

# 基于轨迹挖掘的公交车自组织网络路由机制

张扶桑<sup>1),2)</sup> 金蓓弘<sup>1)</sup> 汪兆洋<sup>1),2)</sup> 胡佳锋<sup>1),2)</sup> 张利锋<sup>3)</sup>

<sup>1)</sup>(中国科学院软件研究所计算机科学国家重点实验室 北京 100190)

<sup>2)</sup>(中国科学院大学 北京 100190)

<sup>3)</sup>(鲁东大学信息与电气工程学院 山东 烟台 264025)

**摘 要** 数据传递是车载自组织网络(VANET)应用不可或缺的组成部分. 在城市中, 公交车网络具有覆盖面广、行驶线路固定等特点, VANET 应用中的数据传递可以通过公交车自组织网络实现, 但公交车网络能否在 VANET 数据传递中发挥其独特的作用取决于对公交车运行规律的挖掘和利用的程度. 文中提出了一种公交车自组织网络的路由机制 Vela, 该机制依托从公交车历史轨迹中挖掘的公交车运行时空规律, 建立了公交车数据传递的概率时空模型, 并给出了基于路段上公交线路相遇的数据传递路径的选择策略. 与已有相关研究相比, Vela 路由机制一方面以路段为粒度刻画了公交车运行的时空规律, 这使得所获得的公交车运行规律兼具准确性和稳定性, 另一方面, 基于公交车运行规律估算路段上数据传递的延迟和可靠性, 能获得尽可能好的服务质量级别的数据传递路径. 实验结果表明采用 Vela 路由机制能实现具有高到达率和低延迟的数据传递, 而且 Vela 路由机制具有较强的可伸缩性.

**关键词** 车辆自组织网络; 基于公交车的路由; 轨迹挖掘; 时间序列分析

中图法分类号 TP393

DOI 号 10.3724/SP.J.1016.2015.00648

## A Routing Mechanism over Bus-Based VANETs by Mining Trajectories

ZHANG Fu-Sang<sup>1),2)</sup> JIN Bei-Hong<sup>1)</sup> WANG Zhao-Yang<sup>1),2)</sup> HU Jia-Feng<sup>1),2)</sup> ZHANG Li-Feng<sup>3)</sup>

<sup>1)</sup>(State Key Laboratory of Computer Sciences, Institute of Software, Chinese Academy of Sciences, Beijing 100190)

<sup>2)</sup>(University of Chinese Academy of Sciences, Beijing 100190)

<sup>3)</sup>(School of Information and Electrical Engineering, Ludong University, Yantai, Shandong 264025)

**Abstract** Data delivery is an indispensable component of applications over Vehicular Ad Hoc Networks (VANETs). In urban scenarios, bus-based networks have distinctive features such as wide coverage and fixed lines so that they can be employed to achieve data delivery in VANET applications. However, whether a bus-based VANET can play its unique role in data delivery depends on how to effectively mine and utilize spatial-temporal rules implied in bus-based networks. This paper proposes a bus-based ad hoc routing mechanism Vela. By mining spatial-temporal rules from bus historical trajectories, Vela builds a probabilistic spatial-temporal model for data delivery and then gives the routing selection strategy based on patterns of bus line encountering on road segments. Compared with existing solutions, the routing mechanism Vela characterizes the spatial-temporal rules of buses in a moderate granularity of road segments, which makes the acquired spatial-temporal rules of buses both accurate and steady. On the other hand, Vela estimates data delivery delay and reliability based on the spatial-temporal rules, thus obtaining the routing paths with the best possible QoS (Quality of Service) levels for data delivery. The

