

# 一种动态和自适应公交到站时间预测方法

谢 玲 李培峰 朱巧明

(苏州大学计算机科学与技术学院 江苏省计算机信息处理技术重点实验室 苏州 215006)

**摘 要** 公交到站时间预测是实现智能化公交信息服务的基础,可靠地预测公交到站时间有利于提高公共交通的服务水平,以吸引更多的城市居民选择公共交通。以某城市公交系统海量的历史数据为基础,建立了基于 SVM 的集合了静态和动态数据的公交预测模型,该模型引入上游路段速度、下游路段最新速度、下游路段最新花时、时间段和路况拥挤程度等动态信息作为模型特征。在此基础上,根据大量公交到站时间历史数据的波动性,提出了一个基于波动性的自适应预测模型。实验结果表明,自适应预测模型优于现有模型,提高了预测的精确度和效率。

**关键词** 公交到站时间,实时预测,动态预测,自适应模型,支持向量机,波动性统计

**中图分类号** TP39 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.1.056

## Adaptive Method of Predicting Arrival Time of Buses on Dynamic Traffic Information

XIE Ling LI Pei-feng ZHU Qiao-ming

(Key Laboratory of Computer Information Processing Technology of Jiangsu Province, School of Computer Science and Technology, Soochow University, Suzhou 215006, China)

**Abstract** Bus arrival time prediction is the foundation of realizing intelligent bus information service. Reliable prediction of bus arrival time is beneficial to improve the public transport service level, so that it attracts more and more city residents to use public transportation. In this paper, based on the massive historical data of a city bus system in real time, SVM (Support Vector Machine) was applied to establish a bus forecasting model on the static and dynamic information. And the speed of upstream, the latest speed of downstream, the latest travel time of downstream, time-of-day, traffic congestion, etc were introduced into our dynamic model. Besides, this paper put forward an adaptive prediction model to improve the efficiency of prediction based on a lot of volatility of bus arrival time historical data. The experimental results show that the adaptive model outperforms those existing static models.

**Keywords** Bus arrival time, Real-time prediction, Dynamic prediction, Adaptive model, SVM, Volatility statistics

## 1 概述

目前,公交到站时间预测技术的研究在智能交通领域逐渐兴起,准确可靠的公交预测模型对城市公共交通的发展起着积极的推动作用,是智慧城市的一个重要应用。在理论研究中,已经提出的模型主要有基于历史数据模型、统计回归理论模型、卡尔曼滤波模型和机器学习模型等。Li<sup>[1]</sup>、李天雷<sup>[2]</sup>基于历史数据模型,以大量的历史数据为基础,通过假设公共交通是一个循环往复的过程来预测公交到站时间。基于历史数据模型原理易懂、操作简单,然而对突发事件的预测效果不理想。回归预测模型<sup>[3]</sup>通过变量之间的关系建立回归方程,根据自变量在预测时段的变化来对因变量进行预测。回归预测模型能够在最大程度上减小因素波动所造成的影响,然而它对应用环境的要求比较高,较大的历史均值偏差将导致预测的不准确。Shalaby<sup>[4]</sup>、Liu<sup>[5]</sup>基于自动车辆定位系统建立

了一个基于 Kalman 滤波的预测模型,该模型具有良好的预测性能,但随着迭代次数的增加其性能会有非常明显的衰退。Park<sup>[6]</sup>、Yu<sup>[7]</sup>、Shen<sup>[8]</sup>、Rahman<sup>[9]</sup>、于滨<sup>[10]</sup>利用机器学习的方法来预测公交到站时间。其中,于等<sup>[10]</sup>使用时间段和天气两个静态特征,路段、当前路段的运行时间和下一路段的运行时间 3 个动态特征,共 5 个特征作为 SVM 的特征输入,实验结果显示,整合了最新公交运行信息的 SVM 可以很好地应对运行过程中的随机延迟。SVM 能够解决小样本、非线性、维数灾难、过学习和局部极小点等问题,具有出色的统计学习能力,近年来常被用于公交到站时间预测。

