|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 重庆邮电大学  CHONGQING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS | | | | |
| 博士学位论文  DOCTORAL DISSERTATION | | | | |
|  | | | | |
|  | | | | |
| **论文题目** | | **重庆邮电大学学位论文** | |  |
|  | | **格式模板** | |  |
|  | |  | |  |
|  | **学科专业** | |  |  |
|  | **学 　号** | |  |  |
|  | **作者姓名** | |  |  |
|  | **指导教师** | |  |  |
|  | **学　　院** | |  |  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 学校代码 | 10617 | | | | | UDC |  | |
| 分类号 |  | | | | | 密级 |  | |
| 学　位　论　文 | | | | | | | | |
| **重庆邮电大学学位论文格式模板** | | | | | | | | |
|  | | | | | | | | |
|  | | | | **某　某** | | |  | |
|  | | | |  | | |  | |
|  | | | | | | | | |
| 指导教师 | | **某某某** | | | | | **教　授** |  |
|  | | **某 某** | | | | | **副教授** |  |
|  | |  | | | | |  |  |
|  | |  | | | | |  |  |
|  | |  | | | | | |  |
|  | |  | | | | | |  |
|  | |  | | | | | |  |
| 申请学位级别 | | **博士** | | | 学科专业 | |  | |
| 专业学位领域 | |  | | | | | | |
| 答辩委员会主席 | | **某某某 教 授** | | | 论文答辩日期 | | **2021年5月20日** | |
| 学位授予单位和日期 | | | **重庆邮电大学** | | | | **2021年6月** | |

|  |  |
| --- | --- |
| **Dissertation Templemplate for Master Degree of Engineering in**  **CHONGQING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS** | |
| A Doctoral Dissertation Submitted to  Chongqing University of Posts and Telecommunications | |
|  | |
| Discipline |  |
| Student ID |  |
| Author |  |
| Supervisor |  |
| School |  |

**重庆邮电大学**

**学位论文独创性声明**

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文中不包含其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在论文中以明确方式标明。本人完全知晓本声明的法律后果由本人承担。

学位论文作者签名：

日期： 年 月 日

**重庆邮电大学**

**学位论文使用授权书**

本人同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。

本学位论文属于 ：

□**公开论文**

□**涉密论文**，保密 年，过保密期后适用本授权书。

（请在以上方框内选择打“**√**”）

作者签名： 导师签名：

日期： 年 月 日

**摘　要**

学位论文是研究生从事科研工作的成果的主要表现，集中表明了作者在研究工作中获得的新发明、新理论或新见解，是研究生申请硕士或博士学位的重要依据，也是科研领域中的重要文献资料和社会的宝贵财富。

为进一步规范我校研究生学位论文撰写格式，提高研究生学位论文质量，参照国家标准《学位论文编写规则》（GB/T 7713.1-2006），结合我校实际，制定本模板。

**关键词：**学位论文，撰写规范，论文模板，重庆邮电大学

**ABSTRACT**

Dissertation /Thesis is postgraduate’s main academic performance to display her/his works of scientific research, which shows the author’s new invention, new theory or new opinion in her/his research. It is the crucial document for the graduate students to apply for degree, and it is also the important scientific research literature and the valuable wealth of society.

In order to further standardize the format of dissertation/thesis writing and improve graduate dissertation/thesis quality, this temolate is formulated with reference to the national standard "Rules for Dissertation Writing" (GB/T 7713.1-2006) and the reality of CQUPT.

**Keywords:**Dissertation/Thesis, Writing Specification,Thesis Template, Chongqing University of Posts and Telecommunications

目　录

[摘　要 I](#_Toc116807345)

[ABSTRACT II](#_Toc116807346)

[图目录 V](#_Toc116807347)

[表目录 VI](#_Toc116807348)

[主要符号表 VII](#_Toc116807349)

[缩略词表 VIII](#_Toc116807350)

[第1章 绪论 1](#_Toc116807351)

[1.1 研究背景及意义 1](#_Toc116807352)

[1.2 国内外研究现状 1](#_Toc116807353)

[1.3 论文研究主要内容 1](#_Toc116807354)

[1.4 论文组织结构 1](#_Toc116807355)

[第一章 绪论 2](#_Toc116807356)

[（一）研究背景及意义 2](#_Toc116807357)

[1. 研究背景 2](#_Toc116807358)

[2. 研究意义 2](#_Toc116807359)

[（二）国内外研究现状 2](#_Toc116807360)

[（三）论文研究主要内容 2](#_Toc116807361)

[（四）论文组织结构 2](#_Toc116807362)

[第2章 论文结构及文字格式 3](#_Toc116807363)

[2.1 本章引言 3](#_Toc116807364)

[2.2 论文结构 3](#_Toc116807365)

[2.3 字数要求 3](#_Toc116807366)

[2.3.1 硕士论文要求 3](#_Toc116807367)

[2.3.2 博士论文要求 3](#_Toc116807368)

[2.4 字体和段落 4](#_Toc116807369)

[2.5 本章小结 5](#_Toc116807370)

[第3章 图表、公式格式和印制要求 6](#_Toc116807371)

[3.1 本章引言 6](#_Toc116807372)

[3.2 图和表格式 6](#_Toc116807373)

[3.2.1 图 6](#_Toc116807374)

[3.2.2 表 7](#_Toc116807375)

[3.3 公式格式 8](#_Toc116807376)

[3.4 印制要求 8](#_Toc116807377)

[3.5 本章小结 9](#_Toc116807378)

[第4章 总结与展望 10](#_Toc116807379)

[4.1 主要结论 10](#_Toc116807380)

[4.2 研究展望 10](#_Toc116807381)

[参考文献 11](#_Toc116807382)

[附录A 各学院中英文名称对照表 14](#_Toc116807383)

[附录B 常见一级学科中英文名称对照表 15](#_Toc116807384)

[附录C 常见专业学位类别中英文名称对照表 16](#_Toc116807385)

[作者简介 17](#_Toc116807386)

[1. 基本情况 17](#_Toc116807387)

[2. 教育和工作经历 17](#_Toc116807388)

[3. 攻读学位期间的研究成果 17](#_Toc116807389)

[3.1 发表的学术论文和著作 17](#_Toc116807390)

[3.2 申请（授权）专利 17](#_Toc116807391)

