

文章编号:1005-0523(2018)01-0046-09

基于预测算法的公交车实时到站时间查询系统

王 茁

(大连科技学院交通运输学院,辽宁 大连 116052)

摘要: 公交车是目前城市中最普遍的一种大众运输工具,随着城市化和机动化的发展,使城市人口和地域不断增加,对公共交通的需求相应快速增长,出行者对公交系统提出了更多更高的要求。公交车辆到站时间是城市智能公共交通系统服务的重要指标参数之一,也是乘坐公共交通出行者最为关注的基本信息之一,为满足公交出行者对所乘线路公交车辆的到站时间可实时查询的需要,介绍城市公交实时到站时间查询系统的设计。在需求分析的基础上,通过概念结构、逻辑结构的设计建立了查询系统的数据模型,提出了一种到站时间的预测算法。该算法将路线划分成站间路段,以实时收集到的车辆进出站信息为基础,从历史纪录中得到通过每个路段的运行时间。通过聚类分析、判别分析的方法确定预测的运行时间,最后通过简单计算得到用户需要的预测到站时间,并且将到站时间通过网络公开展示给系统的用户。

关键词: 公交车辆;到站时间预测;预测算法;查询系统

中图分类号: U491

文献标志码: A

由于天气、交通流、公交需求等因素随时间和空间的变化,使得城市公交车辆不能按照所计划的时间到达公交停靠站点,给公交出行者的出行规划造成诸多不便。为提高公交服务的质量,吸引更多的出行者选择公交出行方式,利用公交车辆装载的自动定位系统的数据,设计有效的算法,估计车辆的实时到站时间,并开发基于互联网的查询系统,可以使公交出行者通过网络了解所要乘坐的公交车辆的实际到站时间,进行合理的公交出行规划,提高出行的效率^[1-4]。通过进一步的研究,本系统还可以移植到手机上,用户可以通过手机查询指定车辆的实际到站时间。

1 城市公交车实时到站时间查询系统的数据库设计

1.1 需求分析

目前我国城市公交车站大多采用传统的导向式站台,提供线路的基本信息,无法为乘客提供确切的车辆到达信息,乘客不能准确估计在车站的等待时间,公交运营服务可靠性无法保障。公交车实时到站查询系统的开发目的:实时预测公交车的到站时间,并将其发布给用户。到站时间查询系统初步设计流程,如图1所示。

从图1可以看到,系统中数据库首先实时从交通管理部门接收公交车出行数据,通过转换数据格式后实时加入数据库,即输入为车辆到站信息表。但用户需要查询到站时间时,通过

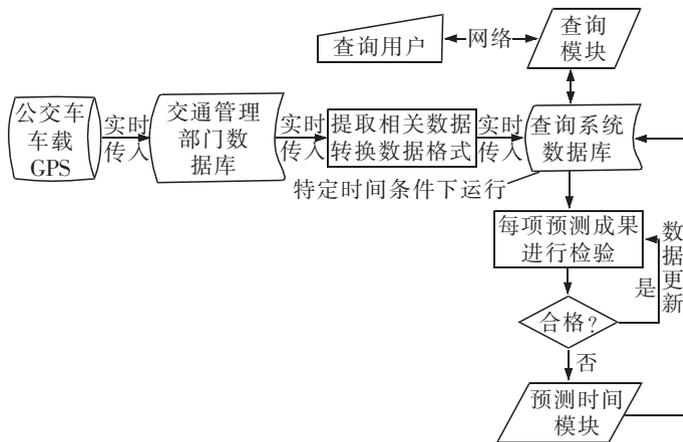


图1 到站时间查询系统流程图
Fig.1 The flow chart of the system

收稿日期:2017-08-31

作者简介:王茁(1987—),女,讲师,主要研究方向为交通规划,交通运输。

网络发布需求信息,由系统查询模块处理后从数据库提取相应数据。最后当特定条件下(预测准确度降低或者到达更新周期),利用预测模块进行数据更新,数据库输出为车辆到站时间。进一步分析系统和表达用户需求,系统的数据流程图如图 2 所示。

1.2 概念结构设计

完成需求分析后,还应该将现实世界的事物转换为机器世界能处理的数字信息,这需要数字化和抽象化。通常概念模型是以 E-R 图的方式表示出来的。E-R 图的设计按照先局部,在总体的思路进行,也就是现根据需求分析的结果,将系统划分为 3 个子系统,按子系统设计分 E-R 图,然后再将分 E-R 图集成,如图 3 所示。

