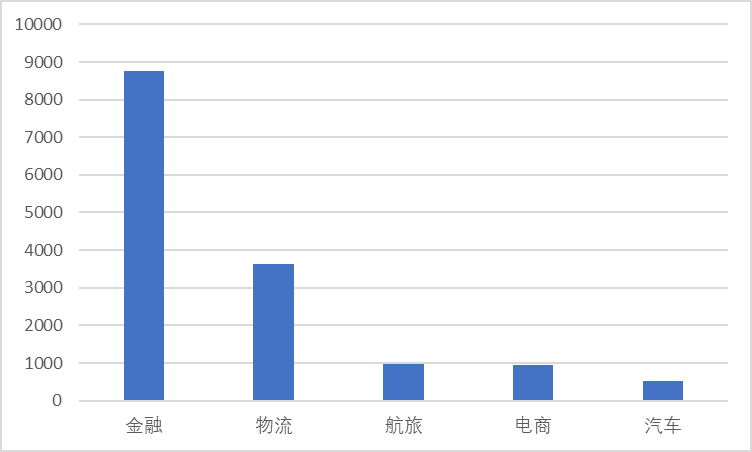


图1-2 2023年数据泄露数量Top5行业

在应对数据安全挑战的同时，**隐私保护计算需求**急剧上升。目前，隐私计算技术市场的年增长率预计将达到 17%，显示出各行各业对于高效且安全的数据处理技术的渴望。联邦学习，作为一种在不同参与者间共同构建模型而不需直接交换数据的技术，已成为研究的热点。使用联邦学习技术，数据泄露风险可降低 30%，同时，由于数据不需要中心化存储，还可以减少中心化存储的成本，据估计可为企业节约高达 15%的年度数据管理预算。

在**医疗**领域，电子病历（EHR）已经成为现实世界医疗数据的一个重要来源，被用于重要的生物医学研究。联邦学习可能是实现EHR数据大规模代表性机器学习的工具。联邦学习是连接医疗机构EHR数据的可行方法，允许医疗机构在保证隐私的情况下分享经验，而不是数据。在这些场景中，通过对大型和多样化的医疗数据集的反复改进学习，机器学习模型的性能将得到显著提高[2]。

在**金融**领域，联邦学习技术的应用正在深刻地重塑数据分析和隐私保护的界限。近年来，人工智能技术处于高速发展期。然而，随着人工智能技术发展进入深水区，出现了越来越多的壁垒和难题，**数据孤岛**就是其中之一。微众银行则顺势而为，其主要将联邦学习技术用于小微企业信贷以及个人贷款的风险管理，旨在解决银行在此类场景下存在的「数据维度缺失」、「数据低频」等问题。实际上，在互联网金融领域中，利用上游数据的概念已经较为常见，但是随着监管越来越严，数据出库的难度日益提升，我们希望通过联邦学习技术搭建一个桥梁，连接银行数据以及其它的互联网数据。

在**交通**领域，由于交通运输行业的不断发展，传统驾驶方式在安全性效率性等方面出现了不可忽视的问题，辅助驾驶技术和自动驾驶技术的需求也在日益增加，其中数据集隐私问题和**算力分配问题**也是一个重大的研究方向。对此可以结合联邦习和辅助驾驶场景两个方面,在保护车辆数据隐私和信息安全的前提下，利用驾驶车辆自身的数据资源，通过联邦学习模式搭建出交通标志识别分布式架构系统，利用联邦学习中多设备协同训练的方式训练预浏模型，在不泄露驾驶员的驾驶数据隐私的前提下，利用车辆计算设备的算力资源为预测模型提供本地训练，辅助驾驶员辦识交通标志，提高驾驶安全性。

本节**概述**了全球数据泄露的严峻现状。在此背景下，隐私保护和数据安全的重要性日益凸显，传统的集中式数据处理模式由于安全漏洞被逐步弃用。联邦学习技术作为一种创新解决方案，允许在不直接交换数据的前提下跨机构共建模型，显著降低数据泄露风险，提高数据处理效率，并在医疗、金融及交通等多个重要领域展现了**广泛的应用潜力和实际效益**。

## 1.2相关工作

**本节主要介绍了联邦学习的隐私保护现状、联邦学习助力隐私保护、基于边缘计算的联邦学习隐私保护的研究现状。**

### 1.2.1联邦学习隐私保护现状

联邦学习是一种旨在通过分散的数据来源进行协作机器学习，以保护用户隐私的技术。主要隐私保护技术包括差分隐私、安全多方计算（MPC），和同态加密。这些技术使得数据在不离开原始存储位置的情况下，依旧能够参与到全局模型的训练中。

差分隐私在联邦学习中的应用数据主要有：

（1）在Google的**Gboard键盘应用**中，使用了联邦学习来改进词汇预测和其他语言处理功能，同时采用差分隐私确保用户输入的保密性。数据显示，在引入联邦学习后，Gboard的语言模型错误率降低了至少15%[3]。

（2）调查显示，在采用联邦学习的企业中，有58%使用**差分隐私**来处理敏感数据[4]。这些应用通常能够减少数据泄漏的风险，并提高数据处理的安全性。

同态加密和安全多方计算的使用情况主要有：

（1）在金融行业中，同态加密被用于安全地在不同银行之间共享与信用评分相关的数据，而不暴露客户的具体信息。据报道，使用同态加密技术可以在不牺牲隐私的情况下提升数据分析的质量和速度[5]。

