**分类号： 单位代码：10422**

**密 级： 学 号：202037055**



（同等学力申请硕士学位）

**论文题目：基于联邦学习的在线教育辍学预测研究**

**作 者 姓 名 唐献**

**培 养 单 位**

**专 业 名 称**

**指 导 教 师**

**合 作 导 师**

**2023年06月01日**原 创 性 声 明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的科研成果。对本文的研究作出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律责任由本人承担。

论文作者签名： 日 期：

关于学位论文使用授权的声明

本人同意学校保留或向国家有关部门或机构送交论文的印刷件和电子版，允许论文被查阅和借阅；本人授权山东大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或其他复制手段保存论文和汇编本学位论文。

(保密论文在解密后应遵守此规定)

论文作者签名： 导师签名： 日 期：

**摘 要**

文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本。

文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本。文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本。

**关键词：**关键词1；关键词2；关键词3；关键词4

**ABSTRACT**

Text text text text text text text text text text text text text text text, text text text text text text text, text text text text text text text text text text text text text text. Text text text text text text text text text text text text text text text, text text text text text text text, text text text text text text text text text text text text text text. Text text text text text text text text text text text text text text text, text text text text text text text, text text text text text text text text text text text text text text. Text text text text text text text text text text text text text text text, text text text text text text text, text text text text text text text text text text text text text text. Text text text text text text text text text text text text text text text, text text text text text text text, text text text text text text text text text text text text text text.

Second paragraph text text text text text text text text text text text text text text, text text text text text text text, text text text text text text text text text text text text text text. Text text text text text text text text text text text text text text text, text text text text text. Second paragraph text text text text text text text text text text text text text text, text text text text text text text, text text text text text text text text text text text text text text. Text text text text text text text text text text text text text text text, text text text text text.

**Key words:** keywords1; keywords2; keywords3; keywords4

**目 录**

[**摘 要** II](#_Toc126947111)

[**ABSTRACT** II](#_Toc126947112)

[1 绪论 2](#_Toc126947113)

[1.1 研究背景与意义 2](#_Toc126947114)

[1.1.1 研究背景 2](#_Toc126947115)

[1.1.2 研究意义 2](#_Toc126947116)

[1.2 国内外研究现状 2](#_Toc126947117)

[1.3 论文研究内容与创新点 2](#_Toc126947118)

[1.3.1 论文研究内容 2](#_Toc126947119)

[1.3.2 论文创新点 2](#_Toc126947120)

[1.4 论文组织结构 2](#_Toc126947121)

[2 相关理论与技术 2](#_Toc126947122)

[2.1 联邦学习介绍 2](#_Toc126947123)

[2.1.1 联邦学习的分类 2](#_Toc126947124)

[2.1.2 联邦学习的架构 2](#_Toc126947125)

[2.1.3 联邦学习的安全机制 2](#_Toc126947126)

[2.2 深度学习介绍 2](#_Toc126947127)

[2.2.1 卷积神经网络 2](#_Toc126947128)

[2.2.2 注意力机制 2](#_Toc126947129)

[2.3 异常点检测技术介绍 2](#_Toc126947130)

[2.3.1 介绍 2](#_Toc126947131)

[2.3.2 异常件检测技术常用算法 2](#_Toc126947132)

[2.4 本章小结 **错误!未定义书签。**](#_Toc126947133)

[3 基于联邦学习的在线教育辍学预测模型方案设计 2](#_Toc126947134)

[4 框架设计 **错误!未定义书签。**](#_Toc126947135)

[5 结论与展望 2](#_Toc126947136)

[**结 论** 2](#_Toc126947137)

[**参考文献** 2](#_Toc126947138)

[**致谢** 2](#_Toc126947139)

[**附 录** 2](#_Toc126947140)

1. 绪论
   1. 研究背景与意义

### 研究背景

随着信息技术的迅速发展以及计算机、手机等终端的普及，在线教育（e-Learning）逐渐成为现代教育方式变革的潮流。在线教育所提供的灵活与便捷的教育方式，逐渐满足现代人对高效、灵活、自主学习的需求。2019年，教育部等多部门发布《关于促进在线教育健康发展的指导意见》指出要充分利用现代信息技术，提供并不断优化在线教育服务，丰富现代学习方式[1]。用户在线学习的过程中，产生大量教育记录数据，例如：习题集完成时间以及正确率、参与课程讨论、查看在线学习资源等学习行为数据。通过对教育大数据进行研究分析从而对平台课程建设者以及教师提供支撑以提高教学效果。在线教育为人们提供了优质丰富的互联网教育资源以及便捷灵活的学习方式，但是，在线教育目前存在教学模式单一、缺乏有效的教学监督以及沟通方式等原因，致使在线教育用户流失现象严重，真正的学习效果受到多方质疑。有研究报告指出，以MOOCs平台为代表的在线教育平均辍学率高达90%以上[2]。因此，如何降低在线教育的辍学率，提高课程的通过率成为在线教育亟待解决的问题。通过对用户的在线学习记录数据分析，提前预测其是否会辍学，在课程学习早期及时发现用户辍学的可能，以便平台或者教师进行有效的干预以激励用户保持学习状态。

现有思路往往从用户在线教育数据中提取特征，利用机器学习算法建模预测用户是否辍学。然而，现实场景下单一的在线教育平台数据样本量不足，建模效果不佳，即数据孤岛问题。尤其是在以深度学习为代表的机器学习算法中，参数数量通常达到了百万级以上，所以样本量应当足够大。为了提高模型预测性能，将平台参与方数据收集共享是一种理想的解决办法。但是，数据集中往往蕴含大量的用户个人敏感信息（例如，用户在线讨论信息、课程浏览信息等），数据的直接共享可能导致严重的个人隐私泄露。为了保护用户隐私，近年来各国制定了相关的法律法规对数据共享进行约束和规范。比如，2012年美国颁布了《网络环境下消费者数据的隐私保护—在全球数字经济背景下保护隐私和促进创新的政策框架》[3]；2016年欧盟通过了《通用数据保护条例》（General Data Protection Regulation， GDPR）[4]；2017年中国通过了《中华人民共和国网络安全法》[5]。这一系列的法律法规对数据共享与发布进行约束与规范。由此可见，若不能解决多方协作进行模型训练导致的隐私泄露问题，这不仅会对个体用户带来伤害，同时对在线教育平台造成一系列经济、法律上的问题。

对于传统的机器学习任务，首先需要将数据分散的数据集中，然后进行模型训练。但是此做法却带来了一些隐私和安全方面的问题，例如数据泄露和个人信息滥用等。为了解决这些问题，联邦学习（Federated Learning）被提出。联邦学习是一种分布式机器学习框架，最早于2016年由谷歌研究人员提出[6]，使得各参与方在不共享原始数据以及加密形态下共同建模。联邦学习的主要特点在于保护隐私保护和数据安全。它允许多个独立参与者协作进行机器学习的模型训练，而不需要将数据全部集中到一个地方，从而提高了数据隐私的安全性。联邦学习通过对参与方之间所共享的数据进行加密或扰动，而原始数据保留在用户本地，达到数据可用不可见的目的，从而避免了数据集共享所带来的隐私泄露风险。

### 研究意义

研究在线教育的辍学问题具有重要意义。一方面有利于教师和平台及时掌握用户的学习状态，发现发现存在辍学风险的用户，并为其提供针对性的指导或干预，提升学习效果。另一方面有利于实现在线教育平台的教育价值以及商业价值，具体可以从以下两个个方面谈起。

第一，有助于教师和平台及时掌握学生的学习行为变化，进而改善课程内容，提升学习效果。教师和平台通过发现学生学习中的潜在规律，进而制定个性化教学方案，改善课程组织方式和讲解内容，考虑如何设置课程时长和如何根据学生层次去安排课程，满足不同学生的需求，提高学生对课程的满意度，从而保证学生的学习效率，进而降低辍学率。此外，辍学预测可以帮助在线教育平台评估课程效果，分析课程缺陷，发现教学方式的不足，从而不断提高教师教学水平以及课程质量，增强课程竞争力。

第二，促进在线教育平台智能化建设，更好地提供付费教学服务。当前的在线教育平台仅提供基础的学习服务，未能充分利用用户的学习数据，缺乏对用户学习状态的深入了解。然而，在线学习过程中，用户产生的大量学习数据，例如课程学习数据、作业完成数据、互动沟通数据等，可以作为智能化建设的有力基础。通过对这些教育大数据进行充分挖掘分析[7]，在线教育平台可以充分掌握用户的学习状态，更好地满足用户的学习需求，提高用户的留存率，从而提高在线教育平台的付费教学服务质量。

* 1. 国内外研究现状

随着在线教育的迅猛发展以及数据获取、存储技术的提高，教育大数据时代已经来临。然而学习者辍学现象始终是教育行业存在的问题，在线教育行业该问题尤为突出。尤其是近年来大规模开发课程的出现，使得部分研究人员开始分析在线教育平台产生的各种学习行为数据，通过对学习数据分析建模，将辍学问题转化为机器学习当中的二分类问题，提前发现用户的辍学风险，供平台和教学者提供合理帮助，以提高在线教育的教学质量。在线教育平台中，有大量用户的学习行为数据，如点击流数据、作业提交情况、成绩信息、论坛讨论参与度等[8]。平台通过挖掘分析学生历史学习数据，可以得知影响辍学的关键因素，并预测学生是否会退学[9]。下面分别介绍在线教育平台的辍学预测的两种主要方法：基于传统的机器学习方法和基于以深度学习为代表的机器学习方法。