公交车在实际的运行过程中会受到很多随机因素的干扰,如:时间段(平高峰)、节假日、天气、交通拥挤程度、客流变化和意外事件等,准确预测公交的到站时间非常困难。为了比较准确地预测公交到站时间,需要对公交行驶过程中各种信息进行实时采集。现有的预测模型考虑的影响因素都偏

到稿日期:2014-03-06 返修日期:2014-06-24 本文受国家自然科学基金(61070123),江苏省自然基金(BK2011282),江苏省高校自然科学基金重大基础研究项目(11KJ520003)资助。

谢玲(1987—),女,硕士生,主要研究方向为自然语言处理,E-mail:20114227052@suda.edu.cn(通信作者);李培峰(1971—),男,副教授,主要研究方向为中文信息处理、分布式信息系统、网格计算;朱巧明(1963—),男,教授,博士生导师,主要研究方向为自然语言处理、Web 信息处理、嵌入式系统。

少,没有多方面考虑交通的复杂性和突发事件,预测模型健壮性不强。本文以大量的公交实时运行数据为基础,考虑多种影响公交到站时间的动态因素,将其量化并建立了基于 SVM 的集合了静态和动态信息的公交到站时间预测模型;同时,根据大量公交到站时间历史数据的波动性,自适应地选择公交预测模型,以在保证准确性的同时提高预测效率。

## 2 基准系统和存在的问题

公交车的运行时间受到很多因素的影响,主要包括道路、车辆、人为、交叉口、突发事件、交通拥挤、时间段、天气、星期、法定假日、地理位置等因素。本文基于 SVM 方法建立动态公交预测模型。

### 2.1 基于 SVM 的动态公交到站时间预测模型——基准系统

本文首先参照于滨<sup>[10]</sup>等人提出的 SVM 公交到站时间预测模型,如图 1 所示。模型选择“时间段”和“天气”两个静态特征来描述公交车运行属性,选择“路段”、“下一个路段最新运行时间”和“当前路段的运行时间”这 3 个动态特征来表述公交的实时运行信息,分别用  $x^1$ 、 $x^2$ 、 $x_{k \rightarrow k+1}^3$ 、 $x_{k+1 \rightarrow k}^4$ 、 $x_{k-1 \rightarrow k}^5$  来表示;确定 SVM 特征向量后,将这 5 个特征作为输入变量,输出变量为公交到站时间。以  $m$  车到达  $k$  站点为例,输入  $(x^1, x^2, x_{k \rightarrow k+1}^3, x_{k+1 \rightarrow k}^4, x_{k-1 \rightarrow k}^5)$  来预测  $m$  车到达  $k+1$  站点的时间,然后更新  $k-1 \rightarrow k$  路段的最新时间,即  $x_{k-1 \rightarrow k}^5 = x_{k-1 \rightarrow k}^5$ 。随着公交到站时间不断更新,后续公交到站时间不断修正,直到公交车运行结束,抵达终点。

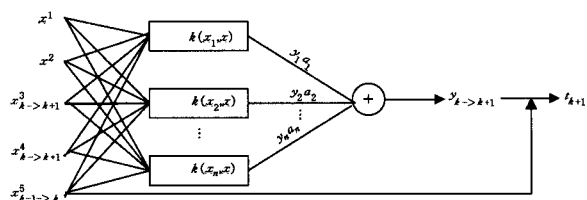


图1 公交到站时间预测模型结构

### 2.2 存在的问题

上述的基准系统虽然在 SVM 中引入了实时信息,但是实时交通信息具有复杂性,在于滨<sup>[10]</sup>等人建立的模型中,仅仅使用了“时间段”和“天气”2 个静态变量以及“路段”、“下一个路段最新运行时间”和“当前路段的运行时间”3 个动态变量来描述复杂的交通状况。实际上,影响公交到站时间的因素有很多,如前面提到的一些静态因素和动态因素。可见,基准系统中的模型所考虑的因素偏少。针对这一问题,本文提出了一个综合考虑多种因素的公交到站时间预测模型。