1.3 逻辑结构设计

逻辑结构设计的内容简单的说就是将概念结构模型设计结果 E-R 图转换为数据库管理系统支持的数据模型。逻辑设计的步骤一般分为 3 步^[12],分别是:

1) 将概念结构转换为关系模式。

首先,将 3 个实体转换为关系模式的结果为:

Bus (busID, routeID, stopNo)

Route (routeID, stopID, stopNo)

Busstop (stopID, stopName)

其次,将车辆—站点间联系转化为关系模式,由于这个联系都是多对多的类型,采用实体分开的方法,结果为:

Case (caseID,busID,stopID,time,dir,week,mode,date,class)

最后,计算过程中可能会有需要记录的数据,故增加两个关系模式。

Temp (runningtime, week, time)

Result (stopID, dir ,...)

2) 设计数据的完整性。已确保主键的实体完整性,由于联系转换时没有合并的方式,因此参照性已经完整不用检查。

3) 对数据模型进行优化。经过这个过程后最终的逻辑数据模型的关系数据模型如下:

Bus (busID, routeID, stopNo)

Route (routeID, stopID, stopNo)

Busstop (stopID, stopName)

Case (caseID, busID, stopID, time, dir, week, mode, date, class)

Temp (runningtime, week, time)

Result (stopID, dir ,...)

1.4 数据库数据字典设计

对每一项数据模型的含义和组成进行叙述便得到了形同数据库的数据字典。数据库数据字典设计如表 1~表 6 所示。

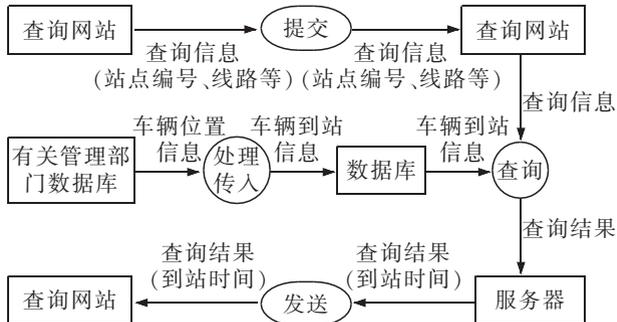


图 2 系统数据流程图

Fig.2 System data flow chart

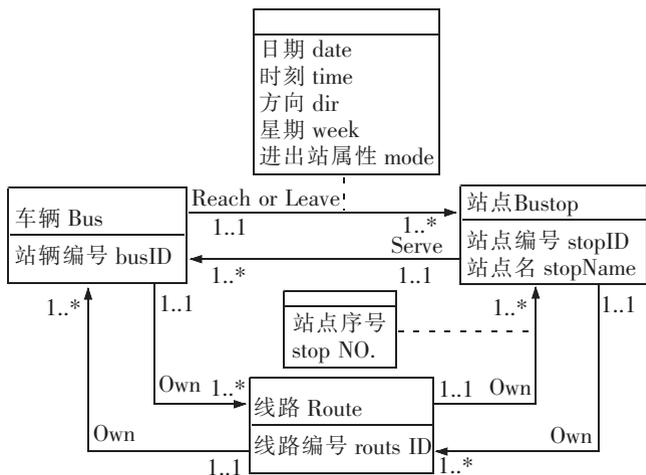


图 3 公交车与公交站点的 ER 图

Fig.3 ER diagram of bus and bus stops

表 1 系统数据字典
Tab.1 System data dictionary

数据表名	含义说明	组成
车辆信息表	记录车辆信息	车辆编号,线路编号
站点信息表	记录站点信息	站点编号,站点名称
进出站信息表	记录车辆进出站时间和方向信息	事件编号,车辆编号,站点编号,时刻,方向,星期,进出站属性,日期等
聚类计算临时表	临时存储聚类分析所需数据	站间运行时间,星期,时刻
判别分析表	记录判别分析结果	站点编号,方向,判别方程系数数据

表 2 车辆信息表数据项定义
Tab.2 The data item definition of the vehicle information table

数据项名	含义说明	别名	数据类型	长度
车辆编号	表示车辆的唯一标志	busID	变长字符型	8
路线编号	表示车辆的线路归属	route	数字整型	3
线路站点序号	—	stopNO	数字整型	2
行车方向	—	dir	布尔型	1

