AIIA

中国人工智能产业发展联盟标准

AIIA xxx（V1.0）—2020

共享学习系统
技术要求

Shared machine learning system technical requirements

2020-xx-xx 发布　　2020-xx-xx 实施

中国人工智能产业发展联盟 发布

目　　次

前　　言

本标准按照GB/T1.1-2009《标准化工作导则 第1部分 标准的结构和编写》给出的规则起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本标准由人工智能产业发展联盟提出并归口。

本标准起草单位：浙江蚂蚁小微金融服务集团有限公司、中国联通股份有限公司、中国信息通信研究院、中国电信股份有限公司、阿里巴巴（中国）有限公司、北京大学、中和农信

本标准主要起草人：李克鹏、朴昕阳、周俊、王磊、王力、陈超超、加雄伟、孙明俊、洪澄、吴秉哲、李楠

共享学习系统技术要求

1 范围

本标准项目用于制定共享学习系统的技术要求，来规范化共享学习的定义、技术框架及流程、技术特性、安全要求。

本标准项目适用于指导共享学习系统的设计、开发、测试、使用、运维管理等。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 22081—2016 信息技术 安全技术 信息安全管理实践指南

ISO/IEC 27033 信息技术 安全技术 网络安全

3 术语和定义

下列术语和定义适用于本文件。

3.1

机器学习 Machine Learning

在历史数据中自动发现规律并利用规律对未知数据进行应用（预测）的算法（技术），它能帮助人利用数据做出更好的决策。

3.2

共享学习 Shared Machine Learning

由多数据提供方参与且在各数据提供方与平台方互相不信任的场景下，平台能够聚合或协助聚合多方数据信息并保护多方数据隐私的学习范式。

3.3

安全多方计算 Multi Party Computation

在一个分布式网络中，多个参与方各自持有来自其他参与方秘密输入，各方希望共同完成对某函数的计算，且要求每个参与方除计算结果外均不能得到其他参与方的任何输入信息。

3.4

可信执行环境 Trusted Execution Environment

可信执行环境是主处理器内的安全区域，在隔离环境中与操作系统的并行运行。可信执行环境保证其中加载的代码和数据在隐私性和完整性方面受到保护。

3.5

远程认证 Remote Attestation

是主机（客户端）通过其向远程主机（服务器）验证其硬件和软件配置的方法。

3.6

数据 Data

本文把所有能够泄露个人用户隐私的信息都称为数据。

4 缩略语

下列缩略语适用于本文件：

MPC　　安全多方计算（Multi Party Computation）

TEE　　可信执行环境(Trusted Execution Environment)

5　共享学习系统概述

共享学习系统是基于数据安全和隐私保护技术，解决多个数据提供方之间进行机器学习时的隐私保护问题。

系统主要包括 TEE（Trusted Execution Environment，可信执行环境）方案和 MPC（Multi Party Computation，安全多方计算）方案。

