

多线路信息融合的公交车行程时间预测算法

马林宏 陈廷伟 郝明 张雷

(辽宁大学信息学院 沈阳 110036)

摘要 针对公交车行程时间预测存在数据稀疏、数据缺失及更新间隔长等问题,提出了一种基于相似路段划分并融合多线路信息的卡尔曼滤波算法。该算法对每条路段的属性特征和空间结构特征进行归一化处理,利用属性特征和空间结构的相似性及 POI(Point of Interest)对交通影响的变化动态地划分相似路段;然后融合相似路段与目标路段上的多条公交线路的数据信息,用相似路段的数据丰富实验数据;最后结合卡尔曼滤波算法动态性高、实时性强等特点建立模型,从而实现短时预测,并对信息进行修正。选取沈阳市 162 线路和 299 线路作为实验线路,各划取一段相似路段进行基础数据采集并进行实验。通过相似路段上的信息来推断数据稀疏或缺失路段的信息,能够缩短数据更新间隔并提高算法预测的实时性及精准性,尤其在早高峰时段,提出的算法模型的绝对平均百分误差达到 13.2%,能达到实时查询的性能需求。

关键词 行程时间,相似路段,卡尔曼滤波,多线路信息,行程时间预测

中图分类号 TP39 **文献标识码** A **DOI** 10.11896/jsjcx.180901764

Bus Travel Time Prediction Algorithm Based on Multi-line Information Fusion

MA Lin-hong CHEN Ting-wei HAO Ming ZHANG Lei

(College of Information Science, Liaoning University, Shenyang 110036, China)

Abstract Aiming at the problems of bus travel time prediction, such as sparse data, lack of data and long update interval, this paper proposed a Kalman filter algorithm based on similar section segmentation and fusion of multi-line information. In this method, the attribute features and spatial structure features of each road segment are normalized, the similar road segments are dynamically divided by using the similarity between the attribute features and the spatial structure and the change of the traffic impact of the POI. Then, the data information of multiple bus lines on similar road segments and target road segments are integrated, and the experimental data are enriched by using the data from similar road segments. Finally, combining the dynamic and real-time characteristics of Kalman filtering algorithm, the model is established to realize short-term prediction and correct the information. In the experiment, 162 lines and 299 lines in Shenyang City were selected as experimental lines, and a similar section was taken for basic data collection and experiments. The information on the similar road sections is used to infer sparse information or missing road sections, thereby shortening the data update interval and improving the real-time performance and accuracy of the algorithm prediction. Especially in the early peak period, the absolute average percentage error of the proposed model reaches 13.2%, which can effectively meet the performance requirements of real-time query.

Keywords Travel time, Similar road, Kalman filter, Multi-line information, Travel time prediction

1 引言

城市交通是智能交通的研究重点,很多学者在智能交通系统方面进行了大量研究。智能手机的普及^[1]更推动了智能交通的发展,人们在预备出行或乘坐公交的过程中,越发地关注行程时间预测的实时性和可靠性。例如,乘客在乘坐公交车时,可以通过手机 APP 或其他公交车设备查看公交车到达目的地的时间,以做出适当的选择。为满足人们的这一需求,

各学者们研究建立了多种模型^[2],从静态的预测模型逐渐转向了根据 GPS、AVL 等建立更具适应性的动态预测模型^[3]。

目前,在这个领域国内外有着较丰富的研究成果。Dhivyabharathi 等用粒子滤波技术^[4]预测公交旅行时间,证明粒子滤波器是预测公交旅行时间的可行工具;Chen 等依据车辆定位数据建立了基于 Link 和 Section 的公交车行驶时间预测模型^[5],实验表明在工作日时段 Section 的模型较好,在非工作日时段两个模型的预测较为相似;Stephen 等基于最小

收到日期:2018-09-18 返修日期:2019-02-25 本文受国家自然科学基金(61174115)资助。

马林宏(1992-),女,硕士生,主要研究方向为智能交通、轨迹数据;陈廷伟(1974-),男,博士,教授,CCF 会员,主要研究方向为智能交通、轨迹大数据,E-mail:twchen@lnu.edu.cn(通信作者);郝明(1992-),男,硕士生,主要研究方向为智能交通;张雷(1993-),男,硕士生,主要研究方向为轨迹数据。