（2）在医疗领域，安全多方计算技术使得不同医疗机构能够合作开发共享模型，如癌症预测模型，同时确保患者数据的隐私不被泄露。这些合作通常能够提高诊断的准确性，例如，合作模型在某些病例的诊断准确率提高了20%以上[6]。

但是如今仍面临一些挑战：

1. 研究指出，高频率的模型更新交换可能导致**高通信成本**，尤其是在移动网络环境下，通信成本可能是计算成本的3至5倍[7]。
2. 在一个涉及多个行业的联邦学习实验中，约有10%的参与节点遇到了恶意攻击，这些攻击尝试通过模拟正常操作来窃取或污染模型数据[8]。

### 1.2.2联邦学习助力隐私保护

联邦学习作为一种新兴的人工智能技术，通过在多个参与方之间分布式地进行隐私保护建模，有效促进了数据隐私保护和数据利用最小化原则的实现。这项技术不仅减少了数据泄露的风险，还在提高建模性能方面发挥了重要作用，尤其是在处理延迟和功耗方面[3]。

目前，联邦学习的应用已经涵盖了金融、零售、互联网/电商/数字媒体、工业和医疗等多个行业，其中金融行业是探索最多的领域。在金融行业，联邦学习技术被应用于营销、信贷风控和反欺诈等场景中，有效应对了数据欠缺和隐私保护的挑战。互联网和电商领域也通过联邦学习技术在广告投放、反作弊和数据分析等方面实现了数据隐私保护和效能提升[3]

尽管联邦学习在隐私保护和数据利用方面取得了显著进展，但仍存在一些挑战。2023年发布的《联邦学习应用安全研究报告》指出，联邦学习技术面临的主要问题包括数据泄露风险、应用安全风险的隐蔽性、协调方带来的不确定性、安全保护与性能要求之间的相互制约，以及安全相关标准的不完善[4]。

联邦学习作为一种新兴的**分布式隐私保护建模方法**，已经吸引了全球的注意和投资。这一技术能够在不同参与方之间共同训练新的数据模型，同时保证所有训练数据不出域，从而促进数据隐私保护和数据利用最小化原则的实现。随着对隐私保护和数据安全要求的提高，联邦学习的研究和应用正在迅速发展。下面介绍相关的研究进展：

（1）**The Federated Learning Portal** 提供了一个关于联邦学习领域的综合性网络平台，追踪与联邦学习相关的书籍、工作坊、标准化工作和其他重要事件。这个平台展示了全球对联邦学习的研究和应用进展，以及正在进行的重要研究项目和即将发布的期刊特刊[4]。

（2）此外，**技术公司如Inpher**正在开发利用密码学的"秘密计算"技术，以支持在保持数据私密、安全和分散的同时进行高级分析和机器学习模型的建立。他们的技术包括全同态加密和安全多方计算，已被应用于金融服务、医疗保健、国防和物联网等多个领域[5]。

随着技术的进一步发展和安全性的提高，联邦学习有望在更多领域得到应用，从而更好地保护个人隐私和数据安全，推动数据共享与分析的进步。

### 1.2.3基于边缘计算的联邦学习隐私保护研究

基于联邦学习的隐私保护是一种通过**分散数据存储**和计算，同时集中模型学习的过程，以最小化数据暴露风险和保护用户隐私的机器学习方法。这种方法能有效减少敏感数据被泄露的可能性，因为原始数据不需要离开其原始位置，只有必要的模型信息（如权重和偏差的更新）在网络中传播。

面对医疗、交通和金融等领域中的数据隐私保护问题，联邦学习技术展现出了其**跨领域**的应用潜力。这些领域共同面临一个核心挑战：如何在不牺牲用户隐私的前提下，充分利用数据的潜在价值来提升服务质量和效率。然而医疗、金融，交通数据极具隐私敏感性，若将用户的原始体征数据直接提交到云数据中心，则在通信、处理以及存储中都存在着隐私泄露风险。这些信息如果被不当处理或泄露，可能会暴露用户的日常习惯和隐私。联邦学习不仅能够提高模型的准确性和效率，还能确保客户的财务数据在本地加密处理，仅将必要的模型信息进行交换，极大降低了数据泄露和滥用的风险。为此，联邦学习作为隐私保护的关键技术被广泛应用[6]。

联邦学习实现了数据不出本地的模型训练过程，但移动设备的计算能力往往有限，不能进行完整的神经网络训练，故需借助**边缘计算**[7,8]的架构，将大部分计算任务卸载到边缘服务器[9-11]。但将移动设备的原始数据全部传给边缘服务器也是不可靠的，因为相比**云数据中心**，边缘服务器面临更多的安全和隐私威胁。另外，联邦学习过程中传输的模型也面临推理攻击等的威胁，这些问题都有待很好的解决。

