课题名称：小二号，黑体，加粗，居中，行距18磅，段前0.5行，段后0.5行。上下各空一行。

基于可传递设备的联邦学习研究

摘 要

四号，黑体，居中。行距18磅。段前0.5行，段后0.5行。

生成函数法是一种测定稳定常数的常用方法。这种方法既可以用于酸稳定常数的测定，又可以用于配合物稳定常数的测定。根据数据处理方式的不同，生成函数法可分为 ……。

空一行

**关键词：**稳定常数，生成函数，电位滴定，酸，配合物

摘要正文300字左右，五号宋体，首行缩进2个汉字符。行距18磅。

五号宋体，加粗。

关键词3～5个，五号宋体。逗号分开，最后一个关键词后面无标点符号。

**Application of Bjerrum Function in Determination of Stability Constants**

换页。英文课题名称：小二号，Times New Roman，加粗，居中，行距18磅，段前0.5行，段后0.5行。上下各空一行。

**ABSTRACT**

四号Times New Roman 居中，段前0.5行，段后0.5行。行距18磅。

Bjerrum function method is a common method to determinate the stability constants. It can determinate the acid stability constants but also the complexes stability constants. According to the different processing methods of data, Bjerrum function method can be subdivided into …….

五号Times New Roman, 首行缩进2个汉字符，行距18磅。

空一行

**Key words：**stability constants, bjerrum function, potential titration, acid, coordination compound

五号Times New Roman，各关键词之间逗号分开，逗号后加一空格。行距18磅。

五号Times New Roman加粗

目 录

换页，上下各空一行；四号黑体居中，目录2字中间空1格，段前0.5行，段后0.5行。行距18磅。

五号宋体(英文Times New Roman)，单倍行距

[1 引 言 1](#_Toc402184259)

[1.1 稳定常数测定的意义 1](#_Toc402184260)

[1.2 稳定常数测定常用的方法 1](#_Toc402184261)

[1.2.1 Gorden法 1](#_Toc402184262)

[1.2.2 Monte Carlo法 1](#_Toc402184263)

[1.3 生成函数法的基本概况 1](#_Toc402184264)

[1.4 本文所作的工作 1](#_Toc402184265)

[2 理论部分 2](#_Toc402184266)

[2.1 生成函数法的基本关系式 2](#_Toc402184267)

[2.2 各种生成函数法的测定原理 2](#_Toc402184268)

[2.2.1 直接计算生成函数法 2](#_Toc402184269)

[2.2.2 分段拟合生成函数法 2](#_Toc402184270)

[2.2.3 半整数生成函数法 2](#_Toc402184271)

[3 实验部分 3](#_Toc402184272)

[3.1 仪器和试剂 3](#_Toc402184273)

[3.1.1 仪器 3](#_Toc402184274)

[3.1.2 试剂 3](#_Toc402184275)

[3.2 溶液的配制及标定 3](#_Toc402184276)

[3.2.1 NaOH标准溶液的配制及标定 3](#_Toc402184277)

[3.2.2 氯化钾离子强度调节剂的配制 3](#_Toc402184278)

[3.3 实验步骤 3](#_Toc402184279)

[4 结果和讨论 4](#_Toc402184280)

[4.1 多元酸体系的结果和讨论 4](#_Toc402184281)

[4.1.1 直接计算法 4](#_Toc402184282)

[4.1.2 半整数法 4](#_Toc402184283)

[4.1.3 分段拟合法 5](#_Toc402184284)

[4.2 氨基酸合铜体系的结果和讨论 5](#_Toc402184285)

[4.3 关于计算方法的讨论 5](#_Toc402184286)

[4.4 关于其他问题的讨论 5](#_Toc402184287)

[5 结论和展望 6](#_Toc402184288)

[5.1 结论 6](#_Toc402184289)

[5.2 展望 6](#_Toc402184290)

[参考文献 7](#_Toc402184291)

[谢 辞 8](#_Toc402184292)

2级标题：五号，黑体(英文Times New Roman)，顶格，序号与题名之间空一格，行距18磅，段前0.5行，段后0.5行。

1级标题：换页，空一行；四号，黑体(英文Times New Roman)，居中，行距18磅，段前0.5行，段后0.5行。

# 1 背 景

## 1.1 可穿戴设备的联邦学习

随着智能手表和智能腕带的普及，便携式可穿戴设备已经被消费者广泛穿戴在皮肤表面还是整合到衣物上。尽管这些设备有望充当机器学习任务的数据源，但针对可穿戴应用的机器学习模型的训练通常需要的数据远远超过每个单独设备收集的最大值，进而引发了对数据共享的需求。出于隐私考虑，一系列政策禁止未经用户许可的数据收集和存储**[1]**。例如，从可穿戴设备收集的大多数数据都受数据保护法规的约束包括欧洲委员会的一般数据保护法规(General Data Protection Regulation)**[2]**和美国的消费者隐私法(Consumer Privacy Act)**[3]** 使得汇总用户数据以进行大规模数据分析变得更加困难。

为了应对上述挑战，联邦学习 (Federated Learning) 作为一种保护隐私的分布式机器学习范例兴起：多个客户端用户训练模型并上传模型权重或梯度服务器，服务器聚合用户上传数据并通过加密算法或者机制设计保护用户各自的私有数据。根据客户端是数据中心还是大量的移动或IoT设备，联邦学习可以分为跨孤岛(cross-silo)和跨设备(cross-device)。其中跨设备联邦学习中的用户大多来自移动设备，面临来自无状态和不可靠客户的挑战。在联邦学习中，由于客户可能会参与模型更新或直接退出，因此它们的模型更新可能涉及不同部分的客户端，因此很难跟踪每个参与者。此外，由于跨设备联邦学习系统的操作主要依赖Wi-fi或更慢的通信网络**[4]**，通信会成为影响联邦学习效率的另一个瓶颈。

由于较低的通信带宽和有限的设备容量，可穿戴设备的联邦学习会更加复杂。联邦学习中数据传输的延迟，可能导致网络丢包，而传统联邦学习采用TCP/IP协议通过重传保证数据完整性，但也导致了联邦学习每次迭代时间变长(用户延迟符合长尾分布，即少数延迟的用户极大地拖慢整体的数据传输进度)，训练效率也因此变低。

一些工作已经开始研究并试图解决相关问题：DeepWear**[4]**提出了一种轻量级的在线调度算法，根据可穿戴设备的状态将训练任务自适应地加载到合适的手持设备上。尽管这种设计适合设备的本地训练，但它不符合联邦学习对隐私的保护需求。 FedHealth**[5]**将迁移学习与联邦学习结合在一起，为可穿戴式医疗保健提供个性化的训练范式。 FedCS**[6]**提出了基于本地资源约束选择客户端更新的方法。这些现有的方案都侧重于客户端选择之后和模型更新聚合期间的数据转移，而忽略了客户端间存在的非独立同分布数据对训练的影响，这也影响了联邦学习系统中当前公平方案和个性化算法的性能。

## 1.2 联邦学习的用户选择

考虑到提高通信效率，近期的大多数工作都提出了基于阈值的用户选择方法，即确定一个阈值选择足够网络容量的客户端。但是这样的用户选择不可避免会导致的数据移位，进而训练出一个偏向于部分网络良好的用户的本地数据的模型。尽管一些研究人员提出了专门针对联邦学习训练过程中公平的聚合方案，在客户选择之初发生的数据转移却仍未被解决。联邦学习的效果受到影响。

当一部分用户始终被排除在联邦学习之外，会让训练出的模型呈现出更加贴合另一部分用户的数据，进而导致公平性问题。仅考虑模型权重的聚合方法无法应对此类挑战**[7][8]**。在这里，本课题总结了联邦学习中产生有偏差模型的常见原因，包括：(1)部分用户的数据被过度学习，(2)部分用户的数据被学习不足，(3)部分用户的数据从未被学习。 针对(1)和(2)，已有部分解决了训练过程偏差的方法，例如Agnostic Federated Learning (AFL)**[9]**和q-FedAvg **[10]**。 作为AFL的拓展, q-FedAvg在公平性和准确性之间进行灵活权衡。尽管这些方法通过减轻训练过程的偏见来弥补参与设备间的表现差异，它们仍无法解决由(3)中的不正确的客户选择引起的模型偏差，AFL的作者也提到了这一点。