收稿日期: 2014-02-27; 最终修改稿收到日期: 2014-10-27. 本课题得到国家自然科学基金(61472408, 61372182)资助. 张扶桑, 男, 1987 年生, 博士研究生, 中国计算机学会(CCF)会员, 主要研究方向为分布式计算、普适计算. E-mail: zhangfusang10@otcaix.iscas.ac.cn. 金蓓弘(通信作者), 女, 1967 年生, 博士, 教授, 博士生导师, 中国计算机学会(CCF)高级会员, 主要研究领域为分布式计算、移动与普适计算、中间件和分布式系统. E-mail: beihong@iscas.ac.cn. 汪兆洋, 男, 1989 年生, 博士研究生, 中国计算机学会(CCF)会员, 主要研究方向为分布式计算、普适计算. 胡佳锋, 男, 1988 年生, 博士研究生, 主要研究方向为分布式计算、普适计算. 张利锋, 男, 1977 年生, 博士, 中国计算机学会(CCF)会员, 主要研究方向为分布式计算、普适计算.

experimental results show that Vela performs much better in terms of delivery ratio and delay, and have the stronger scalability than other solutions.

**Keywords** vehicular ad hoc networks; bus-based routing; trajectory mining; time series analysis

## 1 引言

车联网(Vehicular Ad hoc Network, VANET)是车辆节点按自组织方式组成的一个移动无线通信网络. 基于 VANET 及相关技术(如传感器技术、智能处理技术等), 可以实现包括车辆主动安全、智能交通管理、城市生活服务、应急救援等在内的多种应用. 可以预见, 随着 VANET 的发展, 相较于 3G、WiFi 等无线网络, VANET 在部署代价和使用费用方面的潜在优势, 使得它有望成为智慧城市中开放的数据交换平台, 并在多个领域发挥作用<sup>[1]</sup>.

公交车是城市公共交通系统中的重要交通工具, 公交车在安装了车载无线通信单元(On Board Unit, OBU)后, 便可构成公交车自组织网络. 由于公交车网络具有覆盖面广、行驶线路固定、起始站发车时间间隔相对稳定等特点, 因此依托公交车自组织网络实现数据传递有望成为多种 VANET 应用的基础.

本文研究公交车自组织网络的数据传递问题. 在公交车自组织网络中, 公交车节点的快速运动会使得网络拓扑变化频繁, 通信链路生存周期较短, 数据传递路径不稳定, 网络分区时常发生, 这都会直接导致数据传递服务质量的抖动, 而城市建筑物、各类公共设施等对无线信号传播的阻挡和干扰会进一步加剧网络通信质量的不稳定性. 因此如何挖掘公交车的运行模式来提高数据传递服务质量, 是公交车自组织网络数据路由机制面临的一个挑战. 换言之, 针对公交车网络准确挖掘公交车行驶的时间特征和相遇的空间特征, 并有效应用到数据路由中, 是个值得研究的问题.

本文针对 V2V(Vehicle to Vehicle)通信场景提出了一种数据路由机制 Vela, 用于将数据从某个公交车节点传递至某个目的路段. 该机制基于公交车在路段上的延迟估计和相遇规律来决定数据传递的路径, 它包含以下 3 部分: (1) 为了获得公交车运行的时空规律, 依托公交车在路段上的历史运行时间的统计数据, 通过相关性分析, 建立车辆的路段运行时间的自回归滑动平均模型; 同时挖掘了不同公交线路的公交车在各路段上的相遇规律, 建立相遇

概率模型; (2) 利用上述模型所反映的时空规律, 预测在路段上传递数据的时间和可靠性, 建立了公交车数据传递的概率时空模型; (3) 发起数据传递的公交车基于概率时空图计算到达目的路段的具有最短延迟的前  $k$ (Top- $k$ ) 条路径, 并选择前  $k$  条路径中可靠性最高的路径, 随后, 按此路径路由数据, 若在路由时未能遇到预定的车辆节点, 则重新计算从当前车辆节点到目的地的最优路径, 并继续前述过程. 由于 Vela 机制是在可容忍的数据延迟基础上, 选择可靠性最高的路径进行数据传递, 因而它能为数据传递提供尽可能高的服务质量级别.

本文的主要贡献包含以下几个方面:

(1) 通过对真实公交车轨迹的统计分析, 发现了公交车轨迹的一些基本特征; 运用时间序列分析方法, 给出了预测公交车在路段上的运行时间的模型, 并且通过实验验证了预测方法的有效性.