表 3 站点信息表数据项定义
Tab.3 The data item definition of the site information table

数据项名	含义说明	别名	数据类型	长度
站点编号	表示站点的唯一标志	stopID	变长字符型	8
站点名称	站点名称	stopName	文本	20

表 4 进出站信息表数据项定义
Tab.4 The data item definition of the entry and exit information table

数据项名	含义说明	别名	数据类型	长度
事件编号	表示进出站信息的唯一编号	caseID	字符型	8
车辆编号	表示车辆的唯一标志	routeID	数字整型	3
车辆编号	表示车辆的唯一标志	busID	字符型	8
站点编号	表示站点的唯一标志	stopID	变长字符型	8
时刻	记录事件发生时间	time	时间	—
方向	记录车辆行驶方向	dir	布尔型	1
星期	记录事件发生星期	week	数字整型	1
进出站属性	进站为 0,出站为 1	mode	布尔型	1
月份	事件发生月份	date	日期	—

表 5 聚类计算临时表数据项定义
Tab.5 The data item definition for the temporary table of clustering calculation

数据项名	含义说明	别名	数据类型	长度
站间运行时间	记录上一站到达此站所用时间	running time	时间	—
星期	记录事件发生星期	week	数字字段	1
时刻	记录事件发生时间	time	时间	—

表 6 判别分析表数据项定义
Tab.6 Data item definition for the discriminant analysis table

数据项名	含义说明	别名	数据类型	长度
站点编号	表示站点的唯一标志	stopID	变长字符型	8
方向	记录车辆行驶方向	dir	布尔型	1
判别方程系数数据	记录判别方程系数	Panbie	整型	4

2 城市公交车实时到站时间预测算法

城市公交车实时到站时间查询系统主要考虑实现 3 项查询功能, 分别是最近公交车到站时间查询、无换乘的车辆到站时间预测和有换乘的车辆到站时间预测^[10]。要实现这 3 个功能, 系统分 4 个模块设计, 示意图如图 4 所示。

第 1 为公交车相邻站间行程时间预测模块。它是实现所有功能的基础, 它的目标是求出在指定时间下公交车在两个相邻站之间运行的时间。

第 2 为最近公交车查找模块, 目标是找到距离查询站最近的公交车出发站和出发时刻。

第 3 为无换乘的车辆到站时间预测模块, 即在不考虑换乘的情况下, 在最近车到站时间模块的基础上得到最近车辆到达目标站的耗用时间。

最后为有换乘的车辆到站时间预测模块, 考虑起始站和目的站在不同的线路上, 通过中转站换乘所耗用的等待时间和车辆运行时间。

2.1 公交车相邻站间行程时间预测算法

公交车相邻站间行程时间的预测, 是按照假设构建的, 即在时间和星期不改变的情况下(不考虑天气和突发事件), 城市道路上的交通状况是相似的^[11]。

相邻站间的行程时间可以通过从历史事件找出时间与星期的规律, 由此规律对相邻站间到站时间进行预测。前一个过程在数据挖掘中叫做聚类分析, 是对事前没有类别归属的数据对象按照其内在的某种联系划分类别。而后一个过程被称为是判别分析, 判别分析的基本问题是根据不知所属的观察值 X , 去判别观察值属于哪一组的过程。

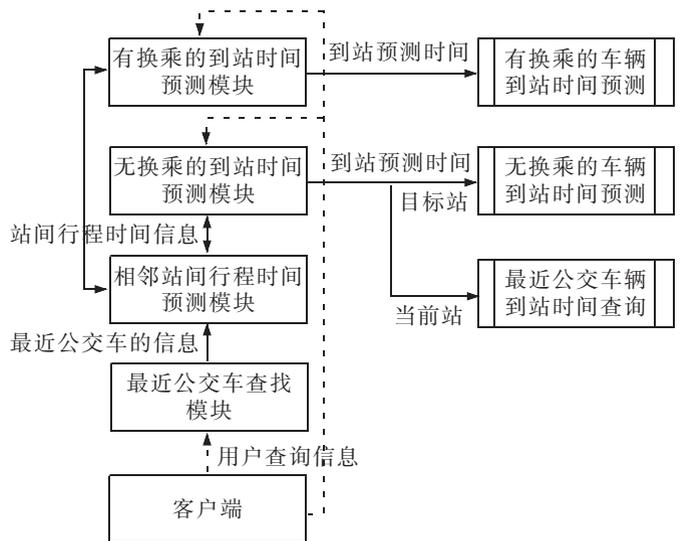


图 4 到站时间查询系统功能实现模块图
Fig.4 Functional implementation module of arrival time query system

