

The Bus Arrival Time Service Based on Dynamic Traffic Information

Jian Dong, Zuyun Wang, Zhihong Chen, Songsong Pang

State Key Laboratory of Software Development Environment, Beihang University, Beijing
Email: zhutongyu@nlsde.buaa.edu.cn, mafajin@nlsde.buaa.edu.cn, tao_m@nlsde.buaa.edu.cn, licong@nlsde.buaa.edu.cn

Received: Aug. 16, 2012; revised: Aug. 23rd, 2012; accepted: Sep. 5th, 2011

Abstract: The bus arrival time (BAT) service is a key service to improve public transport attractiveness by providing users with real-time bus arrival information which can help them to arrange their bus travel schedule intelligently. Thus the technique of real-time bus arrival prediction has become a research hotspot in the community of Intelligent Transport Systems (ITS) nowadays. In this paper, a novel model on bus arrival time prediction is proposed. The model proposes a complete set of programs to solve BAT prediction for large-scale real-time traffic information calculating. It adopts an effective algorithm calculating vehicle's driving direction real-timely. BAT is calculated based on dynamic traffic information and visual prediction is a way to complement when GPS information is not arrived as expected. Experimental results indicate that the model has considerable efficiency in accuracy (over 85.1%) and computational speed (max 5000 GPS records per second).

Keywords: Bus Arrival Time Prediction; GPS Data; Dynamic Traffic Information

基于动态交通信息的公交到站预测服务

董 健, 王祖云, 陈智宏, 庞松松

北京航空航天大学软件环境国家重点实验室, 北京
Email: zhutongyu@nlsde.buaa.edu.cn, mafajin@nlsde.buaa.edu.cn, tao_m@nlsde.buaa.edu.cn, licong@nlsde.buaa.edu.cn

收稿日期: 2012年8月16日; 修回日期: 2012年8月23日; 录用日期: 2012年9月5日

摘 要: 公交到站时间(BAT)服务是提高公共交通吸引力的关键服务之一。服务给使用者提供实时车辆到站信息能够使用户更好的安排自己的公共交通行程。因此, 实时公交到站时间预测技术在智能交通领域已经成为研究的热点。在本文中采用了新的公交到站时间预测模型。该模型提出了一套完整的算法来解决 BAT 预测中大规模的实时交通信息的计算, 采用了高效的算法来实时纠正车辆的行驶方向。动态交通信息作为数据基础支撑 BAT 的计算。同时, 采用虚拟预测来处理 GPS 数据丢失的情况保证了服务提供的稳定性。实验结果表明模型拥有较高的准确率(超过 85.1%)和计算速度(最高处理 5000 条每秒)。

关键词: 公交到站预测; GPS 数据; 动态交通信息

1. 简介

由于不断增长的城市化、人口的大量增长以及人口密度的改变, 交通拥堵已经成为世界性问题。解决问题的方法之一是提供更多的基础设施来满足不断增长的车辆数目。然而该方案不能满足持续增长的交

通需求, 所以其他更符合动态发展的交通需求的方案需要被探索。一个好的方案是通过智能交通系统(ITS)对已存在的设施进行管理。

先进公共交通系统(APTS)是众多对缓解交通拥堵的重要 ITS 应用之一。APTS 通过提高公共交通系

统效率、改善交通服务质量将更多的人转入搭乘公共交通。公交到站预模型做为 APTS 应用之一会向乘客提供详细的公交到站信息并减少乘客在公交站点等待次数。这就要求模型提供包括公交停靠站时间、交叉口旅行时间、路段旅行时间在内的所有公交车辆线路旅行时间。