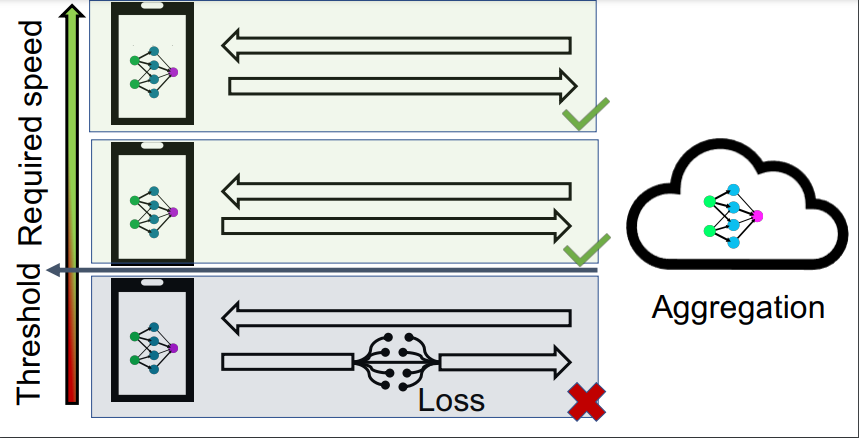


图1. 基于阈值的方法的图示，低于阈值的设备被认为是容易产生

丢包和延迟的，会被排除出联邦学习。

## [1.3 联邦学习的网络延迟](#生成函数法及其优势)

近期的一些工作试图主动丢弃信息以避免延迟等待时间，以提高通信效率来缓解需求与实际网络容量之间的差距。其中一些提议使用有损压缩来减少数据传输量。**[11][12]**使用结构化更新和草图更新对模型更新进行压缩，其主要思想是从有限的空间中学习目标模型的参数分布。**[13]**关注服务器到客户端的通信，并应用了一种更新频率较低的有损压缩方法。**[14]**挖掘了分布式机器学习中的损失容忍极限。这些工作启发了本课题探索跨设备联邦学习的网络数据丢失容忍度。

正文：五号，宋体(英文Times New Roman)，两端对齐，段落首行左缩进2个汉字符，行距18磅，段前0行，段后0行。

过去的一些工作为避免数据包错误和客户端丢失)，跨设备联邦学习系统通常使用传输速度和电池状态作为移动客户端选择的标准。在这种情况下，具有更多数据包错误和丢包的客户端不太可能被选中参与模型聚合。甚至于对处于较差网络条件的用户，其数据分布可能永远不会在训练中呈现(图 2)。

3级标题：五号，黑体(英文Times New Roman)，缩进2个汉字符书写序号，序号与题名之间空一格，行距18磅，段前0.5行，段后0.5行。

## 1.4 联邦学习的模型传输

鉴于以上提到的联邦学习中网络延迟带来的用户选择偏差，一个思路是考虑模型中的计算是不是都是必要？如果不是的话，有没有可能简化模型来减少存储占用和通信开销。一种可行的解决方案是每个用户传输部分重要的模型权重进而减少整体的通信开销。根据传输模型参数的形状和结构，模型传输可以被进一步分成非结构化传输和结构化传输。非结构化(non-structured)传输是指用户单次传输的参数无法构成一个完整的模型，典型的方法有随机遮盖(Random Mask)，每次选取每个权重矩阵的一部分并在若干次迭代后聚合平均；结构化(structured)传输是指上传的模型参数可以构成一个完整的可独立工作的子模型，代表性的如滤波器剪枝(filter pruning)，对于一个卷积神经网络模型，每次上传模型中更加重要的滤波器而不对滤波器进行更细层次的划分。

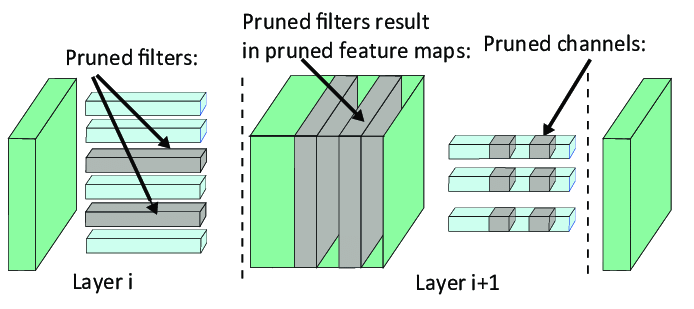


图3. 一个基于滤波器剪枝的结构化传输实例，传输后的模型权重仍构成独立的子模型。

在确定了传输的模型参数是在整个模型中相对更重要的一些之后，如何衡量模型参数的重要性成为另一个问题。本课题借用了网络剪枝 (network pruning)的一系列理论和方法，核心是如何有效地裁剪模型且最小化精度的损失，来指导联邦学习传输过程中重要参数的选择。

一个简单方法就是按**参数(或特征输出)绝对值大小**来评估重要性，直接删去绝对值较小的权重，这类称为基于幅值的权重剪枝(magnitude-based weight pruning)，这一类方法实现简单，也有利于模型的稀疏化，然而它基于参数绝对值越小，其对最终结果影响越小的假设，不适用于数据非独立同分布的任务场景，如迁移学习(Transfer Learning)，跨设备联邦学习 (Cross-device Federated Learning)。

第二种思路就是考虑**被裁减参数对损失函数的影响, 在上世纪90年代初的[15][16]中提出了OBD和OBS这两种需要计算Hessian矩阵的方法，由于Hessian矩阵涉及到计算二阶梯度，十分耗时，在大规模的神经网络中几乎不可能被广泛应用。近年来有一些基于该思路的方法被研究和提出， 通过各种优化算法来快速近似二阶梯度，其中也包括对一阶梯度进行改造。**

**第三种思路是**考虑**对特征输出的可重建性的影响**，目标是最小化裁剪后网络对于特征输出的重建误差。如果对当前层进行裁剪不会对后面的输出产生大的影响，说明裁掉的是不太重要的信息。其中较早的工作是只考虑对后一层网络输出的影响(贪心算法)，后续也有最小化分类网络倒数第二层的重建误差，并将重要性信息反向传播到前面以决定哪些网络结构需要裁剪 (基于学习的算法)。

## 1.5 本课题工作

本课题可以被分为两个阶段。在第一个阶段，本课题重新审视网络状况带给联邦学习的挑战和基于阈值的方法的特性，主要回答以下几个问题：1. 挑战是否被夸大了？ 2. 基于阈值的客户选择方法的缺点是什么？ 3. 是否有更好的解决方案？

具体来说，本课题在第一阶段中做出以下贡献：

**瓶颈** 本课题进行了对联邦学习训练中涉及的客户的进行分析，发现网络限制挑战在某些方面可能被夸大了。但同时本课题也发现由基于阈值的用户选择会引起偏见，这些偏见是多方面的以至于之前被忽视。本课题进一步分析了它对公平聚合(fair aggregation)和模型个性化 (personalization)中目前最佳算法效果的影响作为论证。

**丢包容忍** 本课题通过使用ThrowRightAway(TRA)进行随机丢包来探索联邦学习的容错极限。顾名思义，TRA随机丢弃数据包以加速低带宽设备的上载，从而使所有客户端都可以进行完全公平的选择，同时避免了由重传引起的混乱影响。对于特定算法，TRA改造聚合算法以补偿丢失的信息。

**适用性** 通过将TRA与个性化和公平聚合两大领域内的最新算法分别集成，全面比较其与基于阈值的用户选择算法的区别。实验结果表明，在相同数据传输量下, TRA有更好的效果。本课题将可穿戴设备的联邦学习demo 和 TRA 与两种算法的集成开源：[demo](https://github.com/FL-Wearable/Trainer)和[loss-tolerant federated learning](https://github.com/Greilfang/Loss-Tolerant-Federated-Learning).