## 1.3作品简介

本作品名为**《Fed Privacy Shield》**。本系统基于web端，改进隐私保护和联邦学习算法，设计并实现了一个轻量级边缘智能协同联邦学习隐私计算平台，并在医疗、交通、金融等实际场景中进行了应用。平台提供一种安全、高效的方式，让多个参与方共同训练机器学习模型，同时保护个人数据的隐私。该平台的主要功能包括强大的隐私保护机制，采用加密技术、差分隐私技术等保护数据隐私；建立稳健的联邦学习框架，实现模型聚合、参数更新和模型评估等功能；确保数据安全性，包括防止未经授权的访问和数据泄露风险；提供用户管理和权限控制功能，以确保只有授权用户可以访问和使用平台；实现模型监控和解释功能，增强平台的透明度和可信度；可以实现发布联邦学习任务，查询进展，对比模型，创建模型，管理信息与模型等。

具体来说，系统包括主要多个云数据中心，服务端以及客户端，主要用户为管理员，管理员可通过Web端查看各集群状态及应用部署，发布任务、查看任务进展，创建和修改模型、查看训练报告等。从而实现一个**云边端协同**的联邦学习隐私计算服务。

平台的创新之处在于：

1. 采用多种**创新型联邦学习算法**如Fast-PDMM fed算法，Re-sim的鲁棒聚合算法等，保证了聚合过程中的鲁棒性和适应性，加快模型聚合的收敛速度。
2. 在数据传输的加密过程中采用**自适应差分隐私**，引入半可信端，中心服务器进行聚合后将聚合结果交给**agent半可信端**解密后传回中心服务器进行全局梯度更新，加强了隐私保护。
3. **选择性提交梯度更新**，通过差分隐私训练时对梯度影响衡量梯度数据的隐私度，对于隐私度较低的数据进行梯度压缩。构建一个方便高效、智能精准、隐私保护的基于联邦学习的智能隐私计算系统。

## 1.4创新性综述

平台的创新性主要体现在创新型联邦学习算法、通信传输过程中的流式传输以及多重负载均衡等、基于SSL/TLS双向安全通信与密钥证书认证、自适应差分隐私、引入agent半可信端助力解密过程等。基于以上创新点，本系统构建了一个方便高效、智能共享、安全可信的联邦学习隐私计算平台。

### 1.4.1平台方便高效：

**分布式数据处理方便高效**：本项目允许在数据源位置进行模型的训练，这意味着不需要将敏感数据集中或传输到中心服务器。这种方法减少了数据传输的需求，显著提高了处理速度和效率。

**实时更新方便高效**：系统采用创新型联邦学习算法，能够实时接收来自各参与方的更新，快速反馈到整个模型的改进中，从而实现高效的模型迭代和优化。

**维护与升级简化方便高效**：系统在后端采用反向代理负载均衡，由于模型训练和数据处理不依赖于单一中心服务器，系统的维护和升级可以更加灵活和高效，降低了对中央系统的依赖。

**模型部署快速方便高效**：系统支持快速部署和更新模型到边缘设备，这样可以确保所有设备都运行最新的算法，快速响应市场和环境的变化，使得项目运行更加迅速高效。

**集群管理方便高效**：系统管理员只需登录web端即可查看集群监控大屏，随时查看端设备状态，随时发布训练任务或查看任务进度，还可以查看模型列表，界面划分细致，操作方便高效。

**数据传输方便高效**：本系统后端采用流式传输允许数据分批次传输，每次只处理一小部分数据，显著降低了内存压力。同时使用多层负载均衡，轮询机制用于提高系统的稳定性明显抵抗高并发。

### 1.4.2 平台智能共享

**跨域知识共享智能**：系统支持跨机构或跨领域的知识共享而不泄露个人数据，这有助于创建更加全面和精准的预测模型。

**学习效果智能共享**：本项目整合多个数据源的学习结果，它能够提供更广泛的数据视角，从而训练出泛化能力更强的模型。同时在server端使用新型算法课保证模型的适应性更强。

**异构数据融智能共享**：平台能够处理和融合不同类型和来源的数据，即使数据是No-IIID型，系统提供了相应的鲁棒聚合算法进行使模型能够在更加多样化的数据集上进行训练，提升其精度和鲁棒性。

### 1.4.3平台安全可信

**差分隐私安全可信**：平台采用自适应差分隐私技术，自适应选取差分隐私噪声强度、选取每批的裁剪阈值。选择性提交梯度更新，通过差分隐私训练时对梯度影响衡量梯度数据的隐私度，对于隐私度较低的数据进行梯度压缩。

**模型聚合安全可信**：在联邦学习中，模型的更新是在本地计算后再进行聚合，聚合过程中可以采用Fast-PDMM聚合算法、Re-sim鲁棒聚合算法、基于poseidon的聚合算法等创新型联邦学习聚合算法，既加快了模型的收敛速度，也极大增强了全局模型的适应性和精度。

**数据治理安全可信**：实施强有力的数据治理策略，包括数据分类、数据生命周期管理和合规性检查，从而在整个数据处理过程中保持对隐私的严格控制。

**加密过程安全可信**：引入agent半可信端，中心服务器进行聚合后将聚合结果交给agent半可信端解密后传回中心服务器进行全局梯度更新。保证数传输过程中的隐私保护。

## 1.5 应用前景

基于联邦学习的智能隐私计算平台，是一种新型的数据安全和隐私保护技术，其主要目的是在不泄露各参与方数据隐私的前提下，通过协同合作来共同完成数据分析和模型训练任务。本平台的应用前景广泛，具体包括但不限于以下几个方面：