历史上,研究者采用了诸多方法来进行公交到站预测。如历史和实时基于路径的方法、机器学习方法(神经网络法、支持向量机)、基于路径的模型(卡尔曼滤波)以及静态方法(回归分析)等。Lin 和 Zeng^[1]利用 GPS 定位数据、公交时刻表、时刻表匹配度等信息为 Virguria 和 Blacksburg 地区提出了基于 GPS 定位数据的公交车到站预测模型,但是该算法依赖的历史定位数据间隔较大,且精度不高,适用于郊区车辆稀疏的情况,不适合城市内交通情况变化较大的情形。Wall 和 Dailey^[2]提出使用卡尔曼滤波模型来跟踪车辆的位置,并在 Seattle 的华盛顿地区结合自动车辆定位系统(AVL)和历史数据来预测公交车的到站时间,但是他们并没有将停靠时间作为模型的独立变量。Chenro^[3]建立的模型假设城市的交通状况是周期进行变化的,且在特定路段内历史行程时间与当前行程时间的比值是不变的。在历史数据库的基础之上,建立预测模型,利用实时定位数据调整公交车的预测到站时间。Jeong 和 Rilett^[4]基于德克萨斯州休斯顿市采集的公交车实时 AVL(车辆自动定位系统)数据,在给定交通拥挤实时信息和公交车在每个站点的停靠时间的条件下,建立了基于历史统计的模型、多变量回归模型和人工神经网络模型,通过比较三种模型的平均绝对百分误差发现人工神经网络模型^[5]的预测精度要比历史数据模型和回归模型高^[6]。

Shalaby 和 Farhan^[7]提出了基于卡尔曼滤波方法的公交车行程时间预测模型,他们使用从多伦多城区采集而来的车辆定位数据来进行分析预测,发现卡尔曼滤波方法比历史数据模型,回归模型以及神经网络模型效果更好。他们采集了 2001 年 5 月每个周中 5 天的数据,选择其中 4 天的数据来训练模型,剩余一天的数据用来做测试。此外,他们还提出了一个分开的卡尔曼滤波预测算法^[8],用来计算运行时间和停靠时间。所建立的历史平均模型,回归模型以及神经网络模型都包括了路段行程时间中的停靠时间,也就是

说模型对于停靠时间和运行时间没有分开来考虑。他们将路段定义为在两个检查时间点站点间的距离,每个路段包括 8 个公交站点,因此他们只预测在时间检查点的停靠时间。

温惠英等^[9]“利用灰色理论对影响行程时间序列的各因素进行灰色关联分析,根据灰色关联度^[10]的大小来选取路段行程时间的主要影响因素,并对 Kalman 滤波方法进行改进,研究表明改进后的 Kalman 滤波性能优于传统的 Kalman 滤波方法。Kalman 滤波器模型利用不断逼近的方式获得较高的预测精度,尤其在提前一步预测行程时间时,该模型具有良好的预测性能。但是,其能力却随步骤的增加而不断衰退。

本文提供了一种基于 GPS 的计算车辆旅行时间(包括车辆站点间旅行时间、站点停靠时间和路口延迟)的信息和算法。为了保证基础数据(如线路数据、站点数据、上下行数据)的预处理效率,提出了一套完整的自动寻找车辆线路,核查站点位置、纠正车辆行驶赏析性的方法。同时,本文提出了一个新的搜索匹配路线的算法用来满足实时计算车辆 GPS 信息的需求。另外,虚拟预测模型在本文中被用来处理实时信息丢失情况。所有的模型模块组成了预测车辆到站时间和距离模型。

2. 数据处理

公交到站预测模型每天需要实时处理 18,000 车的 GPS 信息。考虑到郊区和市区道路信息之间的差异,本文选取 944 线路作为研究对象,其同时包含了郊区和市区线路。944 线路是北京使用频率最频繁的线路之一,由 35 个站点和 55 个信号灯路口组成,单程旅行时间为 80 分钟左右。旅行时间相关数据由安装在车辆的 GPS 采集设备提供。944 线路车辆全部安装了 GPS 采集设备。GPS 在线率(GOR)达到 85%。

2.1. 线路和站点数据的选取与纠错

本文中链路端点的坐标被存储在线路文件中。根据车辆方向经过的端点顺序。线路文件为计算精确的车辆位置和到站实时距离提供了计算支持。

链路定义如下:

$$L = \{l = (id, length, sCoor, eCoor, samplingpoints)\} \quad (1)$$

startCoordinate 是链路的起始点坐标。endCoordinate

是链路终止点坐标。samplingpoints 是一个 GPS 点的集合。

线路文件描述如下：

$$BL = BL(B, D, L) \quad (2)$$

L 是一个按链路顺序排序的链路集合。当 $\forall b \in B, d \in D, l \in L$ 一条线路文件 $o(B, D, L)$ 能够从集合中获取。