在第二阶段， 本课题做出如下贡献：

**历史信息** 本课题充分利用了用户在每次上传后的模型信息，并以此来预测网络丢包中丢失权重，探究了避免传统TCP/IP协议数据重传的可行性。本课题还用实验分析了保留的历史信息量和训练效果的关系。

**用户选择** 本课题采用了结合了动量剪枝的用户选择技术，通过选择有代表性的对全局模型更新有促进的用户模型参与聚合，减少了每次模型聚合需要的参与用户数，极大的节约了通讯量，同时也加快了收敛速度。

# 2 相关工作

每章另起一页

在本节本课题重点介绍有效的联邦学习对网络状况提出的要求。此外，本课题还讨论了用于压缩通讯量的现有技术以及利用剪枝技术节约通讯量可行性。通过这些技术的结合，本课题有望构建一个高通信效率和可用性的联邦学习系统。

## 2.1 用于可穿戴设备的机器学习范式

可穿戴设备是直接穿在身上或整合到用户的随身衣服或配件的便携设备。它不仅是一种硬件设备，也通过软件支持以及数据交互、云端交互来实现更复杂功能。比如在日常使用中，如智能手表可以记录诸如用户的心跳、血压，运动方向等数据，进而预测出用户在进行什么运动，并检测用户的活动是否过于激烈。谷歌发布了全新应用于可穿戴设备的Android Wear 2.0系统**[17]**和相关设备以运行全新的设备端智能回复技术。 **[18]**研究了Android Wear的执行效率，通过运行15个基准测试套件，检查了系统的CPU使用率、空闲时间片、线程并行性和微体系结构，发现影响效率的因素和未知的OS设计缺陷。

在机器学习应用方面，**[19][20][21]**强调了在可穿戴无线人体局域网这类敏感网络中部署机器学习(深度学习)模型的主要挑战和问题。 在具体实施上，**[22]**提出了一套完整的基于可穿戴设备的深度学习方法论，其第一次对外部惯性传感器数据中采集的特征和来自一组浅层特征的补充信息进行领域融合(domain adaption)以实现更准确的活动分类，然而并没有解决深度学习模型参数量导致的高昂的传输和计算开销；在细分应用上，**[23]**提出了不过分依赖于特定领域知识帕金森病预测系统；另有相当部分的工作关注使用可穿戴设备对心理疾病 (抑郁**[24]**, 焦虑**[25]**))和生理疾病**[26]**的分类和回归；此外还有摔倒检测**[27][28]**、活动监测**[29]**等应用。这些工作极大地拓宽了可穿戴设备机器学习的使用范围，一次次地提出了最新的SOTA算法，使得模型对可穿戴设备普遍存在的采集数据的误差及数据分布的差异更加鲁棒，精度不断提升。

## 2.2 可穿戴设备的联邦学习

如上节最后所述，可穿戴设备的机器学习在各个领域具有可行性，但这些工作都未能解决一些共同的缺陷：(1). 这些工作都是可穿戴设备的机器学习在特定领域的应用，并没有提出广泛的通用的提高 (2). 这些工作都在提高预测精度，却没有在网络延迟，电池功耗等方面的改进。 (3). 这些方法依赖于源源不断地被采集到的用户数据，却没有保护用户的隐私。其中第(3)点尤其值得注意，可穿戴设备强化了用户对网络的依赖性，这些设备采集健康指数、行为习惯、生活偏好和工作履历痕迹，带来与日俱增的隐私泄露风险，因此可以引入联邦学习的概念解决隐私保护的难题。联邦学习可以通过同态加密(Homomorphic Encryption)**[30]**，差分隐私(Differential Privacy)**[31]**，安全聚合(Secure Aggregation)**[32]**等技术保护隐私，这些算法由于各自的不同特性，都需要更多的聚合轮次或通信量来保证加密后的训练的收敛性。于是解决(3)实现隐私保护又进一步对(2)提出了更高的要求。在预测精度以外，这些缺陷阻碍了可穿戴设备机器学习的实用，也催生了将联邦学习的相关概念与技术引入可穿戴设备机器学习的各种尝试。

FedHealth**[5]**提出了全新的迁移学习的框架，添加新的损失函数项度量全局和用户本地模型的差异，虽然有助于模型对用户本地数据的预测任务的表现，但会让每个用户需要保留一个多余的模型，框架没有充分利用这个历史模型应对网络状况的突然变化。在资源紧张的情况下，这个框架会减小模型复杂性可以同时容纳两个可移植模型，其是否相对于使用单个较复杂模型具有优势还值得研究；Per-FedAvg**[33]**寻找一个特殊的初始化模型，这个模型可以让每个用户用尽可能少的迭代步找到适用于本地模型的最优模型。pFedMe**[34]**在全局损失函数中增加了特殊项限制模型复杂性和鼓励本地模型的个性化，收到了良好的模型个性化效果。本课题也将pFedMe作为目标来研究用户偏向选择对个性化性能的影响。[to add]

这些算法都在各自的研究领域对可穿戴设备的机器学习有所改进，而这些算法在实际中应用还需要考虑到网络延迟等因素对它们效果的影响。例如在移动设备（手机）联邦学习中，如果设备的电量耗尽退出训练和偶发的网络丢包会极大地影响某个算法训练的效果，则该算法也不具备实用的潜力。研究网络状况对联邦学习效果的影响，也是本课题的一大研究内容。

## 2.3 联邦学习的通信压缩

为了应对上述网络延迟对联邦学习的挑战，一系列节约通讯量的方法被提出，它们在保证模型聚合的收敛性的情况下尽可能减少通信量。

一些工作试图选择本地模型和全局模型更接近的一部分用户参与联邦学习训练：MOCHA [12]利用多任务学习（Milti-task Learning）框架来获得联邦学习训练数据的分布特性，其目标是同时训练有关联的模型，通过他们的关系矩阵提取所有用户（任务）之间的关系；Gaia**[35]**通过比较来关注本地更新的重要性，在每次迭代中，每个用户只保留绝对值超过一个预定阈值的更新(被认为是重要的)，其余的被视为不重要的更新并被丢弃。Gaia可以减少网络占用空间，但忽视用户和全局模型的相关性，随着训练的进行每个用户的更新值会不断逐渐，这给预定值的设定造成了困难; 作为对Gaia的改进，CMFL**[47]**考虑到了用户模型和全局模型的相关性，相比Gaia获得了更高的收敛加速比和通讯压缩比，但是其相关性只考虑了模型参数的正负号而不包括具体的数值，由于深度学习模型参数冗余，最优模型大多数参数在0附近，CMFL选择参数符号相似的用户的策略会导致全局模型参数向一个方向更新。 另外CMFL每次迭代需要所有用户计算本地更新，在导致计算开销的同时导致始终有相似的用户被重复选择，不利于全局模型学习到所有用户的不同数据分布，影响模型鲁棒性。在后续章节，本课题将对这些缺陷进行改进。

另一些工作，则试图压缩用户单次上传的参数量，标准是自定义的权重重要性标准。由于本课题借用模型剪枝的一些工作作为重要性的衡量标准，这些工作将在2.4中被详细概述，在本节本课题介绍除此之外的一些衡量根据权重重要性传输的方法。Konecny**[36]** 提出了两种降低上行链路通信成本的方法：(a) 结构化更新(structured update)，直接从有限的数据空间中学习部分权重更新进行低秩矩阵分解或随机遮盖。(b) 草绘更新(sketched update)，每个用户学习完整的模型更新，然后使用量化(quantization)，随机旋转(random rotation)和子采样(subsampling)的组合对其进行通信压缩，这些方法已经在其他工作中被证明有效, **[36]**并没有对联邦学习非独立同分布假设给出更多的改进；对集成学习场景，FedBoost**[37]**提出了通信量与集成规模无关的集成算法；Caldas**[38]**采用Federated Dropout 计算在本地较小的子模型上的局部更新，这些子模型是服务端全局模型的子集，并可以更新到全局模型。这些算法从不同层面提出了联邦学习的通信优化方案，但它们都是静态地确定模型参数重要性，无法准确追踪模型权重随着训练深入的重要性变化情况以获得最大程度的精度提升。在2.4中本课题将介绍一系列基于剪枝的模型权重度量方法，其中基于动量的剪枝算法也由于其能够动态反应模型参数重要性的迁移并对相应的参数进行裁剪，在本课题的算法中被采用。[to add]

