

文章编号: 2096-1472(2017)-04-15-03

基于GPS轨迹数据的公交到站时间预测方法研究

李少伟¹, 曹成涛¹, 杨 骥²

(1. 广东交通职业技术学院智能交通工程技术应用中心, 广东 广州 510650;

2. 广州地理研究所, 广东 广州 510070)

摘要: 公交到站时间预测作为智能公交的核心内容之一, 不仅可以极大地提高公交服务满意度进而提高公交出行率, 而且可为智能公交调度系统的优化提供数据依据。本文将公交GPS数据和公交站点与线路等GIS数据进行融合、处理与分析, 基于公交车轨迹数据和路段在空间关系和时空分布上的分析和处理, 设置一定的离站距离阈值对当前公交站点进行缓冲区分析, 搜索在最近一段时间内停靠该站点的公交车轨迹数据, 获取不同车辆的轨迹信息, 进行公交到站时间预测研究, 建立预测模型, 推算出公交到站时间, 并以经过广州市体育中心公交站的多条公交线路为例验证预测方法的准确性。研究表明:通过模型预测所得多条公交线路的公交到站时间与实际调查记录时间误差较小。

关键词: GPS轨迹数据; 公交到站时间; 缓冲区分析; 数据融合; GIS

中图分类号: TP274 **文献标识码:** A

Research on the Bus Arrival Time Prediction Method Based on GPS Track Data

LI Shaowei¹, CAO Chentao¹, YANG Ji²

(1. Guangdong Communication Polytechnic, Intelligent Traffic Engineering Technology Application Center, Guangzhou 510650, China;

2. Guangzhou Institute of Geography, Guangzhou 510070, China)

Abstract: As one of the core contents of intelligent transit, the bus arrival prediction can not only increase the satisfaction of bus service and the trip rate of public transport, but also provide data basis for the optimization of the intelligent public transportation dispatching system. In this paper, the GPS data of the bus and the GIS data of the bus station and the bus route are integrated, processed and analyzed. Based on the analysis and processing of the bus track data and the road segment in the spatial relation and the spatial and temporal distribution, a certain distance threshold from the station is set to carry out the buffer analysis, search the bus track data of the station in the last period of time, obtain the track information of different vehicles, predict the bus arrival duration, establish the prediction model and calculate the bus arrival time. Multiple bus routes of Guangzhou Sports Center are taken as cases to verify the accuracy of the prediction method. The results show that the time errors between the prediction values and the actual records of multiple bus routes are suitably small.

Keywords: GPS track data; bus arrival time; buffer analysis; data fusion; GIS

1 引言(Introduction)

随着中国城镇化进程的不断加快, 大城市规模和人口不断扩大, 城市居民的出行需求不断增长, 交通拥堵、环境污染及过度能源消耗等负面效应日益突出, 成为经济和社会协调发展的热点问题, 大力发展公共交通不仅可以提升交通出行效率, 而且可以减轻交通拥挤, 减少环境污染, 合理利用资源, 促进社会公平, 实现低碳绿色交通, 但我国城市公共交通在准时性、便捷性等方面不能满足乘客的需求, 导致公共交通吸引率长期偏低, 因此积极开展公交到站时间预测方法的研究, 已成为改善交通压力的重要研究方向: 向出行者

提供公交到站时间信息可以极大地提高公交服务满意度, 从而提高公交出行率; 向管理者提供公交到站时间信息可为公交线路及站点的调整提供依据, 为智能公交系统的实时调度优化提供基础数据。

2 常见预测方法综述(Summary of common prediction methods)

根据预测方法不同, 公交到站时间预测方法主要分为四类: (1) 基于时空变化规律的预测方法: 根据公交行程时间在时间与空间上的变化规律, 遵循已有的历史数据对车辆在未来时间里经过某一路段的行程时间进行预测, 目前这种预

测方法主要有时间序列法、卡尔曼滤波法^[1]、神经网络法^[2]、支持向量机^[3,4]等。(2)基于影响因素的预测方法：通过建立公交行程时间与路段长度、天气情况、交通流量、出行时间等因素之间的数学模型来进行公交到站时间预测，目前主要包括统计回归理论预测方法与数学模型预测方法^[5-7]。(3)基于数据融合的预测方法：通过对多源检测器采集的数据进行综合处理，得到更全面准确的数掘，或结合多种预测方法的结果，提高预测精度、保证预测结果稳定性，目前数据融合方法主要有加权平均法、贝叶斯方法^[8,9]、统计决策理论^[10,11]、神经网络法^[12]等。(4)混合模型预测方法：由于预测算法的各自特点，各种预测模型在展现良好预测性能的同时也存在固有缺点，所以近年来，公交到站时间预测研究朝着多种方案相结合的混合预测模型方向发展。