二乘法的多元回归分析提出了 TBST 模型^[6],推断出 TBST 模型可用于改善人口稠密地区的调度和公交系统规划;基于 MapReduce 聚类和神经网络的思想,Xie 等^[7]提出的预测模型提高了预测精度和预测速度;基于路况相似性,Sun 等提出了一种公交车到站时间机制^[8],提高了预测结果的精度。

尽管城市交通领域取得了较多研究成果,但仍存在一些需要改进之处。有的模型在交通拥堵时预测精度下降,比较适用于历史数据波动较小的情况;有的模型预测的时间较长,对输入因素的独立性要求较高,在实际应用时存在较大的难度,当某些路段存在数据稀疏甚至数据缺失的问题时,便难以对行程时间进行估计。鉴于上述问题,本文提出多线路信息融合的卡尔曼滤波算法预测模型。其利用路网相似性动态地划分相似路段,将相似路段与目标路段上的多条公交线路的数据信息融合,并采用运行速度较快的卡尔曼滤波进行短时预测。

2 公交线路的相似路段划分

在公交车行程时间预测的研究中,已有研究多以单条线路为实验路段,实验数据至少间隔相邻车辆的发车时间,不利于数据更新及实时传输,且车辆 GPS 采集存在数据稀疏或数据缺失的问题,因此本文提出了相似路段的概念。每条道路的交通情况会因 POI^[10-11]活跃度的不同而不同,因此在郑业晴等^[9]提出的基于路网相似性的路段行程时间估计的基础上,本文利用路网间属性和空间结构的相似性,考虑了 POI 的活跃度来动态地划分相似路段,将相似路段上采集到的多条线路公交车行驶信息作为卡尔曼滤波算法的基础数据,且相似路段的划分具有动态性,当行驶中出现临时堵车或数据缺失等状况时可以参考 POI、道路宽度等因素,及时匹配相似路段,将相似路段与目标路段上的多条公交线路的数据信息融合,用相似路段上的信息来推断数据稀疏或缺失路段的信息,即将相似路段上的行驶时速或行程时间等信息及时传输给预测模型,从而缩短更新间隔,提高算法预测的实时性和精准性。

2.1 相似特征的衡量因素

公交车在道路上的行驶速度与道路的通行能力密切相关,道路的不同、交通条件的差异及交通外环境的好坏等都影响着道路的通行能力。在衡量路网相似性特征的过程中,属性特征的表示符号如下: Len 表示站间距离长度; Euc 表示站间两端间的欧氏距离; Cur 表示站间路段的弯曲度, $Cur = Len/Euc$; Dir 表示站间方向; Cro 表示道路横断面分为单幅、双幅、三幅、四幅; Gra 表示道路等级,其分为快速路、主干路、次干路、支路; Wid 表示车道宽度,其分为直线单行 3.0~3.5、直线双行 5.5~7.0、曲线单行 3.8~5.0、曲线双行 7.0~10.0。

结构特征的表示符号如下: $SecIn$ 表示站间入度; $SecOut$ 表示站间出度; POI 表示 POI 影响度。

基于 Esri 的 POI 分类标准,现阶段可将 POI 类型分为餐饮、购物、企事业单位等 17 类,且不同时段不同类别的 POI 活跃度对道路交通有不同的影响,POI 影响度可以根据式(1)计算:

$$ImpDeg = \frac{C_{p_i}}{\sum_{i=1}^n C_{p_i}} * W_{p_i} \quad (1)$$

其中, p_i 表示 POI 类别; C_{p_i} 表示 i 类 POI 的数量; W_{p_i} 表示 i 类 POI 对交通影响的权重; S 表示缓冲区面积。

2.2 相似性度量

本文采用相似度量中的 Jaccard 系数^[12-14]来计算相似程度。Jaccard 系数的公式如下:

$$\Gamma(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (2)$$

其中, $\Gamma(A, B) \in [0, 1]$ 。

为方便基于该方法的实验验证,将每条路段的属性特征和空间结构特征进行归一化处理,即对路段属性信息和空间信息分别进行数值化,将路段对应的路况集合用一维向量表示。 $AttVector$ 表示属性向量符号, $SpaVector$ 表示空间结构向量符号。分别计算属性相似度($AttSim$)和空间结构相似度($StrSim$):