除上述问题之外,在基准系统的模型中,每一段预测都是基于同一个模型,输入的都是同样的特征变量,却忽略了每一段的交通状况都是不一样的。从以往的历史数据中可以看出,前一站与后一站到站时间波动性不大的情况是存在的,波动性很大的情况也同样是存在的,具体的原因可能与路况条件等因素有关。所以,用同一个模型来预测整个公交到站时间显然是不够准确的。而且,随着公交车到站时间的更新,用于 SVM 模型的数据量也不断扩大,每一次更新数据都需要重新建模,导致 SVM 的训练时间和训练次数也会不断增加。因此,本文根据大量公交到站时间历史数据的波动性,提出了一种自适应的选择公交预测模型,该模型在保证公交预测时间准确性的同时,也提高了预测效率。

## 3 动态和自适应公交到站时间预测模型

由于基准系统在选择影响公交到站时间因素方面,考虑因素偏少,本文以大量公交实时数据为基础,结合多种影响公交到站时间的动态因素,将其量化并建立了基于 SVM 的集合了静态和动态信息的公交预测模型。同时针对基准系统在预测时,整个预测过程都使用同一个模型的问题,本文根据大量公交到站时间历史数据的波动性,自适应地选择公交预测模型,以提高预测效率。

### 3.1 动态预测模型

影响公交到站时间的因素有很多,这些因素总体上分为两类,一类是静态因素,另外一类是动态因素。其中静态因素有:

- 天气因素:天气状况的变化影响公交的某些性能参数,对于天气信息的采集可以查询天气预报或者实时记录当天的天气情况。本文设某一天的天气情况为  $w$ ,  $w = \{0, 1, 2, 3, 4\}$ , 其中 0 表示晴天,1 表示小雨,2 表示大雨,3 表示雪,4 表示冰雪。
- 星期因素:工作日和休息日的交通情况是不同的。一周的 7 天分别表示  $d = \{1, 2, 3, 4, 5, 6, 7\}$ 。
- 是否法定假日:法定假日内,客流量往往会急剧变大,交通繁忙,突发性事件发生频率也较高。本文用  $b$  来表示当天是否是法定节日,  $b$  为布尔型的变量,取值为:

$$b = \begin{cases} 1, & \text{true} \\ 0, & \text{false} \end{cases}$$

- 地理位置因素:每段公路路段所处的位置对预测准确性是有一定影响的。交通繁忙路段的交通状况比较复杂,预测难度相对较大。本文结合了百度地图卫星地图来定位当前位置。设某一站点所处的位置  $l = \{1, 2, 3, 4, 5, 6, 7, 8\}$ , 其中 1 表示位于学校附近,2 表示位于商业区附近,3 表示位于医院附近,4 表示位于旅游区附近,5 表示位于居民区附近,6 表示位于政府机构附近,7 表示位于火车站附近,8 表示其他地方。

由于统计设备和实验条件有限,无法获取全方面的数据,并且部分因素是随机的或突发的,无法实时掌握,但是历史同期,相邻路段或者相近时段的交通状况对预测的影响是有内在规律可循的<sup>[11,12]</sup>,因此本文将相邻或相近的行驶速度、时间作为道路因素、车辆因素、人为因素、交叉口因素和突发事件因素等的替代因素,用符号  $v_{\uparrow}$ ,  $v_{\downarrow}$ ,  $t_{\text{new}}$  来表示。动态因素具体有:

- 道路因素:公交车经过的道路等级、车道数、路面、视距、线形、坡度等。
- 车辆因素:车辆本身的性能、最高车速、最大爬坡度、加速性能和制动性能等。
- 人为因素:主要是驾驶员的驾驶技术、习惯以及反应速度等。
- 交叉口因素:由于交叉口红绿灯和排队情况而引起的公交车行驶时间延误占整个公交运行时间相当大的一部分,据统计调查得知交叉口占整个运行时间的比例约 25% ~ 33%。
- 突发事件因素:主要包括突发的交通事故、临时修路以及交通管制等。

• 交通拥挤情况:交通拥挤程度很能说明路段的实时交通情况。若拥堵没有得到及时有效地解决,则会呈现“点一线一面”蔓延的趋势。为了获取交通拥挤程度,本文采用上游速度与上游平均速度差值来表示路况拥挤程度。