由于公交线路数目繁多，如果手工在地图上挑取线路，将是一个巨大的工作。本文提出对车辆行驶链路坐标聚类的方法来自动寻找车辆行驶的路连集。以车辆行驶 GPS 为样本，对车辆 GPS 坐标匹配后聚类，将聚类结果处理后，得到线路路连集。

从官方途径获取的站点坐标与站点实际位置存在偏差。这个偏差会导致公交到站出现 100 米的误差。为了找到精确的站点位置，本文提出了 3 种方法来寻找站点位置并相互修正。

首先，站点坐标信息(SLI)能够从官方数据库中获得。同时，从 google 地图中也能获取到站点的位置坐标。最后，鉴于车辆在站点的车辆行为模式，对车辆行驶 GPS 聚类后，可以得到站点的位置坐标。

2.2. 行驶方向和下一站站数据计算方法及效率优化

在八方达和公交公司的车辆中都配有 gps 设备。实时的位置信息就是从该设备获取。从 gps 设备单元获取的实时数据，大约在 20 s 中一个。包含公司名称、车辆号、车辆 id、上下行、经度、维度、下一站站点等信息。

作为预测到站距离及时间的基础数据源，车辆号，上下行、经度、维度及下一站站点信息是我们计算的核心基础数据。我们需要确保上下行、下一站站点及经度维度数据。经测试八方达公司给予的上下行数据，错误率在 33%。公交公司的上下行数据，错误率在 47%。公交公司的下一站数据错误率 20%。公交集团公司的下一站站点数据无法提供。两家公司的 gps 数据都有偏移。鉴于我们的计算强依赖于以上数据。我们需要确保其准确性。进而保证我们的到站时间预测的准确性。

首先对于 gps 偏移数据，我们将其匹配到路练上。匹配的方式为垂直匹配。计算并获取到垂足。

对于计算上下行数据和下一站数据。我们首先对线路数据进行处理，将线路对应的采样点进行编号。然后我们缓存 5 个实时 gps 数据，并将每一个新传到的数据存储并匹配到采样带上。同时，我们缓存 5 个实时 gps 数据，来作为我们判断车辆行驶方向的依据。对于每一辆车我们都记下匹配的对应线路的采样点。那么通过观测采样点的顺序。如果采样点的为顺序则上行，反之则为下行。通过此种方法我们即能判断线路的上下行。

同时，我们会对站点进行匹配采样点的工作。记录 gps 点对应的采样点序号，并与站点对应的采样点进行比较，以此来得到该 gps 点对应的下一站站序。

在我们的实时系统测试过程中，我们发现。按照以上算法，每 2000 条数据处理需要 3 秒钟。然而鉴于我们的发布频率是 20 秒发布一次到站时间数据。而 10,000 辆公交车每 20 秒会传递回来 9000 多条的数据。并且其他计算会需要 15 秒左右的时间。所以该算法无法满足大数据量的实时计算。

该算法的主要时间复杂度消耗在匹配线路的采样点上。一般一条线路有 1000 个采样点。每次匹配都需要对所有的采样点进行匹配计算，那么对与 n 条记录，每条采样点的规模为 m ，那么时间复杂度就是 $o(n*m)$ 。

我们对采样点进行建模。我们建立一个二维数组来存储我们的采样点。首先我们按照采样点坐标的 x 轴值，对数组进行排序。排序时间复杂度为 $m \log m$ ，gps 点坐标 (x, y) 匹配到采样点的阈值是二者距离小于 100 米。对于给定的 x ，对 x 进行二分查找，查找到差值最小的 x ，以此为基础计算在 $x \pm 0.0001$ 范围内的点(匹配到点小于 100 米)于 gps 点坐标的距离。以此方法来寻找匹配的采样点，算法效率为 $o(n)$ 。提升了 25 倍左右。如果线路全为东西向的路段则效率提升 100 倍。效率提升后，计算 10,000 条需要不到 1 秒的时间，如图 1 所示。