$$AttSim(AttVector(1), AttVector(2)) = J(AttVector(1), AttVector(2))$$

$$StrSim(SpaVector(1), SpaVector(2)) = J(SpaVector(1), SpaVector(2))$$

两路段的杰卡德系数越高,路段通行的能力就越相似,当系数为 1 时,认为两个路段具有等效的通行能力。

在判断路段相似性时,如果单一地判断其中的一种相似性,会使判断的结果不准确,因此要兼顾属性特征和空间结构。在判断相似度的过程中,需要定义一个鲁棒的阈值来衡量 $AttSim$ 或 $StrSim$ 对路段相似度的重要性,然而受到诸多条件的限制, $AttSim$ 和 $StrSim$ 之间的关系会随着路网环境的改变而改变,致使这个关系不稳定且不能在任何情况之下都成立,因此基于同步的方法来判断路段的相似性的难度较大。本文采用分步处理的方法来判断路段之间的相似性,即先对路段进行属性特征上的相似性判断,经过初步筛选后,再对筛选后的路段进行空间结构上的相似性判断,用 $Sim(A-S)$ 表示其杰卡德系数,这样可使得最后判断出的相似路段的相似性更加准确。

$$Sim(A-S) = AttSim + StrSim/2 \quad (3)$$

两路段的 $Sim(A-S)$ 值越大,它们的相似度就越高,若 $Sim(A-S)$ 的值大于给定的阈值,则认为公交站点的路段可以合并成为相似路段。通常,相邻站点的路段极有可能为相似路段,不相邻的站点间的路段也会存在路况相似的情形。依据路况的相似性可将目的路段 L 划分为 i 个相似子路段,相似子路段构成集合 $L = \{L_1, L_2, \dots, L_i\}$ 。路段的相似划分示意图如图 1 所示。

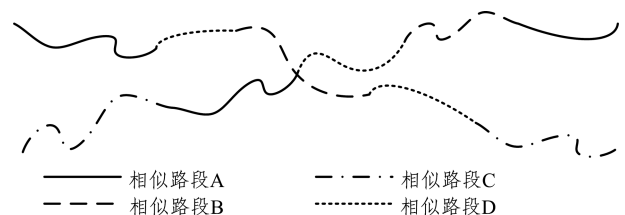


图 1 相似路段划分示意图

Fig. 1 Schematic diagram of similar road sections

将公交车所经过的相似路段抽象为直线段,公交站台和控制交叉路口可以抽象为这条直线段上的节点,则可直线化公交车的行驶路线,如图2所示。

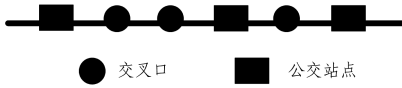


图2 公交车行驶路线示意图

Fig.2 Schematic diagram of bus travel route

3 行程时间预测

用于公交车行程时间的预测模型较多,算法的时间资源和内存资源等因素会导致模型预测的时间较长。近期的算法模型也多是优化经典算法入手的,但应用广泛的BP神经网络、RBF神经网络等算法在建模时需要很大的训练样本,因此其存在一定的局限性,如BP神经网络算法在本质上为梯度下降法,将不可避免地出现“锯齿形现象”和“过拟合”,从而导致算法效率降低、收敛速度较慢。本文对比各种预测模型后发现,卡尔曼滤波是一种最优化的自回归数据处理算法,其不需要存储大量数据,且在观测到新数据时借助系统本身的状态转移方程递推出新的滤波值,算法的运行速度较快,效率较高,适合实时预测。因此,本文在基于相似性动态划分相似路段并融合多线路的数据信息后,采用运行速度较快的卡尔曼滤波进行短时预测。

根据相关研究发现,公交车在周一到周五的早高峰和晚高峰的行程时间比其他时间段的行程时间长。同路段相邻运行的两辆公交车的运行时间差呈现了较为典型的瑞利分布;同天公交车在同时段或邻近时段的行驶速度较为类似,不同天出行的公交车在相似路段的行驶速度也较为类似。

不妨假设车辆到达交叉口信号控制区域的概率服从泊松分布,本文对公交车通过交叉口的延误时间不做深入研究。为避免GPS数据回传周期的影响,采用GPS数据的进出站信号来获取公交车的停靠站时间。

3.1 实验相似路段划分

我们先选取沈阳市162路公交车线路作为本文的实验场景1,以崇山路嘉陵江街站到华商晨报站连续3站之间的路段为例。

162路公交车线路如图3所示。



图3 162路公交车线路图

Fig.3 Map of bus route 162

不同类别的POI在不同时段的活跃度不同,如餐饮类在6:30-8:30、11:00-13:00、17:00-20:00时段对交通的影响较大,因此根据时间的工作特性,将POI影响度分为了早高峰、晚高峰、夜间及其他时段(见表1),我们以POI活跃度作为主要因素来划分路段。