• 时间段情况:不同的时间段,交通状况是不一样的,如早晚高峰的时间段,往往交通拥挤现象严重。其中早上(5:00—11:00)用0来表示,中午(11:00—13:00)用1来表示,下午(13:00—17:00)用2来表示,晚上(17:00—22:00)用3来表示。

综上所述,动态预测模型包括4个静态变量和5个动态变量。其中1—4是天气、星期、法定假日和地理位置4个静态变量;5—9表示上游路段速度、下游路段最新速度、下游路段最新花时、路况拥挤程度和时间段。本文以上述的9个变量为静态和动态特征,建立了基于SVM的集合了静态和动态特征的公交到站时间回归预测模型,如图2所示。该模型不仅引入了4个静态特征,还加入了5个动态特征,较为全面地考虑了各方面的影响因素。

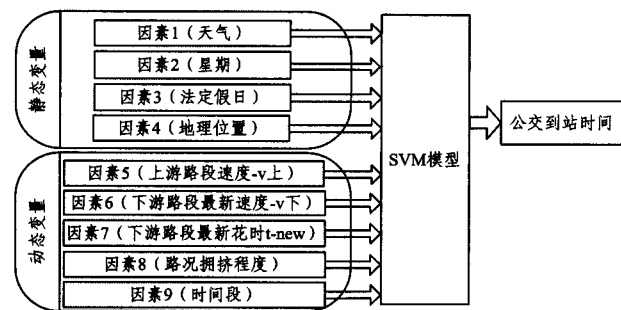


图2 SVM预测模型

### 3.2 自适应预测模型

从以往的历史数据中可以看出,前一站与后一站到站时间波动性不大的情况是存在的,波动性很大的情况同样也是存在的,依据这种现象,本文对每一段公交的到站时间波动性进行分析。如果一个路段的到站时间波动性低于某个阈值,就用到站时间历史数据的平均值作为路段的公交到站时间;否则就采用SVM全局预测模型来预测公交到站时间。由于计算历史数据的平均时间几乎没有时间代价,因此自适应模型可以有效降低预测所花的时间。另外,如果某2段以上路段有着相似的波动性,本文会将其合并成一条路段进行预测,这样的机制同样可以降低预测时间。

公交到站时间自适应预测算法综合考虑了日期、时段和路线的影响。日期可以简单地分为工作日和假期,而如何划分时间单元和路线则是算法的重点。算法的第一部分是时间分段算法。本文以0.5小时为基准单位,当路段的到站时间波动性相似时,将相邻单元合为一个整体单元。这样的时间单元划分保证到站时间预测更加准确。不论是工作日还是假期,时间分段算法同样适用。具体的时间分段算法如下:

1)定义日期类型 $DT$ (工作日或者假期);

2)将一天分别以0.5小时,1小时,1.5小时,2小时,2.5小时,...,24小时为单元划分,从而得到 $T_{0.5}^i, T_1^i, T_{1.5}^i, T_2^i, \dots, T_{24}^i$ ,其中 $T$ 表示到站时间集合,下标表示时间单元,上标表示时间单元总数。例如 $T_{0.5}^i$ 表示日期为 $DT$ 类型下的将一天每隔0.5小时划分的所有时间段集合;

3)对于每个 $T_{tp}^i$ 时间段集合,统计公交到站时间波动性,

用 $\lambda_{tp}^i$ 来表示,其中 $i$ 表示第 $i$ 个时间段, $tp$ 表示时间单元;

4)对于每个时间单元 $tp$ (例如0.5小时时间单元、1小时时间单元、1.5小时时间单元.....),分别计算它们的平均波动性,用 $\overline{\lambda_{tp}}$ 表示;