可信执行环境方案是利用可信执行环境解决多个数据提供方在进行共享学习时的隐私保护问题。

安全多方计算方案是利用安全多方计算技术（包括但不限于：秘密分享，混淆电路，同态加密，不经意传输等）解决多个数据提供方在进行共享学习时的隐私保护问题。

6　基于可信执行环境的共享学习系统

6.1　可信执行环境方案技术框架

共享学习系统的可信执行环境方案技术框架如图 1 所示：

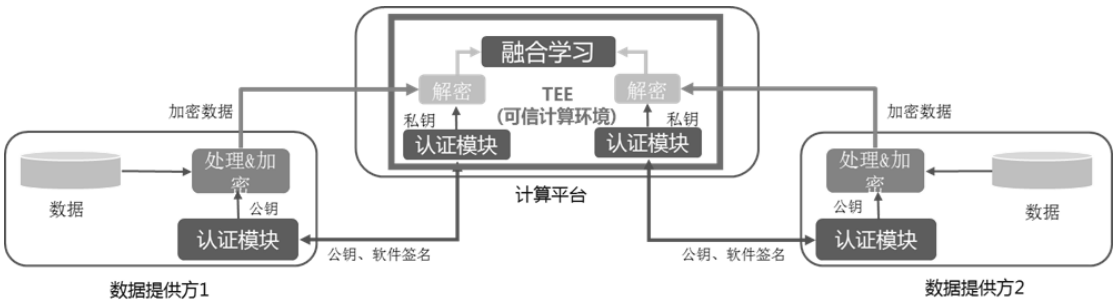

图1 可信执行环境方案技术框架

可信执行环境方案的技术框架主要由计算平台和多个数据提供方组成，计算平台由认证模块、数据解密模块、融合学习模块组成。数据提供方由数据加密模块、认证模块和数据组成。

其中，数据提供方的数据在进行处理后，通过认证模块的公钥加密，再上传到计算平台的可信计算环境中，计算平台的认证模块通过私钥对加密数据进行解密后，发送给融合学习模块，对解密后的多方数据进行融合机器学习，确保隐私信息不会泄露。

6.2　可信执行环境方案功能组件

6.2.1　计算平台

6.2.1.1　概述

计算平台主要由融合学习模块、解密模块、认证模块所组成，这三个模块都位于可信执行环境中。

6.2.1.2　认证模块

计算平台的认证模块负责下发公钥到数据提供方，以及提供私钥给解密模块，私钥用于对数据提供方上传至平台的加密数据进行解密。

计算平台的认证模块负责对运行在可信执行环境的软件代码进行签名，并支持数据提供方对软件代码进行验签。

6.2.1.3 数据解密模块

计算平台的解密模块负责对数据提供方上传的加密数据基于私钥进行解密。

6.2.1.4 融合学习模块

融合学习模块负责基于解密后的数据进行学习。

6.2.2 数据提供方

6.2.2.1 概述

数据提供方主要由数据、处理加密模块、认证模块所组成。数据提供方可以有两个或多个。

6.2.2.2 数据处理和加密模块

数据提供方的数据处理和加密模块用于对数据进行必要的处理，并基于计算平台下发的公钥，对数据进行加密，然后上传到计算平台。

6.2.2.3 认证模块

数据提供方的认证模块用于实现数据提供方与计算平台的远程认证，接收计算平台下发的公钥，并发送给加密模块对数据进行加密。

6.2.2.4 数据

数据提供方将会泄露用户隐私信息的数据经过加密模块加密后，上传到计算平台。

6.3 可信执行环境方案技术流程

在可信执行环境技术方案中，首先各数据提供方处理加密本地数据，并上传到计算平台，其中任何一个数据提供方都可以发起训练任务。之后计算平台创建可信环境，解密接收到的各个数据提供方发送来的加密数据，并基于解密数据进行模型训练，得到共享模型。最后销毁可信环境，以保证数据安全，实现隐私保护。