(2) 以路段为粒度挖掘公交车历史运行轨迹, 获得了公交车运行的时空规律. 与粗粒度(即公交线路级别)规律挖掘相比, 该方法挖掘出的规律的准确性更高; 与细粒度(即公交车级别)规律挖掘相比, 该方法挖掘出的规律的稳定性更高.

(3) 针对公交车自组织网络, 提出了一种新的数据路由机制 Vela, 该机制利用公交车运行时空规律建立数据传递的概率时空图, 并以满足数据传递不同的服务质量为目标选择数据传递的路径, 从而为 VANET 数据传递提供高性能和高可伸缩性的路由.

本文第 2 节介绍相关工作; 第 3 节给出本文要解决的问题, 并对公交车轨迹数据进行了初步分析; 第 4 节给出公交车数据传递的概率时空模型; 第 5 节描述基于概率时空模型的路由机制; 第 6 节是对所提路由机制的实验评估; 最后是全文总结.

## 2 相关工作

近年来, VANET 上数据传递机制的研究非常活跃. VANET 上进行数据传递的基本策略是存储和转发, 即车辆节点携带数据前行, 直到遇到可以转发数据的节点, 而数据的转发策略通常可以采用基于地理位置的贪心策略(如 GPCR<sup>[2]</sup>), 或采用基于

最短数据传递延迟的贪心策略,后者涉及的数据传递延迟可以采用历史交通流数据进行估算(如 VADD<sup>[3]</sup>、TBD<sup>[4]</sup>)、或直接采用实时的交通流数据(如文献[5])、或两者相结合的方法获得(如文献[6])。

在 VANET 数据路由中,车辆的相遇是数据转发的前提条件,因此车辆的相遇关系(如频度、空间分布、相遇间隔等)被广泛挖掘并用于数据路由决策,例如:文献[7]通过共享轨迹构造了车辆相遇图,并基于此进行路由,文献[8]利用马尔可夫链预测车辆相遇间隔时间,并根据邻居节点与目标车辆的相遇间隔时间决定数据传递的下一跳车辆节点。文献[9]采用与文献[8]相同的相遇延迟估计方法,并进一步挖掘了社会层面的移动性,构造自我联系(ego contact)图,这样车辆在缺少邻居车辆与目的车辆的相遇延迟估计时,会选取在自我联系图中占据更重要位置的邻居车辆,进行数据传递。

考虑到公交车网络的特点,有些研究工作挖掘了公交车节点的相遇概率,并基于此设计了公交车网络的路由机制。

MaxProp<sup>[10]</sup>是在真实的校园公交车网络(Umass DieselNet)中使用的路由方法,它使用携带转发和数据包优先级技术传递数据。其中,数据包的优先级是由传递数据到目的地的开销来决定的,而开销又是通过记录两个公交车相遇的可能性来估计的。当公交车遇到另一辆公交车时,转发具有最高优先级且以该公交车为转发节点的数据包。

针对节点之间有周期性相遇规律的这类特殊场景,文献[11]提出了一种数据路由协议 RCM。它首先利用节点的历史相遇信息和先验知识刻画出节点间的相遇时间和概率,然后去掉时间维的约束,使用马尔可夫决策过程中得到消息传递的期望的最小延迟,最终给出基于期望的最小延迟的数据路由路径。文献[11]认为所提的方法适用于公交车网络。但城市的公交车网络受多方面因素的影响,一般很难满足此方法所要求的前提条件,即车辆节点具有稳定的、周期性的相遇规律。

文献[12]假设目的移动车辆的轨迹已知,尝试在无静态基础设施辅助的条件下,利用公交车网络实现 I2V(Infrastructure to Vehicle)型数据传输,它的目标是在满足用户给出的数据传输到达率条件下最小化传输延迟。它首先将公路网模型化为概率状态空间图,运用和文献[11]相同的马尔可夫决策方法得到期望的最小延迟,然后,根据期望的最小延迟和目的移动车辆的轨迹,确定公交车与目的移动车辆交汇的路口以及到该路口的数据路由路径。但该