医疗健康：联邦学习可以使得不同医疗机构在不直接交换患者数据的情况下，共同训练模型，提高疾病诊断的准确性，同时保护患者隐私。例如，各医院可以共享他们的模型更新（而非患者数据）来共同改进疾病预测模型[18，19]。

交通与移动性：联邦学习可以应用于交通流量分析、路况预测等场景，帮助城市管理者优化交通配置，提高道路使用效率，同时保护个人用户的行程数据隐私。[20,21]

金融行业：在金融风险管理、欺诈检测和信用评分等领域，联邦学习能够帮助金融机构在遵守数据隐私法规的同时，提高决策的准确性。金融机构可以合作建立共有模型，而无需共享客户的具体交易数据[22]。

智能制造：制造业公司可以利用联邦学习进行跨工厂的设备故障预测和维护时间优化，通过共享模型学习的结果而非具体的设备数据，从而优化生产效率并降低维护成本[23]。

**扩大和部署该智能隐私计算平台能有效解决多个领域中的数据隐私保护和协同工作中的痛点问题，助力现代化国家建设。**

## 1.6 本章小结

本章主要进行《Fed Privacy Shield》作品概述。

**1.1节**主要介绍了目前的联邦学习在各个隐私计算场景的社会背景，分析了联邦学习在隐私保护领域痛点问题，提出开发基于联邦学习的智能隐私计算平台解决上述问题；

**1.2节**综述了联邦学习隐私保护现状、联邦学习助力隐私保护以及边缘计算和基于边缘计算和联邦学习的隐私保护研究；

**1.3节**对作品进行简介，具体从系统的Web端模块的功能进行了说明。简要概述了作品中采用的技术方案、核心算法与系统结构，以及其在实际应用中的现实意义。

**1.4节**随后，阐述作品创新性，对作品特色关键词“方便高效”、“共享智能”与“安全可信”引入，概述作品表现在系统基于端-边分层的联邦学习训练模型、聚合算法和安全加密的特色与创新点；

**1.5节**介绍作品应用前景，分析了本联邦学习在隐私计算方面的应用。说明了作品的实际使用价值。

# 第二章 需求分析与确定

本章主要包括系统需求分析、功能需求、非功能需求以及应用场景描述。

## 2.1需求概述

目前进行联邦学习训练任务时存在如下痛点问题：参与训练的用户数据隐私安全难以保障、模型性能和收敛速度较慢、可能存在不可靠参与方对模型训练产生破坏以及聚合方案单一不合理。针对以上问题，需要设计一套安全、高效的基于联邦学习的智能隐私计算系统，为联邦学习用户提供有效且有力的技术与安全保障，从而提升联邦学习训练效率与隐私安全，推进联邦学习在医疗任务、车联网、金融服务等方面的应用。

Fed Privacy Shield平台需要设计两个板块，分别是**训练模型用户web端**和**平台管理人员Web端**需求。训练模型用户web端面向用户，主要是进行个人任务管理、执行对比实验、自定义模型与场景、个人信息管理；平台管理人员Web端执行集群监控、任务管理、用户管理、模型管理。

## 2.2功能需求

根据联邦学习与隐私保护的目的与功能需求，本系统主要分为训练模型用户Web端和平台管理人员Web端。其中，训练模型用户Web端主要分为任务中心、对比实验、自定义服务和个人中心；平台管理人员Web主要分为集群监控、任务中心、用户管理和模型管理。

### 2.2.1 训练模型用户Web端需求

训练模型用户Web端分为任务中心模块、对比实验模块、自定义服务模块和个人中心模块，共10个功能点。用例图如图2-1所示。

#### UseCaseDiagram1

图2-1 训练模型用户Web端

#### 2.2.1.1 任务中心

此模块主要用于发布和管理多种场景下的训练任务。

**（1）发布任务**：用户依据自己的需求定制联邦学习任务，可设定多种复杂的参数，如优化算法、迭代次数、隐私加密方案。系统还提供实时的资源评估，确认任务的可行性，预估所需资源和时间。

**（2）进展查询**：用户通过一个交互式仪表板实时监控当前的任务状态，本模块不仅展示任务的基础进度条，还提供资源使用率、模型性能等多维度的实时数据和图表，同时集成了通知系统，可以在关键里程碑达成时通过邮件或短信通知用户。

**（3）专业服务**：针对特定应用场景，如医疗影像分析、金融市场分析和车联网数据处理，提供了完备的高级模型和数据集，以及专门的分析工具和定制模型调优和数据处理服务。

#### 2.2.1.2 对比实验

此模块主要用于执行多种对比性的任务和分析。

**（1）模型对比：**用户可以在不同的数据集上对比多个模型，分析每个模型的响应时间、准确率、资源消耗等关键性能指标。系统会自动生成详细的对比报告，包括性能图表和关键指标的统计分析。

**（2）模型中毒：**本系统提供模型中毒检测工具，用户可以上传模型进行安全性检测，系统将自动识别潜在的中毒攻击，并提供修复建议或者安全增强的策略。用户还可以模拟模型中毒攻击，选择攻击类型，并注入恶意数据或操纵模型参数，模型开始训练后，系统会追踪受攻击模型的学习过程，并将其与正常模型训练的进程作对比，最终生成一份报告。

