**基于联邦学习的睡眠质量预测**

题目

基于联邦机器学习的睡眠质量预测

组员

|  |  |
| --- | --- |
| 姓名 | 学号 |
| 宣朋羽 | 202021080520 |
| 蒋育韬 | 202022080503 |
| 沙泽鑫 | 202022080521 |
| 王亦明 | 202022080502 |
| 张奔 | 202022080531 |

简介

在智能穿戴设备普及的今天，检测睡眠数据轻而易举，并且在智能手表相对强大算力的加持下，可以利用机器学习进行睡眠质量预测，结合多个智能设备的数据进行分布式学习还能进一步提高预测能力。然而个人生理数据极度敏感，为了解决数据隐私问题，我们引入联邦机器学习以在保护数据隐私的前提下提高睡眠质量预测的效果。

开题报告

研究的问题是什么

问题的重要性和意义

前人工作存在的问题

你的解决方案的关键思路

数据集、评价标准

组员名单

主题：应用

研究的问题

1（200王）：联邦机器学习在睡眠质量预测上的应用。主题是应用。参考“简介”。

重要性和意义

2（200蒋）：睡眠质量预测的重要性和意义。

3（400沙）：分布式和隐私保护的意义。分布式：相对于本地机器学习，提高模型质量。保护隐私：伦理、政策、法规，隐私与个人利益、正常生活息息相关，避免隐私泄露就是保护个人权益，避免大数据杀熟。

存在的问题

4（200张）：问题大概如下，分散数据，集中计算：直接上传数据，违反法规，可能泄露。本地学习：数据集太小，模型质量受限，质量不高还会增加训练轮次，增加计算量。

解决方案

5（300宣）：联邦学习，差分隐私

数据集，评价指标

6（100宣）：https://www.kaggle.com/danagerous/sleep-data

MSE，相对于本地模型的额外开销。

参考文献

总共约15个

基于Spark的分布式健康大数据分析系统设计与实现（用了同一个数据集的论文，它是简单分布式，没有隐私保护）

基于联邦学习的教育数据挖掘隐私保护技术探索

联邦学习中的隐私保护研究进展

智能家居中的边缘计算