

Both Chinese (Traditional) and Chinese (Simplified) are available below:

繁體中文:

基於預測編碼的聯邦學習為用戶節省 99%通訊開銷

北卡州立研究團隊發表能令頻寬有限的無線設備進行聯合學習的最新技術

即時發布 / FOR IMMEDIATE RELEASE

隨著機器學習在人們工作生活中的日益普及，數據保護已成為當今社會的重要議題。聯邦學習，一種能夠保護數據隱私的機器學習架構，應運而生。聯邦學習的訓練中需要大量數據傳輸，因此該技術難以廣泛地應用在以無線邊緣設備（edge devices）為主的場景中。為了促進聯邦學習在邊緣設備場景中的部署和運用，利用數據壓縮以減少通訊開銷是一種重要的手段。

聯邦學習是一種由多用戶聯合發起的機器學習模式。每位用戶只使用自身的數據集在本地訓練數據模型，然後將完成的模型上傳至一個共享伺服器。伺服器將用戶上傳的模型融合為整體性能更佳模型，隨後將其傳輸至所有用戶。在不斷循環優化的過程中，每次迭代均能提升模型的性能。

“聯邦學習的優勢在於，它可以在保護用戶數據隱私的前提下提昇機器學習的效果。”該論文的合著者，來自北卡羅萊納州立大學（北卡州立）電機及電腦工程（ECE）的助理教授王秋韡博士（Dr. Chau-Wai Wong）表示，“舉個例子，你可以使用來自多家醫院的患者數據來提升智慧醫療系統的性能，而個體醫院之間無須共享患者數據，極大程度地保障了數據安全。”

許多基於聯邦學習的任務可以透過例如智能手機之類的邊緣設備完成。因此，聯邦學能夠在不犧牲用戶隱私的前提下利用來自邊緣設備的數據。當然，這個方案有其自身瓶頸：模型更新依賴於伺服器與用戶之間的循環傳輸，其通訊開銷十分高昂。假如頻寬受限或者數據量巨大，用戶與伺服器之間的無線通訊可能不暢順，進而拖慢整個訓練過程。

“我們從數十年來的視訊壓縮領域的研究成果中獲得啟發，為聯邦學習專門設計了一種數據壓縮方法，以加快其無線數據傳輸的速度。”王教授在採訪中表示。

王教授及他的團隊設計的方法能夠有效地壓縮用戶端產生的與模型更新有關的數據。在收到這些壓縮數據包之後，伺服器能對其解壓。為了提高壓縮率，研究團隊選擇使用預測編碼 (predictive coding) 和多模式選擇 (mode selection) 技術。這些技術能夠將與模型更新有關的數據量壓縮到原先的 1%。由伺服器傳輸至用戶端的數據則因頻寬充足而無須壓縮。

“透過我們的方法，使用頻寬有限的邊緣設備也能實現聯邦學習。”該論文的第一作者，北卡州立的博士研究生岳鏜 (Kai Yue) 表示。“許多基於人機交互應用的性能將得到提升，例如語音助手。”

該論文 “[Communication-Efficient Federated Learning via Predictive Coding](#)” 刊登於 IEEE 權威學術期刊 IEEE Journal of Selected Topics in Signal Processing。論文的另外兩位作者分別為北卡州立 ECE 教授戴懷宇 (Prof. Huaiyu Dai) 和北卡州立 ECE 博士後金日成 (Dr. Richeng Jin)。該項目的部分資助來自美國國家科學基金會 National Science Foundation (資助編號 1824518)。

简体中文:

基于预测编码的联邦学习为用户节省 99%通信开销

北卡州立研究团队发表能令带宽有限的无线设备进行联合学习的最新技术

即时发布 / FOR IMMEDIATE RELEASE

随着机器学习在人们工作生活中的日益普及，数据保护已成为当今社会的重要议题。联邦学习，一种能够保护数据隐私的机器学习架构，应运而生。联邦学习的训练中需要大量数据传输，因此该技术难以广泛地应用在以无线边缘设备 (edge devices) 为主的场景中。为了促进联邦学习在边缘设备场景中的部署和运用，利用数据压缩以减少通信开销是一种重要的手段。

联邦学习是一种由多用户联合发起的机器学习模式。每位用户只使用自身的数据集在本地训练数据模型，然后将完成的模型上传至一个共享服务器。服务器将用户上传的模型融合为整体性能更佳模型，随后将其传输至所有用户。在不断循环优化的过程中，每次迭代均能提升模型的性能。