[3.3 参与的科研项目及获奖 17](#_Toc116807392)

[致 谢 18](#_Toc116807393)

图目录

[图2‑1 学位论文基本结构 2](#_Toc97198462)

[图3‑1 不同缩放系数 ν 的缩放结果 5](#_Toc97198463)

表目录

[表2‑1 中、英文字号对应关系 3](#_Toc97198602)

[表2‑2 主要文字及段落格式要求 3](#_Toc97198603)

[表3‑1 电流类型对效率的影响 6](#_Toc97198604)

[表3‑2 球队的比赛结果统计表[5] 7](#_Toc97198605)

[表3‑3 学位论文页面设置 8](#_Toc97198606)

主要符号表

|  |  |  |
| --- | --- | --- |
| **符号** | **说明** | **页码** |
| 𝒫Ω(·) | 集合Ω上的投影算子 | 6 |
| *c* | 电磁波的相平面速度 | 10 |

缩略词表

|  |  |  |
| --- | --- | --- |
| **英文缩写** | **英文全称** | **中文全称** |
| CQUPT | Chongqing University of Posts and Telecommunications | 重庆邮电大学 |
| IEEE | Institute of Electrical and Electronics Engineers | 电气和电子工程师协会 |

1. 绪论

## 研究背景及意义

学位论文……

## 国内外研究现状

## 学位论文……

## 论文研究主要内容

学位论文……

## 论文组织结构

本文……

1. 基础理论和相关方法

## 引言

## 联邦学习基础理论

学位论文包括前置部分、主体部分和结尾部分共三大部分，各部分组成及顺序如图2‑1所示。

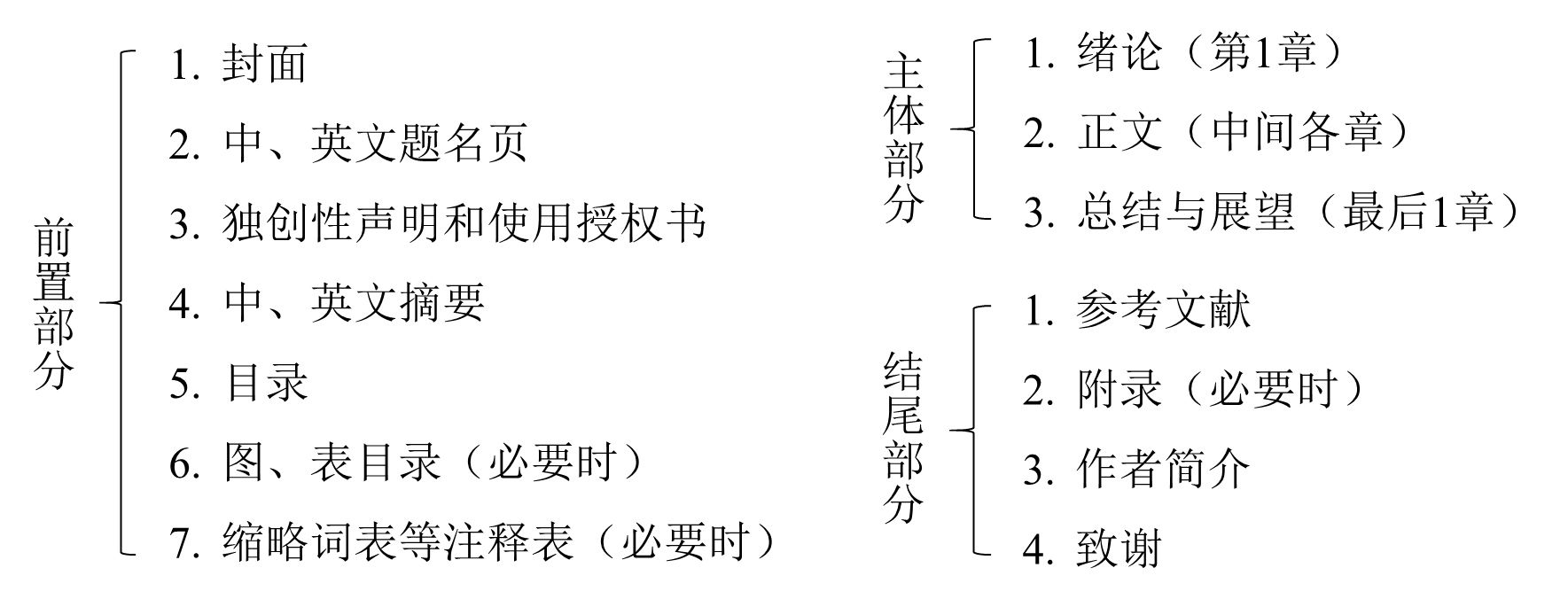


图2‑1 学位论文基本结构

Fig.2-1 Basic Structure of Dissertation

学位论文各部分独立为一部分，每部分应从新的一页开始。

论文的正文（中间各章）是论文的核心部分，一般由标题、文字叙述、图、表格和公式等部分构成。由于涉及的学科、选题、研究方法等有很大的差异，可以有不同的写作表达方式，但应遵循本学科通行的学术规范，必须实事求是，客观真切，准确完备，合乎逻辑，层次分明，简练可读。引用他人研究成果时，应注明出处，不得将其与本人的工作混淆。

## 半监督学习基础理论

### 硕士论文要求

各学科和学部自定。

### 博士论文要求

各学科和学部自定。

## 联邦半监督学习理论框架

学位论文中的中文统一用宋体，数字和英文统一用Times New Roman字体。从中文摘要开始，所有文字段落和标题行间距均取固定值20磅；所有段落按两端对齐、首行缩进2个全角字符方式书写内容。

中、英文混排时，除小数点以及引用的分图序号、公式序号等外，宜使用全角标点符号（逗号、冒号、括号、引号等）；英文段落中，符号使用应遵循英文书写习惯，统一使用半角符号，并规范使用空格。

中、英文字号对应关系如表2‑1所示，主要文字及段落格式要求如表2‑2所示。

表2‑1 中、英文字号对应关系

Table2-1 Chines, English Frontsize Mapping

|  |  |  |  |
| --- | --- | --- | --- |
| 中文字号 | 英文磅数 | 中文字号 | 英文磅数 |
| 二号 | 22 | 四号 | 14 |
| 小二 | 18 | 小四 | 12 |
| 三号 | 16 | 五号 | 10.5 |
| 小三 | 15 | 小五 | 9 |

