

隐私保护下的车辆轨迹聚类方法研究

田冉，高普伦，刘颜星

西北师范大学计算机科学与工程学院，中国兰州市，730070

摘要：车辆轨迹聚类方法作为时空交通数据挖掘的重要工具之一，被广泛地运用于挖掘车辆的行为模式。但是如果直接将原始车辆轨迹数据上传到服务器并进行聚类则具有隐私泄露的风险。因此，针对在保护用户隐私的同时进行车辆轨迹聚类这一挑战，本文提出一种隐私保护下的车辆轨迹聚类框架，并构建了基于变分自编码器和改进K-means算法的车辆轨迹聚类模型（IKV）。在该框架中，客户端计算出车辆轨迹的隐藏变量并上传至服务端，服务端再利用隐藏变量计算出聚类结果，并将分析结果发放给各客户端。IKV的工作流程如下：首先利用历史车辆轨迹数据在变分自编码器上进行训练，当变分自编码器的解码器接近原始数据时，将变分自编码器的编码器部署到车辆端的边缘计算设备；其次由车辆端的编码器计算车辆本地轨迹数据的隐变量，并将隐变量上传至服务器端；最后在服务器端利用改进的K-means算法基于车辆上传的隐变量进行车辆轨迹聚类计算。在三个真实数据集上，我们将IKV与多种基准方法在多个评价指标上进行了比较。在9组性能对比实验中，IKV算法在6组中达到了最优或接近最优的性能水平。同时在9组敏感性分析实验中，该算法在7组实验中展现出了显著的稳定性和对参数变化的鲁棒性。这些结果验证了本文提出的框架不仅适用于注重隐私的生产环境，如拼车服务，而且由于其对聚簇数量的低敏感性，还能适应不同规模的聚类任务。

关键词：隐私保护；变分自编码器；改进K-means；车辆轨迹聚类

<https://doi.org/10.1631/FITEE.2300369>