方法事实上需要使用静态基础设施即 RSU(Road-Side Unit);RSU 必须部署在交汇路口,并由此将数据传递给目的移动车辆,否则要保证用户要求的数据传输到达率,目的车辆与公交车在交汇路口必须恰好相遇并进行数据交换。但由于车辆运行时间的不确定性,这一点很难得到保证。

还有些研究工作从公交车线路间相遇规律入手,分析线路间数据传递的规律,并进而设计基于公交车线路的数据传递机制。

文献[13]提出了基于公交车线路的路由算法 BLER(Bus Line-based Effective Routing)。它首先构建公交车线路图,将公交车线路作为图的顶点,顶点之间的边表示两条公交线路至少相遇一次,边上的权重设置为这两条公交线路的公共路段长度。其次,每辆公交车基于上述公交线路图,计算出到目的公交线路的权重之和即最大的路径,并按此路径转发数据到目的线路。最后,采用往复(Zigzag)过程传递数据到目的公交车,即在目的线路上公交车只向其反向公交车传递数据,直到数据到达目的公交车。文献[14]提出了另一种基于公交车线路的路由算法 R2R,它采用与文献[13]类似的公交线路图,即以公交线路为图的顶点,边表示线路间的相遇关系,但边的权重设置为一对公交线路上公交车的平均相遇时间间隔(即平均相遇频率的倒数)。文献[14]基于该公交线路图,用 Dijkstra 算法计算权重最小路径,并将该路径作为数据传递路径。上述方法较 BLER 对车辆线路相遇规律的衡量更加准确,其路由性能也优于 BLER。

通常,公交车车站会汇聚多条公交线路,文献[15]利用公交线路的这种特性,用固定在车站的 WiFi 通信单元做数据中继,并在此前提下,为两个车站之间的数据传递寻找一条能最大限度按时传递的路径。它首先将公交网络建模成一个随机图:顶点是(车站-线路)对,顶点之间的边表示可以通过同一线路或不同线路的公交车从一个车站到达另一个车站,图上的边还与反映公交车行驶时间的随机变量相关联,随后,在此随机图上通过搜索和剪枝找出一条按时传递概率最高的路径。但该方法中公交车之间不能直接进行数据传递,只能依赖公交车站的 WiFi 通信单元进行数据交换。

公交车网络上数据传递机制与公交车的运行时间关系紧密。到目前为止,也有不少关于公交车的运行时间/到站时间预测方面的研究工作。

文献[16]利用实时的车辆 GPS 定位信息,结合路段长度、车辆速度、车站停留时间等因素,构造了

公交车运行时间估计模型,可以预报到站时间. 这种方法的实时性较好,但由于需要实时获得 GPS 信息,通信开销较大. 文献[17]考虑到公交车运行时间经常受到很多因素影响,如路段长度、交通状况、交叉路口数量、站点乘客数量、天气情况等,建立了包含多变量的回归方程作为预测模型,并根据各个变量在预测时段的变化情况计算得到车辆的运行时间. 此类模型要求各个自变量是相互独立的,而这一点在实际应用中很难满足,因此此类预测模型的应用是有限的. 文献[18]采用  $K$ -means 聚类方法分析公交车运行的历史数据,找出道路状况最相似的前  $K$  个行程记录. 具体来说,文献[18]首先提取历史行程记录的路况特征,聚类各个路段的历史路况评估值. 然后,通过查找与当前时间、位置和路况特征最相似的历史行程记录,计算公交车的到站时间. 但历史交通状况本身并没有明显的聚类特性,在其上应用聚类方法并不能获得真实反映历史路况评估的分类. 文献[19]利用神经网络,通过对历史数据的训练得到预测函数用于公交车运行时间的预测,但神经网络的训练需要较长的时间. 文献[20]将交通流看成时间序列,并使用支持向量机预测车辆长距离行驶所需的时间. 该方法在预测精度上有一定优势,但是在某些参数的确定上(例如:核函数的选取)需要经验或多次尝试. 文献[21]利用公交车乘客智能手机上的多种传感器,收集其感知的信息,然后,结合公交车运行的历史数据,预测公交车的到站时间.