共享学习系统的可信执行环境方案的技术流程如图 2 所示：

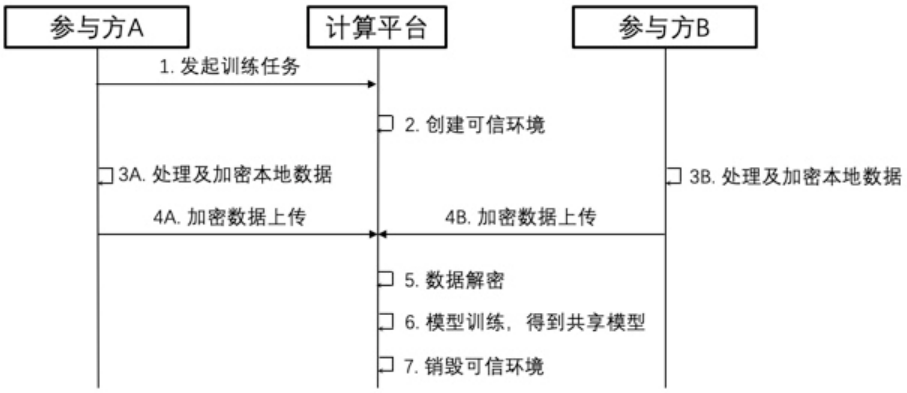

图2 共享学习系统可信执行环境方案流程图

共享学习系统可信执行环境方案的技术流程如下：

1） 参与方发起机器学习训练任务；
2） 计算平台创建可信执行环境；
3） 参与方对于本地数据进行处理和加密；
4） 参与方把密数据上传至计算平台；
5） 计算平台在可信执行环境中对于加密数据进行解密；
6） 计算平台在可信执行环境中对解密后的数据进行融合学习，得到共享机器学习模型；
7） 计算平台销毁可信执行环境，以及数据。
其中 3、4、5、6 可以根据算法需要，循环执行多次

7 基于安全多方计算的共享学习系统

7.1 安全多方计算方案技术框架

共享学习系统的安全多方计算方案的技术框架如下图所示：

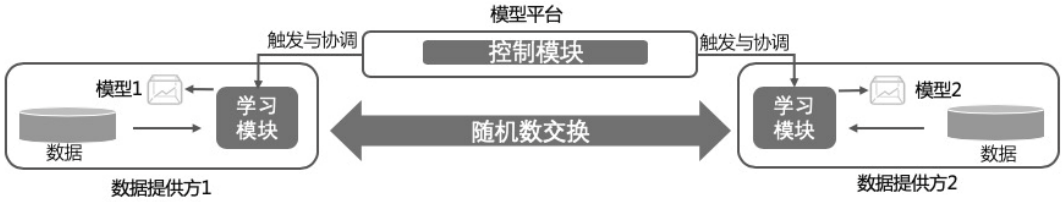

图3 安全多方计算方案技术框架

共享学习系统安全多方计算方案的技术框架主要由模型平台和多个参与方组成，其中模型平台主要包括控制模块，数据提供方由学习模块、数据组成。

其中，数据提供方的学习模块通过交换随机数或加密参数的方式，在模型平台的触发与协调下，进行共享机器学习。各参与方各自部署机器学习模块，各个参与方都可以发起训练任务。模型平台收到训练任务后，进行分解和协调，并下发训练任务到各个参与方。各个参与方读取本地数据到本地的机器学习模块。各参与方之间多次交换随机数或参数，完成共享学习训练，并得到共享模型。