## 2.4 基于剪枝的模型参数重要性度量

目前联邦学习的训练过程需要大量算力、用户内存和电量。数据的非独立同分布又导致联邦学习获得的模型要比普通机器学习训练（即将所有用户数据简单聚集在一起，不考虑隐私泄露）得到的模型更加复杂，每个用户每次迭代传输模型的通讯量随之增加。剪枝是提高推断效率的方法之一，可以高效生成规模更小、内存利用率更高、更容易被压缩的模型，此类技术还包括权重共享(weight sharing)和量化(quantization)。Google**[39]**将具有同等参数量的稀疏大模型和稠密小模型进行性能对比，发现在图像和语音任务上前者表现更佳。

在具体的发展历史上，细粒度剪枝包括对权重连接剪枝和对神经元进行剪枝是最简单，是最早期的剪枝技术。

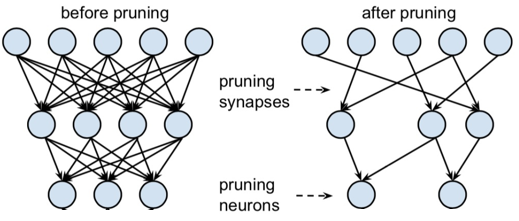


图4. 权重和神经元剪枝实例，剪枝内容包括了连接和神经元

这一类技术可以被划分为以下几个步骤：

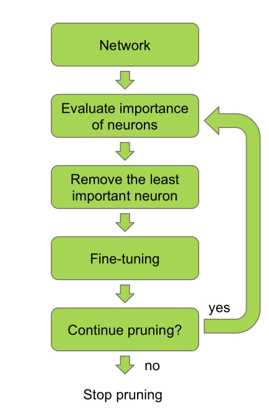


图5. 模型权重（神经元）剪枝步骤

如图5，剪枝中的第一步就是如何评估模型权重重要性，本课题在本节两个最典型的方法代表：基于连接幅度的方法**[40]**和基于损失函数的方法**[41],** 基于权重幅度的方法原理简单，且容易产生容易被压缩的稀疏化模型，但基于其主观地认为权重大就重要性高，只适合无法在模型无法进一步被调优的情况下使用；另一种经典的连接剪枝方法就是基于优化目标，根据剪枝对优化目标的影响来对其重要性进行判断，以最优脑损伤(Optimal Brain Damage, OBD)**[41]**方法为代表，建立了一个误差函数的局部模型并用泰勒展开近似，来预测扰动参数向量对优化目标造成的影响。由于剪枝目标希望找到一个参数集合使其被删除后损失函数增加最小，该方法需要求解损失函数的Hessian矩阵，引起的计算开销在深度神经网络中是不可接受的。

相对于连接权重剪枝，粗粒度剪枝更加实用，可以得到不需要专门的算法支持的精简小模型。 Network Trimming**[42]**通过激活的稀疏性来判断一个通道的重要性，认为拥有更高稀疏性的通道更应该被去除。它使用批正则化(batch normalization)中的缩放因子来裁剪不重要通道进。[to add]。

在联邦学习场景下，每次聚合都要求用户先对本地模型使用自己的数据进行训练获得更新，随后上传更新后的小模型，如果每次用户上传的模型都足够稀疏，则可以更大程度地实现通信压缩。 同时由于可穿戴设备类型各异，每次训练参与的设备数目，每个设备参与的轮次数也各不相同（有的设备可以始终参与训练，另一些可能由于网速或电池耗尽只参与过一次），很难保证数据的总体分布在始终保持稳定，所以在借鉴模型剪枝的参数重要性度量算法时，该算法需要能够在每次上传前能够进行剪枝，使得模型能够做到数据分布敏感且尽可能稀疏。出于这样的需求，在这里本课题引入动量剪枝(movement pruning)**[43]**的技术，其属于细粒度剪枝的一种，因而可以尽可能地做到模型的稀疏化；同时不同于基于幅值剪枝，其不关注模型权重的绝对值而是关注模型参数的变化情况。总结来说，动量剪枝认为：如果权重在训练中值不断远离0，说明其重要性在增加，因而在剪枝过程中应被优先保留。动量剪枝在每一次重要的模型更新后都可以被执行，使模型始终追踪到数据分布的变化并保持足够稀疏。这些优点符合联邦学习的使用场景，因此被本课题的算法采用作为参数重要性的度量。[to add]

# 3 算法

本节详细阐述一个通信高效的联邦学习系统的设计，首先将是研究（1）网络状况如何引起用户选择的偏见以及将（2）在何种程度上影响联邦学习的效果。通过（1）和（2）的分析，进而提出系统设计的解决方案。

## 3.1 网络状况分析

本课题从FCC提供的公开数据集“download and upload speed”中收集了数据。这些数据通过用户电话中的Android和iOS应用程序进行网速测量，它包含来自成千上万名自愿参加者的跟踪记录，记录平均收到的数据包，丢失的数据包和吞吐量。其中吞吐量是作为饱和流期间的速度之和收集的，因而可以被认为是最大速度。本课题将丢失的数据包除以接收到的和丢失的数据包的总和作为数据包丢失量度。 数据集表明90％的用户的数据包丢失率<0.1，76％的用户的上传速度> 2Mbps。可以看出，实际上大多数用户具备参与普通联邦学习系统所需的足够的网络容量。进一步研究发现，用户的上传速度之间差距很大：例如，24%的用户的上传速度<2Mbps，而51％的用户的上传速度> 8Mbps。传输速度是选择客户时的重要指标，并逐渐被工业和学界采用。例如，Openmined将2Mbps设置为默认的客户端选择上传速度阈值。从数据分析中发现，有相当的比例用户因不满足网络阈值而被联邦学习系统始终排除在外。

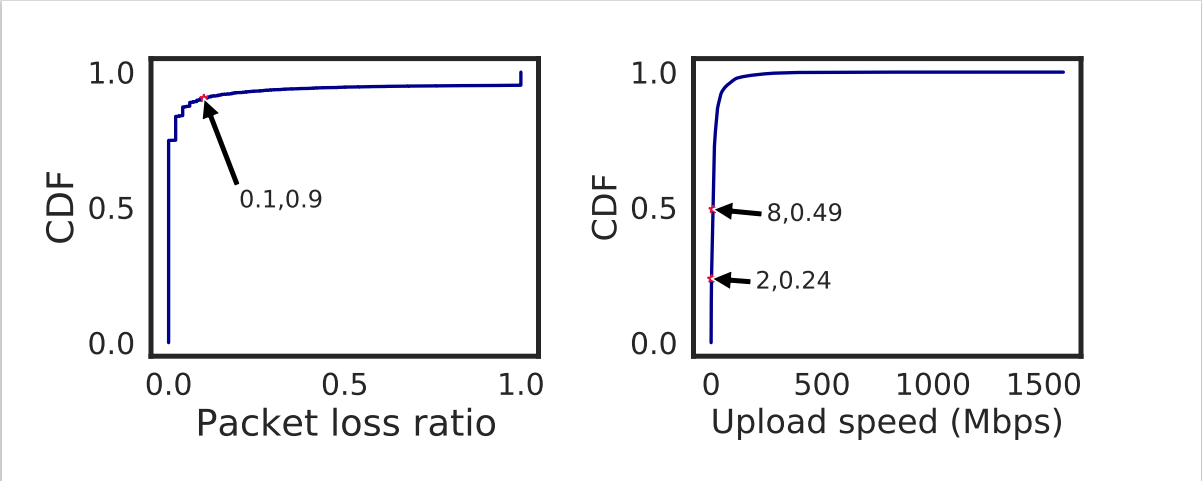


图6. 用户网络状况分析，y轴是用户的累积分布函数 (CDF), x

轴分别是丢包率和上传速度

对这些用户数据的追踪分析表明，大多数移动用户的网络条件并不像大多数相关工作所设想的那样“受限”和“具有挑战性”。但是上载速度的极大差异确实可能导致一个固定阈值将一部分用户始终排除在外，并使得模型产生偏见。