首先对历史到站时间进行聚类分析,将历史的数据记录以到站时间为度量按照“格局及内部对象间的相似程度最大化和格局及对象间相似度最小化”分成若干个类别。每一个类别的数据记录间的相似程度要比不同类别的数据对象间的相似程度要大,可以认为具有相近到站时间的数据记录被分在了同一个类别里,即对数据记录进行类别划分。

然后从类别中寻找规律到站时间,基于判别分析的站间到站时间预测则是通过将查询时输入的时间、星期按照 k-最近邻的方法加入某个类别中,从类别中得到预测的到站时间。

2.2 最近公交车查找算法

在实时状态下,通过更新进出站信息表时,当车辆出发时将该站点编号 stopNo 写入车辆信息表中出站站点编号。这样在寻找某站点的最近车的步骤如下:

- 1) 在线路信息表中由所在站点编号 A 和方向找出上一站编号 B。
- 2) 在车辆信息表中寻找是否最近有车从上一站 B 出发(车辆信息表中出战站点编号=B 的站点编号)。
- 3) 如果寻找到则返回 B 的站点编号 stopNo 和出发时间 time,否则将 B 赋予 A 重复步骤 1,步骤 2,直到找到或者寻找到起始站为止。这样便实现了寻找最近公交车的工作。

2.3 无换乘的车辆到站时间预测算法

无换乘的车辆到站时间预测的流程,如图 5 所示。

首先取得最近车出发的站点编号 stopNo 和出发时刻 time,然后由站点编号 stopNo 和方向 dir 用相邻站间行程时间预测模块预测此站到下一站的行程时间 t ,则下一站的到站时间 T 有下面公式计算得到:

$$T = t - (\text{currentTime} - \text{time}) \quad (1)$$

式中:currentTime 为查询时的时刻。

然后判断下一站是否为目标站,否则继续预测站间得行程时间加入 T_{total} 中,如果到达目标站则输出预测到站时间 T_{total} 。完成公交车到站时间预测

最近公交到站时间与线路上的车辆到达时间原理基本一致,只是运算结束条件的目的站是出行者的出发站和出行者的目的地站的区别。当然这对出行者同样有用,故将它们作为无换乘的到站时间预测模块。

2.4 有换乘的车辆到站时间预测算法

考虑换乘时将车辆的到达时间分为几个部分,起始车辆到达中转站时间,中转站到下一个中转站时间以及中转站到目的站时间。

这三类时间都可以用线路上车辆到达时间预测模块计算,只需要将过程中的时间相加即可。换乘站的选择由出行者输入,即可在系统中实现。

3 基于预测算法的公交车实时到站时间查询系统设计

3.1 系统总体设计

GPS 技术在城市公共交通中,主要应用安装在在出租车和公交车得车辆上,用作车辆跟踪器。可以在监控中心很方便的看到车辆所在的区域和行动路线,达到及时发现运行问题,及时解决问题。另一方面,从 GPS 技术运用在公交车上时,还能记录车辆的到站时间,这就使得车辆到站时间预测可以实现。通过收集车