#### 2.2.1.3 自定义服务

此模块主要用于创建自定义的模型和场景服务。

**（1）创建模型:** 用户可以使用内置的模型构建器，选择不同的算法组件，调整层次结构，定制模型的超参数等，完全个性化定制。此外，系统还支持上传已有模型和进行微调。

**（2）创建服务：**用户能够根据用户的业务需求开发专属算法和服务流程，根据特定背景建立新的应用场景，整合多源数据，配置数据流，以及部署到自定义的硬件环境中

#### 2.2.1.4 个人中心

此模块主要用于进行多种个人信息的操作。

**（1）管理信息：**用户中心提供了全面的个人资料管理，包括但不限于联系信息、认证方式、偏好设置等，以及界面定制和通知偏好。

**（2）历史记录：**用户所有的活动，包括任务提交、查询和下载结果，都会被记录在案。用户可以随时回顾这些记录，系统也会根据这些数据提供定制化的操作建议和改进提示。

**（3）管理模型：**用户可以查看、管理自己创建的所有模型，进行版本控制、模型共享、性能追踪和维护操作，以及对模型进行分类和标记，分享给其他用户或小组，并设置访问权限。

### 2.2.2 平台管理人员Web端需求

Fed Privacy Shield平台管理人员Web端分为集群监控模块、任务中心模块、用户管理模块和模型管理模块，共8个功能点。用例图如图2-2所示。

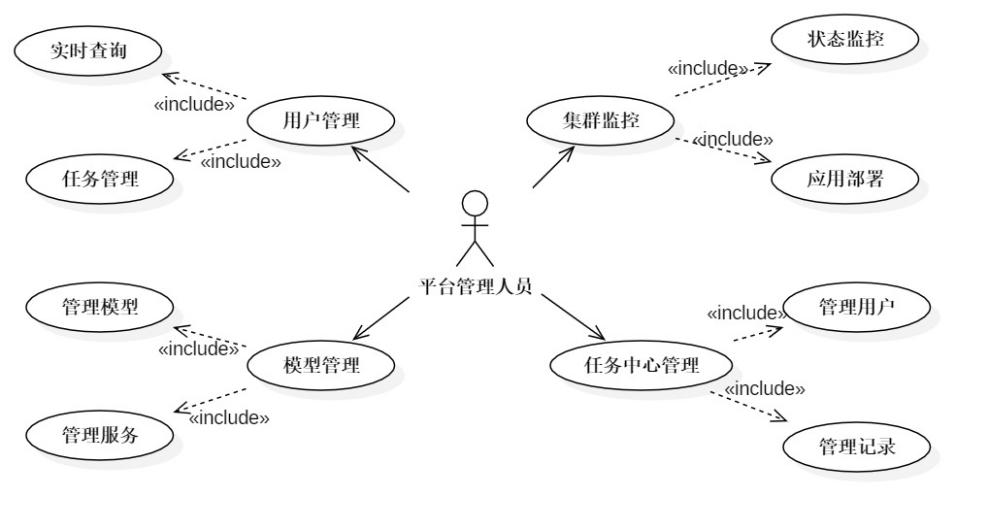


图2-2 平台管理人员Web端

#### 2.2.2.1 集群监控

此模块主要用于监控系统的整体状态和管理所有的服务。

**（1）状态监控：**管理员可以在本模块上查看云数据中心、边缘服务器、以及端设备的状态，如在线状态、IP地址、资源使用情况等，管理员可以迅速识别潜在问题，并进行相关的管理操作，如设备上下线、资源调度等。

**（2）应用部署：**管理员可以管理和监控集群中所有部署的服务，查看应用状态、配置和日志，管理容器生命周期，实现应用的快速迭代和高可用部署

#### 2.2.2.2 任务中心

此模块主要用于管理系统所有的模型训练任务。

**（1）管理用户：**本模块提供了权限和角色管理功能，管理员可以通过用户行为分析，调整用户权限，设置资源配额，还可以追踪用户的活跃度和系统使用情况。

**（2）管理记录：**本模块对所有任务进行详细记录，包括提交时间、执行时长、所需资源和结果。管理员能够查看和审计所有模型训练和任务执行的详细记录，用于生成统计报告，支持系统的决策和优化。

#### 2.2.2.3 用户管理

此模块主要用于管理所有用户信息和分配任务。

**（1）实时查询:**实时查询功能允许管理员跟踪系统中的活跃用户和他们当前的任务状态。通过一个集成的控制面板，管理员可以快速响应用户请求和问题。

**（2）任务管理:**管理员可以直接在控制面板上分配任务给用户，调整任务优先级和资源分配，并监控任务执行情况。

#### 2.2.2.4 模型管理

此模块主要用于管理所有系统模型和服务。

**（1）管理模型：**管理员可以对所有模型进行性能监控和版本控制，比较不同版本的性能，及时更新和优化模型，以适应不断变化的业务需求。

**（2）管理服务：**对于模型相关的所有服务，管理员可以进行状态监控、配置管理和性能优化，保障服务的高效运行和可靠性，服务日志和性能指标有助于诊断问题并优化服务配置。

## 2.3 非功能需求

基于联邦学习的智能隐私计算系统的非功能需求主要包括系统的安全性需求、易用性需求、健壮性需求、可移植性需求及可扩展性需求。

**2.3.1安全性**

基于联邦学习的智能隐私计算系统需要保证用户的原始数据不出本地。采用TLS/SSL传输协议保证信息的机密性、认证性、完整性。采取差分隐私、同态加密等加密方法抵抗成员推理攻击、梯度泄露攻击、合谋攻击，尽可能的保护训练用户的数据隐私。为数据库中敏感数据加密存储。

