

基于 Spark 面向分布式 EEMDN-SABiGRU 模型的乘客热点预测

夏大文¹, 耿建¹, 黄瑞曦¹, 申冰琪¹, 胡杨², 李艳涛³, 李华青⁴

¹ 贵州民族大学数据科学与信息工程学院, 中国贵阳市, 550025

² 贵州交通技师学院汽车工程系, 中国贵阳市, 550008

³ 重庆大学计算机科学学院, 中国重庆市, 400044

⁴ 西南大学电子与信息工程学院, 中国重庆市, 400715

摘要: 针对出租车与乘客之间的供需不平衡问题, 本文提出一种基于Spark的分布式归一化集合经验模态分解和面向空间注意力机制的双向门控循环单元(EEMDN-SABiGRU)模型, 实现乘客热点的精准预测, 旨在降低盲目巡航开支、提高载客效率和实现收益最大化。首先, 提出一种归一化的集合经验模态分解方法(EEMDN), 处理网格中乘客热点数据, 解决非平稳序列问题和数值差异过大造成的预测精度下降问题, 避免EMD本征模态函数(IMF)存在的模态混叠现象。其次, 构建一种基于乘客上下车热点的权重和乘客的空间规律性的空间注意力机制, 捕捉每个网格中的乘客热点特征。再次, 融合一种双向门控循环单元(GRU)算法, 解决GRU仅能获取前向信息而忽略后向信息问题, 提高特征提取的准确性。最后, 在Spark并行计算框架下, 采用真实的出租车GPS轨迹数据, 基于EEMDN-SABiGRU模型实现了乘客热点的准确预测。实验结果表明, 在00网格4个数据集上, 与LSTM、EMDL-STM、EEMD-LSTM、GRU、EMD-GRU、EEMD-GRU、EMDN-GRU、CNN和BP相比, EEMDN-SABiGRU的平均绝对百分比误差、平均绝对误差、均方根误差和最大误差值分别降低了43.18%、44.91%、55.04%和39.33%。

关键词: 乘客热点预测; 集合经验模态分解(EEMD); 空间注意力机制; 双向门控循环单元(BiGRU); GPS轨迹; Spark

<https://doi.org/10.1631/FITEE.2200621>