考虑到在车辆自组织网络中,公交车作为一类公共交通基础设施,利用公交车进行数据传递可以降低数据转发的随机性,减少数据传递延迟的抖动,但这些都需建立在挖掘公交车时空规律的基础上,从已有的相关工作可以看出,公交车个体之间的相遇受诸多因素的影响,周期性特点不明显,导致公交车个体层面的规律不稳定;而基于公交线路的规律反映的是粗粒度的公交线路之间的关系,用它估算数据传递延迟存在不准确的问题,因此我们提出了基于路段上公交线路时空规律的数据传递机制 Vela,该机制有效提高了数据传递的性能,实现了在可容忍的延迟下,数据以可靠性最高的方式传递.

### 3 问题定义及数据分析

#### 3.1 问题定义

两个移动车辆节点利用公交车网络进行数据传

递的过程可以分成 3 步:(1)移动源节点将数据传递到某公交车;(2)利用公交车网络将数据传递到目的节点途径的路段(称目的路段);(3)公交车将数据传递给移动车辆. 若(由于移动车辆节点还未到达目的路段等原因造成的)公交车未能在目的路段上与移动节点实现数据传递,那么公交车在驶离该路段后将以自己为源节点,发起到达目的路段的数据传递请求. 重复此过程,直到携带数据的公交车在目的路段上与移动车辆在通信范围内相遇. 图 1 给出了一个 V2V 通信的实例:移动车辆节点 Veh1 利用公交线路 1~3 上的公交车 A、B、C 将数据传递到移动车辆节点 Veh2.

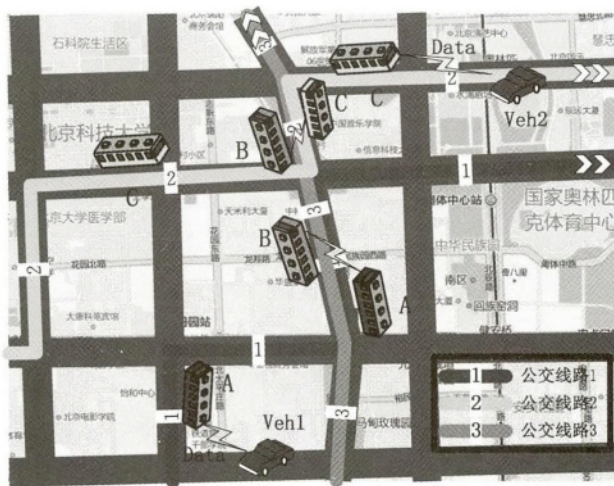


图 1 V2V 场景示意图

上述第(3)步的执行效率与环境紧密相关,可变因素多,随机性强,可能会需要很长时间才能完成数据传递,甚至会出现数据传递失败的情景. 为了缓解数据传递延迟的抖动,提高数据传递的到达率,减少数据传递的开销,可以在目的路段上的公交车站或应用热点位置(Point of Interest, POI)或目的路段路口部署 RSU,由它们存储数据并在目的车辆出现时将数据转发给目的车辆.

本文研究上述场景中将数据从某个公交车节点传递至目的路段的路由机制,该路由机制不仅是实现 V2V 通信的关键,也适用于一些 I2V 或 V2I (Vehicle to Infrastructure) 的通信场景,即源节点或目的节点可以是位于公交车线路旁的静态节点(例如:放置在停车场外的 VANET 节点).

#### 3.2 公交车轨迹数据分析

由于真实的公交车运行会受到实时交通流、道路突发事件等的影响,它的运行从微观上看经常呈现出与预期的调度和运行状况不一致的现象. 为了



确定公交车的运行时间、不同线路的公交车在路段上的相遇是否存在某种宏观的模式,我们基于真实的公交车 GPS 轨迹信息,对公交车运行的时空模式进行了分析。

我们选择了北京市具有代表性的 3 条公交线路,它们是 939 路、944 路和 983 路,其中 939 路每天平均有 20 辆公交车在运行,944 路平均有 60 辆公交车在运行,983 路平均有 42 辆公交车在运行。图 2 是北京的道路图,其中加粗线路是我们关注的公交车线路。这 3 条线路都经过北四环北辰西桥-安慧桥路段,该路段长 1920 m,下文称此路段为采样路段。