表2‑2 主要文字及段落格式要求

Table2-2 Front and Paragraph Format Requirements

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 内容 | 字体 | 字号 | 对齐方式 | 段前距 | 段后距 | 示例或备注 |
| 一级标题 | 黑体 | 小三 | 居中 | 24磅 | 18磅 | 第1章 绪论 |
| 二级标题 | 黑体 | 四号 | 顶格左对齐 | 18磅 | 6磅 | 3.1 设计与验证流程 |
| 三级标题 | 黑体 | 四号 | 顶格左对齐 | 12磅 | 6磅 | 4.4.2 测试方法 |
| 四级标题 | 黑体 | 小四 | 顶格左对齐 | 12磅 | 6磅 | 5.3.1.3 测试结果 |
| 正文 | \* | 小四 | 两端对齐  （首行缩进） | 0磅 | 0磅 | \*未注明字体的，统一按“中文宋体，英文、数字Times New Roman”原则 |
| 页眉 |  | 五号 | 居中 | 0磅 | 0磅 |  |
| 页码 |  | 小五 | 居中 | 0磅 | 0磅 |  |
| 脚注 |  | 小五 | 两端对齐 | 0磅 | 0磅 |  |
| 参考文献 |  | 五号 | 两端对齐  （悬挂缩进） | 0磅 | 0磅 |  |
| 附录 |  | 五号 | \* | 0磅 | 0磅 | \*根据附录形式选择合适的排版方式。 |
| 图片 |  | 五号\* | 居中 | 6磅 | 0磅 | \*图中文字显示大小跟图题文字一致。 |
| 中文图题 |  | 五号 | 居中\* | 6磅 | 0磅 | \*超过一行的图题并非居中，详见[3.2.1](#_图) |

表2-2（续）

Table2-2 (continued)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 内容 | 字体 | 字号 | 对齐方式 | 段前距 | 段后距 | 示例或备注 |
| 英文图题 |  | 五号 | 居中\* | 0磅 | 12磅 |  |
| 表格 |  | 五号 | 居中 | 0磅 | 6磅 | 一般采用三线表样式 |
| 中文表题 |  | 五号 | 居中\* | 12磅 | 0磅 | \*超过一行的表题并非居中，详见[3.2.2](#_表) |
| 英文表题 |  | 五号 | 居中\* | 0磅 | 6磅 |  |
| 图表附注 |  | 五号 | 顶格 | 6磅 | 6磅 |  |
| 公式 |  | 小四 | 居中 | 6磅 | 6磅 |  |
| 公式编号 |  | 小四 | 右对齐\* | 6磅 | 6磅 | \*公式编号前不加引导线，详见[3.3](#_公式格式) |

其他要求：

（1）各级标题不得置于页面的最后一行，即须与下段同页；

（2）两个标题之间无正文时，第二个标题的段前距设置为0磅；

（3）图、表、公式统一采用单倍行距；

（4）只有一、两行文字的，不得单独作为一页内容；除各章最后一页外，中间页面不得出现较大空白；

（5）必要时，可在规定的格式要求基础上适当微调，以利于排版，但显示效果不得与规定的格式要求存在明显差距。

## 数据生成基础方法

本章介绍了……

## 本章小结

1. 基于多方联邦的半监督学习方法研究

## 引言

随着数据隐私保护法规（如GDPR）的日益严格，联邦学习（Federated Learning）技术因其“数据不动模型动”的特性，在医疗影像联合诊断、跨机构金融风控建模、工业物联网设备协同优化等领域展现出重要价值。 例如，在医疗场景中，多家医院可通过联邦学习共享疾病预测模型参数，而无需上传包含患者隐私的原始CT影像数据。然而，实际联邦场景中普遍存在标注数据稀缺的困境：在电商平台中，用户点击行为数据仅有部分被标注为有效兴趣标签；在工业质检场景中，缺陷样本的标注依赖专家经验且成本高昂。半监督学习（Semi-supervised Learning）通过挖掘未标注数据的分布特征，为解决这一困境提供了理论可能，但其与联邦学习的结合却面临矛盾——半监督学习依赖全局数据的一致性假设（如通过一致性正则化约束模型对扰动样本的预测稳定性，或基于伪标签生成实现未标注数据的自训练），而联邦学习的核心原则要求数据严格隔离于本地，导致跨参与方的全局数据分布信息难以被有效捕捉。这一矛盾在多方联邦场景下尤为突出。

尽管现有研究在半监督联邦学习领域取得初步进展，但面向多方纵向联邦场景的解决方案仍存在显著局限。 传统联邦平均算法（FedAvg）基于参与方数据独立同分布（i.i.d.）的强假设，而实际场景中（如多家金融机构联合反欺诈建模时），参与方持有的用户画像数据往往呈现非独立同分布（non-i.i.d.）特性——商业银行主要记录信贷交易特征，电商平台则掌握消费行为特征，导致特征空间与数据分布双重异构。现有半监督联邦方法（如FedMatch）虽在横向联邦（各参与方特征空间同构）中通过参数对齐提升未标注数据利用率，但其依赖共享特征提取器的设计范式，在纵向联邦场景下会因特征空间割裂而失效。这引发出两个关键科学挑战：其一，隐私-性能的尖锐矛盾，如何在避免原始数据交换的前提下，设计跨异构特征空间的半监督知识迁移机制（例如医疗影像联邦场景中，专科医院与社区诊所的CT图像特征维度与分辨率差异显著）；其二，标签分布的极端稀疏性，当仅少数参与方持有标注数据时，传统伪标签传播方法会因未标注方的置信度阈值难以统一设定，导致模型在局部数据分布上过拟合。

针对上述挑战，本章基于提出面向UDD-PU（Unlabeled-Data-Deficient PU）问题的多方纵向联邦半监督学习框架。 该研究聚焦于推荐场景中标注方仅持有正样本、未标注方缺乏标签数据的极端稀疏场景（如电商平台仅掌握用户点击正反馈，物流企业拥有未标注的签收行为数据），通过VFPU（Vertical Federated learning with Positive and Unlabeled data）算法实现三大突破：