表1 不同时段POI影响度

Table 1 POI influence at different time

路段ID	早高峰 (7:00— 9:00)	晚高峰 (17:00— 20:00)	夜间 (20:00— 22:00)	其他 时段
崇山路嘉陵江街至辽宁中医	72	43	51	53
辽宁中医至儿童医院	64	58	48	29
儿童医院至华商晨报	88	74	62	40
崇山西路与淮河街交叉口至崇山中路与怒江街交叉口	67	60	49	34

将崇山路嘉陵江街站至辽宁中医站设为路段A,辽宁中医站至儿童医院站设为路段B,儿童医院站至华商晨报站设为路段C。站间路段A的属性向量和空间结构向量为:

$$AttVector(a) = (Len(a), Euc(a), Cur(a), Dir(a), Cro(a), Gra(a), Wid(a))$$

$$SpaVector(a) = (SecIn(a), SecOut(a), POI(a))$$

站间路段B的属性向量和空间结构向量为:

$$AttVector(b) = (Len(b), Euc(b), Cur(b), Dir(b), Cro(b), Gra(b), Wid(b))$$

$$SpaVector(b) = (SecIn(b), SecOut(b), POI(b))$$

站间路段C的属性向量和空间结构向量为:

$$AttVector(c) = (Len(c), Euc(c), Cur(c), Dir(c), Cro(c), Gra(c), Wid(c))$$

$$SpaVector(c) = (SecIn(c), SecOut(c), POI(c))$$

通过Jaccard系数以及式(3)计算路段的Sim(A-B)值,最终得到实验路段1,如图4所示。

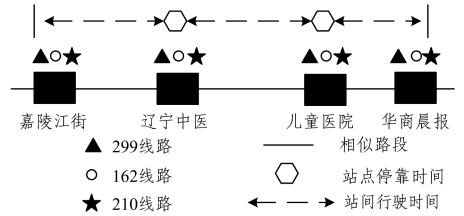


图4 实验路段1的示意图

Fig.4 Schematic diagram of experimental road section 1

通常来讲,短距离内相邻站点的路段成为相似路段的可能性较高,那么,不相邻的站点间的路段是否可以通过计算路况因素来找到与其路况相似的路段,从而丰富实验数据,提高实验结果的精准性。为了验证本文提出的基于相似路段划分的公交车行程时间预测方法的合理性及有效性,本文选取沈阳市299路公交车线路作为此次实验的实验路段,以崇山路嘉陵江街站到华商晨报站,崇山西路与淮河街交叉口至崇山中路与怒江街交叉口之间的路段为实验场景2。

299路公交车线路如图5所示。



图5 299路公交车线路图

Fig.5 Map of bus route 299

为了方便描述,将崇山西路与淮河街交叉口至崇山中路与怒江街交叉口之间的路段设为 S_1 ,崇山路嘉陵江街站到华晨晨报站的路段设为 S_2 。经过考查得出,在路段 S_1, S_2 上行驶都不需经过红绿灯且分别有两个停靠站点,路段 S_1, S_2 在地图上的位置示意图如图 6、图 7 所示。



图 6 S_1 地图位置示意图

Fig. 6 Location of S_1 on map



图 7 S_2 地图位置示意图

Fig. 7 Location of S_2 on map

路段 S_1 的属性向量和空间结构向量为:

$$AttVector(1) = (Len(1), Euc(1), Cur(1), Dir(1), Cro(1), Gra(1), Wid(1))$$

$$SpaVector(1) = (SecIn(1), SecOut(1), POI(1))$$

路段 S_2 的属性向量和空间结构向量为:

$$AttVector(2) = (Len(2), Euc(2), Cur(2), Dir(2), Cro(2), Gra(2), Wid(2))$$

$$SpaVector(2) = (SecIn(2), SecOut(2), POI(2))$$

通过 Jaccard 系数以及式(3)计算路段的 $Sim(A-S)$ 值,可将这两个路段划为相似路段,从而得到实验路段 2,如图 8 所示。

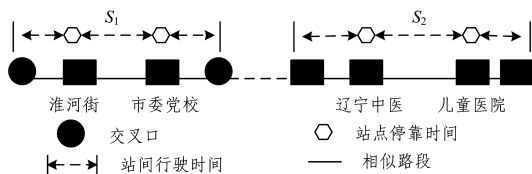


图 8 实验路段 2 的示意图

Fig. 8 Schematic diagram of experimental section 2

对于划分所得到的每个相似路段,假设公交车在路段上行驶时需要通过 i 个站点和 k 个交叉路口,则行程时间为:

$$T = T_r + \sum_{j=1}^k T_c + \sum_{j=1}^i T_s \quad (4)$$

其中, T_r 表示公交车在相似路段的行程时间; T_c 表示交叉路口的延误时间; T_s 表示公交站点的停靠时间。

当面临数据不足或临时堵车等状况时,可以利用 POI、道路宽度等因素及时动态地匹配相似路段,将相似路段与目标路段上的多条公交线路的数据信息融合,用相似路段来推断数据稀疏或缺失路段的信息,把相似路段上的行驶时速或行程时间及时传输过来。

3.2 多线路信息卡尔曼滤波算法

卡尔曼滤波^[15-16]是一种基于动态系统的状态序列求最小方差估计的最优估测算法,它使用状态和观测方程来描述一个系统,不断获取实时数据,并将测量数据用于预测未来状态,递推地消除噪声影响,最后获得修正后的结果。

对应每一个相似子路段 L_i ,有 N_1, N_2, \dots, N_j 共 $j(j \geq 1)$ 条线路的公交车经过,那么子路段 L_i 对应的公交线路集合为 $S = \{N_{i1}, N_{i2}, \dots, N_{ij}\}$ 。不同线路公交车在相似路段经过时,第 k 辆车经过测速点的行驶速度可记为 V_k ,那么近一个时间段内公交车在站间路段 L_i 的行驶速度信息包括该时间段内 m 辆公交车的行驶速度,依照公交车辆经过选定测速点的时间序列,形成集合 $\{V_1, V_2, \dots, V_m\}$ 。

多线路经过路段的示意图如图 9 所示。

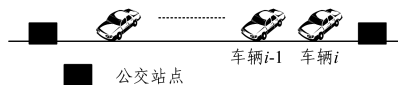


图 9 多线路公交车行程示意图

Fig. 9 Schematic diagram of multi-line bus travel schedule

在相同时段内,某相似路段前面经过的公交车辆的行驶速度对下一时段公交车的行驶速度具有较大的参考作用,使用卡尔曼滤波算法进行下一时段预测,根据 $k, k-1, \dots$ 前车的行驶信息来预测下一状态 $k+1$ 时段公交车的行程时间。

标准卡尔曼滤波系统的状态方程与观察方程如下:

$$x(k) = Ax(k-1) + Bu(k) + w(k) \quad (5)$$

$$Z(k) = Hx(k) + v(k) \quad (6)$$

将上述系统状态方程和观测方程改为状态空间表示形式:

$$V(k+1) = H_0(k)V(k) + \dots + H_{n-1}(k)V(k-n+1) + w(k) \quad (7)$$

其中, $V(k)$ 表示经过某一路段的第 k 辆公交车的行驶速度。将需要预测的此路段的下一辆公交车的行驶速度记为 $V(k+1)$,之前通过该相似路段测速点的前 n 辆公交车的站间行程速度为 $V(k), V(k-1), \dots, V(k-n+1)$ 。若 n 取 4,则模型为:

$$V(k+1) = H_0(k)V(k) + H_1(k)V(k-1) + \dots + H_3(k)V(k-3) + w(k) \quad (8)$$

定义状态向量:

$$A(k) = [V(k), V(k-1), V(k-2), V(k-3)] \quad (9)$$

$$x(k) = [H_0(k), H_1(k), H_2(k), H_3(k)]^T \quad (10)$$

$$y(k) = V(k+1) \quad (11)$$

为方便使用卡尔曼滤波预估状态变量,对公式做如下转换:

$$x(k) = \varphi(k, k-1)x(k-1) + u(k-1) \quad (12)$$

$$y(k) = A(k)x(k) + w(k) \quad (13)$$

其中, $x(k)$ 表示状态向量; $y(k)$ 表示观测向量; $\varphi(k, k-1)$ 表示状态转移矩阵; $A(k)$ 表示观测矩阵; $u(k-1)$ 表示模型噪声,假设它为零均值的白色噪声; $Q(k-1)$ 为协方差; $w(k)$ 为观测噪声且假设为零均值的白色噪声; $R(k)$ 为协方差。