7.2 安全多方计算方案功能组件

7.2.1 模型平台

模型平台主要包括控制模块，用于触发和协调学习训练任务。

7.2.2 数据提供方

7.2.2.1 学习模块

数据提供方的本地学习模块，用于接收模型平台下发的机器学习任务，基于数据以及与其他数据提供方交互的随机数或参数，进行机器学习。

7.2.2.2 数据

数据提供方将数据输入到本地学习模块，进行机器学习。数据提供方会泄漏用户隐私的数据不出本地。

7.2.3 安全多方计算方案的技术流程

在安全多方计算方案下，多个参与方基于既定协议下通过交换不泄露隐私的信息来进行共享学习。

共享学习系统的安全多方计算方案的技术流程如下图 4 所示：

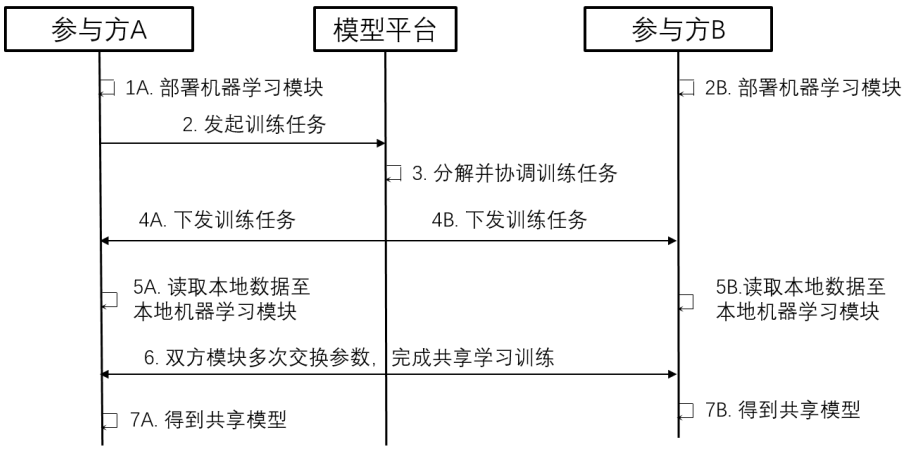

图4 共享学习系统安全多方计算方案流程图

共享学习系统安全多方计算方案的技术流程如下：

1） 数据提供方部署机器学习模块；
2） 数据提供方发起机器学习训练任务；
3） 模型平台对机器学习训练任务进行分解和协调，
4） 模型平台下发训练任务到各个数据提供方；
5） 数据提供方将本地数据读取至本地机器学习模块；
6） 各个数据提供方的机器学习模块进行多次的参数交换，完成共享学习训练；
7） 各个数据提供方得到共享学习模型。

8 共享学习系统技术要求

8.1 基本功能要求

8.1.1 数据管理功能要求

1） 应具备对共享数据的管理功能，包括数据获取、数据处理、数据传输等；
2） 应具备样本对齐、特征对齐功能；
3） 应支持多方安全计算技术、可信执行环境技术、确保数据只能按约定好的行为进行使用，避免数据滥用。

8.1.2 算法组件管理功能要求

1） 应支持至少一个机器学习主流算法，例如：线性回归、逻辑回归、树模型、深度神经网络、图神经网络等
2） 应支持对算法组件的管理，采用多种方式提高算法鲁棒性，增强安全性。
3） 宜支持机器学习模型训练优化，包括指标定义评估、算法策略选择、数据集划分、参数调优等，以提高模型性能和泛化能力；

8.1.3 计算管理功能要求

1） 应提供建模或预测任务的管理能力；

2） 应具备训练任务的分解与调度功能，并对共享学习的任务进行状态跟踪与记录。
3） 应支持权限管理和服务接口规范。

8.2 可扩展性要求

1） 应具备良好的可扩展性，可以根据业务需求方便增加新的功能组件
2） 应支持用户通过 API 接入系统；

8.3 可靠性要求

1） 应确保数据只能按约定好的行为进行使用，避免数据滥用；
2） 应保证系统的可用性，无论输入什么样的样本数据，系统都不会因为错误数据而停机；
3） 宜支持可信集群内各可信环境间的数据同步和持久化，使得可信集群具备支持可信计算和可信存储的能力；
4） 宜支持训练和预测的集群化和跨机房灾备，使服务具备故障转移和容灾能力，提升系统的可用性；
5） 宜具备在出现故障（比如服务器故障、硬盘故障、网络故障、关机、重启等）后系统进行自动容灾恢复的能力，包括数据备份和恢复等。

8.4 兼容性要求

1） 共享学习算法应兼容非共享学习版本机器学习算法，并确保模型效果与非共享学习版本的模型效果基本保持一致
2） 宜支持水平切分和垂直切分，支持模型训练和模型预测，支持集中式共享学习和去中心化共享学习，覆盖广泛的场景
3） 宜支持模型训练和预测中所需要的各种周边辅助算子，包括数据分析，隐私求交等多种数据预处理算子;
4） 宜支持异构硬件平台和不同的操作系统；
5） 对于可信执行环境方案，宜能在主流的可信执行环境上正常运行；