1. 动态负采样机制：通过重复随机采样未标注数据构建伪负样本池，形成平衡训练集（如电商推荐场景中，从未标注的物流签收数据中动态抽取潜在负样本），迭代优化基分类器；

2. 加密特征对齐：在纵向联邦框架下，利用同态加密实现跨异构特征空间的梯度交换（如医疗联合建模中，医院实验室特征与社区诊所影像特征的隐私保护对齐），避免原始数据泄露；

3. 置信度驱动的伪标签协同更新：基于测试集预测分数的累积概率分布（如金融反欺诈场景中，银行通过跨机构交易特征协同识别高风险伪标签），动态筛选高置信度正样本补充训练集，解决标签稀疏性问题。

本章的结构描述如下：3.2节将系统阐述多方联邦半监督学习的问题定义与方法框架，并对方法执行流程和算法进行详细介绍。3.3小节在多个英文数据集上进行实验，验证方法有效性，最后对实验结果进行分析。3.4 小节对本章工作进行总结。

## 问题设置

首先，考虑一个包含\( K \)个数据所有者（data owner）和一个中央服务器（central server）的协作场景。假设第\( k \)个数据所有者持有的数据可表示为矩阵\( \mathcal{D}\_k \)，其中矩阵的每一行对应一个样本，每一列对应一个特征。整个数据集可形式化表示为三元组\( (\mathcal{I}, \mathcal{X}, \mathcal{Y}) \)，其中\( \mathcal{I} \)表示样本ID空间，\( \mathcal{X} \)表示特征空间，\( \mathcal{Y} \)表示标签空间。在传统纵向联邦学习（Vertical Federated Learning, VFL）框架中，通常要求至少有一个参与方拥有完整的样本标签。然而在实际业务场景中，由于商业竞争和数据隐私保护的限制，获取完整标注的数据集往往面临极大挑战。

为便于理解，我们以三个数据所有者（参与方A、B、C）的典型场景为例展开分析。这些参与方各自持有敏感数据，需要在保证数据隐私安全的前提下进行协作建模。假设样本类别包含正样本（positive）和负样本（negative）两类，且三方在样本ID空间上存在部分重叠。特别地：

- 参与方A拥有正样本集合\( P \)，该集合仅包含标注明确的正类数据

- 参与方B和C持有未标注数据集合\( U \)，其中排除了参与方A已拥有的样本

- 三方需联合训练推荐模型，从\( U \)中识别可靠的正样本（即模型预测为可信正类的样本集合\( R \)）

由于\( U \)的真实标签未知，传统纵向联邦学习算法无法直接应用，其根本原因在于没有任何参与方持有完整的标签信息。一种潜在的解决思路是采用半监督学习技术，例如PU学习（Positive-Unlabeled Learning）。然而，经典PU学习方法要求同时访问正样本集\( P \)和未标注集\( U \)，而本场景中仅有参与方A持有\( P \)，参与方B、C仅能提供\( U \)的部分数据，这导致现有方法存在直接应用的局限性。

由此产生了一个新型学习问题——未标记数据缺失的PU学习（Unlabeled-Data-Deficient PU Learning, UDD-PU）。具体而言，当参与方A（如电商平台）需要为其推荐业务开发联合建模服务时，由于其他参与方（如金融机构、物流企业）无法提供完整的未标注数据，传统PU学习框架中的假设条件不再成立。这一问题的核心挑战在于如何在多方数据隔离且标签分布不完整的条件下，实现可靠正样本的有效识别与推荐服务的协同优化。

## 基于多方联邦的半监督学习方法

本节将系统阐述未标记数据缺失的PU（UDD-PU）学习问题设定，并提出基于正样本与未标记数据的纵向联邦学习（VFPU）算法。我们将详细阐述VFPU算法如何为UDD-PU学习问题提供解决方案。

为应对UDD-PU学习问题，本章提出VFPU算法，该算法将纵向联邦学习框架与PU学习技术相结合。VFPU旨在解决参与方A未标注样本不足的问题，并提升推荐模型性能。本文提出的基于VFPU的推荐流程包含三个核心步骤：数据预处理、加密样本对齐和VFPU算法执行。VFPU的主要目标是从未标注数据集中识别可靠正样本，这些样本可使模型深入理解正例特征属性，从而为参与方A提供更精准和个性化的产品推荐。同时，我们确保多方数据隐私保护和协同合作。图1展示了推荐流程示意图，具体步骤将在后续章节详细阐述。

### 数据预处理与加密样本对齐

(1) 数据预处理

我们对参与方A、B、C持有的数据应用多种预处理技术，包括数据清洗、标准化和特征编码。具体而言：

分类特征：采用独热编码（One-Hot Encoding）处理，将离散型变量转换为二进制向量形式。

数值特征：使用标准化缩放（Z-score标准化）进行归一化，消除量纲差异。

(2) 加密样本对齐

完成数据预处理后，三方通过以下两步安全执行样本对齐：

步骤1：参与方B与C对齐样本ID空间，仅保留双方共有的样本，剔除未对齐样本。处理后，B与C共享相同样本但持有不同特征。

步骤2：参与方A与C对齐样本ID空间，不删除任何样本。对齐样本（即同时存在于A和C数据集中的样本）在C中被标记为1（正样本），未对齐样本标记为-1（未标注样本）。

通过上述流程，参与方C将同时持有正样本和未标注样本，从而将UDD-PU推荐问题转化为参与方B与C之间的垂直联邦PU学习问题。

隐私保护机制：采用基于Blind RSA的隐私集合求交协议（PSI），确保各参与方在计算数据集交集时不泄露样本隐私信息。该协议通过盲签名技术实现加密状态下的集合运算，满足《数据安全法》对敏感信息处理的要求。

### 基于正样本与未标注数据的纵向联邦学习

VFPU算法的核心目标是在保护数据隐私的前提下，从包含未标注样本的数据集中高效识别可靠的正样本。该算法创新性地将经典PU（Positive-Unlabeled）学习技术与垂直联邦学习框架相结合，其中主要融合了"两步法"（two-step technique）[14]和"PU装袋法"（PU bagging method）[13]两种关键技术。本节将按照算法1（此处需补充算法编号）的流程对VFPU算法进行详细阐述，其核心流程如图3.2（需补充图示编号）所示。

