面向更快更好的联合学习：一种特征融合方法

*姚欣、黄天池、吴成雷、张瑞霄、孙立峰*

清华大学

计算机科学与技术系

中国北京

# 摘要

联合学习支持在由大量现代智能设备（如智能手机和物联网设备）组成的分布式网络上进行设备培训。然而，在这种情况下，主要的优化算法，即联邦平均，存在着通信开销大和不可避免的性能下降的问题，特别是当本地数据以非IID方式分布时。本文提出了一种特征融合方法来解决这一问题。通过将局部模型和全局模型的特征集合起来，可以以较少的通信开销获得更高的精度。此外，特征融合模块为新进入的客户机提供了更好的初始化，从而加快了收敛过程。在流行的联合学习场景下的实验表明，基于特征融合机制的联邦学习算法在准确度和泛化能力上都优于基线，同时减少了60%以上的通信轮数。

*索引术语*-联合学习、特征融合、通信代价、泛化

# 1.    简介

手机、可穿戴设备和物联网设备在现代生活中扮演着重要角色。这些设备上的智能应用越来越流行，如智能个人助理、机器翻译、键盘输入建议等，这些应用通常使用预先训练好的模型，对客户端进行正向推理，缺乏灵活性和个性化。与此同时，智能边缘设备正在生成大量有价值但对隐私敏感的数据，这些数据有可能改进现有模型。为了充分利用设备上的数据，传统的机器学习策略需要从客户端收集数据，在服务器上训练一个集中的模型，然后将模型分发给客户端，这给通信网络带来了沉重的负担，并暴露出很高的隐私风险。

最近，一系列被称为联合学习（FL）[1，2，3]的工作可以直接在分布式网络上进行设备上的培训。主导算法被称为fed-

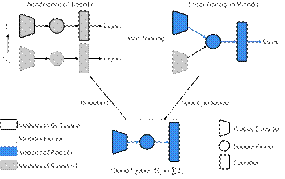


图1。FedFusion的训练迭代。在客户端的训练过程中，通过特征融合模块将局部和全局特征抽取器串联起来。

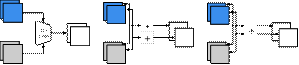
*速率平均*（FedAvg）[3]，它使用自己的本地数据对客户机执行分布式培训，并将这些模型聚合到中央服务器中，以避免数据共享。通过这种方式，FedAvg以一种有效的通信方式缓解了隐私问题。然而，进一步的研究[4,5]指出，与其他因素（如计算成本）相比，通信成本仍然是FL的主要限制因素，如果使用病理性非IID数据训练模型，则FedAvg的准确性将显著下降。

针对上述问题，本文提出了一种基于特征融合机制的联邦学习算法FedFusion。通过引入特征融合模块，在特征提取阶段完成后，我们从局部模型和全局模型中提取特征，而不需要额外的计算开销。这些模块使得每个客户机的培训过程更加高效，并且更加有针对性地处理非IID数据分发，因为每个客户机都将学习最适合自己的特征融合模块。

最后，我们的贡献如下：

|  |
| --- |
| 978-1-5386-6249-6/19/$31.00©2019 IEEE ICIP 2019 |

•据我们所知，这是第一篇将特征融合机制引入外语在线设备培训程序的论文。



          （a） 变频器（b）多功能（c）单

图2。三种类型的特征融合模块。融合算子实际上是一个映射函数：R→R×H×W*F*2C×高×宽*C*

•提出的特征融合模块以高效和个性化的方式将本地和全局模型中的特征进行聚合。

•在流行的FL数据集上的实验表明，所提出的FedFusion在精度和泛化能力方面都优于基线，同时减少了num-

通信轮次误码率超过60%。

# 2.    相关工作

联合学习是由McMahan等人提出的。[3] 为了解决大规模分布式智能设备的分散训练问题，而不直接访问隐私敏感数据。

考虑到通信成本仍然是制约外语教学的主要因素，人们已经进行了一些研究。Konecnˇy等人。[2] 在客户机到服务器通信环境中提出结构化和草图更新。Yao等人。[5] 在设备培训过程中引入了额外的约束，目的是在拟合本地数据的同时，集成更多来自其他客户的知识。Caldas等人。[6] 提出了联合丢弃来训练客户端上的子集，并将有损压缩[7]扩展到服务器到客户端通信。

# 3.    方法

在这一部分，我们将首先介绍所提出的特征融合模块，然后给出我们的具有特征融合机制的联邦学习算法（FedFusion）。

3.1条。特征融合模块

特征融合模块的详细架构如图2所示。

具体地说，输入图像通过局部特征提取和全局特征提取分别转化为两个特征空间，特征映射为（x），Eg（x）∈R×H×W*十埃尔如埃尔C*.