5)选择 $\overline{\lambda_{tp}}$ 值最小的 $tp$ 作为最终的时间划分单元,这样将一天按照 $tp$ 小时划分为不同时间段单元。

自适应公交到站时间模型的第一部分(时间分段算法)获得时间划分单元 $tp$ ,进而将一天以 $tp$ 小时为单位进行划分。自适应模型的第二部分是路线分段算法。该算法是以时间分段算法为基础的。首先,确定当前时间 $T$ 所属 $tp$ 划分的时间段,用 $T'$ 表示。这样获得 $T'$ 时间段下各路段到站时间波动性;第二步,将预测路线划分为不同的路段,当相邻路段到站时间波动性变化较小时,将其整合为一个路段。对于当前时间 $T$ ,路段分段算法将整个路线划分为相邻的路段,具体算法如下:

1)初始化路段集合 $S$ ,表示为 $S=\{S_1, S_2, \dots, S_i, \dots, S_n\}$ ,其中 $n$ 表示路段总数, $S_i$ 表示第 $i$ 段路段。初始化 $CS$ 为空集合,将 $S$ 集合元素赋值给 $CS$ ;

2)确定当前的时间 $T$ 所属 $tp$ 划分下的时间段 $T'$ ;

3)获得 $T'$ 时间段下的各路段 $S_i$ 的到站时间波动性,用 $u_i$ 表示 $S_i$ 的波动性;

4)重复5)和6);

5)对于相邻的路段( $S_i, S_{i+1}$ ),计算 $u_i$ 与 $u_{i+1}$ 之差,用 $d_{i,i+1}$ 表示,其中下标表示相邻路段;

6)如果 $d_{i,i+1}$ 大于或者等于阈值 $\delta$ ,则将 $S_i, S_{i+1}$ 添加到 $CS$ 集合中;否则将 $S_i, S_{i+1}$ 作为一个整体路段来预测,记作 $S_i$ ,然后重新计算第 $i$ 段路段的到站时间波动性。其中阈值 $\delta$ 的调整通过实验结果来确定,其取值范围为0到1;

7)直到 $CS$ 集合不能再重新整合为止;

8)返回 $CS$ 集合包含的所有路段的分段结果。例如, $CS=\{\{S_1, S_2\}, \{S_3, S_4, S_5\}, \dots, \{S_i, \dots, S_n\}\}$ 表示路段 $S_1, S_2$ 和路段 $S_3, S_4, S_5$ 作为一个整体来预测。

## 4 实验结果及分析

为了验证本文提出模型的精确性和有效性,选取A路公交线路的数据来对其进行验证。本文选择A路公交从起点S到终点D这段路线,其中这一段路线经过商业区和旅游区,并且通向火车站,人流量大,交通状况比较复杂,以站点为节点分为XG、GS、SB、BQ、GH 5段公交路段。

如前所述,本文提出的方法包含9个输入特征:地理位置、天气、星期、法定假日、同一公交上游路段的速度、不同公交下游路段最新速度、下游路段最新花时、所属时间段和路况拥挤程度。本文采用的数据是A路公交到达每个站点的实时时间,记录了从2012年12月26日至2013年1月31日(法定假日为12月30日、12月31日、1月1日)的每班公交到达每个站点共37天的时间数据,总共1286组。本文选取前1026组数据作为机器学习的训练集,后260组数据作为验证数据。实验硬件环境为Intel(R) Core(TM) i7-2600 CPU,内存16G;软件环境为MATLAB R2012b。为了进一步对模型精度和适应度进行评价,本文选择2种误差指标对预测结果进行对比分析,其计算方法如下:

• 均方根误差(RMSE)为:

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^N(T_{true}(t)-T_{pred}(t))^2} \quad (1)$$

• 平均相对误差绝对值(MAPE)为

$$MAPE=\frac{1}{N}\sum_{i=1}^N\frac{|T_{true}(t)-T_{pred}(t)|}{T_{true}(t)} \quad (2)$$

其中,  $t$  表示公交当前出发时间;  $N$  表示用来预测公交到站时间的数据个数;  $T_{true}(t)$  表示实际的公交到站时间;  $T_{pred}(t)$  表示预测的公交到站时间。

#### 4.1 预测模型性能分析

本文首先以基准系统的 2 个静态变量(时间和天气)和 3 个动态变量(路段、当前路段运行时间和下一路段运行时间)为基准来采集数据,对其工作进行重现并将其作为基准系统。接着,在基准系统基础上加入新的静态和动态特征,量化的具体方法已在上节有过详细的描述。表 1 是优化的特征模型与基准系统模型总体性能的对比。