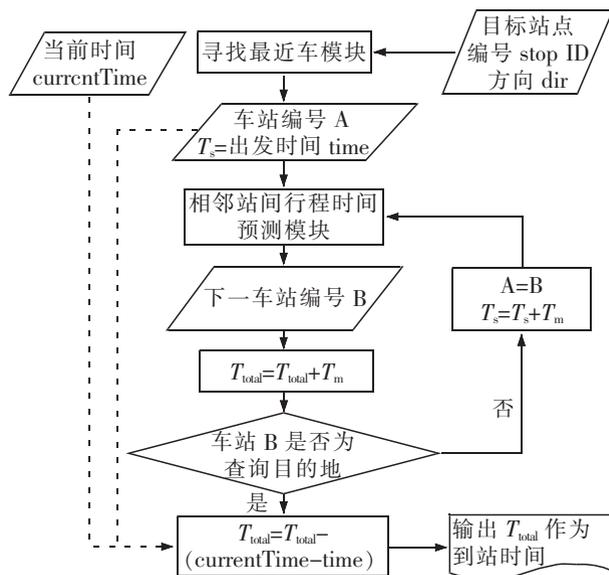


图 5 无换乘车辆到达时间预测流程

Fig.5 Arrival time prediction process for non-transfer bus

辆历史的到站时间信息,通过网络服务器,与用户交互查询。这样通过公交车到站时间预测查询系统即可以在服务器端非常方便地对公交车的某公交线路上前时间下的到站时间进行预测,最主要的是可以在客户端通过网络实现对车辆到站时间进行远程查询和分析,从而在一定程度上使公交车的信息发布和查询更加数字化、网络化和智能化^[12]。

明确系统的需求,要求该系统要实现以下功能:

- 1) 预测到站时间。在系统的客户端,可以实现由输入的公交车站点、出行方向的信息查询最近一辆公交车的预测到站时间。
- 2) 预测无换乘的线路上的车辆运行时间。在系统的客户端,由公交出发站点、公交目的站点,乘坐线路和到达出发站的时间查询出车辆运行时间。
- 3) 预测有换乘的车辆运行时间。由道路连接属性或者乘车人的意愿确定换乘路钱,查询整个出行的耗时。对有多个换乘方案的通过耗时确定最优方案。

3.2 系统架构设计

本系统基于 Internet/Intranet 标准,采用 MySQL 数据库管理系统、PHP 服务器内嵌语言,以及 web 服务器 Apache 等技术构建的实现公交实时到站时间查询的功能。查询系统是基于 B/S 的三层结构,如图 6 所示。

从表示层、应用程序逻辑层和数据层搭建逻辑系统,在具体实现中三层的结构被分别划分为客户端、应用服务器和数据服务器。使用标准 Web 浏览器的客户端,如 Internet explorer 和 Netscape,主要功能仍然是管理用户界面、应用程序逻辑、数据请求、向服务器发送请求,并接受服务器处理结果。应用服务器的功能是集中的管理核心资源,处理客户端的发送请求,是系统资源在客户端和服务器上合理配置,使用 Apache 服务器实现的线路查询系统。空间数据库服务器使用 MySQL 数据库来组织、存储和管理线路、站点,以及用于地理的空间数据和属性数据^[6-9]。

基于 B/S(浏览器/服务器)的三层结构体系的设计将数据的存储、管理与应用分开,可方便地实现数据库系统对各类数据的组织、管理、应用和发布,既能够使系统的稳定性和可扩展性得到提高,同时又保证了数据的安全性^[13]。

3.3 系统功能实现

3.3.1 查询系统主体实现流程

数据来源:目前系统采用外部静态导入,系统已导入的数据包括 GPS89(哈尔滨市 89 路公交车 4 月 19 日至 5 月 17 日车辆 GPS 进出站数据),GPS63(哈尔滨市 63 路公交车 5 月 5 日至 5 月 24 日车辆 GPS 进出站数据)。

1) 将数据库的相关数据导入 MySQL 中,导入后作 k-means 聚类分析,得出每条记录得所属种类,一起存入车辆进出站信息表。然后在数据库中根据需要人工构建车辆信息表和站点信息表并输入车辆信息和站点信息。

2) 当用户端发送查询信息(包括线路、出发站点、方向、目的站点和出发时间)时,由服务器接受并将他们分别存入变量 route, dir, stop, endstop 和 time 中。然后将其直接显示给用户。

3) 之后开始寻找下一个站点编号。这里先查找 stop 的站点序号 stopNo,根据 dir 计算下一站点编号,并将其赋给 stopNo。接着利用寻找最近车模块查找最近车的出发时间 starttime 和出发站点编号 stopID。