## 3.2 网络状况的影响

紧接3.1的分析，本节继续研究网络状况给各领域的联邦学习算法带来的影响。

**定义1（合格客户）**合格客户是指满足联盟学习聚合所需的网络阈值的客户。

**定义2（合格比率）**合格比率是合格客户在所有客户中所占的比例。

在基于阈值的设定当中，只有一定合格比例的合格用户才有可能被选中参与模型聚合。由于始终有一些用户不满足阈值要求，系统只会选择合格用户，这是偏见的产生的原因。为了让评估更加全面，本课题在实验中测试不同合格比例 (100%, 90%, 80%, 70%)。 更具体的本课题观察这种模型选择的限制对模型聚合(aggregation)，公平性(fairness)，个性化的影响(personalization)。为了体现联邦学习需要处理的数据的非独立同分布，本课题引入了一个名叫synthetic的数据集，其有两个参数α和β来控制数据的差异性（α和β越大，数据差异性也越大）。

**聚合(Aggregation)** 首先本课题检查了用户的偏差选择对普遍的模型聚合算法的影响，本课题选取了目前联邦学习领域最广泛最通用的FedAvg**[44]**算法作为测试对象。如图3所示，较小的合格比率对模型性能具有较高的影响。如果使用如图6中的2 Mbps作为选择阈值，则模型精度将降低10％左右。

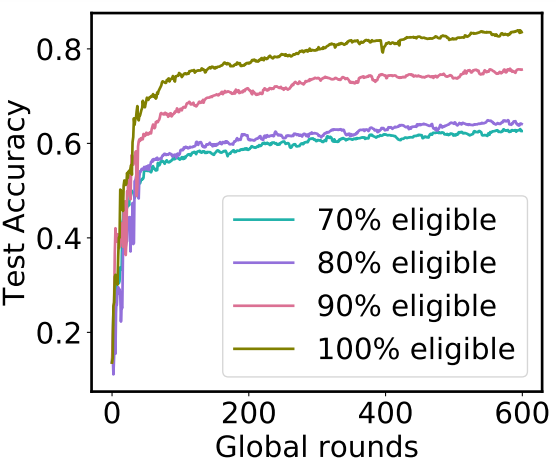


图7. 偏差选择对聚合算法(FedAvg)的影响，数据集是Synthetic(0.5,0.5)。

**公平性** 如1.2节所述，现有方案可解决用户参与训练不足和用户参与训练过多两者的差距，但无法应对一些从未参加过训练的用户的出现。为了验证该论点，本课题使用q-FedAvg**[10]**的代码和默认参数重现q-FedAvg的评估。本课题调整训练样本数据在每个设备上的分布（从i.i.d数据到非i.i.d数据），以全面测试由有偏见的客户选择导致的准确性和公平性的下降。表1显示，由于i.i.d和非i.i.d数据分布的偏向选择，q-FedAvg的性能受到影响。非i.i.d数据在准确性和公平性方面比i.i.d数据表现出更大的性能下降。

**个性化** 现有的方法要么通过迁移学习**[5]**，要么通过损失函数来衡量局部和全局模型的查一性。在资源密集型情况下，前提学习会减小模型的大小以使设备可以同时容纳两个可相互移植的模型，其是否比在单个较大模型的优势需要研究。 Per-FedAvg **[33]**寻找一种初始共享模型，客户可以根据自己的数据通过一些梯度下降轻松适应。pFedMe**[34]**在全局训练的损失函数中增加了Moreau Envelop且表现优秀。本课题将pFedMe作为基准算法。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dataset** | **TH** | **Average** | **Best 10% / Worst 10%** | **Variance** |
| **Synthetic(i.i.d)** |  | 72.47% | 91. 85% / 43.19% | 179 |
| **√** | 68.67% | 94.25% / 36.30% | 245 |
| **Synthetic(0.5,0.5)** |  | 66.21% | 98.30% / 22.51% | 536 |
| **√** | 52.81% | 99.79% / 0 | 1350 |
| **Synthetic(1,1)** |  | 64.17% | 100% / 7.67% | 937 |
| **√** | 55.24% | 100% / 0 | 1439 |

表1. 有偏见的客户选择对公平性（q-FedAvg）的影响。阈值(Threshold, TH)表示是否考虑70％的合格比率。Best 10% / Worst 10% 表示精度最好（最大）的10%的用户的平均精度。

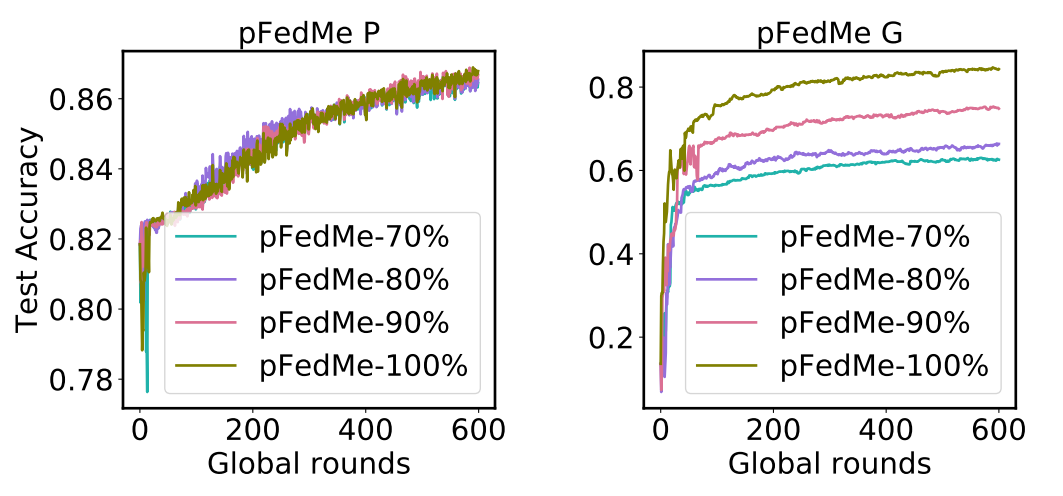


图8. 有偏见的客户选择对pFedMe的个性化表现的影响。

P表示用户各自的个性化模型评估本地数据的平均精度，

G表示全局模型评估数据集的精度。

数据集为Synthetic(0.5,0.5)。

如图8所示，pFedMe在其个性化模型中显示了其对网络丢包的鲁棒性。但是全局模型的表现会以合格比率的降低而显著衰减。pFedMe通过更多的计算和功耗成本实现了个性化模型性能的鲁棒性：与大多数在用户本地更新之前选择用户的方法不同，pFedMe让所有客户都进行本地更新，因此其个性化模型的表现收到全局模型表现的影响较小，代价是要更多的计算和客户端用户端更新。对Per-FedAvg等只在全局聚合后使用更新的模型，低合格比率会明显降低性能。

**总结** 由基于阈值的用户选择引起的偏见会严重影响联邦学习的聚合，个性化和公平性的表现。因此，需要一种允许用户公平参与的可以容忍网速延迟和丢包的替代方案。

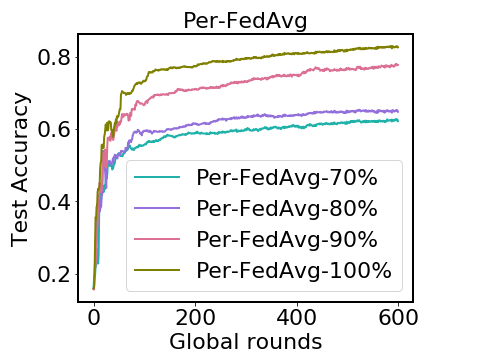
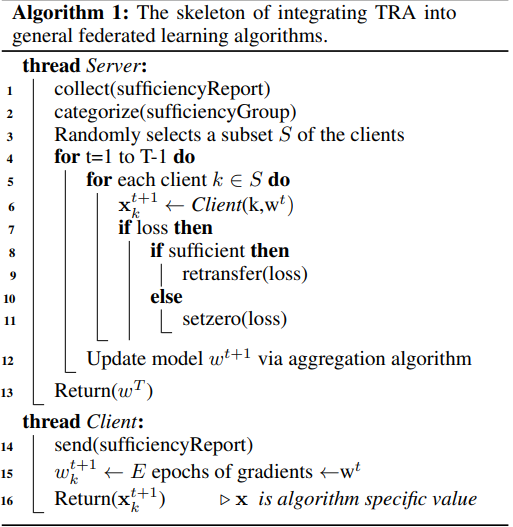