### 基于传统的机器学习辍学预测研究现状

传统的机器学习方法将用户辍学问题转变为分类问题，算法模型包括：逻辑回归（Logistic regression，LR）、决策树（Decision Tree，DT）、线性（Support Vector Machines，SVM）、梯度提升树（Gradient Boosting Decision Tree，GBDT）[10][11][12]等，该类方法通过建立具备可解释性的算法模型，提高模型预测精度。Liang[13]等人考虑到用户的个人身份信息、课程属性信息以及在线学习行为数据，使用梯度提升决策树（GBDT）算法预测用户辍学风险，可以预测到用户将来10天的辍学概率。Hagedoorn[14]等人则通过比较不同机器学习的分类算法，预测用户未来一周是否辍学，最后实验证明逻辑回归模型精度略高于随机森林以及AdaBoost算法。Vitiello[15]等人在edx平台课程中，利用用户原始学习行为特征，使用决策树建模预测辍学问题。从已有研究可以看出，点击流数据是在线教育辍学预测使用较为广泛的数据，尤其是在以MOOCs为代表的在线教育平台中[16]。点击流数据特征非结构化，并含有大量的冗余信息，一方面增加算法模型复杂度，另一方面也影响到了模型的算法性能。Goel[17]等人考虑到点击流数据应用的广泛性，利用数据挖掘技术从点击流数据中挖掘关键学习行为特征，并送入逻辑回归模型用于预测辍学。Chui[18]等人提出了一种基于训练向量的简化支持向量机（RTV-SVM），通过删除冗余向量以减少训练时间和支持向量，同时保证了模型预测精度。Jin[19]等人考虑到用户的学习行为数据包含丰富的特征信息，设计一种特征提取方法从用户每周学习行为提取特征，采用支持向量回归（SVR）模型作为辍学预测模型，该模型相较于随机选择参数的SVR模型拥有更好的预测性能。Walk、Dass[20]等人研究了2016~2020年Open edX平台的数据，提出一种基于随机森林（RF）的算法预测用户辍学情况，并且使用Shapely值对RF中的关键特征以及特征相互作用进行了解释。PARK[21]等人通过收集用户在线学习日志，例如：用户个人属性、学习行为数据、课程属性信息等，利用随机森林算法查找影响辍学的关键因素。研究显示与辍学息息相关的因素是作业的分值与提交时间、登录次数与间隔、电子资料学习频率等。

相关研究表明，特征工程对传统的机器学习辍学预测模型精度影响较大，输入数据的质量直接决定了模型精度上限。因此从原始在线学习行为数据中识别并提取有效特征是研究关键[22]。Gelman[23]则提出基于非负矩阵分解的特征提取方法，进一步发现若干个影响辍学行为的关键特征。Bote-Lorenzo[24]等人所设计的基于相关性的特征选择算法，可以同时识别关键性特征、冗余特征。Qiu[25]等人则提出一个带有特征选择的集成框架（FSPred）来预测辍学，FSPred框架包含特征生成、特征选择以及辍学预测模块，FSPred框架通过用户在若干天内的在线学习数据生成细粒度的特征，并通过特征选择模块得到最优特征子集，并作为逻辑回归算法模块的输入，最后进行辍学预测，实验结果表明该方法与其他辍学预测方法具有相当的精度结果。Ai[26]等人设计一种在线教育辍学预测系统，该系统可以从点击流数据自动提取学习特征，并且基于设计的特征综合选择策略有效筛选出与辍学相关的关键特征。相关实验表明，该模型可以有效提高辍学预测精度。Jin[27]等人定义与样本距离、标签同时相关的最大邻域概念，研究样本的初始化权重计算、实现算法，进一步使用智能优化算法优化样本初始权重，最后将加权训练样本训练出的分类器作为预测模型。相关实验表明，该方法可以显著提高辍学预测模型性能。

### 基于深度学习的辍学预测研究现状

后疫情时代来临，在线教育平台用户数量迅猛增加。随之产生的在线学习数据呈现出高维、动态、非线性等特点，为辍学预测带来一定挑战。近年来，以深度学习为代表的机器学习算法神经网络层数多，适应性较强，在多个领域表现优异[28]，并在教育领域的辍学预测方面被广泛研究。Wang[29]等人利用CNN自动提取特征的优势，与循环神经网络（RNN）结构从而构建端到端的预测模型。实验数据表明，该方法获得模型精度与基于特征工程的方法相当。Qiu[30]等人利用CNN构建端到端的算法模型，通过利用不同时间窗口从而对数据进行变换，自动提取特征，获得更好的特征表示，该方法显著提升了模型精度。在XuetangX平台的学习行为特征进行实验表明，模型精度达到86.75%。Feng[31]等人提出的方案，利用数据平滑技术得到不同数据的特征值，考虑到用户属性信息、课程属性信息以及在线学习信息，结合CNN以及注意力机制方法构建的模型，提高辍学性能，实验结果表明精度达到86.71%。Chen[32]等人提出一种新型混合模型，该算法结合决策树（DT）以及极限学习机（Extreme Learning Machines，ELM），通过决策树选择与辍学相关的重要特征并确定特征权重，最后映射到ELM中来优化ELM结构。该方案相较于上述方案，不需要迭代训练便可取得较好的预测效果。Yin[33]等人为提取特征，将单一特征映射到高维空间，并使用注意力机制（Attention Mechanism）以及卷积神经网络（CNN）重构特征，从而获得更好的关键特征表示。Lai[34]等人构建了一个基于Broad Learning System模型，该模型首先将输入数据映射为特征节点层，通过特征节点层激活，进而生成增强节点层，最后结合特征层和增强层进行线性变换，最后的输出层用于辍学预测。在KDD CUP 2015提供的数据集上进行实验，实验显示该方法相较于主流方法显著减少训练时间，同时拥有更好的训练精度。Sahin M[35]等人构建了一种基于自适应神经网络和模糊推荐系统的ANFIS模型，该方案通过发挥模糊推荐系统以及神经网络的优势，获得了较高的预测精度。

在线教育学习数据具有典型的时间序列特征，一些研究人员利用点击流数据构建辍学预测模型。Wang[36]等人充分考虑到时间特征信息，根据时间间隔对学习行为数据加权，提出基于E-LSTM的辍学预测模型。Zheng[37]等人从在线教育平台中提取时序特征，利用分类回归树算法（CART）过滤得到关键特征并计算特征权重，基于特征权重对时序数据进行加权并作为CNN网络的输入，用于辍学预测。Qian Fu[38]等人首先使用CNN提取局部特征，并使用核策略建立特征关系。其次，将提取出的高维特征向量作为长短期记忆网络的输入，以获取包含时序信息的特征向量。再次，使用注意力机制得到特征向量中每个维度的权重。最后，将特征向量送入分类器得到用户辍学的概率。Mubarak[39]等人为了提高模型预测精度，在提取特征步骤完成后，考虑到漏报、误报等误分类代价，在损失函数中加入代价敏感技术，提出利用CNN和长短期记忆网络的CONV-LSTM模型。Wang[40]等人则综合考虑用户短期在线学习行为特征以及长期的变化模式，用于预测辍学倾向。主要包含以下两个阶段：第一，信息提取。即从在线学习日志中提取数据。第二，信息处理。设计一个基于卷积残差递归神经网络（CRRNN）用于辍学预测。Xu[41]等人则根据用户在相邻时间段内具有相似的学习行为特征，基于上述原因提出Lie组区域协方差矩阵提取学习行为的局部相关特征信息，进而构建带有CNN网络来提取更抽象层次的学习特征用于辍学预测。

* 1. 论文研究内容与创新点

### 论文研究内容

第一，考虑到不同在线教育平台的局部数据集属性相同，样本分布在不同在线教育平台，符合样本重叠较少，特征重叠较多的横向联邦学习场景，此外，机器学习模型卷积神经网络拥有较高的分类效果，因此提出基于横向联邦学习的在线教育辍学预测模型研究方法，其中机器学习模型采用卷积神经网络（CNN）。

第二，考虑到在线教育平台与离线教育平台拥有用户的不同属性，即同一用户的不同属性分别属于不同的局部数据集，符合样本重叠较多，特征重叠较少纵向联邦学习场景。这里提出基于线性纵向联邦学习的在线教育辍学预测模型研究方法，此方法的核心是使用同态加密保证计算过程的安全性，使用差分隐私保证计算结果的安全性。相对于已有的分别基于同态加密、差分隐私的解决方案，解决了在现实生活中难以找到可信第三方的问题，增强了隐私保护效果，并通过敏感度上界计算和扰动方法优化，降低噪音规模，保证了模型精度。

### 论文创新点

本文创新点在于：

第一，针对多方协作进行模型训练提出的去中心化的联邦学习架构，相较于现有的工作，具有以下优势：该联邦学习架构可以验证聚合结果的完整性，首次利用异常点检测技术检测腐败服务器。

该方案架构可以验证聚合结果的完整性，首次利用异常点检测技术检测腐败服务器，解决了腐败服务器破坏聚合结果以致模型难以收敛的问题。完整性验证挑战在于被攻击的腐败服务器会恶意破坏聚合结果并下发给参与方，参与方难以采取有效措施去判断当前聚合结果的有效性。现有聚合结果的完整性验证工作较少，在去中心化的联邦学习架构的完整性验证现有工作中[42]，其方案设计并不能准确检测腐败服务器是否对聚合结果的恶意破坏，此外现有工作方案上传的部分数据含有个体敏感信息，未进行隐私处理，存在隐私泄露风险。

第二，设计一种针对用户ID编码加密的方法，在保护ID隐私的前提下完成样本ID对齐，采用差分隐私技术扰动敏感数据，完成针对同一个体的所有记录的统计信息求解。该方案相对于现有工作的优势体现在通过对用户ID编码加密，再由shuffler打乱，从而减少多方样本对齐的时间复杂度，提高计算效率。

第三，提出基于同态加密和差分隐私技术的线性纵向联邦学习隐私保护方法。此方法的核心是使用同态加密保证计算过程的安全性，使用差分隐私保证计算结果的安全性。相对于已有的分别基于同态加密、差分隐私的解决方案，本发明解决了在现实生活中难以找到可信第三方的问题，增强了隐私保护效果，并通过敏感度上界计算和扰动方法优化，降低噪音规模，保证了模型精度。

* 1. 论文组织结构

第一章为绪论。主要介绍了在线教育的发展背景以及现阶段存在的辍学率问题、数据孤岛问题，进而引出本文工作，即基于联邦学习的在线教育辍学预测模型研究。然后分别介绍了在线教育、联邦学习的国内外研究现状，最后介绍本文的研究内容和创新点。

第二章为相关介绍。首先介绍了联邦学习的分类、架构以及本文将要用到的安全技术，例如：差分隐私、加法秘密共享等。然后介绍了第三章使用的机器学习模型卷积神经网络结构以及注意力机制，最后介绍异常点检测算法，从而为本文研究提供可靠的理论支撑。