3 基于GPS轨迹数据的公交到站时间预测(Bus arrival time prediction based on GPStrack data)

随着智能公交系统的快速发展，基于GPS的自动定位系统、GIS技术、无线通信技术得到了更加广泛的使用，利用公交GPS轨迹数据进行数据挖掘，已经成为一种预测公交到站时间的新方法。

3.1 GPS数据处理与融合分析

本文实验数据来源于广州市公交GPS数据，公交车在回传数据时，受到特殊地形、恶劣天气的影响，或是无线网络通信质量不佳，都会对回传数据的质量造成影响，所以在数据融合分析之前需要对数据进行预处理以保障数据的精确性，从而避免对预测造成不必要的负面影响，主要采取两个步骤：(1)GPS数据清理，去除漂移点。轨迹数据的经纬度坐标是最重要的基础数据，但由于模块定位过程中受到大型建筑物的遮挡，以及其他事物的干扰，会出现很多漂移点而对于严重偏离所属线路的情况，会对预测精度造成重大影响，因此必须去除。本文通过叠加GPS轨迹点和公交线路图层，在地图上非常直观地表现出了车辆行驶轨迹偏离所属线路的程度，通过设置可变的阈值，控制轨迹数据中坐标的精度。(2)GPS数据丢失插值处理。由于城市中高楼耸立，GPS数据传输过程中会出现丢失的问题，造成GPS定位时间间隔过久，直接影响提取出交通数据的准确性。本文采取插值处理的方法补充空白点，运用前后两个数据点的平均值作为插入数值。

数据融合分析是将车辆GPS定位数据、公交车辆数据、公交线路数据、公交站点数据、公交线路站点关系、运营计划数据等进行融合，通过GPS空间信息和站点空间信息进行空间分析，结合融合后的数据获得公交车GPS的时空轨迹数据。

具体关联如图1所示。

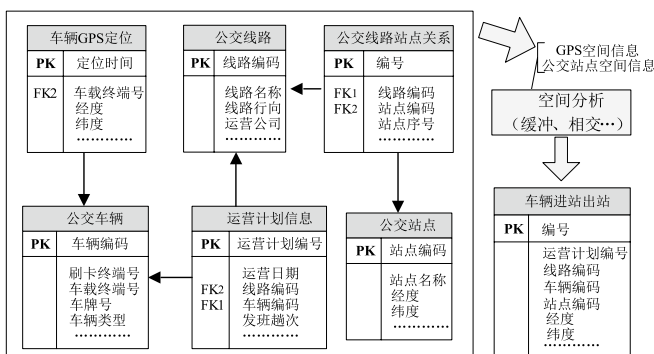


图1 融合数据各数据表之间的关联关系

Fig.1 Association of seven data sources

3.2 公交车停站时刻获取

为了获取更准确的公交车停站时刻，需要充分了解公交车运行规律：公交从始发站开往终点站，中途经过多个停靠站，公交车总是从站台的上游减速进入停靠站，然后在站内停靠片刻等待乘客上下车，然后起步加速离站，进入站台的下游路段。公交车经过每一个站台都要经历减速进站和加速离站的过程。

根据公交站点与公交车空间关系和时空分布特征，设置一定的离站距离阈值对公交站点进行缓冲区分析，搜索在最近一段时间内停靠该站点的公交车轨迹数据，获取不同车辆的轨迹信息，根据公交车和停靠站点的坐标信息，可得到当前时刻公交车离站距离，将当前线路公交车速度、离站距离及时间信息进行叠加，可以得到公交车停站过程图。

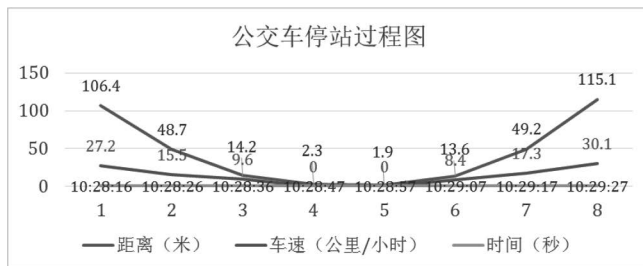


图2 公交车停站过程图

Fig.2 Bus stop process chart

以广州体育中心公交车站为例，测得某路公交车到站的车速和离站距离随时间变化规律如图2所示，由图可知，公交车先减速到站后再加速离站，离站距离由大变小再由小变大，由于车辆静止时GPS接收器计算精度和外界多径效应引入的干扰，位置数据不是同一值而是小范围值，因此可以将公交车停站时的最小离站距离范围设置在GPS定位误差范围内，将该范围内第一个最小车速(车速的计算依赖于GPS定位数据，由于GPS的定位误差，公交车停站时的最小车速值应是接近0的小范围值)作为停站时的车速，即取距离和车速曲线左侧