然后将（x）和（x）嵌入到融合特征空间中，其中（El（x），Eg（x））∈R。*F埃尔如FC*×高×宽

算法1特征融合服务器联合学习：

1： 初始化*G*0

2： 每轮r=0，1，2。。。做

3： ←随机抽样客户*高级米*

4： 对于每个客户机∈Sr do并行*t*

5： Ltr+1←客户端（Gr）6：结束

第七章：

8： 结束

客户：取整客户*rt*

1号文件：//是分类器*C*

2： 每批做x（y）

3： 计算L（C◦F（El（x），Eg（x）），y）

4： 反向传播更新*标高、F、C*

5： 结束

第六章：返回服务器



*卷积和多项式相乘*运算符（）被实现为1×1的卷积-*Fconv公司*

行动，

*Fconv公司*（El（x），Eg（x））=W conv（Eg（x）| | El（x））（一）

其中，Ris是学习的权重矩阵，| |表示沿通道轴连接特征映射的操作。*W变频器*∈2C×C

*多个*算子（）引入学习权向量∈Rand计算局部和全局特征映射之间的加权和，*功能λC*

*功能*（El（x），Eg（x））=λEg（x）+（1−λ）El（x）（二）

其中，首先将加权向量广播成×H×W的形状，然后按元素乘以特征映射，如图2（b）所示。*λC*

*单身*运算符（）使用学习的权重标量，并计算局部和全局特征映射之间的加权和，*F单λ*

*F单*（El（x），Eg（x））=λEg（x）+（1−λ）El（x）（三）

其中全局和局部特征映射分别按和1−λ缩放，然后按元素相加，如图2（c）所示。*λ*

3.2条。基于特征融合机制的联合学习

所提出的FedFusion的典型训练迭代如图1所示。

在回合开始时，我们保留了全局模型（）的特征提取器，而不是像在*我如*

初始化后的FedAvg。

|  |
| --- |
| （a） （b）（c）（d）  图3。在不同的设置下，整体准确度与沟通曲线呈圆形。（a） 和（b）：CIFAR10的人工非IID分区。（c） ：用户特定的非IID分区，通过对MNIST应用不同的置换来实现。（d） ：CIFAR10的IID分区。multi、single和conv都有相应的融合算子，而没有融合算子 |

在训练过程中，被冻结，并附加了一个特征融合模块。介绍了3.1。实际上，表示FedAvg。*如*

通过一轮正向推理可以记录全局特征映射。换言之，附加的特征融合模块带来了有限的额外计算成本。*如*

在设备上训练完成后，将结合特征融合模块的局部模型发送到中心服务器进行模型融合。对于多个和单个操作，我们使用指数移动平均策略来平滑更新。

FedFusion的伪代码如算法1所示。

# 4.    实验

在本节中，我们将首先介绍实验装置（第。4.1），然后显示几种不同设置下的结果（第4.2、4.3和4.4节）。

4.1条。实验装置

数据集集合

我们使用MNIST[8]和CIFAR10[9]作为实验的基本数据集。我们进一步提出了三种类型的数据划分来基准FL算法，例如我们的FedFusion和原始的FedAvg。

第一种是人工非IID划分，它是通过分割现有的IID数据集来实现的，以满足FL场景，在以前的FL研究中常用[1,2,3,6,10]。在这个分区中，单个客户机通常有总数据类的子集。例如，大多数客户机在[3]中最多有两位数的MNIST。