(1) 初始样本集合构建

VFPU算法采用迭代优化的执行策略，整体包含\\( M \\)次主迭代过程。每个主迭代周期\\( m \\in \\{1, ..., M\\} \\)由三个核心阶段构成：

1) 随机采样阶段：通过\\( T \\)轮次的数据采样构建训练子集

2) 联合训练阶段：多方协作训练联邦模型

3) 预测评估阶段：对未标注样本进行可靠性评估

在每个迭代周期开始时，算法根据参与方C提供的标注信息动态构建正样本集\\( P\\_m \\)与未标注样本集\\( U\\_m \\)，其数学定义为：

\\\[

P\\_m = \\{i | \\mathcal{Y}^C\\_i = 1, \\ i \\in \\mathcal{I}\\_C\\}

\\\]

\\\[

U\\_m = \\{i | \\mathcal{Y}^C\\_i = -1, \\ i \\in \\mathcal{I}\\_C\\}

\\\]

式中，\\( \\mathcal{I}\\_C \\)表示参与方C的样本ID空间，\\( \\mathcal{Y}^C \\)为对应的标签空间，\\( i \\)为样本唯一标识符。值得注意的是，这种动态集合构建机制使得算法能够随着迭代过程不断优化样本选择策略，逐步提升正样本识别的准确率。

（建议补充图示说明）可在此处插入样本集合构建过程的示意图，展示参与方C中正样本（标注为1）与未标注样本（标注为-1）的分布情况，以及迭代过程中样本集合的动态更新过程。

（2）采样、训练与预测

如图1所示，在第\\( m \\)次迭代的第\\( t \\)轮（\\( t \\in \\{1,2,...,T\\} \\)）采样中，通过自助采样法（bootstrapping）\[13\]从无标签样本集\\( U\\_m \\)中生成伪负样本集\\( N\\_m^t \\)。其数学表达式为：

\\\[

N\\_m^t = \\{\\text{从 } U\\_m \\text{ 中随机选取 } |P\\_m| \\text{ 个样本}\\}, \\tag{2}

\\\]

其中\\( |P\\_m| \\)表示正样本集\\( P\\_m \\)的样本数量。

由于无标签样本的真实类别未知，\\( N\\_m^t \\)被视为伪负样本集，可能同时包含真实负样本与正样本。通过从\\( U\\_m \\)中抽取与\\( P\\_m \\)等量的样本，可构造与正样本集规模平衡的伪负样本集。

在训练过程中，将\\( P\\_m \\)与\\( N\\_m^t \\)合并为二分类训练集，用于训练纵向联邦学习模型，使其学习正负样本的区分能力并应用于后续预测任务。

自助采样法是一种有放回的随机抽样技术。通过该方法，VFPU能够生成多样化的平衡训练集，从而提升模型的泛化能力、减少潜在偏差，并增强推荐系统的整体性能。

未被自助采样选中的样本称为袋外样本（out-of-bag samples）。袋外分数表示袋外样本被预测为正类的概率。因此，袋外样本集\\( O\\_m^t \\)可通过从\\( U\\_m \\)中剔除\\( N\\_m^t \\)得到，其数学表达式为：

\\\[

O\\_m^t = U\\_m - N\\_m^t.

\\\]

随后，参与方C对\\( N\\_m^t \\)、\\( P\\_m \\)和\\( O\\_m^t \\)进行加密并发送至其他参与方（以参与方B为例）。参与方B与C根据接收到的三组样本ID分别构建训练集与测试集，具体定义为：

\\\[

\\mathcal{D}\\_{train}^{K} = \\{(i, x\\_i, y\\_i) | i \\in P\\_m \\text{ 或 } i \\in N\\_m^t \\};

\\\]

\\\[

\\mathcal{D}\\_{test}^{K} = \\{(i, x\\_i, y\\_i) | i \\in O\\_m^t \\},

\\\]

其中\\( \\mathcal{D}\\_{train}^{K} \\)为二分类训练集，\\( \\mathcal{D}\\_{test}^{K} \\)为测试集，\\( K \\in \\{B, C\\} \\)表示参与方编号，\\( x\\_i \\in \\mathcal{X} \\)为特征向量，\\( y\\_i \\in \\mathcal{Y} \\)为标签。

当参与方B和C完成各自训练集与测试集的构建后，二分类问题将转化为垂直联邦训练与预测任务。此时，基估计器作为各参与方的机器学习模型，需适配垂直联邦学习框架进行使用。

如算法1第5-11行所示，内部循环本质上是一个装袋（bagging）过程。该过程包含采样、训练、预测三个核心步骤，经过T轮迭代后形成集成模型。每轮迭代中的操作均相互独立，因此可采用并行计算技术加速处理，从而显著降低算法整体时间开销。

理解垂直联邦学习（VFL）的一般训练流程至关重要[12]。整个过程可分为四个关键步骤，具体展示了在保护隐私前提下，如何基于多方数据训练基估计器：

加密初始化：服务器创建加密密钥对，并向参与方B和C分发公钥

中间结果交换：参与方B和C对梯度与损失计算所需的中间结果进行加密后交互传输

加密计算：参与方计算加密梯度并添加随机掩码，同时生成加密损失值，将加密结果发送至服务器

解密更新：服务器解密梯度与损失值后返回，参与方去除掩码并更新本地模型参数

为支持上述训练流程，研究者已提出多种适用于垂直联邦框架的隐私保护机器学习算法[12]，主要包括：

逻辑回归（LR）[45,46]

随机森林（RF）[47]

梯度提升决策树（GBDT）[45]

XGBoost（XGB）[48,49]

LightGBM（LGB）[39]

本文将通过应用不同类型的基估计器，系统评估推荐模型的性能表现。图2展示了垂直联邦学习的参数更新流程，其中虚线箭头表示加密数据传输，实线箭头表示明文通信，椭圆区域标注了各参与方的本地计算任务。

（3）可靠正样本识别

在第\( m \)次迭代中，采样、训练及预测过程共执行\( T \)次。完成全部\( T \)轮次后，收集足够信息以确定未标记样本集\( U\_{m} \)中所有样本的概率集合\( P\_{m} \)，该概率表征各样本被视为正类的可能性，可用于后续决策（如筛选可靠正样本并更新训练集）。

为计算完整概率集合\( P\_{m} \)，需为每个未标记样本\( u(u \in U\_{m}) \)计算其概率值\( P\_{m}(u) \)。具体地，\( P\_{m}(u) \)由第\( m \)次迭代中所有\( T \)轮次下\( u \)的\*\*出袋分数（Out-of-Bag Score）\*\*之和，除以\( u \)在第\( m \)次迭代中作为出袋样本出现的总次数，其计算公式为：

\[

P\_{m}(u) = \frac{\sum\_{t=1}^{T} S\_{m}^{t}(u)}{\sum\_{t=1}^{T} I(u \in O\_{m}^{t})}.