图9. 有偏见的客户选择对Per-Fedavg的影响，数据集是Synthetic(0.5,0.5)。

## 3.3 ThrowRightAway

在本节中，本课题提出了一种基于阈值的方案的替代，以解决由于偏向选择而导致的性能下降, 这个算法的思路被本课题进一步使用，在系统设计中，该算法将被进一步改进。

**[45]**表明，一定程度的数据丢失在分布式学习系统中不一定有害。通过经验评估，其发现机器学习算法可以容忍有限数据丢失(10%–35%)。受此启发，本课题探索跨设备联邦学习系统中的损失容限。本课题提出了ThrowRightAway(TRA)方案，即使设备的网络容量不满足需求，也可能在联邦学习中被选到，服务器也接受所有客户端作为合格的参与者。更新上载期间的重新需求和不希望的丢包率。 TRA轻便易于实现，可以被集成到不同种类的联邦学习算法中以增强其性能。算法1显示了TRA和通用联邦学习算法集成的框架。



算法1. ThrowRightAway算法的框架，在setzero的具体实现可以根据系统需求进行修改，在后续的系统设计中，可以进一步定制。

在选择开始时，每个客户端都会将其网络条件与预设标准进行比较，并向服务器发送包含用户参与训练是否可行的关键信息，例如0和1分别表示不可行和可行。在收集了参加的客户的参与训练可行性信息后，服务器根据信息将候选客户分为**充足**(sufficient)和**不足**(insufiicient)。服务器会随机选择多个客户端（选择标准与所属组无关）发送全局模型。客户经过本地训练后会传回更新。在检测到数据丢失时，如果客户端属于**充足**组，服务器将发送重新传输通知，否则将直接预测丢失数据。其余过程遵循通用的联邦学习流程。

本课题根据不同的数据丢失发生的用户类型估计重传时间：对**充足用户**，在不影响聚合速度的情况下进行重传只需要很少的时间。但是对**不足用户**，重传的延迟可能会很严重。出于这样的考虑，TRA不会重新传输丢失的数据包，而是丢弃此类延迟/丢失的数据包，并记录数据丢失。上传完成后，TRA使用丢失记录和其他用户同位置上传参数重新计算样本空间实现自动聚合。在初步试验中，丢失数据用以下的公式被预测：

和   分别表示n个**充足用户中**第i个用户上传的模型权重和m个不足用户中第j个用户上传的权重，参数r考虑到了所有用户的平均丢包率（r在初步实验中被赋固定值）。本课题记的数学期望 ,在假设模型权重的数值大小分布大致相同的情况下，由于**不足用户**有一定几率丢包，被丢弃的数据会先被置0，则 ,很显然,这也等于没有丢包（理想情况下）模型权重的期望。

为了体现TRA易于整合和良好的性能，本课题将会将其整合入q-FedAvg和pFedMe，具体数据会在实验评估部分展示。

## 3.4 用户选择算法 MAFL

3.3节详细阐述了当网络延迟导致数据丢失时，系统的应对措施：（1）传输-重传 （2）丢弃-预测，并给出了一个预测的算法实例。这两种措施都是在数据丢失时试图去消除它的影响，而本节则试图从用户选择的角度出发，选择状况最好最有利于联邦学习模型训练的若干具**有代表性**的用户，这能够极大的减少通信量。如果由于不能满足网速要求而导致上传的模型权重里有大量被预测的有偏差的权重，则会带来噪音影响收敛，所以直觉上说，有利于联邦学习训练的**有代表性的**用户应该是网速良好可以满足要求的，所以有效的用户可以能根本上减少数据丢失和延迟。

### 3.4.1. 理论支持

减少联邦学习的通信开销主要有两种思路：（1）减少单次更新的上传数据量。（2）减少更新的次数。主要有不同于3.3用丢包预测来减少（1），而有效的用户选择算法通过选择有代表性的用户来减少（2），这要求参与的用户的模型尽可能接近聚合后前后全局模型的更新量。

本课题提出Movement Aware Federated Learning (MAFL) 的用户选择算法 通过动态识别无关的本地模型的更新，并将其从上传模型序列中排除。这种简单的方法是有效且通用的。首先，因为联邦学习是在边缘设备上执行，训练更新会根据这些设备上的数据分布的不同有区别。鉴于数据的非独立同分布性，一部分模型更新会产生偏见甚至会被认为是离群值。整合这些有偏见的更新会将全局模型推向与最优模型相反的方向排除这些异常更新，排除这些异常更新不会对联邦学习的训练造成不利影响反而可以避免不必要的数据传输；这类方法通常适用于广泛的联邦学习。

同步更新方案。基于**[34]，[44]，[45]**，本课题假设一个同步更新方案在联邦学习的每次训练迭代中进行的操作：

1. 用户用本地数据在独立训练一个本地模型，比如Gboard的点击历史**[46]**
2. 所有用户均上传其模型更新（梯度更新）到云端中央服务器进行聚合。
3. 中央服务器汇总收到的本地更新（通常通过平均）来获取全局更新，用全局更新改进全局模型。

训练目标 首先形式化地表述训练过程，把所有参与训练的用户写成, 第k个用户有自己的数据集 一个全局预测模型在所有D个客户间被传递和共享。目的是找到最佳模型参数 能够最小化全局损失函数。

其中 是第k个用户在其数据上的损失函数。在实践中，本课题采用FedAvg作为联邦优化算法。每个用户通过随机梯度下降 (Sochastic Gradient Descent) 在每次训练迭代中计算批次梯度并用其更新模型。随后每个用户上传更新后的本地模型，服务器将这些模型聚合。当迭代t次时，模型参数可以被写为：

其中和 分别是第k个用户在其本地数据上的梯度和方差，是一次更新，设总的迭代次数是，MAFL的核心思想就是从上式中的所有u中找到与全局模型更新最相关的。

**如何预测模型参数** 找到相关模型更新的前提是能够感知到全局模型更新的分布。如前文所述，Gaia**[35]**提供提供了一种简单的解决方案，这种方案仅通过判断模型更新的幅度是否大于一个预设的阈值来判断是否要上传该模型。本课题首先探索这样的思路是否能移植到联邦学习场景中，答案是否定的，我们在下方给出了理由和有效的方法应具备的条件。

1. 首先，移动端联邦学习有大量用户并且数据不满足独立同分布，一些用户由于数据较多，每次迭代更新量(Update)较大；此外更新幅度也取决于设定的学习率和训练是否深入（早起轮次的更新幅值要比训练后期大得多）因此很难找到一个合适的可以在训练中持续可用的阈值。**有效的方法应该能够通过用户间的比较得到选出合适的用户，如果引入阈值，则阈值的设定应和其他的超参数无关。**
2. 其次，由于全局模型是大量客户端本地模型的聚合，本地更新是否有助于全局模型更新不能简单地通过其更新幅度判断。大量的客户进行小规模的同方向直接决定了全局模型的更新。在基于幅值的方法中，这样的更新直接被忽略。CMFL**[47]**考虑到了更新的正负号，因而在方向上更加精确。**有效的方法应该准确预测模型更新的方向和幅值，其中方向更重要。**

由于在每个迭代轮次中，全局模型的权重会被下发到用户，用户可以通过两次下发的模型判断模型的更新方向。

唯一的问题是: 服务器无法在聚合用户模型前就得到这轮模型的更新方向，在这里本课题采用上一轮的更新方向代替这一轮，即每次用户端收到服务器下发的模型时，先计算更新方向并作为下一次聚合后模型的更新防线，依据是每次迭代不会有太多参数的更新方向发生变化**[47]**，本课题的模型相关性的度量方法也会减小这种近似导致的误差(3.4.2节)。

**衡量更新相似度** 给定在当前迭代中的全局更新和用户的一个本地更新，本课题需要选择一个衡量更新相关性的适当指标。训练更新本质上是模型的梯度向量参数，本课题比较两个更新（本地更新和全局更新）并计算有多少个参数在两次更新中有接近的动量(movement)，用参数总数对结果进行归一化得出一个百分比本课题将使用此百分比以衡量本地更新的相关性。和的相似度可表示为：