第二种是特定于用户的非IID分区，其中不同客户机上的数据通常具有相似的类，但遵循不同的分布。这通常用于多任务学习研究[11,12,13]。

最后一个是IID划分，这是一个简单但必要的划分来评估FL算法[3,6]。

模型

对于MNIST数字识别任务，我们使用与FedAvg相同的模型[3]：一个CNN具有两个5×5的卷积层（第一层有32个通道，第二个有64个通道，每个表1）。不同设置下FedFusion和FedAvg的收敛精度。（a-d）对应于图3中的值。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | （一） | （二） | （三） | （四） |
| 费达夫 | 89.89 | 92.21 | 95.20 | 80.01 |
| 丰盛+单身 | 89.77 | 92.32 | 95.25 | 80.85 |
| 融合+多重 | 90.51 | 92.78 | 95.43 | 82.95 |
| FedFusion+Conv公司 | 90.11 | 92.53 | 95.79 | 82.15 |

接着是ReLU激活和2×2 max池），一个具有512个单元的完全连接层（具有ReLU激活和随机丢失），以及最终的softmax输出层。

对于CIFAR10，我们使用具有两个5×5卷积层的CNN（两个都有64个通道，每个通道后面都有一个ReLU激活和3×3最大池，步长为2），两个完全连接的层（第一个有384个单元，第二个有192个单元，每个单元都有一个ReLU激活和随机退出）和一个最终的softmax输出层。

4.2条。人工非IID分区

在人工非IID场景下，我们使用3e-3的学习率，每轮指数衰减系数为0.985，我们所有的FedFusion方法（使用不同的融合算子）和比较的FedAvg。它们的收敛行为如图3（a）和（b）所示，而最终的收敛精度如表1所示。

多算子表示融合的曲线总是高于其他算子，这意味着它在较少的通信开销下获得了更高的精度。使用conv的FedFusion的精度在开始时也提高得更快，但没有达到更好的收敛点。FedFusion与single和FedAvg的性能相对较差。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 94% | | 95% | |
| 子弹 | 减少 | 子弹 | 减少 |
| 费达夫 | 100 | （参考） | 256 | （参考） |
| 丰盛+单身 | 87 | 13.0% | 227 | 11.3% |
| 融合+多重 | 78 | 22.0% | 201 | 21.5% |
| FedFusion+Conv公司 | 34 | 66.0% | 92 | 64.1% |

这样的结果显然是由于多重融合算子。如前所述，在人工非IID场景中，大多数客户端都有总类的子集。多操作员允许客户机上的模型选择有助于本地数据的特征映射。相反，FedAvg不提供表2。达到一定精度里程碑的沟通轮数。FedAvg被视为基线，并列出了通信轮数的减少。选择和单一操作员没有提供足够的调整空间。

4.3。用户特定的非IID分区

为了模拟特定于用户的非IID分区，我们对每个客户机上的MNIST应用不同的置换，这就是之前的一些研究中所说的置换MNIST[14,15]。我们使用2e-3的学习率，每轮指数衰减系数为0.99。

达到特定准确度里程碑的通信轮数（此处为94%和95%），以及与FedAvg相比通信轮数的减少，如表2所示。结果表明，与人工非IID分区不同的是，conv的FedFusion在很大程度上导致了边界的改变。使用conv的FedFusion实现了最佳性能，同时将通信轮数减少了60%以上。在特定于用户的非IID分区中，客户机上的数据具有相似的类，但遵循不同的分布。conv算子具有更好的能力来整合来自局部和全局模型的特征映射，换句话说，就是来自其他客户端和数据分布的知识。值得注意的是，用户特定的“非IID分区”更接近于实际的FL场景，因此改进在这种情况下更有意义。

此外，我们还研究了以往外语研究中常被忽视的模型的泛化能力。图4示出了新进入的客户端达到收敛的本地时间。我们可以看到，当一个新的客户端加入一个现有的FL系统时，FedFusion和conv提供了比其他算法更好的初始化，从而加快了收敛过程。

4.4条。IID分区

IID划分是评价FL算法的一个简单而必要的划分。如果一种策略不能处理这种划分，那么它的有效性就值得怀疑。

如图3（d）所示，使用multi和conv的FedFusion以较少的通信成本实现更高的精度。在最终收敛精度方面，采用multi和conv的FedFusion比其他方法有了显著的提高。

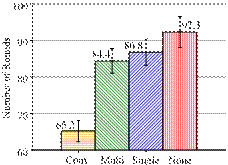


图4。为新进入的客户端达到收敛的本地时间段数。

对特征融合算子做了简要的总结：多算子在局部和全局特征映射之间提供了灵活的选择，并且具有更强的可解释性。权重向量中的条目说明了对应通道中全局特征映射的比例。当数据类中存在空白时，多操作员将学会选择最有用的特征映射。conv算子更善于将全局模型和局部模型的知识进行集成。如果客户机上的数据具有相似的类但遵循不同的分布，conv操作符的性能会更好。实验结果表明，单一算子的改进很小，不应在实际应用中采用。*λ*