图 2 北京的公交车线路实例

每辆公交车每天从早晨 6 点至晚间 22 点每 20 s 报告一次自己的 GPS 轨迹信息,内容包括时间、公交车车牌号、公交线路号、公交车的经度和纬度、车辆当前的速度和方位角、上下行编号、下一站编号。表 1 给出了公交车报告的 GPS 轨迹信息的一个实例。

表 1 公交车 GPS 轨迹信息举例

日期时间	2013-03-01 21:49:36
车牌号	京 G31279
公交线路号	939 路
经度/度	116.494115
纬度/度	40.057195
速度/(km/h)	29
方位角/度	192
上下行编号	下行
下一站编号	13

我们从 2013 年 3 月连续 4 周的公交车 GPS 轨迹数据中分析了采样路段上公交车的运行情况。

首先对公交线路的 GPS 轨迹信息进行预处理,包括修正漂移的 GPS 点数据,插入丢失的 GPS 点数据,过滤掉非采样路段的 GPS 点信息;其次,根据

公交车车牌号,获得各个车辆的 GPS 点信息;然后,对每辆车按照上下行信息,提取每次经过采样路段的 GPS 点序列(一辆车一天会多次往返始发站与终点站)。最后,基于上述数据,统计公交车在运行时间和相遇关系上的特性。

图 3 给出了 3 条公交线路从 2013 年 3 月 4 日(周一)到 3 月 10 日(周日)在采样路段上的通行时间。可以看到,从周一至周五的不同时段,公交车行驶通过采样路段的时间有明显的波动,早晨 7 点到 9 点采样路段的通行时间明显较长,晚上 18 点左右有程度较弱的晚高峰。周六、周日的运行时间较为平稳。从图 3 数据还可以得出,虽然周一至周五公交车通过采样路段的运行时间波动明显,但是遵循相同的变化模式。另外,周六、周日的运行时间平稳,运行时间的标准差仅为 11.7 s。

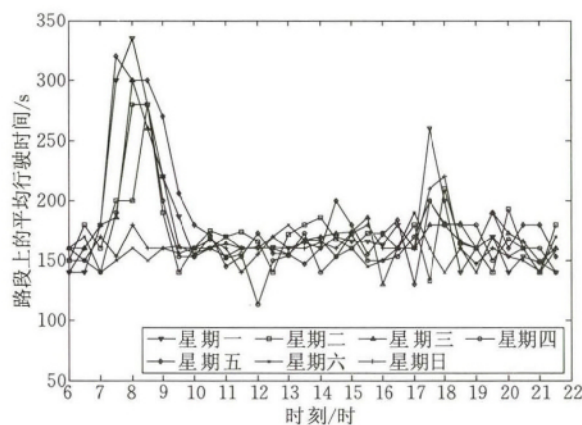


图 3 采样路段上的公交车平均行驶时间

在本文中,两辆公交车相遇是指在某个时间区间内一辆公交车在另一辆公交车的通信半径内,而公交线路 A 在某路段上与公交线路 B 相遇的概率定义为在指定的时间区间内线路 A 的公交车与线路 B 的公交车在该路段上相遇的次数与线路 A 的车辆通行该路段的次数的比值。这个概率反映了公交线路 A 的车将数据成功转发到公交线路 B 的概率。

我们分析了 944 路公交车与 939 路公交车在采样路段上的相遇概率。图 4 给出了 2013 年 3 月连续四周的工作日里(即 4 日~8 日,11 日~15 日,18 日~22 日,25 日~29 日)两条线路的车辆平均相遇概率值。为了检验上述数据是否存在周期性,我们对连续四周工作日的相遇概率组成的数据,通过傅里叶变换求得数据的瞬时频率,然后在频域上计算其功率谱,功率谱结果如图 5 所示。从功率谱上可以明显地看到出现尖峰,这说明该数据序列具有周期性。