4) 转换当前时间 time 和 starttime 的时间为时间戳格式,并计算它们之间的时间差。

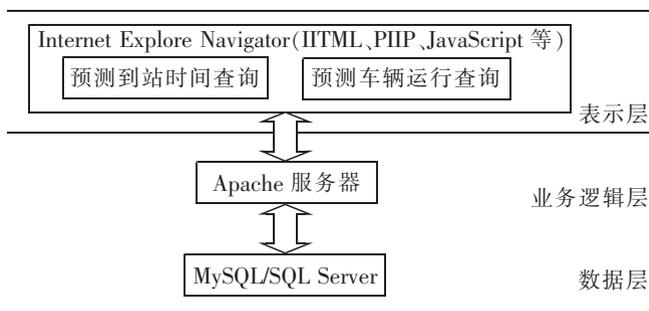


图 6 查询系统构架

Fig.6 The architecture of query system

5) 由 starttime, stopNo, dir, route 计算到下一站的运行时间,利用 runtimetemp 累计运行时间。

6) 判断结束条件,是否下一站为查询站或者为查询目的站。否则重复上面第3)~第5)步,直到结束条件成立停止计算输出结果。

3.3.2 客户端系统应用流程

本公交车查询系统的功能主要包括:最近公交车到站时间、单线路车辆到站时间查询、有换乘的车辆到站时间查询。网络查询界面如图7所示。

查询时用户根据所查内容。在查询界面中选择型输入信息。点击提交键,系统会自动返回所查信息。下面分别对实现的功能进行查询说明。

1) 最近公交车到站时间查询。最近公交车到站时间查询是本系统的核心,当公交查询者在 APP 查询界面上输入相关的信息,系统将调用相关算法进行计算,找出最近的公交车辆,继而进行到站时间预测^[14]。当然为了方便站台上的乘客了解到站时间也可以在站台的站牌上安装显示屏,由服务器向显示器发送调整信号,显示器按照时间流逝递减,当公交车到站时,系统会从新计算时间,再将调整后的时间发送,使得显示具有实时更新的特点,增加准确性和稳定性^[15-17]。

当要进行最近公交车到站时间查询时,需要用户在界面中选择线路、行车方向。目的站点和查询时间可以默认不选。点击查询按钮,则出现查询结果界面如图8所示。由此用户可以清楚地看到所查信息为63线路上,开往达江小区方向,到达哈工大站的最近车辆刚刚从第10站出发,到达哈工大站需要等待5分钟。用户可以选择等待。

2) 无换乘车辆到站时间查询。而进行无换乘车辆到站时间查询(图7)时,需要输入目的站,点击查询按钮。则出现查询结果界面如图9所示。

由于输入了目的地站点是达江小区,故显示结果为增加了该车辆到达沿线站点所需要的运行时间(这里为了表示过程将中途的预测到站时间全部显示出来)。这样乘客就可以了解到在这一站候车并乘坐该车辆到达目的站一共需要花费的时间。用户默认情况下,系统查询时间为当前时间。当用户并不想以当前时间作为查询时间时,可在出发时间处填写出发时间。这样系统会更改变查询出发时间,从而实现用户查询当天未来时间的车辆到站时间。

此外,对于有换乘的车辆到站时间可以由上面功能叠加实现,将有换乘的到站时间分为其实线路到达中转站时间,中转站到达中转站时间和中转站到达目的站时间,分别用系统进行预测然后相加即可得到。



图7 系统查询界面
Fig.7 Interface of query system



图8 查询结果界面
Fig.8 Interface of query result



图9 单线路到站时间查询结果
Fig.9 Arrival time query result of single-line bus

3.4 实时到站时间查询系统成果检验

实时到站时间查询系统中的到站时间是根据实时时间信息加上相邻站间运行时间得来,由于目前没有实时传入对到站时间直接进行误差分析不能实现。而相邻站间的预测运行时间是预测到站时间的基础,因此它的误差可以间接反映最终预测值的误差^[18]。

从公交 GPS 数据库选取 2016 年 5 月 20 日—24 日的记录,每日数据中随机选取 50 个数据记录作为样本,各样本通过系统产生相应预测运行时间,进行标准差分析。由分析结果可知最大误差为 60%,平均误差时间为 40 s,预测时间差最大为 68 s,误差均方差为 23.2。相对于预测时间来说误差较大,但对于实际情况来说,满足要求。而且由于系统会引进实时信息修正,最终结果可以接受。

4 结语

本公交到站时间查询系统的设计是以 webGIS 的 B/S 三层体系结构为基础,采用 html 和 PHP 等技术完成客户端和服务端的设计,从而实现城市公交查询系统的计算。运用 MySQL 数据库达到空间数据和属性数据一体化的存储效果作为数据库服务器。早前研究多采用静态计算,不仅算法复杂,预测结果也不令人满意。这些不符合公交运行的特点,同时与人们在等待公交车时仅能提供目的地出发站点和出发时间实际情况也不相符。本系统在查询功能上实现了查询结果实时反馈给客户端的显示功能,使车辆到站时间随着车辆的进出站而实时更新,不仅信息获取更直观清晰,且对公交车辆的实时位置可以估计。