### 2.3.2 易用性

训练模型用户与平台管理人员均可直接在浏览器中即可访问Web系统。训练端仅需要在训练时间保持程序运行、网络连接。系统的UI界面美观简洁、信息丰富；云端在发起训练请求操作时，训练、聚合、传输速度较快，速度受加密影响较小。对用户的输入及系统输出遵循规范文档，当有错误发生时能给出错误提示窗口。

### 2.3.3 健壮性

系统要求保护掉线用户隐私并且不影响其他成员聚合结果。系统能够很好地自动负载均衡，提供基础的故障排查处理功能，故障节点能够及时修复，确保数据一致性。系统定期进行数据库备份，不可控灾难发生后，可以恢复大部分数据。控制系统版本，定期备份系统配置和代码，确保系统可以回滚到之前的状态。

### 2.3.4 可扩展性

为了应对系统业务扩展和更新需求，系统设计时应设计足够的接口。例如，预留接口用于添加新的模型训练方法或处理用户掉线等功能。设计第三方接口以方便外部接入，扩大联邦学习的应用范围。当系统需求更新时，能在较短时间内完成对更新需求的实现，并及时完成测试工作、上线，并给出相关文档。

### 2.3.5 可移植性

将系统配置信息独立化，以便根据不同环境需求进行灵活配置和调整，实现系统定制与移植。定义标准化的接口和协议，以确保系统不同组件之间的互操作性和通信顺畅，从而提高系统的可移植性。丰富云端可以选择的训练、聚合、加密方案，用以适应各种训练场景。

## 2.4 应用场景

系统主要应用于医疗、车联网、金融等场景，利用多方隐私数据合作学习、训练模型。

### 2.4.1 医疗任务场景

在推进智慧医疗的过程中，病症、病理报告、检测结果等病人隐私数据常常分散在多家医院、诊所等不同地区不同类型的医疗机构，Fed Privacy Shield平台使机构间可以在数据不出本地的情况下，通过多方合作建立预测模型，能够更准确地预测癌症、基因疾病等，解决医疗领域局部机构样本少，数据质量低的问题。

有研究人员曾收集5978位患者并分析其5年期间的前列腺癌活检的数据，从中选择2426位患者数据训练模型。使用基于联邦学习的隐私计算系统训练的模型较之前传统方式的LR和SBT模型准确率分别提高3.41％和2.22％。[26]研究人员还通过采集30位志愿者活动，获得10299个样本。并模拟数据集呈现孤岛的形式，将志愿者数据分散到不同计算节点上。通过实验发现应用联邦学习隐私计算系统训练的模型与传统深度学习模型相比，前者的准确率平均高于后者5.3％左右。

用户可以利用任务中心模块发布和管理定制的联邦学习任务，该功能让用户设定各种参数，如优化算法和隐私加密方案，以提高医疗影像诊断的准确性。通过这种个性化的任务设置，医疗专业人员可以针对不同类型的医疗影像数据定制化模型训练流程，从而实现更精准的诊断结果。针对医疗数据的高度敏感性，用户可以使用对比模块比较不同医疗影像分析模型的性能，包括响应时间、准确率等关键指标。通过对模型性能的全面对比分析，医疗专业人员可以选择最优秀的模型来处理医疗影像数据，确保数据的隐私安全和精准性。

用户还可以使用Fed Privacy Shield平台将患者数据与保险等相关行业数据相结合，以提升医疗行业数据在其他行业的应用度和影响力。[24]

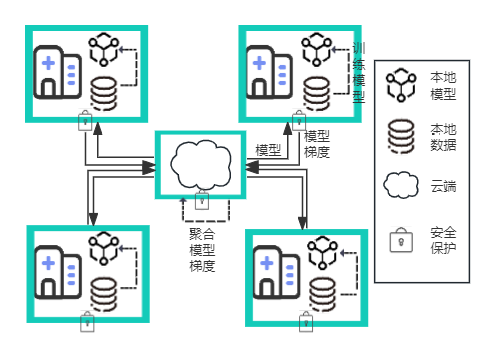


图2-3 联邦学习医疗领域应用场景

### 2.4.2 车联网

在车联网中，可以利用面向联邦学习的隐私计算系统进行车辆之间数据交流。据预测，到2025年，所有的车将会通过各种形式进行互联；到2035年，路面上75%的车将会是自动驾驶车。在此背景下，Fed Privacy Shield平台可以利用车载设备产生的路况信息数据结合车辆自身环境感知能力进行训练，使车辆用户获得更准确的综合判断。这种基于联邦学习的车联网技术不仅提升了数据的利用效率，充分挖掘了数据的价值，还保护了个人隐私，为车联网产业的可持续发展提供了重要支持。

在车联网领域，Fed Privacy Shield平台支持用户创建专为车联网设计的模型，例如交通流量预测模型。通过自定义服务模块，用户可以选择不同的算法组件、调整模型层次结构，并优化响应时间和数据处理效率。由于联邦学习10轮通信内便可达到约80%的全局准确度，用户可以将通信轮数限制在100轮以内，用较小的轮次缓解计算负担，提高实时性。[11]这种个性化的模型设计可以帮助车联网领域的专业人员更好地应对不同交通场景下的需求，提高交通管理的效率和准确性。用户还可以利用对比实验模块在不同数据集上比较交通流量预测模型的性能，分析每个模型的优势和劣势。这种对比分析可以帮助用户选择最适合实际应用的模型，从而提高交通流量预测的准确性和实用性。