表示动量，如果两次更新的同一位置参数动量正负号相同，即, 则，否则为-1。

**动量(mov)** 在讨论了衡量相似度度量的逻辑和用户选择的主要流程后，现在给出的定义和理论依据**[43]**。

为了简便，令  代表模型的更新矩阵（实际中矩阵维度可以是其他形状的），为了确定中哪个权重需要被裁减，我们要确定它们各自的重要性分数，因此我们引入一个和维度相同的分数矩阵**,** 我们将动量作为更新重要性的量度，则和有如下关系：

其中的具体计算是从模型权重矩阵 中得出的，不同于幅值剪枝(magnitude pruning) 关注绝对值较大的权重，动量剪枝试图保留那些在更新中远离0的权重，因为不断远离0说明随着训练的深入，该权重变得越来越重要。

我们定义一个掩盖矩阵以及动量剪枝输出的权重矩阵，则有, 是神经网络的输入**。**我们将的表达式分拆，可以得到。由于在被剪枝后网络最后一层的梯度大部分是0，可以用straight-through estimator**[48]**直接跳过剪枝层的梯度，即,设损失函数对的求梯度如下：

在这样的定义下，尽管剪枝会留下大量梯度为0的参数，但是得分矩阵会随着更新而更新。

在动量剪枝中，损失函数随着的变化有标准形式： , 将其代入上式(因为每次更新的变化很小将其省略):

上式给出了的定义和其与的关系，我们进一步分析，根据该式，因为随着训练的进行，会逐渐变小**，**如果对于参数，其重要性在增大，则需要<0。在以下两个情况下会满足：

* 且
* 且

**结论** 在动量剪枝中，重要的参数是在训练过程中逐步远离0的参数，其重要性用度量。对第次迭代，用每个用户本地进行梯度更新的更新量**替代**，则可以得到的表达式。

因为不能在聚合前得到，如前所述，用替代，得到：

这就是的定义，我们用其代入得到两次更新的相关性。

# 4 实验评估

## 4.1 多元酸体系的结果和讨论

### 4.1.1 直接计算法

三级以下单独占行标题采用A.、B.、C.和a.、b.、c.…或(1)、(2)、(3)和①、②、③…。黑体五号。标题序号缩进2个汉字符。序号和标题之间空一格。

五号，黑体(英文Times New Roman)，行距18磅，段前0行，段后0行。

……。

### 4.1.2 半整数法

A. 半整数法的求解过程

a. 可以直接得到半整数的情况

从表4.1可见，当为0.5和1.5时的pH分别是5.32和2.74，即……。

b. 无法直接得到半整数的情况

在用半整数生成函数法直接求解时，会遇到这样的问题……。表4.4列出了滴定过程中乙二酸溶液的各主要物理量的部分数据。从表4.4可见，……。

图表可横版

表序写在表题左方不加标点，空一格写表题，表题末尾不加标点，表格逐章编序，表序必须连续，如表4.4表示第四章的第四个表。表题：小五，宋体(英文Times New Roman)，居中置于表上方，行距18磅，段前0行，段后0行。

……

(1)表格上下与正文之间各空一行；(2)采用三线表，两端与页面对齐；

(3)表中文字：小五，宋体 (英文Times New Roman)，行距18磅，段前0行，段后0行。

……。

表4.4 草酸的部分数据列表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *V*/mL | *E*/mV | pH | [H+]/(mol·L-1) |  |
| …… | …… | …… | …… | …… |
|  |  |  |  |  |
| 1.00 | 2.62×102 | 2.22 | 6.07×10-3 | 1.14 |
| 1.50 | 2.60×102 | 2.27 | 5.42×10-3 | 1.10 |
|  |  |  |  |  |
| …… | …… | …… | …… | …… |

续表4.4

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *V*/mL | *E*/mV | pH | [H+]/(mol·L-1) |  |
| …… | …… | …… | …… | …… |
| …… | …… | …… | …… | …… |
| …… | …… | …… | …… | …… |

(1)若表格分页，则该表第2页的表题省略，但表头(即“*V*/mL┄”所在行)应重写，并在表右上方加注“续表X.X”；

(2)“续表X.X”的格式：小五，宋体(英文Times New Roman)，行距18磅，段前0行，段后0行，右空2格。



(1)图居中，上下与正文之间各空一行；

(2)图中文字：小五，宋体(英文Times New Roman)，行距1倍，段前0行，段后0行。

图4.1 乙二酸 ≥1.15数据段曲线及其拟合曲线(实线--实际曲线，虚线--拟合曲线)

针对这种情况，可以采用多项式拟合的方法求解。以乙二酸为例，在Excel中，选取1.28<<1.15之间的数据，以为横坐标，pH为纵坐标，做-pH曲线(见图4.1)，并添加趋势线，选择相关系数*R*2最接近1的多项式作为拟合方程，……。

B. 半整数法的计算结果

利用半整数生成函数法，对各种多元酸三次平行实验数据分别进行处理，并求平均值，结果见……。

(1)插图应有图序和图题，全文插图以章分组编序号，图序必须连续，不得重复或跳缺。如图4.1表示第四章的第一幅图。

(2)图题：小五，宋体(英文Times New Roman)，居中置于图下方，行距18磅，段前0行，段后0行。

C. 半整数法计算结果的讨论

……。

### 4.1.3 分段拟合法

……。

## 4.2 氨基酸合铜体系的结果和讨论

……。

## 4.3 关于计算方法的讨论

## 4.4 关于其他问题的讨论

# 5 结论和展望

段落号：(1)…。

(2)…。

(3)…。

## 5.1 结论

(1)生成函数法可以分为直接计算生成函数法、分段拟合生成函数法及半整数生成函数法。这三种方法有如下特点：①……；②……；③……。

(2)本文运用三种不同生成函数法，测定了多元酸和氨基酸合铜配合物的稳定常数，得到了……。

(3)三种生成函数法中无论哪一种方法，对待测酸或配合物稳定常数的大小均有一定的要求，如……。

……。

## 5.2 展望

(1)生成函数法理论可靠，计算方便，但……。

(2)在生成函数法的应用中，还有以下问题有待研究和解决：①……；②……。

……。

段内层次号 ：①…；②…。

根据配位(或酸碱)反应的平衡关系，可以得到生成函数[6]与配合物(或酸)的各级累积稳定常数及溶液中游离配位剂(或氢离子)的平衡浓度之间的关系：

 或  (2.1)

(1)公式应另起一行，正文中的公式、算式或方程式等应编排序号，公式的编号用圆括号括起，序号标注于该式所在行(当有续行时，应标注于最后一行)的行末。公式按章节顺序编号，如(2.2)表示第二章的第二个公式。公式序号必须连续，不得重复或跳缺。重复引用的公式不得另编新序号。公式和编号之间不加虚线。

(2)公式：五号，宋体(英文Times New Roman)，使用公式编辑器，居中，式号右对齐。

(3)公式行行距1.5倍，段前0行，段后0行。

(4)计算式及正文中涉及的物理量和变量，应为斜体；元素符号、记号及运算符号，应为正体。

式(2.1)中，是配合物(或酸)的生成函数；*βj*(*j*=1,2,…,*n*)是配合物(或酸)的各级累积稳定常数；[L](或[H+])是溶液中游离配位剂(或氢离子)的平衡浓度(其中，L的电荷已略)。

将式(2.1)写作以下形式，可以得到生成函数法测定配合物(或酸)稳定常数的基本关系式：

 或  (2.2)