2.3. GPS 和动态交通信息

本文研究内容是基于北京动态交通信息服务系统，由 CENNAVE 组织构建。这个系统能够处理 140,000 浮动车和 18,000 量公交车并且每 2 分钟发布一次实时路况。发布信息包括：拥堵状况、旅行时间、

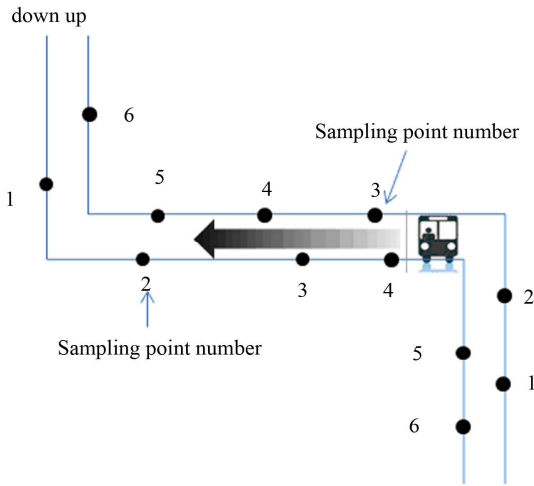


Figure 1. Model used to judge up and down direction
图 1. 判断车辆行驶方向

每个路链的状态。这个研究也使用了北京的历史交通信息。历史交通信息每 2 分钟发布一次并每周更新。

3. 预测模型

公交到站时间预测是一个非常复杂并且受诸多因素干扰的问题。这个模型提高了预测的准确率并且针对用户计算了多种场景的数据。整个预测模型被分为两类，第一类是短途预测，模型会预测车辆下游到 3 站的到站时间。第二类是长距离预测，系统会提供 3 站以外的站点的车辆到站时间预测。这两模型将在下面分别介绍：

3.1. 短途预测

当预测站点对于预测车辆在 3 个站点区间以内时，当前行驶时间被分为 3 个部分——路链旅行时间、站点滞留时间、信号灯延迟时间。整个行驶的部分又被分为更多的小区间。旅行时间、站点滞留时间和信号灯行驶时间在每一个子部分都被重新计算。当车辆经过 k^{th} 子部分。让 $t(k)$, $x(k)$, $d(k)$ and $l(k)$ 分别代表总体的旅行时间，路链行驶时间，站点滞留时间，信号灯站点滞留时间。路链行驶时间，站点滞留时间，新号灯滞留时间如图 2 所示。