参考文献:

- [1] 周雪梅,杨晓光,王磊. 公交车行程时间预测方法研究[J]. 交通与计算机,2002(6):12-14.
- [2] AGAFONOV A A, CHERNOV A V, SERGEEV A V. Using satellite monitoring and statistical data to predict arrival time of city public transport[J]. Pattern Recognition & Image Analysis, 2015, 25(3): 385-388.
- [3] 涂颖菲,陈小鸿. 法国城市旅客手机信息服务系统[J]. 城市交通. 2007, 5(2): 86-90.
- [4] 韩 印,杨兆升,胡坚明. 公共交通智能化调度基础数据预测方法研究[J]. 公路交通科技, 2006(6): 140-143.
- [5] 嵇绍娟. 基于 Net 的综合管理平台接处警信息系统设计与实现[D]. 苏州:苏州大学,2013.
- [6] 康丽珠. 基于 Web 的企业竞争情报系统中权限构件及短信系统的设计与实现[D]. 昆明:昆明理工大学,2008.
- [7] 陈国俊,杨晓光,刘好德,等. 基于动态百分位行程时间的公交到站时间预测模型[J]. 公路交通科技,2009(26): 102-106.
- [8] 陈海燕,刘珍. 基于 Google Maps 的地理信息标注系统的设计与实现[J]. 湖南科技大学学报,2009,24(3): 87-90.
- [9] 李天雷,王爱民. 基于 GPS 数据的公交行程时间计算与预测系统[D]. 吉林:吉林大学,2009.
- [10] RAMAKRISHNA Y, RAMAKRISHNA P, LAKSHMANAN V, et al. Use of GPS probe data and passenger data for prediction of bus transit travel time[C]//transportation Land Use, Planning, and Air Quality Congress, 2008: 124-133.
- [11] SIMROTH A, ZAHLE H. Travel time prediction using floating car data applied to logistics planning[J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(1): 243-253.
- [12] 陶佩枫. 城市公交查询系统的设计与实现[D]. 长沙:中南大学,2008.
- [13] 杨亭亭. 基于时空可达的城市公交出行信息系统[D]. 哈尔滨:哈尔滨工业大学,2010.
- [14] 徐健华. 基于 WebGIS 赣州市公交查询系统的研究[D]. 赣州:江西理工大学,2008.
- [15] LI JINGLIN, GAO JIE, YANG YU, et al. Bus arrival time prediction based on mixed model[J]. China Communications, 2017, 14(5): 38-47.
- [16] CATS O, LOUROS G. Real-time bus arrival information system: an empirical evaluation[J]. Journal of Intelligent Transportation Systems, 2016, 20(2): 138-151.
- [17] TAN C W, PARK S, LIU H, XU Q, et al. Prediction of Transit Vehicle Arrival Time for Signal Priority Control: Algorithm and Performance[J]. IEEE Transactions on Intelligent Transportation Systems, 2008, 9(4): 688-696.
- [18] 阳衡,刘成良. 公交车行驶安全智能监控系统开发与设计[D]. 上海:上海交通大学,2008.

Real-time Bus Arrival Time Query System Based on Prediction Algorithm

Wang Zhuo

(Faculty of Transportation Engineering, Dalian Institute of Science and Technology, Dalian 116052, China)

Abstract: Buses are the most popular means for urban mass transportation. With the development of urbanization and motorization, the population of cities and the scale of regions have been on the rise, which has brought greater demands for public transportation and higher requirements for public traffic system. Bus arrival time, as one of the important parameters in urban intelligent public transport system service, is the basic information that public transport travelers are most concerned about. To meet the needs of the bus travelers to query the bus arrival time in real time, this paper explores the design of real-time arrival time query system of city buses. Based on the requirement analysis, the data model of the query system was established through the design of conceptual and logical structure. Then, the prediction of arrival time algorithm was put forward, which divided the route into the station sections and obtained the running time of each section from the historical records based on the data of bus arrival and departure collected in real time. The predicted running time was confirmed by clustering analysis and discriminant analysis. Finally, the prediction of arrival time for users was obtained by simple calculation and the arrival time was displayed to the system users through the network.

Key words: bus; prediction of arrival time; prediction algorithm; query system