底部对应的时刻作为公交车停站时刻的替代值！

假设当前某个时间段*T*内通过公交站点*Z_i*(*i*为公交车下一站停靠的站点序号, *m*为乘客候车站点序号, *i*、*m*均为正整数且根据当前公交线路经过站点先后顺序排序, $1 < i < m$)的公交车为*C_j*(*n*均为正整数, *n*表示该时间段内当前公交站点以一定离站距离进行缓冲区分析得到的车辆总数, *j*为对应公交车辆编码的排序号, $1 \leq j \leq n$), 则当前公交车的轨迹为*G_{ij}*(公交车辆编码是唯一值, 因此*G_{ij}*具有唯一性), 从轨迹曲线里取距离和车速的左侧底部对应的时刻*t_{ij}*作为公交车停站时刻的替代值。

3.3 站间行驶时间预测

鉴于同一时间段有多辆不同线路公交车经过该站点, 通过上述方法获取该车辆的公交停站时刻替代值, 则相邻站点间的时间间隔等于同一辆车下一站点的公交停站时刻替代值减去上一站点的公交停站时刻替代值, 由于同一时间段有多辆公交车经过相邻站点, 因此同一时间段的站间行驶时间是一个集合, 取集合的算术平均值, 即可得到相邻站点间的行驶预测时间。

假设当前某个时间段*T*内通过相邻公交站点*Z_i*和*Z_(i+1)*的公交车*C_j*的站间行驶时间为*S_{ij}*:

$$S_{ij} = t_{(i+1)j} - t_{ij} \quad 1 < i < m \quad (1)$$

则该时间段内相邻公交站点*Z_i*和*Z_(i+1)*的站间行驶预测时间为*S_i*:

$$S_i = \sum_{j=1}^n s_{ij} \times \frac{1}{j} \quad 1 \leq j \leq n \quad (2)$$

3.4 公交到站时间预测

将目标车辆距离当前站点的到站时间进行分拆: 目标车辆距离最近站点的到站时间及后续各相邻车站的站间行驶时间总和, 即为公交到站时间, 公交到站时间预测示意图如图3所示。



图3 公交到站时间预测示意图

Fig.3 Sketch map of bus arrival time prediction

假设公交车距离下一站点的距离为*d*(由公交车坐标和站

点坐标计算得到), 公交车所在站点区间距离为*D*(由相邻站点坐标计算得到), 公交车到达乘客候车站点的时间为*Y*, 已知当前公交车下一站点为*Z_i*, 乘客候车站点为*Z_m*, 则公交到站时间预测值*Y*:

$$Y = S_{(i-1)} \times \frac{d}{D} + S_i + S_{(i+1)} \dots S_{(m-1)} \quad 1 < i < m \quad (3)$$

4 案例分析(Case analysis)

本文以广州市体育中心站为例, 将体育中心站作为候车点, 通过实际调查和预测模型计算出相应的公交到站时间, 具体实施方法如下: (1)采用定点调查方式: 选择体育中心站, 在周末下午时间做公交到站时间调查, 记录多趟公交线路到站时间。(2)通过公交线路, 查询距离站点最近公交车, 从数据库中提取出相关车辆的公交车轨迹数据: 位置、途径站点、速度、时间等相关信息。(3)利用本文提出的预测模型对提取的公交车轨迹数据进行处理、分析, 推算出每趟公交线路的公交到站时间。

实验结果如下: 针对体育中心站到站的车辆做了12份记录, 得到公交线路及其到站时间记录12条; 通过推算模型计算出的相同公交车的到站记录12条, 用累计误差除以到站累计时间, 所得准确率为 $1 - (30/149) = 79.87\%$, 取得较好的预测效果, 具体情况如表1所示。

表1 推算结果与实际情况对比图

Tab.1 Comparison between the calculated results and the actual situation

公交线路	130	233	245	263	280	540	545	551	78	89	810	B25	时间累计
到站时间	16	10	11	15	17	10	9	14	8	13	12	14	149
预测时间	19	14	12	18	19	13	8	11	10	15	16	12	167
时间误差	3	4	1	3	2	3	1	3	2	2	4	2	30

5 结论(Conclusion)

为改善城市公共交通在准时性、便捷性等方面的不足, 提升公共交通吸引力, 本文将公交GPS数据和公交站点与线路等GIS数据进行融合、处理与分析, 基于公交车轨迹数据和路段在空间关系和时空分布上的分析和处理, 进行公交到站时间预测研究, 建立模型, 推算出公交到站时间, 弥补了现有公交数据记录的不足。以广州市体育中心公交站为例对推算方法进行验证, 所得公交到站时间的准确率79.87%, 取得较好的预测效果, 具有一定的实用价值。公交到站时间预测提

(下转第8页)