第三章为基于横向联邦学习的在线教育辍学预测算法。待补充

第四章为基于线性纵向联邦学习的在线教育辍学预测算法。待补充

第五章为总结和展望

1. 相关理论与技术
   1. 联邦学习介绍

联邦学习（Federated Learning）是一种分布式机器学习技术，它允许多个边缘设备或服务器共同参与模型的训练。谷歌研究人员于2016年首先提出联邦学习框架[43]。该框架允许每个设备都拥有本地的数据样本，并通过在各自的设备上运行迭代来更新模型参数。这种方法能够有效地利用大量分散在不同地点的数据，并且可以在保护用户隐私的同时进行训练。联邦学习主要应用在数据隐私和数据分布不均等问题上，传统的集中式机器学习模型存在数据隐私泄漏的风险，而联邦学习通过将数据存储在边缘设备上，避免了数据的集中存储和传输，从而降低了数据隐私泄露的风险。

联邦学习定义了一个机器学习的框架，这个框架使得各个数据拥有者在不交换本地数据的情况下进行协作。这个模型实际是虚拟地将各方数据聚合到一起，在联邦机制下，各参与者的身份和地位相同，可建立共享数据策略。由于数据不发生转移，因此不会泄露用户隐私或影响数据规范。达到保护数据隐私的目的，满足合法合规要求。联邦学习优化的目标函数表示如公式2-1、2-2所示：

其中N表示参与方数目，为用户i赋予的权重，实际应用中常根据用户i所包含的样本数目占据总样本数目的比重确定。表示用户i局部数据集的经验损失函数，表示用户i的对应样本j的经验损失函数。

### 联邦学习的分类

联邦学习根据用户以及数据集的特点，分为横向联邦学习（horizontal federation learning）、纵向联邦学习（vertical federation learning）以及迁移联邦学习（transfer federation learning）。假设参与方拥有局部数据集，样本ID集合记为为，属性特征（attribute）集合记为，标签记为。

横向联邦学习指的是多个用户拥有相似的数据集，每个用户都在本地训练自己的模型，然后将训练好的模型参数上传到服务器上，服务器会将多个用户的参数进行聚合，最终得到一个全局的模型。如图2-1所示：横向联邦学习适用于多用户数据集相似或相同的情况。例如，不同地区的银行，业务类似，但是用户重叠较少。参与方和数据分布表示如下：

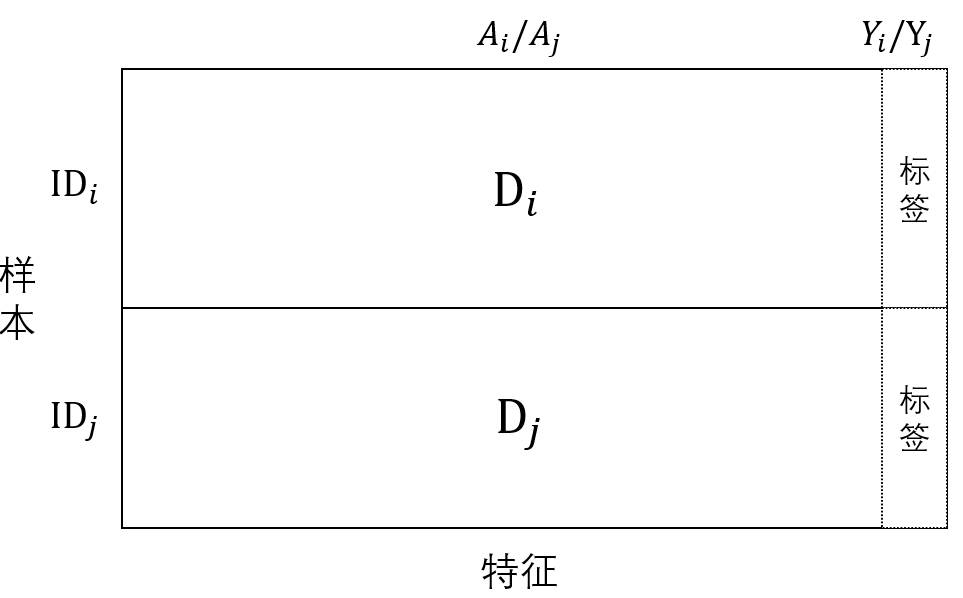


图2-1 横向联邦学习示意图

Fig. 2-1 Schematic diagram of horizontal federated learning

纵向联邦学习指的是一个用户拥有多个不同来源的数据集，每个数据集都在本地训练自己的模型，最终将多个数据集训练出来的模型参数上传到服务器进行聚合，得到一个全局的模型。纵向联邦学习适用于一个用户拥有多个不同来源的数据集的情况。如图2-2所示：

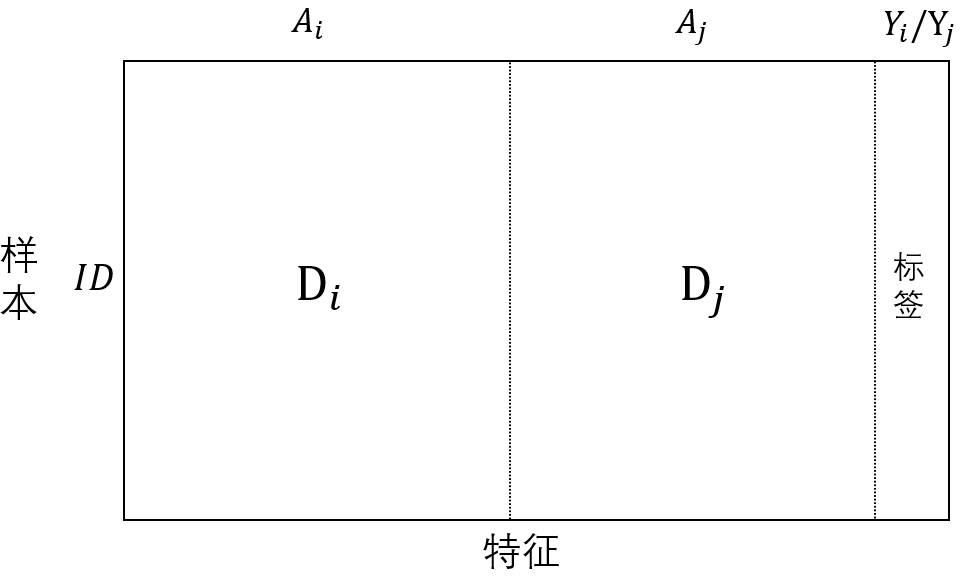


图2-2 纵向联邦学习示意图

Fig. 2-2 Schematic diagram of vertical federated learning

迁移联邦学习指的是从一个领域的数据集中训练出来的模型可以被迁移到另一个领域，这种方法可以使用户在没有大量本地数据的情况下训练出高质量的模型。如图2-3所示：

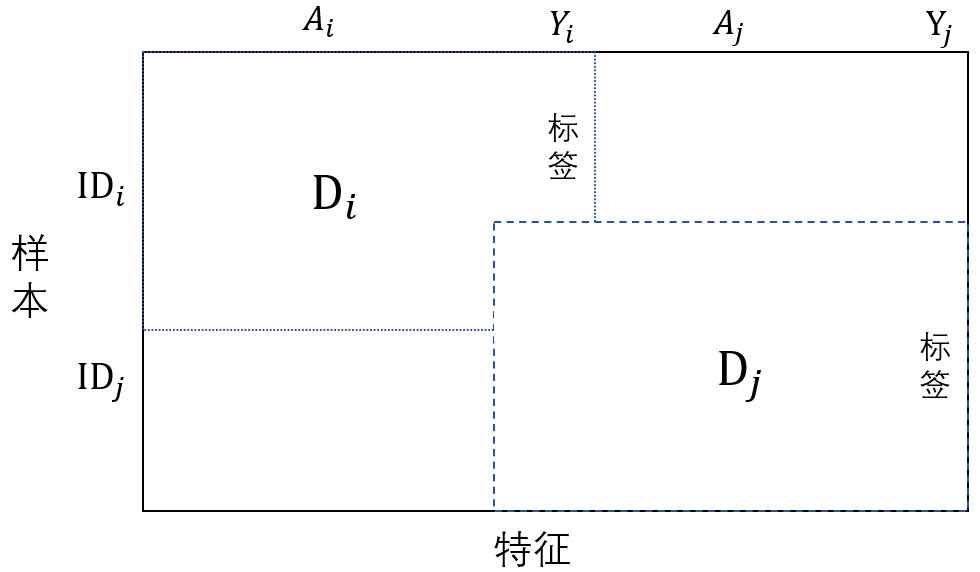


图2-3 迁移联邦学习示意图

Fig. 2-3 Schematic diagram of transfer federated learning

### 联邦学习的架构

1. **中心服务器架构**

该架构[44][45]主要由两部分组成：半可信的中心服务器和参与训练的用户节点。服务器具有充足的计算资源和计算能力，负责模型参数的分发、用户设备的调度以及全局模型的聚合更新，充当一个负责协调的第三方的角色。用户节点主要完成基于本地数据集的计算任务，根据接收到的模型参数进行更新迭代。用户与用户之间一般不会直接进行通信，参数传递只发生在用户与服务器之间。执行流程如下：

第一：参数服务器将当前的全局模型参数分发给各个用户节点。

第二：用户节点在本地数据集上训练模型并更新模型参数。

第三：用户节点将更新后的模型参数发送给参数服务器。

第四：参数服务器收集所有用户节点的更新后的模型参数，并对这些参数进行聚合更新，得到新的全局模型。如图2-4所示：

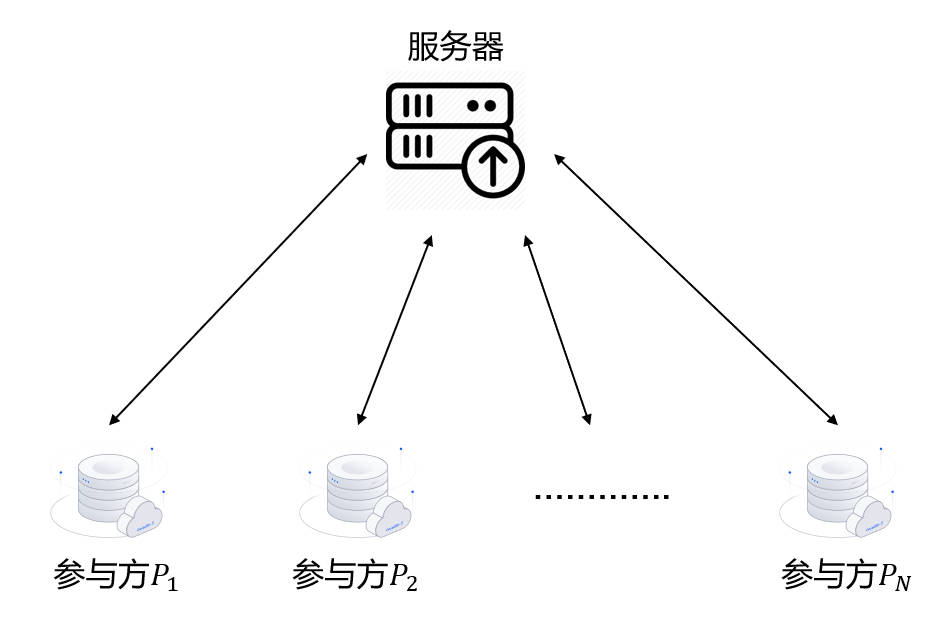


图2-4 中心服务器架构

Fig. 2-4 Central server architecture

但是中心化场景下由于用户设备的数据分布可能不同，模型训练可能存在数据偏差的问题。其次，由于用户设备的连接状态可能不稳定，模型训练可能存在网络中断的问题。最后，由于参数服务器承担了大量的计算负载，可能存在单点故障的风险和较差的可拓展性能去中心化架构。