3.1.1. 车辆行驶时间计算

路网交通流的平均速度可以通过以公交车为数据源的实时路况发布系统中获取。所以，每条路链的旅行时间能够被计算出来。车辆位置能通过 GPS 每

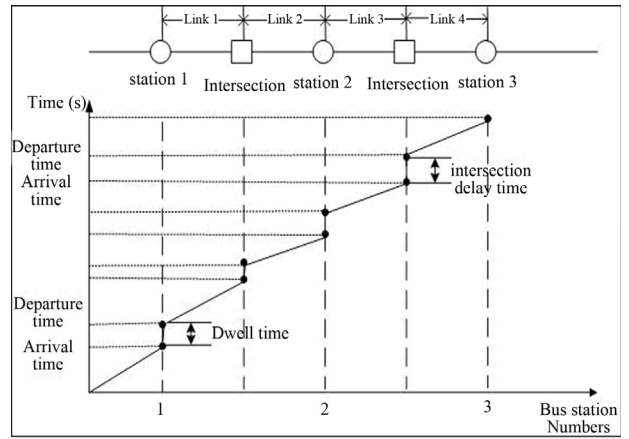


Figure 2. Bus traveling track
图 2. 公交行驶轨迹

20 秒定位一次。预测旅行时间能通过将每条路链的平均旅行时间加总来获取。预测行驶时间不包含站点的停留时间和信号灯停留时间。所以这部分时间需要被补偿计算。

假定不同子区间的行驶时间能够如下表示

$$T_i = \sum_{i=0}^N \frac{L_i}{V_i} \quad (5)$$

T_i 是在 N 个路链上的总体旅行时间。 L_i 是 i^{th} 路链的长度。 V_i 是 i^{th} 路链的平均旅行时间。

3.1.2. 计算站点滞留时间

本文中，只有站点滞留时间在车辆停靠站时间中被考虑。假定在一个子区间中站点滞留时间如下表示：

$$T_h = \frac{\sum_{k=1}^N (T_{dk} - T_{ak})}{N} \quad (6)$$

$$T_i = \frac{\sum_{w=1}^M T_{hw}}{N} \quad (7)$$

T_h 是 i^{th} 站点的平均滞留时间。 T_{dk} 是 K^{th} 的离开时间。 T_{ak} 是 K^{th} 的到达站点时间。 N 是每小时经过站点的车辆数目。 T_i 是 M 周每天这段时间的平均速度。 T_{hw} 是 W^{th} 周。到达时间，离开时间计算完毕。

站点滞留时间可以由下表示

$$T_w = \sum_{i=j}^k T_i \quad (8)$$

T_w 是车辆到达目标站点的所有停靠站时间。

3.1.3. 交叉口延误时间计算

交叉口延误时间能够通过公交到站服务系统实时获取。历史趋势库用以提供历史交叉口延误时间计算。

交叉口延误时间数据库存储在数据立方体中。公交到站服务系统可以获取任何时间段的路口延迟时间。

整体的交叉口延误时间的预测可以通过对每个交叉口的延误时间加总获得。

$$T_d = \sum_{j=0}^k T_j \quad (9)$$

T_d 是交叉口延误时间对于从当前车辆位置到目的站点。 T_j 是 j^{th} 交叉口的平均延误时间。

3.2. 长途预测

在长途预测中，由于有路程远、耗时长、不可遇见性较多的特点。路况信息因此变化较大。如 944 线路路程最长 3 个小时，在 3 个小时内路况可能从高峰期进入平峰期。所以简单依靠路况来计算车辆到达时间已经不能符合要求。

对与长途预测，到达目的站点数大于 3 个的情况。我们采取短途预测时间与历史规律相结合的方法。

我们对每条线路的每个站点之间(如 2 站到 7 站，3 到 9 站)的车辆旅行时间建立模型。采用数据立方体方式来存储我们的历史数据。即对不同星期特征日和不同时间段建立不同的历史趋势库。

经研究发现，相邻站点之间的旅行时间没有特别稳定的旅行时间规律。但是在长距离相距多站点之间存在较稳定的历史规律。

长途预测时间(LBAT)可以如下表示：

$$LBAT = STT + \text{Time}(\text{startStation} + 2, \text{endstation}) \quad (10)$$

3.3. GPS 点缺失，虚拟预报

在公交到站预测时间服务中，对 gps 质量依赖较大。但是 gps 质量本身无法得到大幅度提高。经测试，所有线路的 gps 质量大约在 80%左右。也就是说 20 s 一个 gps，每 5 个点会丢失一个。那么在实际服务过程中，就会出现应有预报信息错乱丢失。鉴于此，需要根据已有 gps 点进行虚拟预报。

对于所有在线车辆，我们实时维护一个统计表。存储每一辆车的公交线路信息(BOI)。每辆车维护最新

的 5 个 BOI。如果 gps 丢失我们将分以下几种情况来处理：

3.3.1. Gps 点丢失。车辆未进站。

由于 gps 点是每 20 秒一个。所以，当 gps 点丢失时，对于虚拟 gps 点虚拟 GPS 信息点(VGI)，到站时间则为 $\text{TimePredict} = \text{BOI.TIMEPREDICT} - 20$ 。