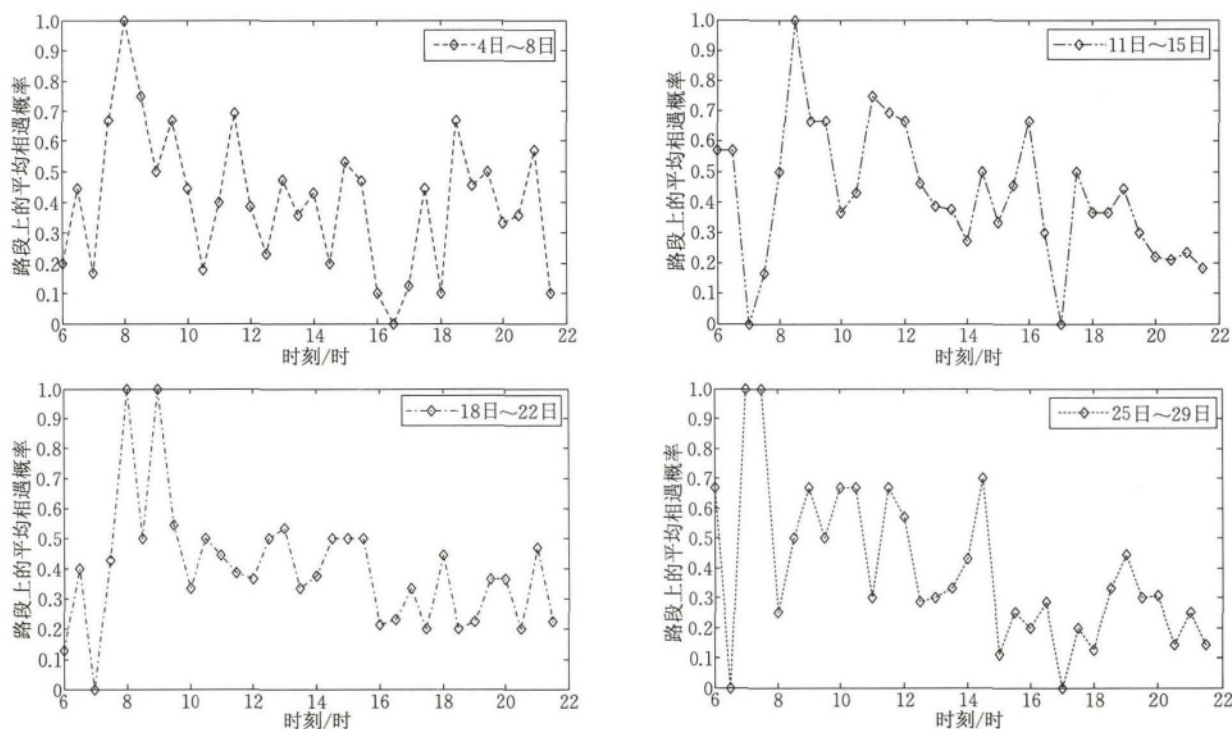


图4 采样路段上两条公交线路的相遇概率

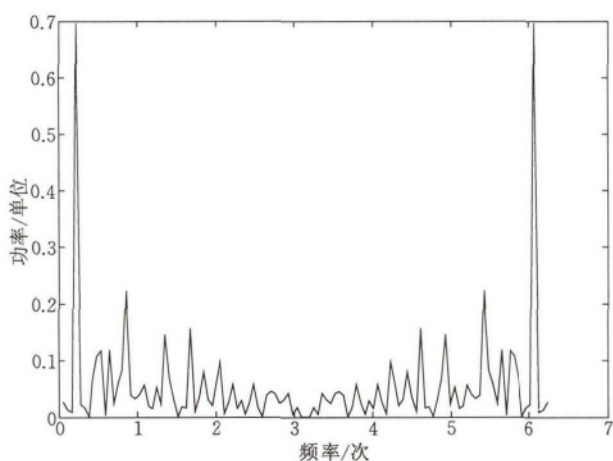


图5 连续4周工作日的相遇概率功率谱图

上述初步的数据分析结果说明了公交车运行存在着规律, 基于此, 本文挖掘了公交车运行的规律, 并将之应用于数据路由中。

## 4 公交车数据传递的概率时空模型

### 4.1 概率时空模型概述

我们将公交车网络模型化为概率时空图  $G = \langle V, E \rangle$ , 其中  $V$  表示顶点集合, 每个顶点代表经过某个路段的一条公交线路, 为方便起见, 一个顶点若代表的是经过路段  $e$  的公交线路  $A$ , 那么用  $A_e$  表示;  $E$  代表边的集合, 若顶点  $A_e$  与  $B_e$  ( $A_e, B_e \in V$ ,