1. **去中心化架构**

由于中心化联邦学习架构存在的上述缺点，提出去中心化架构[46]。在该架构中，没有中心服务器，所有用户节点之间直接进行通信和交换参数，用户节点之间通过像P2P网络这样的网络拓扑结构进行通信。在这种网络拓扑中，每个用户节点都可以作为服务器和客户端，根据需要与其他节点进行通信。因此可以避免单点故障的风险和提高系统的可拓展性。此外具有更高的隐私性。在中心服务器架构中，所有用户的参数都会被汇集到中心服务器上，这增加了黑客隐私攻击的风险。而在去中心化的联邦学习中，用户节点之间通过对等网络进行通信。这意味着用户之间没有明显的客户端-服务器关系，而是相互独立的。在这种架构中，用户可以自由地加入和离开网络，并且可以在网络上进行自由的数据交换。如图2-5所示：

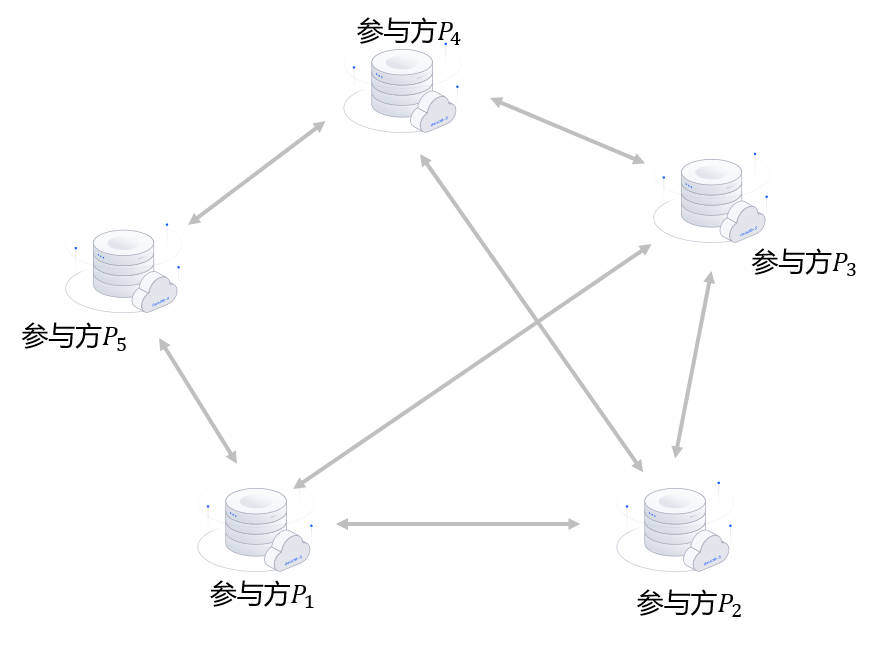


图2-5 去中心化服务器架构

Fig. 2-5 Decentralized server architecture

去中心化的联邦学习架构的缺点在于，它需要更多的协调和同步，因为用户之间直接进行通信。这可能会导致更多的网络延迟和错误。此外，由于数据和计算分散在各个节点之间，网络带宽和计算资源的利用率可能会变得低下。

### 联邦学习的安全机制

联邦学习过程中，参与方所共享的模型参数、梯度等可能导致隐私泄露风险。例如，在医疗领域中，如果个人的健康数据被泄露，可能导致严重后果。此外，恶意参与方可能会试图破坏联邦学习的过程，例如通过提供伪造的模型参数来破坏模型收敛。为了解决上述问题，研究人员使用多种安全技术来保障训练过程的安全性。这里主要介绍本工作用到的两种安全技术。

1. **差分隐私**

差分隐私[47]（Differential Privacy）采用一种噪音随机扰动机制，使得当单条记录改变后，输出的分布变化很小。例如对于相差一条记录的相邻数据集，查询他们获得相同输出的概率值比较接近，使得即使用户获取输出结果，也无法推断输入数据来自哪一方，从而为个体敏感信息提供隐私保护。在差分隐私领域中，相邻数据集定义为若数据集有且仅有一条记录不同，其余记录均相同，则被成为相邻数据集。

基于相邻数据集，差分隐私被定义如下：对于相邻数据集，随机化算法,以及任意的输出，算法满足-差分隐私，当且仅当：

全局敏感度定义如下：对于一个算法函数，是相邻数据集，公式为：

上述符号表示范数。用来衡量隐私保护强度。越小，隐私保护强度越高，对应的数据效用性也越低。噪音量的大小取决于隐私预算以及全局敏感度。

差分隐私可以通过两种典型机制拉普拉斯机制[47]、高斯机制[48]来实现。为保证数据模型的可用性，提供更多的隐私预算。机器学习领域中，往往采用松弛版的差分隐私[49]。定义如下：

表示失败概率。当等于0时，即满足严格的差分隐私，但是实际应用中，为了满足更好的建模效果，往往令，而是取一个较小的值。

1. **秘密共享**

秘密共享[50]（Secret Sharing）是一种用于保护秘密信息安全的技术。它将一个秘密信息拆分成多个份额，并分配给不同的参与者管理。只有在若干个参与者一同协作的情况下，才能恢复原始的秘密信息。这样，即使某一个参与者被黑客入侵，秘密信息也不会被泄露。秘密共享技术在联邦学习中可以用来保护模型参数和梯度的隐私。本研究工作使用一种高效的加法秘密共享方案，介绍如下。

秘密持有者将秘密x拆分为，并将秘密x的份额保存在方，任意少于个份额均无法恢复秘密x。对于(n,n)门限的高效秘密共享，即其中。基本流程如下：

秘密共享Sharing：将秘密x、份额数据m作为输入，输出m个关于秘密x的份额。

第一：随机选取m-1个随机数作为m-1个份额

假设，随机选取m-1个份额

第二：计算第m个份额

秘密恢复Reconst：输入m个秘密份额，恢复秘密x：

* 1. 深度学习介绍

深度学习是近年来机器学习领域对数据进行表征学习的重要算法，深度神经网络通常指隐藏层较多的神经网络，能够为复杂非线性系统提供建模，因此提供更高的抽象层次，提高模型能力。根据神经网络结构可以分为特征分类的全连接神经网络（FCNN）、计算机视觉领域的卷积神经网络（CNN）以及时序特征提取的循环神经网络（RNN）。这里重点介绍本文研究将要用到深度学习网络，包括卷积神经网络、注意力机制。其中卷积神经网络可以自动进行特征提取，从而避免了复杂的特征提取过程。注意力机制对输入特征考虑不同权重参数，从而获取更具描述性的特征信息。

### 卷积神经网络

卷积神经网络（Convolutional Neural Networks, CNN）的结构包括输入层、卷积层、池化层、全连接层以及输出层五个部分[51]。卷积层用于对输入特征进行降维和特征提取，无论输入特征多大，卷积层参数规模本身都是固定的。池化层的作用是降维，减小模型规模，降低过拟合概率，并且具备旋转不变性。全连接层往往位于网络的后层，整合之前学习到的局部信息，将捕获到的特征信息映射到样本标记空间，从而实现分类。

卷积神经网络的经典结构是LeNet-5模型[52]，共分为5层，如图2-6所示，分别为卷积层C1、池化层S2、卷积层C3、池化层S4以及全连接层C5，它在 MNIST 数据集上的识别精度高达99.2%。