利用式(2.2)，可以……。

# 

注：期刊若只有期，没有卷，则可以省略卷号，如参考文献[1]示例。若只有卷，没有(或不分)期，则可以省略期号，如参考文献[2]示例。

# 参考文献

期刊中析出的文献

1. Lim W Y B Luong N C Hoang D T et al. Federated learning in mobile edge networks: A comprehensive survey[J]. IEEE Communications Surveys & Tutorials 2020 22(3): 2031-2063.
2. Custers B Sears A M Dechesne F et al. EU personal data protection in policy and practice[M]. TMC Asser Press 2019.
3. de la Torre L. A guide to the california consumer privacy act of 2018[J]. Available at SSRN 3275571 2018.
4. Xu M Qian F Zhu M et al. Deepwear: Adaptive local offloading for on-wearable deep learning[J]. IEEE Transactions on Mobile Computing 2019 19(2): 314-330.
5. Chen Y Qin X Wang J et al. Fedhealth: A federated transfer learning framework for wearable healthcare[J]. IEEE Intelligent Systems 2020 35(4): 83-93.
6. Nishio T Yonetani R. Client selection for federated learning with heterogeneous resources in mobile edge[C]//ICC 2019-2019 IEEE International Conference on Communications (ICC). IEEE 2019: 1-7.
7. Lin F P C Brinton C G Michelusi N. Federated Learning with Communication Delay in Edge Networks[J]. arXiv preprint arXiv:2008.09323 2020.
8. Karimireddy S P Kale S Mohri M et al. SCAFFOLD: Stochastic controlled averaging for federated learning[C]//International Conference on Machine Learning. PMLR 2020: 5132-5143.
9. Mohri M Sivek G Suresh A T. Agnostic federated learning[C]//International Conference on Machine Learning. PMLR 2019: 4615-4625.
10. Li T Sanjabi M Beirami A et al. Fair resource allocation in federated learning[J]. arXiv preprint arXiv:1905.10497 2019.
11. Konečný J McMahan H B Yu F X et al. Federated learning: Strategies for improving communication efficiency[J]. arXiv preprint arXiv:1610.05492 2016.
12. Dong Y Zhao P Yu H et al. CDC: Classification Driven Compression for Bandwidth Efficient Edge-Cloud Collaborative Deep Learning[J]. arXiv preprint arXiv:2005.02177 2020.
13. Caldas S Konečny J McMahan H B et al. Expanding the reach of federated learning by reducing client resource requirements[J]. arXiv preprint arXiv:1812.07210 2018.
14. Xia J Zeng G Zhang J et al. Rethinking transport layer design for distributed machine learning[C]//Proceedings of the 3rd Asia-Pacific Workshop on Networking 2019. 2019: 22-28.
15. LeCun Y, Denker J S, Solla S A. Optimal brain damage[C]//Advances in neural information processing systems. 1990: 598-605.
16. Hassibi B, Stork D G. Second order derivatives for network pruning: Optimal brain surgeon[M]. Morgan Kaufmann, 1993.
17. Liu R, Lin F X. Understanding the characteristics of android wear os[C]//Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services. 2016: 151-164.
18. Liu R, Lin F X. Understanding the characteristics of android wear os[C]//Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services. 2016: 151-164.
19. Al‐Turjman F, Baali I. Machine learning for wearable IoT‐based applications: A survey[J]. Transactions on Emerging Telecommunications Technologies, 2019: e3635.
20. Beniczky S, Karoly P, Nurse E, et al. Machine learning and wearable devices of the future[J]. Epilepsia, 2021, 62: S116-S124.
21. Miotto R, Wang F, Wang S, et al. Deep learning for healthcare: review, opportunities and challenges[J]. Briefings in bioinformatics, 2018, 19(6): 1236-1246.
22. Ravi D, Wong C, Lo B, et al. A deep learning approach to on-node sensor data analytics for mobile or wearable devices[J]. IEEE journal of biomedical and health informatics, 2016, 21(1): 56-64.
23. Kubota K J, Chen J A, Little M A. Machine learning for large‐scale wearable sensor data in Parkinson's disease: Concepts, promises, pitfalls, and futures[J]. Movement disorders, 2016, 31(9): 1314-1326.
24. Ghandeharioun A, Fedor S, Sangermano L, et al. Objective assessment of depressive symptoms with machine learning and wearable sensors data[C]//2017 seventh international conference on affective computing and intelligent interaction (ACII). IEEE, 2017: 325-332.
25. Halim Z, Rehan M. On identification of driving-induced stress using electroencephalogram signals: A framework based on wearable safety-critical scheme and machine learning[J]. Information Fusion, 2020, 53: 66-79.
26. Inan O T, Baran Pouyan M, Javaid A Q, et al. Novel wearable seismocardiography and machine learning algorithms can assess clinical status of heart failure patients[J]. Circulation: Heart Failure, 2018, 11(1): e004313.
27. de Quadros T, Lazzaretti A E, Schneider F K. A movement decomposition and machine learning-based fall detection system using wrist wearable device[J]. IEEE Sensors Journal, 2018, 18(12): 5082-5089.

(1)所有引用的期刊需写出完整刊名。按论文中参考文献出现的次序，用阿拉伯数字自然编号，序码加方括号，顶格书写。

(2)五号，宋体(英文Times New Roman)，行距18磅，段前0行，段后0行。

(3)参考文献不少于10篇，其中外文文献不少于2篇 (这是最低要求。各学院可以根据本学院情况制定数量要求)。

1. Özdemir A T, Barshan B. Detecting falls with wearable sensors using machine learning techniques[J]. Sensors, 2014, 14(6): 10691-10708.
2. Baig M M, GholamHosseini H, Moqeem A A, et al. A systematic review of wearable patient monitoring systems–current challenges and opportunities for clinical adoption[J]. Journal of medical systems, 2017, 41(7): 1-9.
3. Gentry C. A fully homomorphic encryption scheme[M]. Stanford: Stanford university, 2009.
4. Dwork C, Kenthapadi K, McSherry F, et al. Our data, ourselves: Privacy via distributed noise generation[C]//Annual International Conference on the Theory and Applications of Cryptographic Techniques. Springer, Berlin, Heidelberg, 2006: 486-503.
5. Bonawitz K, Ivanov V, Kreuter B, et al. Practical secure aggregation for privacy-preserving machine learning[C]//proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. 2017: 1175-1191.
6. Fallah A, Mokhtari A, Ozdaglar A. Personalized federated learning: A meta-learning approach[J]. arXiv preprint arXiv:2002.07948, 2020.
7. Dinh C T, Tran N H, Nguyen T D. Personalized federated learning with Moreau envelopes[J]. arXiv preprint arXiv:2006.08848, 2020.
8. Hsieh K, Harlap A, Vijaykumar N, et al. Gaia: Geo-distributed machine learning approaching {LAN} speeds[C]//14th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 17). 2017: 629-647.
9. Konečný J, McMahan H B, Yu F X, et al. Federated learning: Strategies for improving communication efficiency[J]. arXiv preprint arXiv:1610.05492, 2016.
10. Hamer J, Mohri M, Suresh A T. FedBoost: A Communication-Efficient Algorithm for Federated Learning[C]//International Conference on Machine Learning. PMLR, 2020: 3973-3983.
11. Caldas S, Konečny J, McMahan H B, et al. Expanding the reach of federated learning by reducing client resource requirements[J]. arXiv preprint arXiv:1812.07210, 2018.
12. Zhu M, Gupta S. To prune, or not to prune: exploring the efficacy of pruning for model compression[J]. arXiv preprint arXiv:1710.01878, 2017.
13. Han S, Pool J, Tran J, et al. Learning both weights and connections for efficient neural network[C]//Advances in neural information processing systems. 2015: 1135-1143.
14. LeCun Y, Denker J S, Solla S A. Optimal brain damage[C]//Advances in neural information processing systems. 1990: 598-605.
15. Liu Z, Li J, Shen Z, et al. Learning efficient convolutional networks through network slimming[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 2736-2744.
16. Sanh V, Wolf T, Rush A M. Movement pruning: Adaptive sparsity by fine-tuning[J]. arXiv preprint arXiv:2005.07683, 2020.
17. McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Artificial Intelligence and Statistics. PMLR, 2017: 1273-1282.
18. Xia J, Zeng G, Zhang J, et al. Rethinking transport layer design for distributed machine learning[C]//Proceedings of the 3rd Asia-Pacific Workshop on Networking 2019. 2019: 22-28.
19. B. McMahan and D. Ramage, “Federated learning: Collaborative machine learning without centralized training data,” Google Research Blog, 2017.
20. Luping W, Wei W, Bo L I. CMFL: Mitigating communication overhead for federated learning[C]//2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS). IEEE, 2019: 954-964.
21. Yoshua Bengio, Nicholas Léonard, and Aaron C. Courville. Estimating or propagating gradients

through stochastic neurons for conditional computation. ArXiv, abs/1308.3432, 2013.



# 谢 辞

正文内容

出自内心，有感而发。正文：五号，宋体(英文Times New Roman)，两端对齐，段落首行左缩进2个汉字符，行距18磅，段前0行，段后0行。