

# 联邦无监督表示学习

张凤达<sup>1</sup>, 况琨<sup>1</sup>, 陈隆<sup>1</sup>, 游兆阳<sup>1</sup>, 沈弢<sup>1</sup>, 肖俊<sup>1</sup>,  
张寅<sup>1</sup>, 吴超<sup>2</sup>, 吴飞<sup>1</sup>, 庄越挺<sup>1</sup>, 李晓林<sup>3,4,5</sup>

<sup>1</sup>浙江大学计算机科学与技术学院, 中国杭州市, 310027

<sup>2</sup>浙江大学公共管理学院, 中国杭州市, 310027

<sup>3</sup>同盾科技, 中国杭州市, 310000

<sup>4</sup>中国科学院基础医学与肿瘤研究所, 中国杭州市, 310018

<sup>5</sup>杭州灵思智康科技有限公司, 中国杭州市, 310018

**摘要:** 为利用分布式边缘设备上大量未标记数据, 我们在联邦学习中提出一个称为联邦无监督表示学习(FURL)的新问题, 以在没有监督的情况下学习通用表示模型, 同时保护数据隐私。FURL提出了两个新挑战: (1) 客户端之间的数据分布转移(非独立同分布)会使本地模型专注于不同的类别, 从而导致表示空间的不一致; (2) 如果FURL中客户端之间没有统一的信息, 客户端之间的表示就会错位。为了应对这些挑战, 我们提出带字典和对齐的联合对比平均(FedCA)算法。FedCA由两个关键模块组成: 字典模块, 用于聚合来自每个客户端的样本表示并与所有客户端共享, 以实现表示空间的一致性; 对齐模块, 用于将每个客户端的表示与基于公共数据训练的基础模型对齐。我们采用对比方法进行局部模型训练, 通过在3个数据集上独立同分布和非独立同分布设定下的大量实验, 我们证明FedCA以显著的优势优于所有基线方法。

**关键词:** 联邦学习; 无监督学习; 表示学习; 对比学习

<https://doi.org/10.1631/FITEE.2200268>