# 5.    结论

联合学习的通信开销很大，是一个亟待解决的问题。在本文中，我们试图从减少沟通次数的角度做一些改进。我们提出了一种新的具有特征融合模块的FL算法，并在流行的FL设置中对其进行了评估。实验结果表明，该方法在减少通信次数60%以上的同时，获得了较高的精度。此外，我们发现FedFusion为新进入的客户机提供了更好的通用性。

未来的工作可能包括将我们的算法扩展到更复杂的模型和场景，以及将通信轮数减少策略与其他类型的方法（如梯度估计和压缩）相结合。

# 6.    确认

感谢匿名评论者的宝贵反馈。本课题得到国家重点研发计划（2018YFB1003703）、国家自然科学基金（61521002）和北京网络多媒体重点实验室（Z161100005016051）的资助。

# 7.    参考文献

[1] Jakub Konecnˇy、Brendan McMahan和Daniel Ra-`mage，“联邦优化：数据中心以外的分布式优化”，arXiv预印本arXiv:1511.03575，2015年。

[2] Jakub Konecnˇy，H Brendan McMahan，Felix X Yu，`Peter Richtarik，Ananda Theertha Suresh，和Dave´Bacon，“联合学习：提高沟通效率的策略”，arXiv预印本arXiv:1610.054922016年。

[3] Brendan McMahan、Eider Moore、Daniel Ramage、Seth Hampson和Blaise Aguera y Arcas，“从分散数据中高效学习深层网络”，人工智能与统计，2017年，第1273-1282页。

[4] 恩贞、吴承恩、金贤勋、姬宏公园、梅迪本尼斯和金成莲，

“设备机器学习的通信效率：非iid私有数据下的联邦蒸馏和扩充”，arXiv预印本arXiv:1811.11479年2018年。

[5] 姚欣，黄朝峰，孙立峰，“双流联合学习：降低沟通成本”，视觉通信与图像处理（VCIP），2018年。IEEE，2018年，第1-4页。

[6] Sebastian Caldas，Jakub Konecny，H Brendan McMa-ˇhan和Ameet Talwalkar，“通过减少客户资源需求扩展联合学习的范围”，arXiv预印本arXiv:1812.072102018年。

[7] Ananda Theertha Suresh，Felix X Yu，Sanjiv Kumar和H Brendan McMahan，“有限通信下的分布式平均估计”，第34届机器学习国际会议论文集第70卷。吉姆勒。org，2017年，第3329-3337页。

[8] Yann LeCun、Leon Bottou、Yoshua Bengio和Patrick´Haffner，“基于梯度的学习应用于文档识别”，IEEE论文集，第86卷，第11期，第2278-23241998年。

[9] Alex Krizhevsky和Geoffrey Hinton，“从微小图像中学习多层特征”，技术代表，Citeseer，2009年。

[10] 赵悦，李孟，赖良珍，纳文素达，达蒙·西文和维卡斯·钱德拉，“非iid数据的联合学习”，arXiv预印本arXiv:1806.005822018年。

[11] Virginia Smith，Chao Kai Jiang，Maziar Sanjabi，和Ameet S Talwalkar，“联合多任务学习”，神经信息处理系统进展，2017年，第4424-4434页。

[12] 陈飞，董振华，李振国，何秀强，“联合元学习推荐”，arXiv预印本arXiv:1802.078762018年。

[13] Sebastian Caldas，Peter Wu，Tian Li，Jakub Konecnˇy，`H Brendan McMahan，Virginia Smith和Ameet Talwalkar，“Leaf:联邦设置的基准”，arXiv预印本arXiv:1812.010972018年。

[14] 伊恩·J·古德费罗、梅迪·米尔扎、达肖、亚伦

Courville和Yoshua Bengio，“基于梯度的神经网络中灾难性遗忘的实证研究”，arXiv预印本arXiv：1312.62112013年。

[15] Friedman Zenke、Ben Poole和Surya Ganguli，“通过突触智能进行持续学习”，载于2017年国际机器学习会议，第3987-3995页。