“联邦学习的优势在于，它可以在保护用户数据隐私的前提下提升机器学习的效果。”该论文的合著者，来自北卡罗来纳州立大学（北卡州立）电气和计算机工程（ECE）的助理教授王秋韡博士（Dr. Chau-Wai Wong）表示，“举个例子，你可以使用来自多家医院的患者数据来提升智慧医疗系统的性能，而个体医院之间无须共享患者数据，极大程度地保障了数据安全。”

许多基于联邦学习的任务可以通过例如智能手机之类的边缘设备完成。因此，联邦学能够在不牺牲用户隐私的前提下利用来自边缘设备的数据。当然，这个方案有其自身瓶颈：模型更新依赖于服务器与用户之间的循环传输，其通信开销十分高昂。假如带宽受限或者数据量巨大，用户与服务器之间的无线通信可能不畅顺，进而拖慢整个训练过程。

“我们从数十年来的视频压缩领域的研究成果中获得启发，为联邦学习专门设计了一种数据压缩方法，以加快其无线数据传输的速度。”王教授在采访中表示。

王教授及他的团队设计的方法能够有效地压缩用户端产生的与模型更新有关的数据。在收到这些压缩数据包之后，服务器能对其解压。为了提高压缩率，研究团队选择使用预测编码（predictive coding）和多模式选择（mode selection）技术。这些技术能够将与模型更新有关的数据量压缩到原先的 1%。由服务器传输至用户端的数据则因带宽充足而无须压缩。

“通过我们的方法，使用带宽有限的边缘设备也能实现联邦学习。”该论文的第一作者，北卡州立的博士研究生岳锴（Kai Yue）表示。“许多基于人机交互应用的性能将得到提升，例如语音助手。”

该论文“[Communication-Efficient Federated Learning via Predictive Coding](#)”刊登于 IEEE 权威学术期刊 IEEE Journal of Selected Topics in Signal Processing。论文的另外两位作者分别为北卡州立 ECE 教授戴怀宇（Prof. Huaiyu Dai）和北卡州立 ECE 博士后金日成（Dr. Richeng Jin）。该项目的部分资助来自美国国家科学基金会 National Science Foundation（资助编号 1824518）。

-shipman-

編輯鈞鑒：附該研究摘要。 / **编辑钧鉴**：附该研究摘要。 / **Note to Editors**: The study abstract follows.

《透過預測編碼提升聯邦學習的通訊效能》 / 《通过预测编码提升联邦学习的通信效能》 / “Communication-Efficient Federated Learning via Predictive Coding”

作者: 岳鏜/岳锜 (Kai Yue)、金日成/金日成 (Richeng Jin)、王秋韡/王秋韡 (Chau-Wai Wong) 以及戴懷宇/戴怀宇 (Huaiyu Dai)，北卡羅萊納州立大學/北卡罗来纳州立大学 (North Carolina State University)

發表於/发表于: Jan. 13, 2022, IEEE Journal of Selected Topics in Signal Processing

DOI: 10.1109/JSTSP.2022.3142678

摘要: [繁体] 透過聯邦學習，用戶能夠在將數據限制在本地的前提下聯合其他用戶共同訓練模型。在無線移動設備的應用場景下，因為功率和頻寬的限制，通訊開銷往往是聯邦學習的瓶頸。以往的工作使用諸如量化和稀疏化的數據壓縮方法來減少開銷。本文提出一種基於預測編碼的聯邦學習方案。該方案在所有用戶間共享預測函數，並讓每個用戶壓縮和傳輸基於參考訊號的殘差向量。在每輪通訊中，該方案基於數據率—失真的成本選擇預測函數和量化函數，並透過熵編碼進一步降低冗餘。相較其他方案，該方案不僅能減少 99% 的通訊成本，而且能得到更優的模型。

[简体] 通过联邦学习，用户能够在将数据限制在本地的前提下联合其他用户共同训练模型。在无线移动设备的应用场景下，因为功率和带宽的限制，通信开销往往是联邦学习的瓶颈。以往的工作使用诸如量化和稀疏化的数据压缩方法来减少开销。本文提出一种基于预测编码的联邦学习方案。该方案在所有用户间共享预测函数，并让每个用户压缩和传输基于参考信号的残差向量。在每轮通信中，该方案基于数据率—失真的成本选择预测函数和量化函数，并通过熵编码进一步降低冗余。相较其他方案，该方案不仅能减少 99% 的通信成本，而且能得到更优的模型。