表 1 模型性能对比

| 误差指标           | 基准系统  | 优化模型  |
|----------------|-------|-------|
| 均方根误差-RMSE(秒)  | 67.79 | 54.02 |
| 平均相对误差绝对值-MAPE | 0.33  | 0.28  |

为了进一步说明优化后的特征模型与基准系统性能的差异,本文在 XG、GS、SB、BQ、QH 5 个路段上分别对基准系统和优化模型进行实验,并统计了基准系统和优化模型的 RMSE 和 MAPE 值,最后得到的性能对比如图 3 所示。

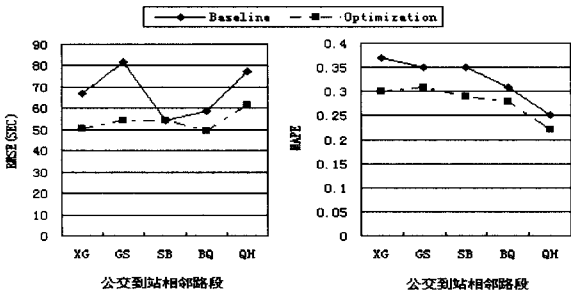


图 3 Baseline 和 Optimization 性能对比

从表 1 和图 3 可以看出,本文实现的优化模型性能与基准系统相比,均方根误差和平均相对误差都有一定程度的提高,可见本文新加入的静态和动态特征的确提高了公交到站时间预测的精确度。

#### 4.2 新特征对预测模型贡献度分析

为了更加明显地看出新加入的特征对基准系统优化的效果,本文在基准系统的基础上分别将不同公交下游路段最新速度、路况拥挤程度、地理位置、星期和法定假日 5 个新的特征一一加入,进行实验。分别统计各新特征加入后预测的 RMSE 和 MAPE 值,最后得到的实验结果如表 2 所列。

表 2 加入新特征的实验结果

| 误差   | Base+v <sub>下</sub> | Base+%拥 | Base+静 1 | Base+静 2 | Base+静 3 |
|------|---------------------|---------|----------|----------|----------|
| RMSE | 53.04               | 54.16   | 53.6     | 51.05    | 51.73    |
| MAPE | 0.31                | 0.32    | 0.31     | 0.29     | 0.3      |

说明:  $v_{下}$ :不同公交下游路段最新速度; %拥:路况拥挤程度; 静 1:地理位置; 静 2:星期; 静 3:法定假日

结果显示,在基准系统中,加入各个新特征后,基准系统预测误差有明显的下降,可见新加入的特征都在一定程度上提高了系统预测精确度,同时各特征的加入对系统预测精确度的贡献程度相当,这些新特征本身之间的相互依赖程度较

小。为了充分说明各新特征之间相互依赖较小的问题,本文对这些特征进行组合后加入基准系统中,结果发现其组合后对系统预测精确度的贡献程度相当,可见各新特征之间没有太大相关性。

#### 4.3 自适应模型性能分析

为了验证本文提出模型的精确性和有效性,选取某城市 A 路和 B 路公交从起点 S 到终点 D 这段路线 XH 的实时数据进行验证。其中 A 路有 1286 组数据, B 路 515 组数据。把前 1400 组数据(A 路 1000 组, B 路 400 组)作为 SVM 的训练集,后 401 组数据(A 路 286 组, B 路 115 组)作为验证数据。另外,统计每天每段公交到站时间的方差  $\sigma$ , 详情见图 4。