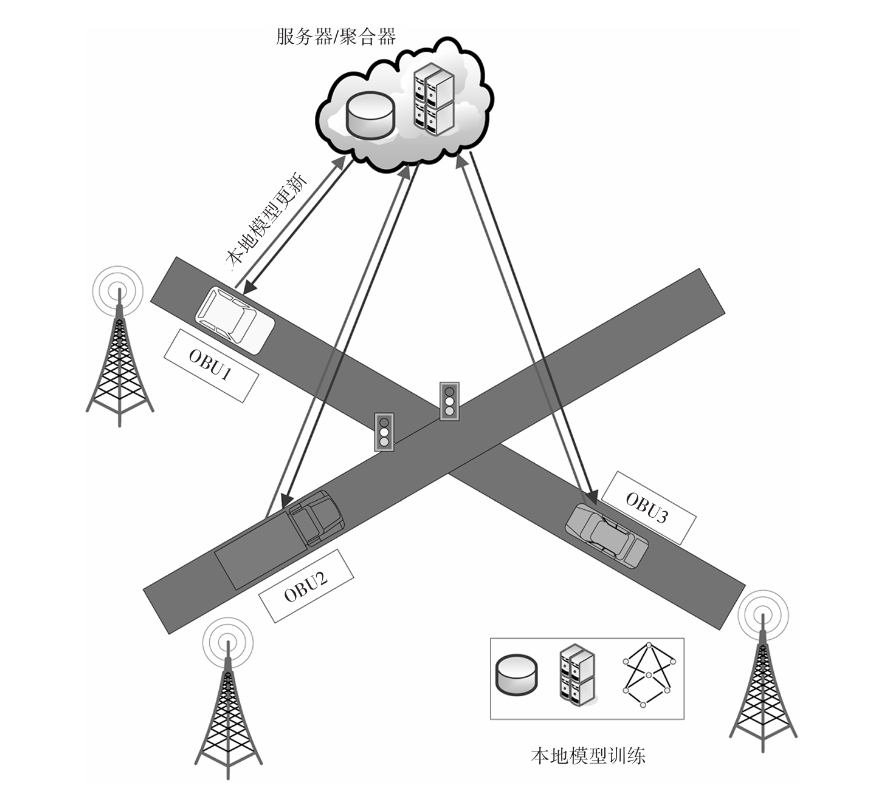
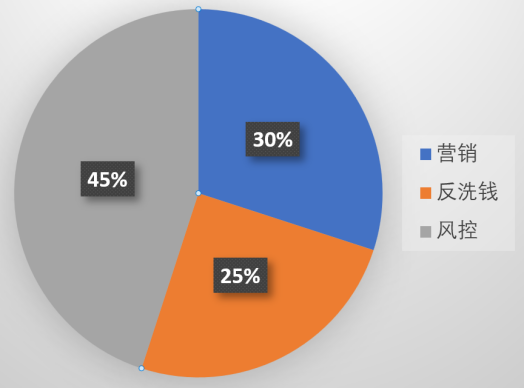


图2-4 联邦学习在车联网应用场景

### 2.4.3 金融服务

银行和金融机构在金融服务中积累了大量的客户数据，包括个人信息、交易记录、信用评分等敏感数据。这些数据对于金融机构提供个性化的金融服务、风险管理和反欺诈至关重要。利用Fed Privacy Shield平台实现共享风险模型可提高金融机构的风险管理能力，降低潜在损失，并为客户提供更安全的金融服务。联邦学习还可用于共享反欺诈算法，提高识别欺诈的准确性和效率，构建更安全和可靠的金融生态系统。采用Fed Privacy Shield平台，旨在保护数据隐私安全的前提下实现联合建模，提升行业竞争力。

据统计，目前联邦学习在金融领域风控环节的应用研究和试点较多，约有45%应用于风控、30%应用于营销、25%应用于反洗钱。[25]

Fed Privacy Shield平台为金融领域提供了金融市场分析的任务管理功能。用户可以根据自身需求定制金融分析任务的参数，包括设备数量、优化算法等，以提供更精准的金融预测和分析。这种个性化的任务定制可以帮助金融从业人员更好地应对市场波动，提高决策的准确性和效率。用户还可以利用对比实验模块测试不同金融模型的性能，从而选择最佳模型用于实际金融市场应用。这种对比实验能帮助金融从业人员更好地了解不同模型的优缺点，从而做出更明智的决策，提高金融市场分析的准确性和效果。 图2-5 金融场景联邦学习应用比例