\]

其中，若样本\( u \)在第\( m \)次迭代的第\( t \)轮采样中未被选为出袋样本，则\( S\_{m}^{t}(u) = 0 \)。指示函数\( \sum\_{t=1}^{T} I(u \in O\_{m}^{t}) \)的取值为：当样本\( u \)属于出袋样本集\( O\_{m}^{t} \)时返回1，否则返回0。

基于概率值\( P\_{m}(u) \)，对所有未标记样本进行降序排列，排名靠前的样本因具有更高的概率被判定为真实正样本，可被选为可靠正样本。第\( m \)次迭代中识别的可靠正样本集\( R\_{m} \)可表示为：

\[

R\_{m} = \{ \text{选取概率值排名前 } |U\_{m}| \times \theta \text{ 的样本} \}.

\]

具体实现步骤如下：

- 步骤一：根据概率值对\( P\_{m} \)中的样本进行非递增排序；

- 步骤二：从排序结果中选取前\( |U\_{m}| \times \theta \)个样本，其中\( \theta \)为人工设定的比例参数，表示可靠正样本的采样率。

在选定可靠正样本集\( R\_{m} \)后，需更新参与方\( C \)数据集中对应样本的标签：将\( R\_{m} \)中的样本添加至正样本集合，同时从未标记数据集\( U\_{m} \)中移除。该操作可形式化表示为：

\[

\mathcal{Y}^C\_r = 1, \quad r \in R\_{m}

\]

其中，集合\( R\_{m} \)中的每个样本\( r \)均被赋予标签1（正类）。此过程将导致未标记数据集\( U\_{m} \)的样本数量逐步减少。

当算法完成全部\( M \)次迭代后，最终可靠正样本集\( R \)由各次迭代生成的\( R\_{m} \)集合取并集得到，即：

\[

R = \bigcup\_{m=1}^{M} R\_{m}.

\]

基于最终可靠正样本集\( R \)，参与方A可针对\( R \)中的每个样本定制推荐策略。通过利用高置信度的正样本信息，能够显著提升推荐系统的准确性与相关性。此步骤是算法的核心目标，确保从海量未标记数据中筛选出对推荐性能优化最具价值的样本，从而增强系统的实际应用效果。

## 实验分析

本节详细阐述VFPU算法的实验设计与评估过程。首先概述实验所用数据集及参数配置，其次明确指导实验的研究问题框架，最后针对各研究问题汇报实验结果并进行深入讨论。

### 数据集介绍

### 参数设置

### 实验结果

(3-1)

本文……

(3-2)

式中，为欧式距离；*d*为邻域半径，即个体之间的邻接距离。

本文……

## 本章小结

涉密学位论文的印刷、制作、传递、存档等，须符合国家、学校相关保密要求。学位论文一律左侧装订。

学位论文页面设置如表3‑3所示。

表3‑3 学位论文页面设置

(mm)

Table3-2 Page setting graph of Dissertation(mm)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 纸张规格 | 页边距 | | 页眉边距 | 页脚边距 |
| 左、右 | 上、下 |
| A4 (210×297) | 30 | 30 | 20 | 20 |

中文摘要之前的前置部分（封面、中、英文题名页、独创性声明和使用授权书），采用单面印刷。

从中文摘要开始，采用双面印刷。

中文摘要及之后的前置部分，包括中文摘要、ABSTRACT、目录、图目录（如有）、表目录（如有）、主要符号表（如有）、缩略词表（如有），在双面印刷时，若某部分页数为奇数，则该部分最后一页单面印刷。例如：若“摘要”只有1页，则其页码是“Ⅰ”，第“Ⅰ”页纸的背面为空白（无页眉或页码）；“ABSTRACT”用新的一张纸印刷，页码从“Ⅱ”开始。

从第1章第1页开始，至论文最后1页，所有页面均双面印刷。例如：若第1章的最后1页为第17页，则第2章的第1页在第17页的背面印刷，页码为“18”（页眉是“重庆邮电大学博士（硕士）学位论文”）。

一次性双面打印整本学位论文技巧：除用于打印的版本外，电子版论文中一律不得出现空白页。论文打印建议使用PDF格式。为方便一次性双面打印，有时可在单面印刷的部分（如封面、中、英文题名页、独创性声明和使用授权书），或者双面打印只有1页的某部分内容（如摘要、ABSTRACT等）后插入1页空白页，该空白页不编排页眉页码；论文中出现的页码应前后连续，不得中断。

# 总结与展望

## 主要结论

本文主要……

## 研究展望

更深入的研究……

# 参考文献

1. 中华人民共和国国家质量监督检验检疫总局,中国国家标准化管理委员会. 学位论文编写规则: GB/T 7713.1-2006[S]. 北京: 中国标准出版社, 2007: 17-20.
2. 中华人民共和国国家质量监督检验检疫总局,中国国家标准化管理委员会. 科技报告编写规则: GB/T 7713.1-2014[S]. 北京: 中国标准出版社, 2007.
3. 中华人民共和国国家质量监督检验检疫总局,中国国家标准化管理委员会. 信息与文献 参考文献著录规则: GB/T 7714.1-2015[S]. 北京: 中国标准出版社, 2015: 1-18.
4. 教育部国家语言文字工作委员会. 通用规范汉字表[M]. 北京: 语文出版社, 2018.
5. 汪继祥. 作者编辑手册[M]. 北京: 科学出版社, 2004.
6. 王晓琰, 殷建芳, 王晓峰, 等. 关于连续出版会议论文著录格式的探讨[J]. 学报编辑论丛, 2019: 162-165.