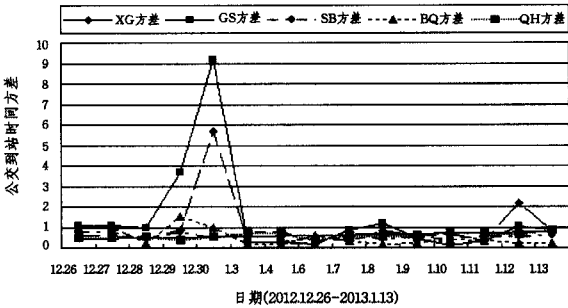


图 4 公交到站时间方差统计

由图 4 可以看出方差  $\sigma$  总体分布在 0—1 范围内,特殊的时间如 12.29—1.1 这 3 天,  $\sigma$  超出了 1, 这是因为这几天是法定假日,人流量大、交通拥堵,所以造成了波动性比较大。由于方差  $\sigma$  总体分布在 0—1 范围内,因此本文确定阈值  $\delta$  的范围为 [0, 1]。如前所述,当公交到站时间方差很小时就采用到站时间的历史数据平均值作为路段的公交到站时间(即静态预测);当到站时间方差很大时,采用动态建模,即从这段路线的起始点到终点,以站点为节点,分成若干路线,迭代预测。这和以往模型的不同点在于,在整个公交预测过程中,公交经过某个站点时,系统根据以往数据的历史波动性自适应地选择预测模型。如果多个路段存在相似的波动性,则将它们作为一个整体来预测。在路段分段算法中,阈值的确定是根据实验结果进行动态调整的。随着  $\delta$  的增大, RMSE 和 MAPE 值相应减小,预测花费时间却急剧增大。实验结果表明,当  $\delta$  取 0.5 时,预测精确度和预测花费时间的加权达到了最优。表 3 给出了对 XH 路段 3 种公交进行预测的结果。

表 3 不同模型公交预测实验结果

| 模型  | RMSE(S) | MAPE (%) | 预测花时 |
|-----|---------|----------|------|
| 静态  | 130.52  | 6.87     | 0.25 |
| 动态  | 54.83   | 3.32     | 2.37 |
| 自适应 | 42.79   | 2.23     | 1.73 |

从表 3 可以看出,本文自适应模型的 MAPE 和 RMSE 值优于静态和动态模型,即本文提出的模型比静态和动态模型的预测准确度要高。自适应模型既引入了动态信息同时又考虑到不同日期、不同时间段到站时间波动性和相似性,根据波动性细粒度划分路段,重新组合路段。虽然在预测时间上,本文的预测模型高于静态预测模型,但相比于精确度的提高,这样的预测耗时是可以接受的。自适应模型的精确度比静态模型平均提高了 57.9%,比动态模型提高了 22%,不仅保证了预测的准确性,同时也减少了训练次数和训练时间,从而提高了预测效率。

(下转第 289 页)

供了思路,进一步工作将重点研究限界传递相似度图界限的自适应调整方法。

## 参考文献

- [1] Wille R, Yahia S, Nguifo E, et al. Formal Concept Analysis as Applied Lattice Theory[M]//Concept Lattices and Their Applications. Springer Berlin Heidelberg, 2008, 4923: 42-67
- [2] Ning L, Guanyu L, Li S. Using formal concept analysis for maritime ontology building[C]//2010 International Forum on Information Technology and Applications (IFITA). 2010, 2: 159-162
- [3] Sánchez D, Batet M, Isern D, et al. Ontology-based semantic similarity: A new feature-based approach[J]. Expert Systems with Applications, 2012, 39(9): 7718-7728
- [4] Li D, Du J, Yao S. Research on Computer Science Domain Ontology Construction and Information Retrieval[J]. Knowledge Engineering and Management, 2011, 123: 603-608
- [5] Kang X, Li D, Wang S. Research on domain ontology in different granulations based on concept lattice[J]. Knowledge-Based Systems, 2012, 27: 152-161
- [6] Maarek Y S, Berry D M, Kaiser G E. An information retrieval approach for automatically constructing software libraries[J]. IEEE Transactions on Software Engineering, 1991, 17(8): 800-813

(上接第 256 页)

**结束语** 本文提出的自适应模型引入了上游路段速度、下游路段最新速度、下游路段最新花时、时间段和路况拥挤程度等动态信息作为模型特征,其性能优于静态预测模型和动态预测模型。而且自适应预测模型是基于日期、时间段和站点波动相似性进行路段分段组合预测的,在保证预测准确性的基础上,提高了预测效率。但是本文提出的预测模型仍然有许多需要改进的地方,如新特征的加入对系统性能的提高并没有达到令人满意的程度,而且特征向量多数趋于离散化;自适应选择方法的改进以及如何降低预测时间,这些都是我们下一步研究的主要方向。