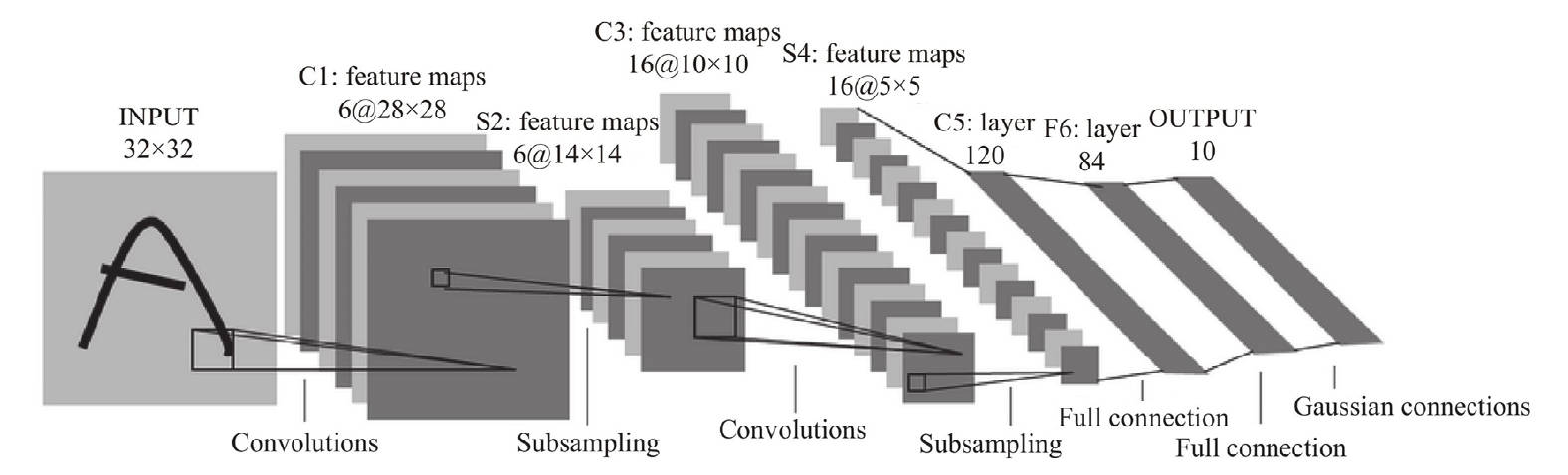


图2-6 LeNet-5网络结构示意图

Fig. 2-6 Schematic diagram of LeNet-5 network structure

1. **卷积与池化操作**

卷积操作以及池化操作如图2-7所示：

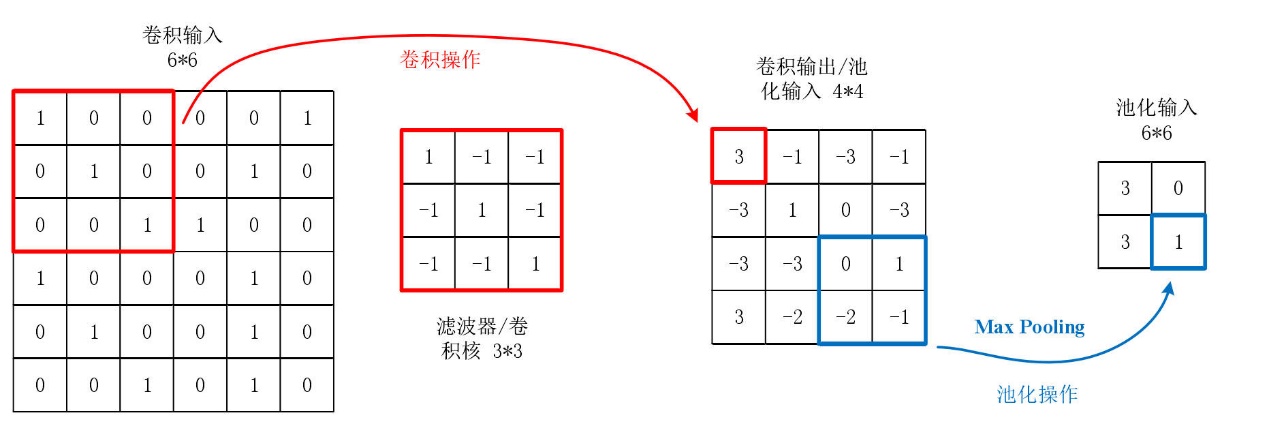


图2-7 卷积池化操作示意图

Fig. 2-7 Schematic diagram of convolution and pooling operations

从图中可以看出，卷积操作的目的是特征提取，利用卷积核在特征矩阵上滑动，对应值相乘相加得到对应输出。常见的卷积操作包括same padding、valid padding。same padding的输入层尺寸等于输出层尺寸，对于矩阵边界进行填零补充。valid padding没有填零操作，输出维度通过固定计算公式得到，具体可见上图。卷积层后面通常会连接池化层，池化层的操作是降低卷积层输出的特征维度，具体有两种方式。其一，maxpool。如图所示，即取滤波器区域中的最大值作为输出；其二是avgpool，即取取滤波器区域中数据均值作为输出

1. **激活函数**

由于卷积操作属于线性操作，所以卷积神经网络使用了激活函数。下面介绍常见的激活函数包括Sigmoid函数、Tanh函数以及ReLU函数。

**Sigmoid函数。**Sigmoid函数是一个单调递增、有界可微函数，函数处处可导、平滑连续。物理意义上最接近神经元。函数表达式如下：

函数输出值范围（0，1）。该函数优点是连续函数，便于求导。同时输出范围有界，可以作为输出层。缺点是在0和1附近梯度接近0，导致反向传播的梯度较小，网络模型参数无法更新，即梯度消失现象。

**Tanh函数。**即双曲正切函数。函数表达式如下：

函数输出值范围（-1，1），优点是解决了Sigmoid函数的zero-centered输出问题问题。实际应用中Tanh会比sigmoid更好，但是仍存在梯度消失问题。

**ReLU函数。**ReLU函数是目前使用最广泛的激活函数，在神经网络模型训练中有着较好的表现。函数表达式如下：

该激活函数将所有负值均映射到0，即激活值为负值时，梯度为0，故只有少量神经元会被激活，即稀疏激活。函数优点是计算高效、收敛较快。在解决梯度爆炸以及梯度消失问题上明显好于Sigmoid函数、Tanh函数[53]。缺点是函数非连续可微，给基于梯度的优化带来一定问题。

### 注意力机制

注意力机制（Attention Mechanism）是一种机器学习领域的数据处理方法，它在处理大量数据时会忽略非关键性信息，有选择性地关注重要信息[54]。相关实验[55]证明此方法可以帮助模型准确高效地获取有用的信息，提高处理效率。

在自然语言处理领域中，使用注意力机制可以更好理解语句含义[56]。例如，在翻译处理任务中，模型需要使用注意力机制识别源语句中的重要信息，并将其翻译成目标语言。在图像识别领域中，模型使用注意力机制在海量图像信息中找到重要特征[57]，如人脸、物体等。

注意力机制通常由一个注意力层来实现。该层通过计算出每个输入的权重值，即表示对应输入的重要性。这些权重值会被用来计算加权平均值，得到最终的输出。计算权重值的方法有点积注意力（dot-product attention）、缩放点积注意力（scaled dot-product attention）、多头注意力（multi-head attention）等。点积注意力是一种简单的注意力机制，它将输入和查询向量做点积，然后用softmax函数转化为权重。缩放点积注意力是对点积注意力的改进，它将点积结果除以输入向量的维度的根号，这样可以防止点积值过大导致softmax函数的值过小。多头注意力是对点积注意力和缩放点积注意力的进一步改进。它采用了多个不同的注意力头来提取不同的注意力信息，再将这些信息拼接在一起。这样可以更好地考虑不同的关注点，并得到更好的结果。

* 1. 异常点检测技术介绍

### 介绍

异常点检测所研究对象为数据集中占比很小的一部分数据实例。目前关于异常（离群点）没有一个明确统一的定义，通常指与数据集中其他大部分数据具有足够大偏差的数据实例。学术界普遍认同Hawkins给出关于异常定义如下：与绝大部分数据相比，异常看起来与众不同。这让人们怀疑异常的产生机制与大部分其他数据不同[58]。从上述定义可以看出来异常的特点：异常在数据集的整体分布中所占比例较小，正常簇群对应的数据实例占绝大多数。而异常检测的目标是在数据集中检测出不符合预期生成机制的少量数据实例。

根据数据范围可以将数据异常分为全局异常点、局部异常点以及异常点聚类。如图2-8给出二维空间中数据异常的示例。其中，数据集可以划分为两个主要簇和，并且数据实例、、远离两个主要簇。根据Hawkins的定义，它们可以被划分为异常。除此之外，还有一种情况是异常点聚类，例如，仍被看作异常。

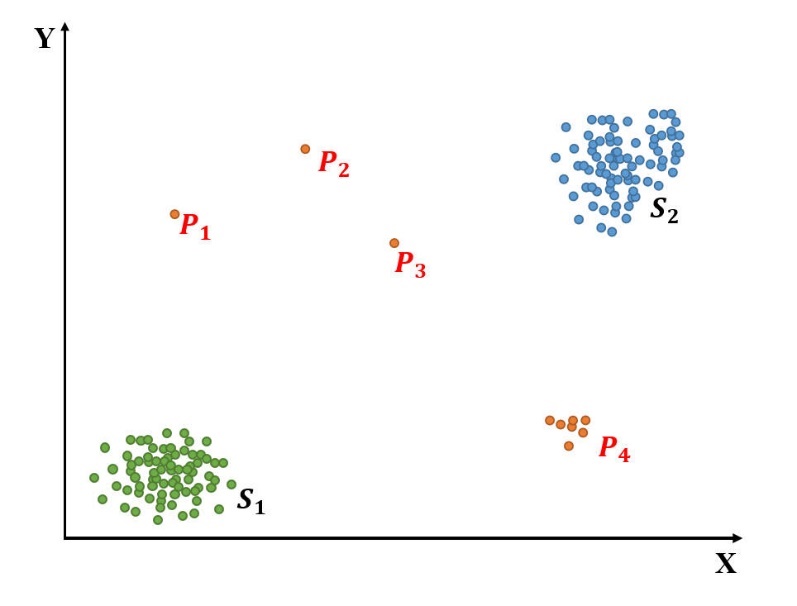


图2-8 二维数据异常示意图

Fig. 2-8 Schematic diagram of two-dimensional data abnormality

与噪声的产生机制不同，大部分异常都有着合理的解释。根据其生成原因大致可以分为以下几种情况：

第一、数据采集、记录、测量等过程出错。客观因素主要包括机械仪器故障、外界自然环境变化、设备异常等，主观因素主要包括技术人员人为操作失当致使采集或数据的生成机制不符合预期，导致数据异常。

第二、自然异常值。由于某种极端因素的影响，例如极端环境、生物界的基因突变等，致使数据自然变异。从概率统计角度看，所收集的目标数据集的数据实例整体会服从某种分布，但在上述极端因素的影响，会有少部分数据实例以极低概率不服从该分布，这些数据实例即为异常。

第三、部分非正常行为或者特殊行为。从数据来源角度讲，某些数据实例来源与数据集中其他数据实例来源有所不同。例如，利用漏洞和安全缺陷进行的网络攻击、金融领域的恶意欺诈行为等。这些行为通常会导致所生成的数据实例在某些属性上发生巨大变化。

第四、新颖性异常行为。例如科学界新的知识工程规则的发现、新的物质发现等，该类异常往往对相关领域具有一定的促进作用。

### 异常件检测技术常用算法

根据数据集是否存在标签，将异常点检测算法分为全监督异常检测、半监督异常检测以及无监督异常检测算法。其中全监督、半监督异常检测算法需要大量预先标签化的数据，实际应用中获取成本较大。而无监督异常检测不需要对数据进行标签化，所以在实际使用中最为广泛。在无监督异常检测的数据集中，包含正常数据以及异常数据，并且前者数据远大于后者数据。无监督异常检测通过训练正常数据，得到算法模型，并计算得到异常分类分数，通过将测试集输入模型得到的异常评分来判断测试数据是否为离群点。

常用异常检测算法基本思想是使用正常样本数据训练模型获取阈值，进而去判断测试数据是否异常。异常检测算法分为基于概率统计模型、基于聚类、基于距离等检测方法[59]。本文简单介绍两种经典的异常点检测方法。