## 2.5 本章小结

本章主要进行系统需求分析。

**2.1节**中概述系统需求。分析联邦学习训练时存在的问题，描述了隐私计算系统部署后的现实意义，说明了系统中的两类用户并据此提出系统总体功能需求。

**2.2节**中详细描述系统的功能需求。从训练模型用户Web端和平台管理人员Web端详细描述了系统的8个模块和18个功能点的需求。

**2.3节**详细描述系统非功能性需求。分别从系统安全性、易用性、健壮性和可扩展性和可移植性五方面分析系统需求。

**2.4节**描述了系统的应用场景。分别详细描述了Fed Privacy Shield平台在医疗任务场景、车联网场景、金融服务场景的应用。

# 参考文献

1. Thomas, R., & Wilkinson, B. (2022). "Data Breaches: Time for a Rethink on Protection Strategies", Journal of Information Security and Applications, Elsevier.
2. Xu, J., Glicksberg, B.S., Su, C. et al. (2021). Federated Learning for Healthcare Informatics.
3. McMahan, H. B., et al. "Communication-Efficient Learning of Deep Networks from Decentralized Data." Artificial Intelligence and Statistics. PMLR, 2021.
4. Li, T., Sahu, A. K., Talwalkar, A., and Smith, V. "Federated Learning: Challenges, Methods, and Future Directions." IEEE Signal Processing Magazine, 2021.
5. Aono, Y., et al. "Privacy-Preserving Deep Learning: Revisited and Enhanced." Applications of Privacy-Preserving Distributed Machine Learning, 2022.
6. Bonawitz, K., et al. "Towards Federated Learning at Scale: System Design." arXiv preprint arXiv:1902.01046, 2023.
7. Kairouz, P., et al. "Advances and Open Problems in Federated Learning." Foundations and Trends in Machine Learning, 2023.
8. Zhu, L., and Han, S. "Deep Leakage from Gradients." NeurIPS, 2024.
9. 中国信息通信研究院. "联邦学习应用安全研究报告 (2023年)." Journal of Physics: Conference Series. IOP Publishing, 2023, 1952(4): 042141.
10. Goebel, R., Yu, H., Faltings, B., Fan, L. & Xiong, Z. (Eds.). "Trustworthy Federated Learning." Journal of Physics: Conference Series. IOP Publishing, 2023, 1952(4): 042141.
11. 邬忠萍,郝宗波,王文静,刘冬.结合联邦学习和增强学习的车联网数据差分隐私保护[J].汽车技术,2023(11):56-62
12. Xia X, Chen F, He Q, et al. Online collaborative data caching in edge computing[J]. IEEE Transactions on Parallel and Distributed Systems, 2020, 32(2)：”281-294.
13. Zwolenski M, Weatherill L. The digital universe: Rich data and the increasing value of the internet of things[J]. Journal of Telecommunications and the Digital Economy, 2014, 2(3): 47.1-47.9.
14. Bellavista P, Foschini L, Scotece D. Converging mobile edge computing, fog computing, and iot quality requirements[C]//2017 IEEE 5th international conference on future internet of things and cloud (FiCloud). IEEE, 2017: 313-320.
15. Zhang Shan, He Peter, Suto K., et al. Cooperative Edge Caching in User-Centric Clustered Mobile Networks[J]. IEEE Transactions on Mobile Computing, 2017, 17(8): 1791-1805.
16. Zhao Yongli, Wang Wei, Li Yajie, et al. Edge Computing and Networking: A Survey on Infrastructures and Applications[J]. IEEE Access, 2019, 7: 101213-101230.
17. Qadir J., Abajo B. S. D., Khan A., et al. Towards Mobile Edge Computing: Taxonomy, Challenges, Applications and Future Realms[J]. IEEE Access, 2020, 8: 189129-189162.
18. Sheller, M. J., Reina, G. A., Edwards, B., Martin, J., & Bakas, S. (2020). "Multi-Institutional Deep Learning Modeling Without Sharing Patient Data: A Feasibility Study on Brain Tumor Segmentation", in Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Springer, Cham.
19. Brisimi, T. S., Chen, R., Mela, T., Olshevsky, A., Paschalidis, I. C., & Shi, W. (2020). "Federated Learning of Predictive Models from Federated Electronic Health Records", International Journal of Medical Informatics.
20. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2021). "Federated Machine Learning: Concept and Applications", ACM Transactions on Intelligent Systems and Technology (TIST).
21. Long, Y., Lu, Q., Xu, K., Liu, L., & Cai, Y. (2022). "Federated Learning for Credit Scoring: Empirical Analysis and Performance Comparison", Journal of Financial Data Science.
22. Lu, Y., Huang, X., Dai, Y., Maharjan, S., & Zhang, Y. (2022). "Federated Learning for Fault Diagnosis in Industrial IoT: A Case Study", IEEE Transactions on Industrial Informatics.
23. Karnouskos, S. (2021). "Federated Learning in Distributed Industrial Systems: Applications, Challenges, and Opportunities", IEEE Industrial Electronics Magazine.
24. 聂文静,信伦,李帜.联邦学习在医疗信息化中的应用研究[J].医学信息学杂志,2022,43(10):12-16
25. 陈琨,李艺,王国赛,时代,杨祖艳.联邦学习在金融行业的应用分析[J].征信,2021,39(10):29-36
26. Liu Z, Chen Y, Zhao Y, et al. Contribution - aware Federated Learning for Smart Healthcare[ C]. Stanford: Proceedings of the 34th Annual Conference on Innovative Applications of Artificial Intelligence (IAAI -22)，2022.
27. 牛志生，沈谢尔曼，张勤宇，唐玉良.沉浸式驾驶体验的空地一体化车载网络[J].物联网学报， 2021， 1（2）：17-27.