通过以上步骤,计算得到 $\hat{x}(k)$,此路段下一时刻站间的行驶速度为:

$$V(k+1) = A(k)\hat{x}(k) \quad (14)$$

那么,在相似路段的行程时间 $T(k+1)$ 为:

$$T(k+1) = \frac{S_{\text{路段距离}}}{V(k+1)} \quad (15)$$

公交车的总路段行程包括 n 个相似路段, 公交车在路段上行驶需要通过 i 个站点和 k 个交叉路口, 则总行程时间为:

$$T_{\text{总}} = \sum_{j=1}^n T_r + \sum_{j=1}^k T_c + \sum_{j=1}^i T_s \quad (16)$$

其中, T_r 表示公交车在相似路段的行程时间; T_c 表示交叉口的延误时间; T_s 表示公交站点的停靠时间。

在使用卡尔曼滤波算法进行预测时, 可以选用不确定的实验数据来确定其参数^[17-18]。

4 仿真与验证

4.1 模型预测

根据已知的公交车站点间的路段距离和先前多条公交线路的站间路段的行驶速度, 可得到站间的运行时间。基于多线路的卡尔曼滤波算法以历史数据行程时间为基准, 以近期得到的公交车速度为观测值, 通过计算卡尔曼滤波方程组获得目标状态, 利用基于全局信息的最优估计值对公交车行程时间进行动态修正。

为了验证本文提出的基于相似路段划分的公交车行程时间预测方法的合理性及有效性, 我们选取沈阳市 162 路、210 路、299 路公交线路作为此次实验的实验路段, 在 2018 年 3 月连续 5 个天气晴朗的工作日进行数据采集, 采集时段集中 8 点到 20 点之间, 使用 GPS 设备跟车采集 162 路、210 路、299 路 3 条线路在对应时段的行驶速度, 同时记录公交车在交叉口及站点的延误时间, 初步得到实验所需的基础数据。将前 4 天采集的数据作为模型预测数据, 将得到的实验结果与第 5 天采集的实际公交车行程时间做对比。

途径崇山路嘉陵江街站、华商晨报站、崇山西路与淮河街交叉口、崇山中路与怒江街交叉口的公交线路及站间距离如表 2 所列。

表 2 相似路段及对应公交车线路

Table 2 Similar routes and corresponding bus lines

相邻公交站 ID	站件间隔/m	途径公交线路
崇山路嘉陵江街至辽宁中医站	1180	162,205,210,299
辽宁中医至儿童医院站	460	162,205,210,213,245,299
儿童医院至华商晨报站	410	162,213,227,245,279,281,299
崇山路淮河街至市委党校	395	162,205,299,228
崇山西路与淮河街交叉口至崇山中路与怒江街交叉口	747	288,299,205

通过含有 GPS 装置设备的车辆进行跟车采集, 通过信息采集得到 162 路、210 路、299 路这 3 条公交线路在其所对应时段的行驶速度, 如表 3 所列。

表 3 公交车 162 路、210 路、299 路站间的行驶速度

Table 3 Road speeds among stations 162,210 and 299

日期	7	8	9	10	...	16	17	18
162 星期一	16.8	15.2	20.8	22.7	...	23.7	13.8	14.2
210 星期一	17.8	14.8	20	23.7	...	20.7	13.1	12
299 星期一	17.5	14.2	18.2	23.6	...	20.5	13.1	12.8
162 星期二	18.6	18.4	23.1	24.5	...	23.5	14.2	15.1
210 星期二	17.9	15.6	20.7	24.3	...	21	13.6	12.8
...
162 星期五	17.3	16.7	22.1	24.7	...	24.2	13.5	13.8
210 星期五	18.2	15	20.5	24	...	20.5	13.5	12.1

4.2 实验结果与分析

在对实验结果进行分析时, 为提高评价预测结果的准确性, 引入了最常用的误差指标, 即绝对平均百分误差^[19-20] (MAPE), 其公式如下:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|t_i - t|}{t} * 100\% \quad (17)$$