对于 VGI 的到站距离。将上一个 BOI 的点匹配到路练上，计算 BOI 点到路练终点的距离的时间。如果如果大于 20 s。则取

$$\text{Distance} = (\text{tjam.length}/\text{tjam.time}) * 20\text{s}$$

如果小于 20 s。则 $\text{time} = \text{Etjam.time}$ 直到 $> 20\text{s}$

$$\text{Distance} = \text{BOI.dis} - \text{disToLinkEnd} + \text{Etjam.length}$$

如果计算出的 distance 小于 0 则说明车辆过了上次预报的站点。需要重新定位下一站。并重新计算到站距离和时间。

站点定位到下一站。

$$\text{Distance} = \text{dis} + \text{站点间距}$$

匹配路练计算方法：

1) 到路练终点的旅行时间：

$$T = ET;$$

2) 路练旅行时间；

3) 路练起点到路练终点的旅行时间。

如果 distance > 0 但是预测时间小于 0 则按照上述步骤重新定位重新计算。

3.3.2. GPS 点丢失，车辆 LBOI 在站内。

在实时预测中，维护一个站点延误时间表。对应每一个站点有一个延误时间。如果车辆进入站点 100 米范围内。则算进站。则不按上述方法计算到站距离。按以下方法计算：

站点定位到下一站：

$$\text{Distance} = \text{站点间距}$$

$$\text{Time} = \text{站点延误} + \text{路练旅行时间}$$

3.3.3. GPS 点丢失，车辆即将到达终点且进站。

如果下一站为终点站但没有进站，按照正常方式预测。若下一站为终点站且进站则预测即将进站。在 BOI 表中去掉该车。

4. 计算验证

绝对百分比误差(MAPE)，线路绝对百分比误差

(MAPEL), 绝对误差(MAE)做为评价指标。具体计算方法如下:

$$MAPE = \frac{1}{N} \sum_{k=1}^N \frac{|t(k) - t_m(k)|}{t_m(k)} * 100$$

$$MAPEL = \left(\frac{1}{N} \sum_{k=1}^N \frac{|t(k) - t_m(k)|}{t_m(k)} \right) * 100 \quad (11)$$

$$MAE = \frac{1}{N} \sum_{k=1}^N |t(k) - t_m(k)|$$

$t_m(k)$ 是测试车辆的到达时间。N 是测试次数。

以上测试方案使用 944 车辆的测试数据。结果如下所示。

当目标站点在 3 站以内的时间。会采用前文叙述的短途预测模型。结果如图 3 所示。在这个实例中, 很容易发现当车辆距离站点越近预测时间越加准确。最大误差为 50 s。

当车辆距目标站点大于 3 站时, 采用长途预测方案。测试站点达到 17 站。结果如图 4 所示。图 5 所示为, 同一站点最近到达车辆的实际到达时间和预测到达时间。可以观察到预测结果和测试指标基本一致。

三种方式 MAPE 的结果分别是 13.3%, 6.8%, 12.9%。

PMape 如表 1 所示。

5. 结语

预测和测试的结果揭示了交通流数据能够较好

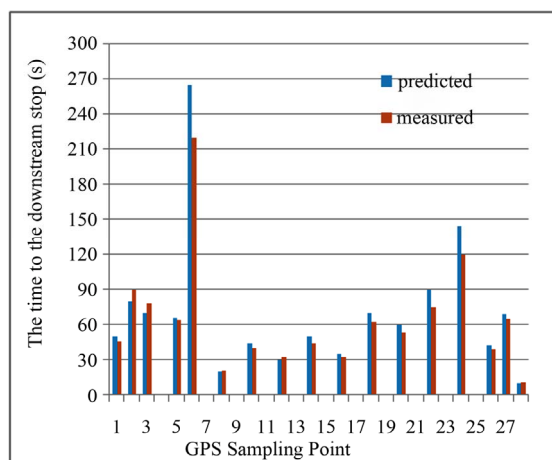


Figure 3. Comparison of the predicted and measured time for the 20 operating buses
图 3. 比较预测和实际测量时间