1. **kNN算法**

kNN（kth-Nearest Neighbor distance）算法主要思想是比较目标数据点与其前k个最近邻居的距离来判断当前数据点是否异常[60][61]。具体可以划分为以下几个步骤：首先，预先计算出数据集中数据对象之间的距离；其次，按照距离分别计算出每个数据对象的第k个最近邻居；再次，按照距离将所有数据对象的第k个邻居降序排序；最后设定合适的阈值，确定前n个数据对象作为异常点。

kNN算法优点是思想简单，准确率高，对离群点不敏感，适合处理多分类问题。缺点是当样本不均衡时，对稀有类别预测准确率低。此外，当特征数较多的时候，计算复杂性和空间复杂性较高，无法给出如决策树的规则，可理解性较差。

1. **LOF算法**

LOF算法（Local Outlier Factor，局部离群因子检测方法）[62]是一种基于密度的离群点检测方法，属于无监督学习类型。其基本思想是通过计算得到数值score，该数值可以反应样本的异常密度。score表示目标数据点周围的数据点所处位置的平均密度与目标数据点所处位置的密度比值。score越大于1，表示目标数据点所处位置密度越小于周围数据点所处位置的密度，该数据点越可能为离群点。

1. 基于横向联邦学习的在线教育辍学预测模型方案设计

文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文

1. 基于线性纵向联邦学习的隐私保护方案设计
   1. 引言

本章介绍一种基于线性纵向联邦学习的隐私保护方案，用于纵向联邦学习中的广义线性模型，如常见的线性回归、逻辑回归和泊松回归等。在纵向联邦学习场景中，由于同一样本的不同属性分布在不同的节点中，在计算损失函数时，需要多方交互中间结果，存在隐私泄露的风险。现有工作思路分别是基于同态加密（Homomorphic encryption，HE）、差分隐私（Differential Privacy，DP）技术的解决方案，主要存在以下问题：基于同态加密的方案中，一方面现实场景可信的第三方难以找到，另一方面即使能找到可信第三方（假设由权威机关，如政府来扮演），由于梯度中包含用户的属性信息，所以第三方解密后发送给各方的梯度存在隐私泄露的风险[1]。基于差分隐私的解决方案中，由于敏感度上界计算以及扰动方式原因，致使噪音规模较大，影响建模效果[2]。因此，本方案提出基于同态加密和差分隐私（HE-DP）技术的线性纵向联邦学习隐私保护方法。此方法的核心是使用同态加密保证计算过程的安全性，使用差分隐私保证计算结果的安全性。相对于已有的分别基于同态加密、差分隐私的解决方案，本方案解决了在现实生活中难以找到可信第三方的问题，增强了隐私保护效果，并通过敏感度上界计算和扰动方法优化，降低噪音规模，保证了模型精度。

* 1. 方案设计

本方案涉及到三类实体，数据应用方（在纵向联邦学习中是含有label的一方，称为guest）、数据提供方（称为host）和第三方（主要来辅助多方联合建模，在纵向联邦学习场景中提供分发公私钥、加解密服务等等，自身既不提供数据，也不使用数据，称为arbiter方）。纵向联邦学习场景中，各个数据拥有方在每一轮迭代过程中需要交换加密中间结果以计算梯度，同时从第三方接收到扰动后的梯度信息进而本地更新模型，从而保证每一个数据拥有者的隐私保护需求。广义线性模型（Generalize Linear Model）[3]是线性模型的扩展，通过联结函数建立响应变量的期望值和线性组合的预测变量之间的关系。常见的逻辑回归、线性回归和泊松回归都是广义线性模型的特例。

在纵向联邦学习场景中，这里采用通用描述方法：guest是数据应用方，在线性纵向联邦学习中指的是含有数据标签的一方，而且建模流程一般由guest方发起。host是数据提供方。arbiter主要来辅助多方联合建模的。纵向联邦学习中用来分发公私钥、加解密服务等等，它本身既不提供数据，也不使用数据。建模最终的目标是得到两个子模型，对于guest方(为了公式描述方便，以下也叫做A方)而言，是子模型；对于host方（为了公式描述方便，以下也叫做B方）而言，是子模型。这也意味着当使用模型进行预测的时候，需要联合两个子模型和才可以进行预测。后续依次介绍系统与威胁模型、单方场景的逻辑回归损失函数和梯度，进而扩展到多方场景。

### 系统与威胁模型

这里假设数据拥有者和第三方都是半诚实的，即“诚实但好奇”的。他们会认真执行协议，但可能会利用自己的背景知识、接收到的中间结果或最终计算梯度信息推断（其他）局部数据集中个体敏感信息。另外，本发明假设数据拥有者之间、数据拥有者与第三方之间存在合谋，因为需要防止合谋攻击。

### 单方场景逻辑回归

这里单方场景下逻辑回归的损失函数定义为式（4-1）。

梯度表示为式（4-2）。

其中：模型参数，样本数据。这里采用SGD方法更新模型参数，故上述公式中n表示样本数据量。表示损失函数，为损失函数对模型参数的梯度。

### 多方场景逻辑回归

当使用同态加密技术时，由于同态加密不支持log运算，需将式（4-1）损失函数形式近似转化为多项式才可以，梯度公式可以近似的写成式（4-3）。为了方便起见，令：。于是损失函数梯度公式如式（4-4）：

在线性纵向联邦学习（使用同态加密技术）的场景[1]下，guest方（为了公式描述方便，以下也叫做A方）拥有数据，以及数据标签Y。host方（为了公式描述方便，以下也叫做B方）拥有数据，所以完整的。在同态加密状态下，例如当为加密状态下时，则表示为：，后续以此类推。本方案延续上述表述。本方案主要包括两部分：其中第一部分是加密样本对齐，第二部分是模型训练部分。其中加密样本对齐指的是将两个表示同一实体的不同样本对应起来，这两个样本一般位于两个局部数据集内，这里采用的是RSA加密和哈希算法的解决方案[4]。经过此算法，双方的样本便达到一致。例如：初始的时候，A方拥有数据集，B方拥有数据集。经过样本对齐后，双方的对齐样本为。如图4-1所示：

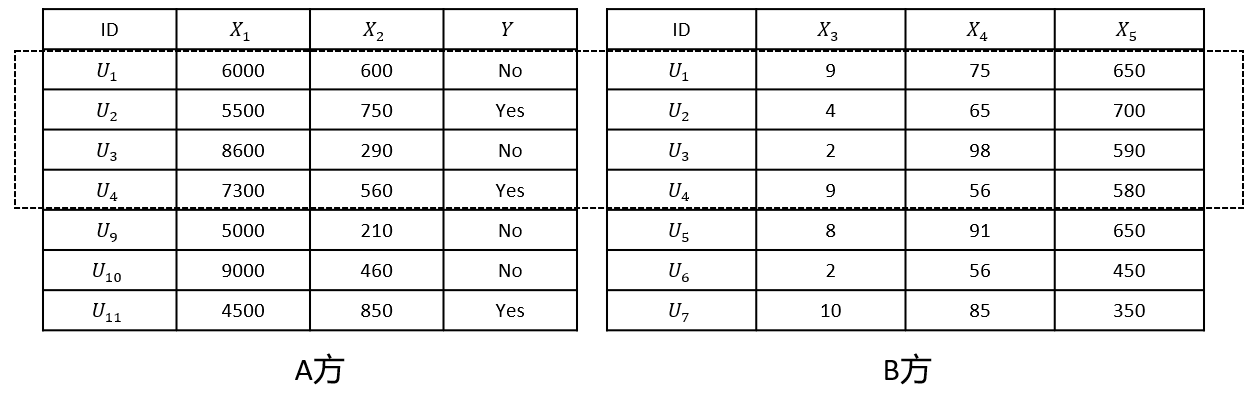


图4-1 样本对齐

Fig. 4-1 Sample Alignment

模型训练具体分为如下几个步骤，完整算法流程图见图4-2。

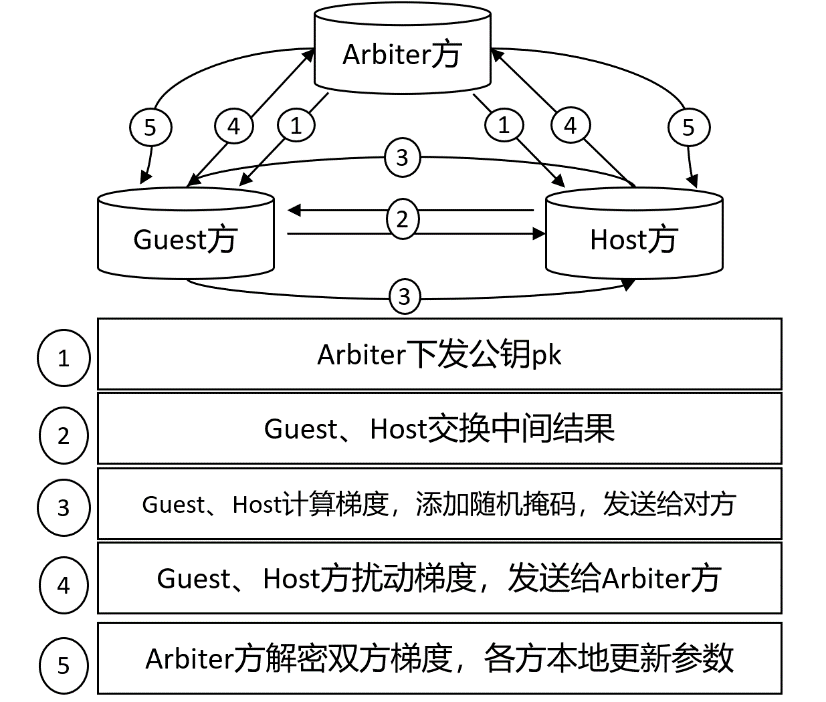


图4-2 算法流程图

Fig. 4-2 Algorithm flow chart

1. **密钥分发**

arbiter方使用同态加密机制生成公私钥，分别为和。然后将公钥分别发送给guest方、host方（后续为了公式描述方便，分别称之为A方、B方）本地。

1. **交换中间密文结果**

首先，B方本地计算中间结果：，使用从arbiter方接受的公钥加密后，结果为：，然后将此密文结果发送给A方。如图4-3所示。其次，A方本地计算中间结果：，这里将当中的偏置b也放在A这一方，所以这里的计算过程中包括了偏置b。首先A方计算出完整的，其次A方获得完整的，其中表示当前批次的第条数据，最后将此结果发送给B方。