8.5 性能要求

1） 在 TEE 方案中，宜支持远程双向认证，可信密钥同步等可信执行环境方案的集群化技术，以便将支持将单机可信环境扩展成为可信集群，提升集群的计算能力
2） 在安全多方计算方案中，宜支持训练和预测节点的集群化扩展，以提升系统的服务能力。

8.6 易用性要求

1） 应支持直接进行机器学习训练,并提供一套完备的 API,可以直接通过 API 发起各种训练任务；
2） 在 TEE 方案中，应通过提升 SDK，降低用户的接入成本；
3） 在 MPC 方案中，应通过提供一键式部署脚本，降低用户的部署成本
4） 宜提供易用性的开发框架，使用户在开发业务逻辑时，不需要关心分布式化的逻辑；
5） 宜提供应用运行时动态修改配置的服务，并提供图形化的集中化管理界面；
6） 在 TEE 方案中，宜支持采用基于心跳的 Enclave(可信环境具有高访问权限的私有内存区域)动态升级机制，让用户尽可能少地参与 Enclave 升级过程，使 Enclave 的升级（新建 Tunnel、灰度验证、下线 Enclave、失败回滚）对用户尽可能的透明，提升用户体验。

9 共享学习系统安全要求

9.1 身份鉴别要求

1） 应具备对接入共享学习系统用户的身份鉴别功能；
2） 应支持远程认证，支持用户远程确认运行在可信环境具有高访问权限的私有内存区域中的代码是否符合预期；
3） 宜支持对同一用户采用两种或两种以上组合的鉴别技术（口令验证、邮箱验证、短信验证等）实现用户身份鉴别。

9.2 访问控制要求

1） 应具备对接入共享学习系统内的用户数据操作进行权限验证的功能；
2） 当会话空闲超过 30 分钟，系统应要求用户重新验证或重新激活会话；
3） 宜支持对不同用户进行细粒度访问控制。
4） 在 TEE 方案中，宜基于可信执行环境、远程认证技术，搭建数据提供方可信赖的数据授权系统，通过技术层面而非第三方信赖来保证平台不会窥探、篡改数据提供方的数据。

9.3 安全审计要求

1） 应具备对接入共享学习系统内的用户数据操作进行日志记录和日志审计的功能；
2） 应保存用户的操作日志。

9.4 数据安全要求

1） 应具备对敏感数据（比如个人信息、商业数据等）进行加密传输和存储的功能；
2） 应支持将数据的传输限制在特定授权节点间；
3） 应保证会泄露用户隐私的数据不泄露给其他数据提供方、协调方或用户。
4） 应确保特征、样本等数据的保密性、完整性和可用性，确保不被未授权用户非法获取。
5） 各数据提供方交互的信息不能包含以任何形式能够定位或被大致定位到个体的隐私数据；
6） 应使用安全的传输协议或安全的传输通道，保证数据传输链路的安全可靠，防止被攻击；
7） 应支持对可信执行环境的销毁；
8） 在多方安全计算方案下，应保证泄露用户隐私的数据不出本地，多数据提供方之间仅交换随机数或参数；
9） 对涉及个人信息的操作，应符合《GB/T 35273-2017 信息安全技术 个人信息安全规范》的要求

附　录　A　共享学习的使用场景（规范性附录）

A.1　智能风控

基于共享学习，可以实现数据融合、联合建模以及模型发布一体化方案，实现大数据风控能力。在数据提供方丰富变量的融合建模下，在具备用户端授权，隐私数据受保护的前提下，提升风控效果。

A.2　智能营销

共享学习可以提供精准权益策略，提高风险识别率的安全合规共享环境。比如在车险场景下，共享学习可以显著提升车险的差异化权益能力。在通过购险前的用户授权条件下，帮助保险公司制定更好的销售策略。通过 “从人”（从车主信息）因素能够细分不同风险的用户，对车主进行精准画像和风险分析，实现‘千人千面’的精准权益策略。