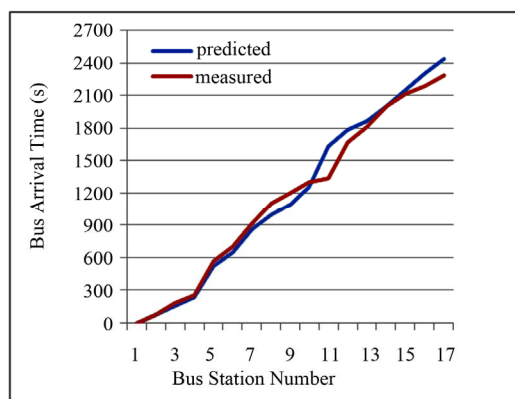


Figure 4. Comparison of the predicted and measured time for the 17 bus stations

图 4. 对比同一车次经过 17 个站点的到达时间和预测时间

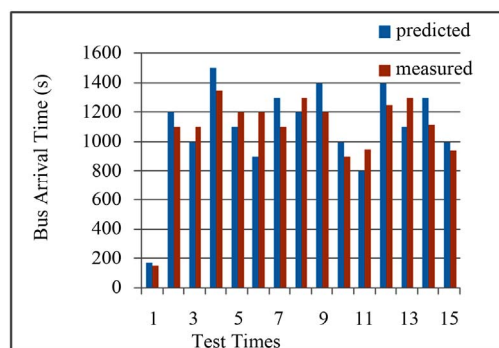


Figure 5. Comparison of the predicted and measured time for 15 times on one station

图 5. 比较同一站点情况 15 次预测和实际到达时间对比

Table 1. PmapE result
表 1. PmapE 结果

Line name	Direction	PMape	MAE < 30 s
977	0	95.3%	86.3%
945	1	94.8%	86.1%
988	1	93.2%	85.3%

的预测链路旅行时间因为交通流对路况变化很敏感。虽然当有异常数据流的情况下, 交通流数据更加高级, 但历史数据 GPS 和交通流数据在现代交通条件中展现了相似的效率。在相对意义里, 传统模型不能够敏锐的处理行驶过程中的变化问题。

当获取交通流数据可行的时候, 本文提出的基于交通流数据建立的公交到站模型会展现出良好的效率和预测准确度。在众多影响车辆运行的条件中, 本文使用了交通流数据和日期相关的方案来预测车辆旅行时间。包括如乘客需求等其他影响车辆旅行时间的因素将会提高结果的准确性。这将会是以后研究的方向。

6. 致谢

本文由公共交通导航与出行位置服务系统及应用示范支撑。(No. 2012AA12A207)。

参考文献 (References)

- [1] W. H. Lin, J. Zeng. An experimental study of real-time bus arrival time prediction with GPS data. *Transportation Research Record*, 1999, 1666: 101-109.
- [2] Z. Wall, D. J. Dailey. An algorithm for predicting arrival time of mass transit vehicle using automatic vehicle location data. *Proceedings of Transportation Research BOARD 78th Annual Meeting*, Washington DC, 10-14 January 1999.
- [3] L. Vanajakshi, S. C. Subramanian and R. Sivanandan. Short-term prediction of travel time for Indian traffic conditions using buses as probe vehicles. *Transportation Research Board, the 87th Annual Meeting (CD-ROM)*, Washington DC, 13-17 January 2008: 18.
- [4] R. Jeong, L. R. Rilett. Bus arrival time prediction model for real-time applications. *Transportation Research Record: Journal of the Transportation Research Board*, 2005, 1927: 195-204.
- [5] M. Chowdhury, A. Sadek, Y. Ma, N. Kanhere and P. Bhavsar. Applications of artificial intelligence paradigms to decision support in real-time traffic management. *Transportation Research Record: Journal of the Transportation Research Board*, 2006, 1968: 92-98.
- [6] J. S. Yang. Travel time prediction using the GPS test vehicle and Kalman filtering techniques. *Proceedings of the American Control Conference*, Portland, 8-10 June 2005: 2128-2133.
- [7] A. Farhan, A. Shalaby and T. Sayed. Bus travel time prediction using GPS and APC. *ASCE 7th International Conference on Applications of Advanced Technology in Transportation*, Cambridge, 5-7 August 2002.
- [8] F. W. Cathey, D. J. Dailey. A prescription for transit arrival/departure prediction using automatic vehicle location data. *Transportation on Research Part C: Emerging Technologies*, 2003, 11(3): 241-264.
- [9] H. Liu, H. J. van Zuylen, H. V. Lint and M. Salomons. Urban arterial travel time prediction with state-space neural networks and Kalman filters. *Transportation Research Board, Annual Meeting (CD-ROM)*, 2006, 1968: 99-108.
- [10] R. P. S. Padmanaban, L. Vanajakshi. Estimation of bus travel time incorporating dwell time for APTS applications. *IEEE Intelligent Vehicles Symposium*, Xi'an, 3-5 June 2009: 955-959.