其中, n 为预测次数; t_i 为模型预测的公交到达时间; t 为公交车的真实到达时间。

式(17)中, MAPE 是相对比率, 在误差评估上较为客观。MAPE 的相关评估说明如表 4 所列。

表 4 MAPE 评估说明

Table 4 MAPE evaluation instructions

MAPE/%	说明
<10	高精度的预测
10~20	良好的预测
20~50	合理的预测
>50	不准确的预测

我们将实验结果与实际采集的公交车行程时间进行计算, 得到实验路段 1 和实验路段 2 的绝对平均百分误差 (MAPE), 根据实验结果来判断预测的精准性, 实验结果如图 10、图 11 所示。

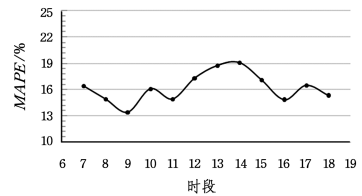


图 10 路段 1 的实验结果

Fig. 10 Experimental results of road section 1

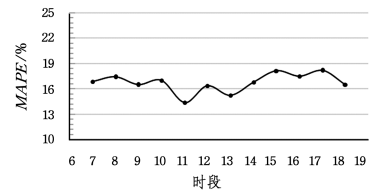


图 11 路段 2 的实验结果

Fig. 11 Experimental results of road section 2

根据路况对公交线路进行路段划分, 丰富了实验数据, 其中公交发车时间越相近, 相似路段上的公交车行程时间的实时性就越高。依据实验路段 1 和实验路段 2 的绝对平均百分误差可以得出, 基于该方法的公交车行程时间预测具有良好的准确性。

本文提出利用属性特征和空间结构的相似性, 考虑 POI 的活跃度动态地划分相似路段, 然后在数据预处理阶段融合多线路的数据信息从而推断缺失路段或数据稀疏情况下的数据信息, 将处理好的数据提供给多种预测模型进行行程时间预测。从短时预测的角度出发, 本文采用了卡尔曼滤波算法。为了突出划分相似路段可以提高实验预测结果的实时性和精准性, 使用了未划分相似路段的经典算法 BP 神经网络和历史平均算法作为本文算法模型的参照。利用 3 种预测模型对所选取的实验路段 1 进行行程时间预测。将采集的数据作为卡尔曼滤波预测、神经网络预测、历史平均算法 3 种模型的数

据资源,从而得到实验数据并将实验结果与实际采集的公交车行程时间进行计算比较,实验数据的小数点部分采取四舍五入的取值方式。实验路段 1 的预测结果的绝对平均百分误差(MAPE)如图 12 所示。

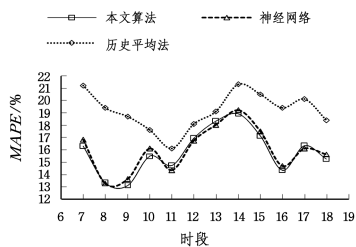


图 12 实验预测对比结果

Fig. 12 Comparison results of experiment prediction

由图 12 可知,基于相似路段划分的卡尔曼滤波算法对实验路段具有更高的精确度,如 7 点至 9 点卡尔曼滤波算法的预测值与实际测量值的 MAPE 平均值为 14.24%,神经网络算法的预测值与实际测量值的 MAPE 平均值为 14.567%,相比神经网络算法,该方法的精度提高了大约 0.33%。在 17 点至 19 点,卡尔曼滤波算法的预测值与实际测量值的 MAPE 平均值为 15.303%,神经网络算法的预测值与实际测量值的 MAPE 平均值为 15.47%,相比神经网络算法,在晚高峰时期该方法的精度提高了大约 0.17%。大量实验的结果表明,本文提出的算法能够对公交车行程时间进行良好的预测,与其他预测模型相比其预测精度较高,也提高了实验预测的实时性,能有效达到实时查询的性能需求。

结束语 公交车的行程时间预测是目前的研究热点之一,可以给出行的乘客们提供实时准确的行程时间,给乘客们带来了极大的便利。本文将公交线路划分为多个相似路段,且将多线路公交车数据信息融合在一起,丰富了实时数据,然后结合卡尔曼滤波算法动态性高、实时性强等优点,对公交车行程时间进行了预测,并引入了其他两种最为常用的公交车行程预测模型与本文算法进行实验对比,更直接地体现了卡尔曼滤波方法模型具有较高精确度的特点。实验结果表明,与现在较为流行的神经网络模型预测以及最常用的历史平均法相比,本文提出的融合多线路信息的卡尔曼滤波算法在高峰期的预测精度的优势明显。虽然在理论和实验中对线路进行相似路段划分体现了一定的价值,但目前相似路段是相对静态的方法,而且受笔者自身水平的限制,针对交叉口的停留时间只是做了简化处理,后续可以对不同类型的信号灯以及信号灯安装位置进行进一步的研究,并在公交车停靠时间方面进一步对停靠时间序列进行研究。

参 考 文 献

[1] HONG C, MIAO L. Mining smartphone data for app usage prediction and recommendations: A survey [J]. *Pervasive and Mobile Computing*, 2017, 37: 1-22.