**顺序编码制参考文献表的著录格式和示例**

**期刊论文示例：**[序号] 作者. 文题[J]. 期刊名, 年, 卷(期): 起-止页码.

1. WU D, YAN J, WANG H, et al. Social Attribute Aware Incentive Mechanism for Device-to-Device Video Distribution[J]. IEEE Transactions on Multimedia, 2 017, 19(8): 1908-1920.
2. HEWITT J A. Technical Services in 1983[J]. Library Resource Services, 1984, 28(3): 205-218.
3. 李炳穆. 理想的图书馆员和信息专家的素质与形象[J]. 图书情报工作, 2000(2):5-8.

**会议论文示例：**[序号] 作者. 文题[C]. 会议名, 会议地, 会议年: 起-止页码.

1. BAERGAMASCO F, ALBARELLI A, COSMO L, et al. Adopting an Unconstrained Ray Model in Light-Field Cameras for 3D Shape Reconstruction[C]. IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 3003-3012.
2. 中国力学学会. 第3届全国实验流体力学学术会议论文集[C]. 天津: 南开大学出版社,1990:10-12.

**普通图书示例：** [序号] 作者. 书名[M]. 译者. 版本. 出版地: 出版者, 出版年, 起-止页码.

1. 竺可桢. 物理学[M]. 北京: 科学出版社, 1973, 56-60.
2. 汪昂. (增补)本草备要[M]. 石印本. 上海: 同文书局, 1912,31-32 .
3. ROOD H J. Logic and Structured Design for Computer Programmers[M]. 3rd ed. Watertown: Brooks/Cole Thomson Learning, 2001,105-116.
4. CRAWFPRD W, GORMAN M. Future Libraries: Dreams, Madness, & Reality[M]. Chicago: American Library Association, 1995, 20-21.
5. 罗杰斯. 西方文明史：问题与源头[M]. 潘惠霞, 魏婧, 杨艳, 等译. 2版. 大连: 东北财经大学出版社, 2011: 15-16.

**学位论文示例：**[序号] 作者. 文题[D]. 授位单位所在地: 授位单位, 授位年, 起-止页码.

1. 李娜芬. 障碍环境中Swarm突现计算模型研究及行为控制[D]. 重庆: 重庆邮电大学, 2013,11-47.

报纸文章示例：[序号] 作者. 文题[N]. 报纸名, 出版日期 (版面数).

1. 顾春. 牢牢把握稳中求进的总基调[N]. 人民日报, 2012-03-31 (3).
2. 丁文祥. 数字革命与竞争国际化[N]. 中国青年报, 2000-11-20(15).
3. 张田勤. 罪犯DNA库与生命伦理学计划[N]. 大众科技报, 2000-11-12(7).

**报告示例：**[序号] 作者. 文题[R]. 出版地: 出版者, 出版年.

1. 冯西桥. 核反应堆压力容器的LBB分析[R]. 北京: 清华大学核能技术设计研究院, 1997.
2. U. S. Department of Transportation Federal Highway Administration. Guidelines for Bandling Excavated Acid-Producing Materials, PB 91-194001[R]. Springfield:U. S.Department of Commerce National Information Service, 1990.
3. World Health Organization. Factors Regulating the Immune Response: Report of WHO Scientific Group[R]. Geneva: WHO, 1970.

**专利示例：**[序号] 发明人. 专利名: 专利号[P]. 授权日期.

1. 张凯军. 轨道火车及高速轨道火车紧急安全制动辅助装置: 201220158825. 2[P]. 1912-04-05.
2. 河北绿洲生态环境科技有限公司.一种荒漠化地区生态植被综合培育种植方法: 01129210.5[P/OL]. 2001-10-24[2002-05-28]. http://211.152.9.47/sipoasp/zlijs/hyjs-yx- new.asp? recid=01129210.5&leixin.

**标准示例：**[序号] 发布单位. 标准名: 标准号[S]. 出版地: 出版者, 出版年: 起-止页码.

1. 国家技术监督局. 国际单位制及其应用: GB 3100-1993[S]. 北京: 中国标准出版社, 1994: 3-6.
2. 国家技术监督局. 有关量、单位和符号的一般规则: GB/T 3101-1993[S]. 北京: 中国标准出版社, 1994: 13-20.
3. 中国国家标准化管理委员会. 信息与文献 参考文献著录规则: GB/T 7714.1-2015[S]. 北京: 中国标准出版社, 2015: 1-18.

**电子文献示例：**[序号] 作者. 文题[文献类型标识/文献载体标识]. 出版地: 出版者, 出版年: 起-止页码 (更新或修改日期) [引用日期]. 获取或访问路径. 数字对象唯一标识符.

1. CLERC M. Discrete Particle Swarm Optimization: a Fuzzy Combinatorial Box[EB/OL]. (2010-07-16) [2022-02-28]. http://clere.maurice.free.fr/pso/Fuzzy\_Discrere\_PSO/Fuzzy\_DPSO.htm.
2. 江向东. 互联网环境下的信息处理与图书管理系统解决方案[J/OL]. 情报学报, 1999, 18(2): 4 [2022-02-28] .
3. 萧钮. 出版业信息化迈人快车道[EB/OL]. (2001-12-19) [2022-02-28]
4. CHRISTINE M. Plant Physiology: Plant Biology in the Genome Era[J/OL]. Science, 1998, 281:331-332[2022-02-28].
5. METCALF S W. The Tort Hall Air Emission Study[C/OL]//The International Congress on Hazardous Waste, Atlanta Marriott Marquis Hotel, Atlanta,Georgia, June 5-8, 1995: impact on human and ecological health[2022-02-28].
6. TURCOTTE D L. Fractals and Chaos in Geology and Geophysics[M/OL]. New York: Cambridge University Press, 1992[2022-02-28].

**著者-出版年制参考文献表的著录格式和示例**

顾春, 2012. 牢牢把握稳中求进的总基调[N]. 人民日报, 03-31 (3).

河北绿洲生态环境科技有限公司, 2001. 一种荒漠化地区生态植被综合培育种植方法: 01129210.5[P/OL]. http:// 211.152.9.47/sipoasp/zlijs/hyjs-yx-new.asp? recid=01129210.5&leixin.

李炳穆, 2000. 理想的图书馆员和信息专家的素质与形象[J]. 图书情报工作, (2): 5-8.

尼葛洛庞帝, 1996.数字化生存[M]. 胡泳, 范海燕, 译. 海口: 海南出版社.