## 参考文献

- [1] Li W, Koendjibiharie W, de M Juca R C, et al. Algorithm for estimating bus arrival times using GPS data[C]//Proceedings of the 5th IEEE International Conference on Intelligent Transportation Systems. Singapore, 2002: 868-873
- [2] 李天雷. 基于 GPS 数据的公交行程时间计算与预测系统[D]. 长春: 吉林大学, 2009
- [3] Li F, Yu Y, Lin H, et al. Public bus arrival time prediction based on traffic information management system[C]//2011 IEEE International Conference on Service Operations, Logistics, and Informatics (SOLI). 2011: 336-341
- [4] Chung E H, Shalaby A. Expected time of arrival model for school bus transit using real-time global positioning system-based automatic vehicle location data [J]. Journal of Intelligent

- [7] Kumar C A. Knowledge discovery in data using formal concept analysis and random projections[J]. International Journal of Applied Mathematics and Computer Science, 2011, 21(4): 745-756
- [8] Formica A. Semantic web search based on rough sets and fuzzy formal concept analysis[J]. Knowledge-Based Systems, 2012, 26: 40-47
- [9] Formica A, Missikoff M. Concept similarity in SymOntos: an enterprise ontology management tool[J]. The Computer Journal, 2002, 45(6): 583-594
- [10] Formica A. Ontology-based concept similarity in formal concept analysis[J]. Information Sciences, 2006, 176(18): 2624-2641
- [11] Biere A, Cimatti A, Clarke E M, et al. Bounded model checking [J]. Advances in computers, 2003, 58: 117-148
- [12] Bayati M, Shah D, Sharma M. Maximum weight matching via max-product belief propagation[C]//Proceedings of the 2005 IEEE International Symposium of Information Theory. 2008, 54: 1763-1767
- [13] Yellin D M. Speeding up dynamic transitive closure for bounded degree graphs[J]. Acta Informatica, 1993, 30(4): 369-384
- [14] Sankowski P, Mucha M. Fast dynamic transitive closure lookahead[J]. Algorithmica, 2010, 56(2): 180-197
- [15] 舒虎, 崇志宏, 倪巍伟, 等. X-Hop: 传递闭包的多跳数压缩存储和快速可达性查询[J]. 计算机科学, 2012(3): 144-148

Transportation Systems: Technology, Planning, and Operations, 2007, 11(4): 157-167

- [5] Liu H, Zuylen H V, Lint H V, et al. Predicting urban arterial travel time with state-space neural networks and Kalman filters [C]//Transportation Research Board, Annual Meeting (CD-ROM). Washington, 2006: 99-108
- [6] Park J, Chen Z, Kiliaris L, et al. Intelligent vehicle power control based on machine learning of optimal control parameters and prediction of road type and traffic congestion[J]. IEEE Transactions on Vehicular Technology, 2009, 58(9): 4741-4756
- [7] Yu B, Lam W H K, Tam M L. Bus arrival time prediction at bus stop with multiple routes [J]. Transportation Research Part C: Emerging Technologies, 2011, 19(6): 1157-1170
- [8] Shen X, Chen J. Study on prediction of traffic congestion based on LVQ neural network [C]//International Conference on Measuring Technology and Mechatronics Automation. 2009: 318-321
- [9] Rahman H A, Marti J R, Srivastava K D. Road traffic forecasting through simulation and live GPS-Feed from intervehicle networks [C]//Global Humanitarian Technology Conference (GHTC). 2012: 36-40
- [10] 于滨, 杨忠振, 林剑艺. 应用支持向量机预测公交车运行时间 [J]. 系统工程理论与实践, 2007, 27(4): 160-164
- [11] 孙玉砚, 刘燕, 周新运, 等. 基于路况相似性的城市公交车到站时间预测机制[J]. 软件学报, 2012, 23(zk1): 87-99
- [12] 曹琛荔. 基于神经网络的智能公交信息服务系统研究[D]. 武汉: 武汉理工大学, 2011