[2] XIANG H Y, PENG X W. Research Status and Development Trend of Bus Arrival Time Prediction [J]. *Traffic Information and Safety*, 2014, 32(4): 57-61.

[3] BAI C, PENG Z R. Prediction of Bus Travel Time Based on Dy-

namic Model [J]. *Journal of Computer Engineering and Applications*, 2016, 52(3): 103-107.

[4] DHIVYABHARATHI B, ANIKUMAR B, LELITHA V, MANOJ P. Particle Filter for Reliable Bus Travel Time Prediction under Indian Traffic Conditions [M]. Berlin: Springer International Publishing, 2017: 2199-9287.

[5] CHEN G J, YANG X G, AN J, et al. Bus-arrival-time prediction models link-based and section-based [J]. *Journal of Transportation Engineering*, 2012, 138(1): 60-66.

[6] STEPHEN A, ERROL N, MELISSA F A, et al. Optimization of transit total bus stop time models [J]. *Journal of Traffic and Transportation Engineering (English Edition)*, 2016, 3(2): 146-153.

[7] XIE F, GU J H, ZHANG S Q, et al. Bus arrival time prediction model based on MapReduce clustering and neural network [J]. *Computer Applications*, 2017, 37(S1): 118-122.

[8] SUN Y Y, LIU Y, ZHOU X Y, et al. Urban bus arrival time prediction mechanism based on traffic similarity [J]. *Journal of Software*, 2012, 23(1): 87-99.

[9] LAN T, YU M, XU Z B, et al. Temporal and Spatial Variation Characteristics of Catering Facilities Based on POI Data: A Case Study within 5th Ring Road in Beijing [J]. *Procedia Computer Science*, 2018, 131: 1260-1268.

[10] SI Y, ZHANG F Z, LIU W Y. An adaptive method for POI recommendation based on check-in and temporal features [J]. *Knowledge-Based Systems*, 2017, 128(15): 59-70.

[11] ZHENG Y Q, ZHU X Y, ZHANG F M, et al. Road trip time estimation based on road network similarity [J]. *Application Research of Computers*, 2018, 35(6): 59-66.

[12] SATU E S. Graph clustering [J]. *Computer Science Review*, 2007, 1(1): 27-64.

[13] ZHANG X L, FU Y Z, CHU P X. Application of Jaccard Similarity Coefficients in Recommendation System [J]. *Computer Technology and Development*, 2015, 4: 158-161.

[14] SERGIO J, FABIO A G, ALEXANDER G. Mathematical properties of soft cardinality: Enhancing Jaccard, Dice and cosine similarity measures with element-wise distance [J]. *Information Sciences*, 2016, 367-368: 373-389.

[15] MARIA V K, JULIA V T. The UD-based Approach for Designing Pairwise Kalman Filtering Algorithms [J]. *IFAC-Papers On Line*, 2017, 50(1): 1619-1624.

[16] WU Z Z, FU M Y, XU Y, et al. A distributed Kalman filtering algorithm with fast finite-time convergence for sensor networks [J]. *Automatica*, 2018, 95: 63-72.

[17] ABDALLAH S, PER S. A methodology for using Kalman filter to determine material parameters from uncertain measurements [J]. *Materials Discovery*, 2015, 2: 1-15.

[18] LOU T, WANG L, SU H S, et al. Desensitized cubature Kalman filter with uncertain parameters [J]. *Journal of the Franklin Institute*, 2017, 354(18): 8358-8373.

[19] PHILIP H F. A note on the Mean Absolute Scaled Error [J]. *International Journal of Forecasting*, 2016, 32(1): 20-22.

[20] SUNGIL K, HEEYOUNG K. A new metric of absolute percentage error for intermittent demand forecasts [J]. *International Journal of Forecasting*, 2016, 32(3): 669-679.