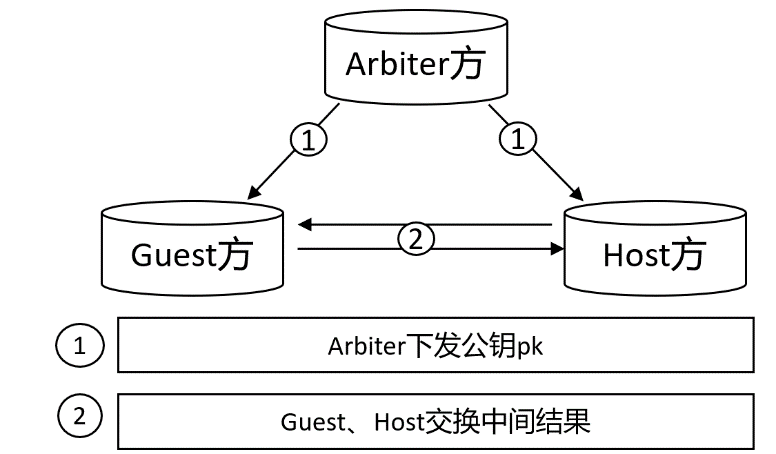


图4-3 交换中间密文结果

Fig. 4-3 Exchange of intermediate ciphertext results

1. **计算子梯度，添加加密随机掩码并交换，防止合谋攻击**

合谋攻击具体指的场景如下：当A计算出密文梯度后，如直接发给B方，B扰动此密文梯度后再将此结果发送给arbiter方。arbiter方解密后获知明文结果，如果arbiter方和B方合谋的话，便会推知A的隐私数据。所以这里A方必须在自身的密文梯度上添加加密随机掩码，然后发送给B方，同样，B方也会做出相同的处理。对于A、B方而言，计算出的密文梯度分别为式（4-5）、式（4-6）。

1. **扰动梯度数据信息并发送至arbiter**

这里由于是在两个子梯度上添加噪音，为减小噪音规模，需要分别计算双方梯度的最大敏感度。在这里假设目标函数满足利普希茨连续条件(Lipschitz continuous)、-光滑()且是-强凸(m-strongly convex)[2]。变量说明见表4-1。

表4-1变量符号说明

Table. 4-1 Variable symbol description

|  |  |  |
| --- | --- | --- |
| Serial Number | Symbol | Notation |
| 1 |  | epoch |
| 2 |  | iteration |
| 3 |  | Learning rate |
| 4 | k |  |
| 5 |  | mini-batch |
| 6 |  | Lipschitz constant |
| 7 | 、 |  |

线性模型中梯度线性可分。这里令：表示损失函数对的导数，见式(4-7)。

A方收到B发送过来的B的密文梯度信息，在此梯度上扰动数据。此时A方所确定的的最大敏感度见式(4-8)。

根据差分隐私后置处理性质可知，可求出B方梯度的最大敏感度，由A方生成高斯噪音，扰动此梯度信息，并发送给arbiter方。B方收到A方发送过来的密文梯度信息，在此梯度上扰动数据。此时B方所确定的的最大敏感度为式(4-9)。同样根据差分隐私的后置处理性质可知，可求出A方梯度的最大敏感度，然后B方生成高斯噪音，扰动此梯度信息，然后发送给arbiter方。

1. **arbiter解密梯度并下发。**

arbiter方收到A、B双方发来的密文梯度，然后使用私钥sk进行解密，后将梯度发给对应的参与方。A、B双方收到解密后梯度，利用自身之前本地保留的加密随机掩码还原梯度信息。由于这里梯度经过扰动，所以双方均无法通过此梯度信息推断对方的属性信息。最后A、B双方利用梯度本地更新模型。

重复上述的步骤二至五，直到达到指定的迭代次数或者模型收敛。

* 1. 实验与分析

本节中分别给出实验环境及设置、实验结果以及分析。

### 实验环境与设置

* 1. **对比算法**

针对纵向联邦学习场景中的线性模型，现有工作的参考文献主要有[1][2]，主要解决思路分别是基于同态加密、差分隐私技术。基于同态加密的工作中，需要设定可信第三方协调密钥分发等任务，训练过程中需要参与方进行交叉运算，Paillier加法同态技术用于保护隐私数据，私钥则由第三方持有。但是该方案一方面现实场景可信第三方难以找到，另一方面梯度中包含用户的属性信息，所以第三方解密后发送给各方的梯度存在隐私泄露的风险。基于差分隐私的解决方案提出HDP-VFL，由于敏感度上界计算以及扰动方式原因，致使噪音规模较大，影响建模效果。本工作将与上述二者作为对比。

* 1. **数据集与参数设置**

本实验采取公开可用数据集Adult[5]和Breast-Cancer[5]。Adult数据集是一个从1994年美国人口普查数据库中抽取的数据集，记录个数为32560，每条记录拥有15个属性。Breast-Cancer数据集是一个来自威斯康星州共包含569个恶性或者良性肿瘤细胞样本的数据集，每条记录包含32个属性。表4-2列出了数据集的关键统计信息。由于数据集Adult、Breast-Cancer含有缺失值情况，数据属性包含离散类型、枚举类型，所以需要对数据进行离散化、二元化等预处理。不失一般性，实验需要将数据集随机垂直拆分，使得不同的属性划分结果得到相同的规律。

表4-2 数据集Adult和Breast-Cancer关键统计信息

Table. 4-2 Dataset Adult and Breast-Cancer key statistics

|  |  |  |
| --- | --- | --- |
| 数据集 | 记录条数 | 属性个数 |
| Adult | 32560 | 15 |
| Breast-Cancer | 569 | 32 |

* 1. **评价指标**

接收者操作特征ROC（Receiver Operator Characteristic）曲线的横轴为FPR（False Positive Rate），纵轴为TPR（False Positive Rate），分别称之为假阳率和真阳率。FPR表示预测错误的正样本占整体负样本比例，公式见(4-10)：

TPR即为召回率Recall。ROC曲线通过坐标图形描述分类器的真阳性率和假阳性率关系，设定阈值后，计算出FPR、TPR，进而得到二维曲线上对应坐标，绘画出曲线。分析人员根据ROC曲线判断异常点检测算法模型优劣。ROC曲线越往左上角凸，说明模型分类效果越好。在比较不同的分类模型性能时，可以将ROC曲线下面积（AUC，Area Under Curve）作为评价模型优劣的指标。AUC值的范围在[0,1]，其值越接近1表示ROC曲线越靠近点(0,1)，模型分类效果越好。

本方案以AUC作为重要评价指标，AUC值与样本类别比例无关，经常作为样本不均衡的数据集分类评价标准。

* 1. **运行环境**

本方案实验环境如表4-3所示。

表4-3 实验软硬件环境

Table. 4-3 Experimental hardware and software environment

|  |  |
| --- | --- |
| 环境 | 配置 |
| 操作系统 | CentOS Linux release 7 |
| 处理器 | Intel(R) Xeon(R) Gold 6226R CPU @2.90GHz |
| 运行内存 | 16G memory |
| 硬盘空间 | 500G hard disk |
| 实验平台 | Python 3.6 + Numpy + Paillier |

### 实验结果及分析

实验结果如图4-4所示。从结果看出，基于同态加密的方案（HE）模型性能最好，在数据集Breast、Adult的AUC值分别接近98%、85%左右，并且不受隐私预算的影响。但该方案现实场景需要可信第三方协调，同时该方案每一轮迭代的梯度含有隐私泄露的风险。基于差分隐私的方案（DP）尽管解决了可信第三方问题以及隐私泄露的风险，但由于噪音较大影响建模效果，当隐私预算为1.00时，在数据集Breast、Adult的AUC值才在75.5%、73.1%左右。与上述两种方案对比，本方案可以在保证降低用户隐私泄露风险的基础上，同样可以达到较好的分类效果。当隐私预算为1.00时，在数据集Breast、Adult的AUC值达到91.2%、77.9%左右。用户可以根据需求灵活调节以便于在隐私保护强度和建模效果取得均衡。

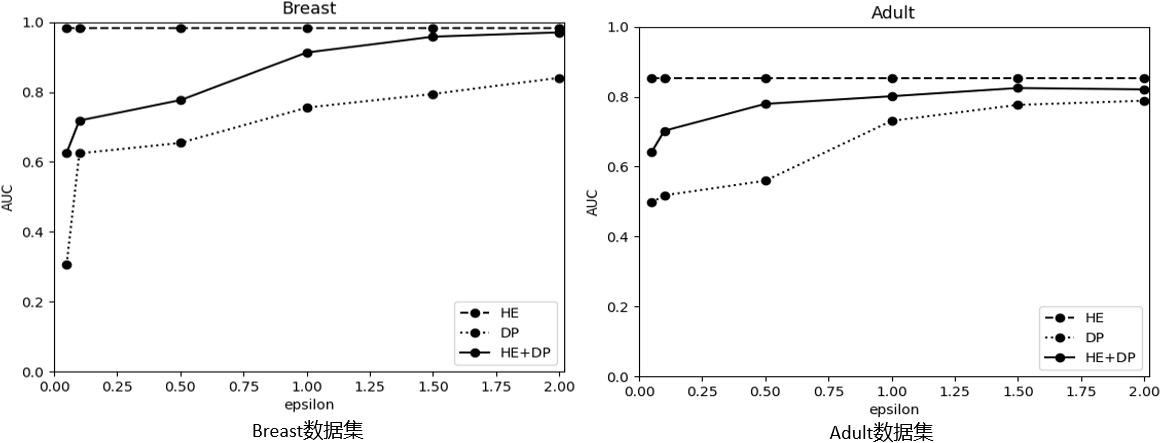


图4-4 算法在不同隐私预算下性能

Fig. 4-4 Algorithm performance under different privacy budgets

* 1. 本章小结

本章介绍一种基于同态加密和差分隐私的线性纵向联邦学习隐私保护方案。与现有的解决方案相比，具有以下优势：第一，使用同态加密保证了计算过程的安全性，差分隐私技术和加密随即掩码保证了计算结果的安全性。第二，解决了现实中难以找到可信的arbiter方的困难问题。