全国信息与文献标准化技术委员会, 2007. 学位论文编写规则: GB/T 7713.1-2006[S]. 北京: 中国标准出版社, 17-20.

全国信息与文献标准化技术委员会, 2015. 信息与文献 参考文献著录规则: GB/T 7714-2015[S]. 北京: 中国标准出版社, 1-18.

汪冰, 1997. 电子图书馆理论与实践研究[M]. 北京: 北京图书馆出版社: 16.

BAERGAMASCO F,ALBARELLI A, COSMO L, et al. 2015. Adopting an Unconstrained Ray Model in Light-Field Cameras for 3D Shape Reconstruction[C]. IEEE Conference on Computer Vision and Pattern Recognition,Boston, USA, 3003-3012.

DOWLER L, 1995. The research university’s dilemma:resource sharing and research in a transinstitutional environment[J]. Journal of library administration, 21(1/2): 5-26.

KENNEDY W J, GARRISON R E, 1975a. Morphology and genesis of nodular chalks and hardgrounds in the Upper Cretaceous of southern England[J]. Sedimentology, 22: 311-386.

KENNEDY W J, GARRISON R E, 1975b. Morphology and genesis of nodular phosphates in the Cenomaman of South-east England[J]. Lathaia, 8: 339-360.

# 附录A 各学院中英文名称对照表

|  |  |  |
| --- | --- | --- |
| **序号** | **中文名称** | **英文名称** |
| 01 | 通信与信息工程学院 | School of Communications and Information Engineering |
| 02 | 计算机科学与技术学院 | School of Computer Science and Technology |
| 03 | 自动化学院/工业物联网学院 | School of Automation / School of Industrial Internet of Things |
| 04 | 先进制造工程学院 | School of Advanced Manufacturing Engineering |
| 05 | 光电工程学院/重庆国际半导体学院 | School of Optoelectronic Engineering / Chongqing International Semiconductor College |
| 06 | 软件工程学院 | School of Software Engineering |
| 07 | 生物信息学院 | School of Bioinformatics |
| 08 | 理学院 | School of Science |
| 09 | 经济管理学院/现代邮政学院 | School of Economics and Management / School of Modern Post |
| 10 | 传媒艺术学院 | School of Media and Arts |
| 11 | 外国语学院 | School of Foreign Languages |
| 12 | 国际学院 | International College |
| 13 | 网络空间安全与信息法学院 | [School of Cyber Security and Information Law](http://english.cqupt.edu.cn/info/1026/1872.htm) |
| 14 | 马克思主义学院 | School of Marxism |
| 15 | 体育学院 | School of Physical Education |
| 16 | 创新创业教学学院/继续教学学院 | School of Innovation and Entrepreneurship Education / School of Continuing Education |

# 附录B 常见一级学科中英文名称对照表

|  |  |  |
| --- | --- | --- |
| **代码** | **中文名称** | **英文名称** |
| 0301 | 法学 | Science of Law |
| 0305 | 马克思主义理论 | Theory of Marxism |
| 0701 | 数学 | Mathematics |
| 0702 | 物理学 | Physics |
| 0710 | 生物学 | Biology |
| 0711 | 系统科学 | Systems Science |
| 0802 | 机械工程 | Mechanical Engineering |
| 0803 | 光学工程 | Optical Engineering |
| 0804 | 仪器科学与技术 | Instrumentation Science and Technology |
| 0808 | 电气工程 | Electrical Engineering |
| 0809 | 电子科学与技术 | Electronic Science and Technology |
| 0810 | 信息与通信工程 | Information and Communication Engineering |
| 0811 | 控制科学与工程 | Control Science and Engineering |
| 0812 | 计算机科学与技术 | Computer Science and Technology |
| 0831 | 生物医学工程 | Biomedical Engineering |
| 0835 | 软件工程 | Software Engineering |
| 0839 | 网络空间安全 | Cyberspace Security |
| 1201 | 管理科学与工程 | Management Science and Engineering |

# 附录C 常见专业学位类别中英文名称对照表

|  |  |  |
| --- | --- | --- |
| **代码** | **中文名称** | **英文名称** |
| 0351 | 法律 | Law |
| 0452 | 体育 | Physical Education |
| 0551 | 翻译 | Translation and Interpreting |
| 0854 | 电子信息 | Electronic and Information Engineering |
| 0855 | 机械 | Mechanical Engineering |
| 0858 | 能源动力 | Energy and Power Engineering |
| 0861 | 交通运输 | Transportation Engineering |
| 1251 | 工商管理 | Business Administration |
| 1256 | 工程管理 | Engineering Management |
| 1351 | 艺术 | Art |

# 作者简介

## 基本情况

张某某，男，重庆人，1993年8月出生，重庆邮电大学XX学院XX专业2018级博士研究生。

## 教育和工作经历

2010.08～2014.06 重庆邮电大学光电工程学院，本科，专业：电子科学与技术

2014.08～2015.06 华为，技术研究工程师

2015.08～2018.06 重庆邮电大学光电工程学院，硕士研究生，专业：电子科学与技术

2018.08～2022.06 重庆邮电大学通信与信息工程学院，博士研究生，专业：信息与通信工程

## 攻读学位期间的研究成果

### 发表的学术论文和著作

1. **ZHANG M** ,XX, XX, et al. XXXX[J]. Future Generation Computer Systems, 2020. (SCI期刊)
2. **ZHANG M** , XX, XX. XXXX[J]. International Journal of Machine Learning and Cybernetics, 2021, 12(9): 2543–2557. (SCI期刊)
3. XX, **张某某**, XX. XXXX [J]. 计算机学报, 2022. (已录用)
4. XX, XX, XX,**张某某**等. XXXX[M]. 科学出版社, 2021. (专著)

### 申请（授权）专利

1. **张某某**, XXX, XXX等. 专利名称: 专利号[P]. 授权日期.

### 参与的科研项目及获奖

格式：XXX项目, 项目名称, 起止时间, 完成情况, 作者贡献.

1. 国家自然科学基金重点项目, XXXX (No.000000), 2017.01-2020.12, 参与.
2. 重庆邮电大学博士研究生人才培养项目, XXXX (No.000000), 主持.
3. XXX, **张某某**, XXX等. 科研项目名称. 重庆市科技进步三等奖, 获奖日期.

# 致 谢

感谢老师、同学们的关心、支持和帮助！