1. 结论与展望

文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文本文

**结 论**

**参考文献**

1. 教发[2019]11号，教育部等十一部门关于促进在线教育健康发展的指导意见[Z].
2. Liyanagunawardena T R, Parslow P, Williams S. Dropout: MOOC participants’ perspective[J]. 2014.
3. 周辉, 孟兆平, 敖重淼, 等. 网络环境下消费者数据的隐私保护——在全球数字经济背景下保护隐私和促进创新的政策框架[J]. 网络法律评论, 2013 (1): 193-218.
4. Voigt P, Von dem Bussche A. The eu general data protection regulation (gdpr)[J]. A Practical Guide, 1st Ed., Cham: Springer International Publishing, 2017, 10(3152676): 10-5555.
5. 孙佑海. 网络安全法: 保障网络安全的根本举措——学习贯彻《 中华人民共和国网络安全法》[J]. 中国信息安全, 2016 (12): 30-33.
6. Hard A, Rao K, Mathews R, et al. Federated learning for mobile keyboard prediction[J]. arXiv preprint arXiv:1811.03604, 2018.
7. Romero C, Ventura S. Educational data mining and learning analytics: An updated survey[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2020, 10(3): e1355.
8. Mozhaeva G, Maslova D, Yakovleva K. Correlation of MOOC Students’ Behavior Patterns and Their Satisfaction with the Quality of the Course[C]//“New Silk Road: Business Cooperation and Prospective of Economic Development”(NSRBCPED 2019). Atlantis Press, 2020: 623-629.
9. Goel Y, Goyal R. On the effectiveness of self-training in MOOC dropout prediction[J]. Open Computer Science, 2020, 10(1): 246-258.
10. Ardchir S, Talhaoui M A, Jihal H, et al. Predicting MOOC dropout based on learner’s activity[J]. International Journal of Engineering & Technology, 2018, 7(4.32): 124-126.
11. Moreno-Marcos P M, Muñoz-Merino P J, Maldonado-Mahauad J, et al. Temporal analysis for dropout prediction using self-regulated learning strategies in self-paced MOOCs[J]. Computers & Education, 2020, 145: 103728.
12. Mubarak A A, Cao H, Zhang W. Prediction of students’ early dropout based on their interaction logs in online learning environment[J]. Interactive Learning Environments, 2022, 30(8): 1414-1433.
13. Liang J, Yang J, Wu Y, et al. Big data application in education: dropout prediction in edx MOOCs[C]//2016 IEEE second international conference on multimedia Big data (BigMM). IEEE, 2016: 440-443.
14. Hagedoorn T R, Spanakis G. Massive open online courses temporal profiling for dropout prediction[C]//2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI). IEEE, 2017: 231-238.
15. Vitiello M, Walk S, Helic D, et al. User Behavioral Patterns and Early Dropouts Detection: Improved Users Profiling through Analysis of Successive Offering of MOOC[J]. J. Univers. Comput. Sci., 2018, 24(8): 1131-1150.
16. Wen Y, Tian Y, Wen B, et al. Consideration of the local correlation of learning behaviors to predict dropouts from MOOCs[J]. Tsinghua Science and Technology, 2019, 25(3): 336-347.
17. Goel S, Sai Sabitha A, Choudhury T, et al. Analytical analysis of learners’ dropout rate with data mining techniques[J]. Emerging Trends in Expert Applications and Security: Proceedings of ICETEAS 2018, 2019: 583-592.
18. Chui K T, Fung D C L, Lytras M D, et al. Predicting at-risk university students in a virtual learning environment via a machine learning algorithm[J]. Computers in Human Behavior, 2020, 107: 105584.
19. Jin C. MOOC student dropout prediction model based on learning behavior features and parameter optimization[J]. Interactive Learning Environments, 2020: 1-19.
20. Dass S, Gary K, Cunningham J. Predicting student dropout in self-paced MOOC course using random forest model[J]. Information, 2021, 12(11): 476.
21. PARK H, SEOKWON K, LEE S. Exploring Factors Predicting Student Dropout in Online Learning:Using Random Forest Model[J]. Journal of Educational Technology, 2022,38(1): 297-332.
22. Li H, Lynch C F, Barnes T. Early prediction of course grades: models and feature selection[J]. arXiv preprint arXiv:1812.00843, 2018.
23. Gelman B, Revelle M, Domeniconi C, et al. Acting the Same Differently: A Cross-Course Comparison of User Behavior in MOOCs[J]. International Educational Data Mining Society, 2016.
24. Bote-Lorenzo M L, Gómez-Sánchez E. Predicting the decrease of engagement indicators in a MOOC[C]//Proceedings of the Seventh international learning analytics & knowledge conference. 2017: 143-147.
25. Qiu L, Liu Y, Liu Y. An integrated framework with feature selection for dropout prediction in massive open online courses[J]. IEEE Access, 2018, 6: 71474-71484.
26. Ai D, Zhang T, Yu G, et al. A dropout prediction framework combined with ensemble feature selection[C]//Proceedings of the 2020 8th International Conference on Information and Education Technology. 2020: 179-185.
27. Jin C. Dropout prediction model in MOOC based on clickstream data and student sample weight[J]. Soft Computing, 2021, 25: 8971-8988.
28. Alzubaidi L, Zhang J, Humaidi A J, et al. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions[J]. Journal of big Data, 2021, 8: 1-74
29. Wang W, Yu H, Miao C. Deep model for dropout prediction in MOOCs[C]//Proceedings of the 2nd international conference on crowd science and engineering. 2017: 26-32.
30. Qiu L, Liu Y, Hu Q, et al. Student dropout prediction in massive open online courses by convolutional neural networks[J]. Soft Computing, 2019, 23: 10287-10301.
31. Feng W, Tang J, Liu T X. Understanding dropouts in MOOCs[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33(01): 517-524.
32. Chen J, Feng J, Sun X, et al. MOOC dropout prediction using a hybrid algorithm based on decision tree and extreme learning machine[J]. Mathematical Problems in Engineering, 2019, 2019.
33. Yin S, Lei L, Wang H, et al. Power of attention in MOOC dropout prediction[J]. IEEE Access, 2020, 8: 202993-203002.
34. Lai S, Zhao Y, Yang Y. Broad learning system for predicting student dropout in massive open online courses[C]//Proceedings of the 2020 8th International Conference on Information and Education Technology. 2020: 12-17.
35. Şahin M. A comparative analysis of dropout prediction in massive open online courses[J]. Arabian Journal for Science and Engineering, 2021, 46(2): 1845-1861.
36. Wang L, Wang H. Learning behavior analysis and dropout rate prediction based on MOOCs data[C]//2019 10th International Conference on Information Technology in Medicine and Education (ITME). IEEE, 2019: 419-423.
37. Zheng Y, Gao Z, Wang Y, et al. MOOC dropout prediction using FWTS-CNN model based on fused feature weighting and time series[J]. IEEE Access, 2020, 8: 225324-225335.
38. Fu Q, Gao Z, Zhou J, et al. CLSA: A novel deep learning model for MOOC dropout prediction[J]. Computers & Electrical Engineering, 2021, 94: 107315.
39. Mubarak A A, Cao H, Hezam I M. Deep analytic model for student dropout prediction in massive open online courses[J]. Computers & Electrical Engineering, 2021, 93: 107271.
40. Wang X, Guo B, Shen Y. Predicting the at-risk online students based on the click data distribution characteristics[J]. Scientific Programming, 2022, 2022.
41. Xu C, Zhu G, Ye J, et al. Educational data mining: dropout prediction in XuetangX MOOCs[J]. Neural Processing Letters, 2022, 54(4): 2885-2900.
42. 董业, 侯炜, 陈小军, 等. 基于秘密分享和梯度选择的高效安全联邦学习[J]. 计算机研究与发展, 2020, 57(10): 2241-2250.
43. Hard A, Rao K, Mathews R, et al. Federated learning for mobile keyboard prediction[J]. arXiv preprint arXiv:1811.03604, 2018.
44. Smola A, Narayanamurthy S. An architecture for parallel topic models[J]. Proceedings of the VLDB Endowment, 2010, 3(1-2): 703-710.
45. Li M, Andersen D G, Park J W, et al. Scaling distributed machine learning with the parameter server[C]//11th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 14). 2014: 583-598.
46. Hegedűs I, Danner G, Jelasity M. Gossip learning as a decentralized alternative to federated learning[C]//Distributed Applications and Interoperable Systems: 19th IFIP WG 6.1 International Conference, DAIS 2019, Held as Part of the 14th International Federated Conference on Distributed Computing Techniques, DisCoTec 2019, Kongens Lyngby, Denmark, June 17–21, 2019, Proceedings 19. Springer International Publishing, 2019: 74-90.
47. Dwork C, McSherry F, Nissim K, et al. Calibrating noise to sensitivity in private data analysis[C]//Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3. Springer Berlin Heidelberg, 2006: 265-284.
48. McSherry F, Talwar K. Mechanism design via differential privacy[C]//48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07). IEEE, 2007: 94-103.
49. Dwork C, Roth A. The algorithmic foundations of differential privacy[J]. Foundations and Trends® in Theoretical Computer Science, 2014, 9(3–4): 211-407.
50. Shamir A. How to share a secret[J]. Communications of the ACM, 1979, 22(11): 612-613.
51. 李炳臻, 刘克, 顾佼佼, 等. 卷积神经网络研究综述[J]. 计算机时代, 2021, 4: 8-12.
52. LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
53. Xu B, Huang R, Li M. Revise saturated activation functions[J]. arXiv preprint arXiv:1602.05980, 2016.
54. Treisman A M, Gelade G. A feature-integration theory of attention[J]. Cognitive psychology, 1980, 12(1): 97-136.
55. Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning[J]. Neurocomputing, 2021, 452: 48-62.
56. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv:1409.0473, 2014.
57. Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13713-13722
58. Hawkins D M. Identification of outliers[M]. London: Chapman and Hall, 1980.
59. Pang G, Shen C, Cao L, et al. Deep learning for anomaly detection: A review[J]. ACM Computing Surveys (CSUR), 2021, 54(2): 1-38.
60. Knorr E M, Ng R T. Finding intensional knowledge of distance-based outliers[C]//Vldb. 1999, 99: 211-222.
61. Knorr E M, Ng R T, Tucakov V. Distance-based outliers: algorithms and applications[J]. The VLDB Journal, 2000, 8(3): 237-253.
62. Breunig M M, Kriegel H P, Ng R T, et al. LOF: identifying density-based local outliers[C]//Proceedings of the 2000 ACM SIGMOD international conference on Management of data. 2000: 93-104.

**致谢**

text

**附 录**

text