$A \neq B$ ) 之间存在边, 则该边表示不同公交线路  $A, B$  途经同一路段  $e$ ; 若顶点  $A_{e_i}$  与  $A_{e_{i+1}}$  之间存在边, 该边表示同一公交线路  $A$  从一个路段  $e_i$  行驶到相邻的另一个路段  $e_{i+1}$ . 图6给出了概率时空模型的一个例子, 图6中, 跨越两个大圈的边表示不同公交线路经过同一路段, 大圈内顶点之间的边表示一公交线路从一个路段到达相邻的下一个路段。

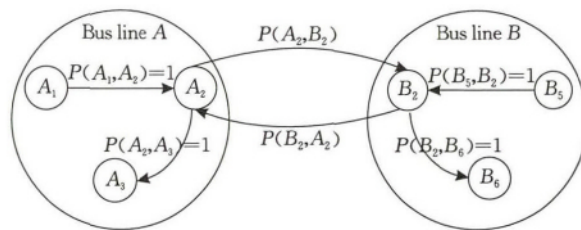


图6 概率时空图举例

我们用  $P(A_e, B_e)$  表示公交线路  $A$  和  $B$  在路段  $e$  上的相遇概率. 本文的 4.4 节给出了车辆相遇概率的估计方法。

由于通过同一条公交线路的相邻路段进行数据转移总是成功的, 所以有

$$P(A_{e_i}, A_{e_{i+1}}) = 1.$$

我们用上述相遇概率表示图  $G$  上通过边  $(A_e, B_e)$  及  $(A_{e_i}, A_{e_{i+1}})$  进行数据传递的概率  $P$ .

另一方面, 数据沿一公交线路  $A$  从一个路段  $e_i$  被携带到下一个路段  $e_{i+1}$  的延迟  $T(e_i, e_{i+1})$  可以通过挖掘公交车的轨迹数据获得, 4.2 节给出了具体

的延迟估计方法. 我们用上述延迟表示图  $G$  上通过  $(A_{e_i}, A_{e_{i+1}})$  进行数据传递的权重. 若在同一时刻, 有同一公交线路的多辆车经过此路段, 那么取它们的平均延迟作为权重. 由于我们认为在同一个路段的不同公交车之间进行数据传递的时间可以忽略, 所以, 图  $G$  上边  $(A_e, B_e)$  的权重为 0.

下面用一个例子来说明上述数据传递模型, 图 7 给出了一个简化的曼哈顿街区类型的路网, 该路网由 9 个路口和 12 个路段构成, 其中共有 3 条公交线路, 分别为 A、B、C, A 经过的路段是 1-4-9-11, B 经过的路段是 2-4-9-12, C 经过的路段是 1-2-5-10-11-12-8-3-1. 根据图 7 的道路拓扑和公交线路, 可以构造出如图 8 所示的概率时空图. 图 8 中, 大圈代表 3 条公交线路, 大圈内顶点间的边对应着同一公交线路上的相邻路段, 其顶点间的转移概率为 1.

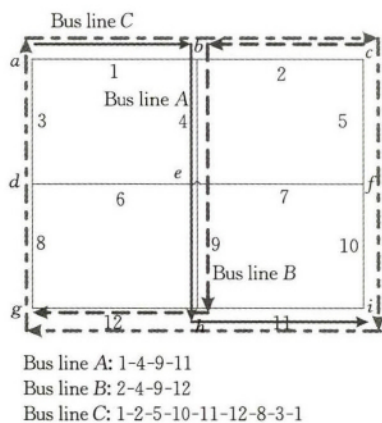


图 7 简化的曼哈顿街区类型路网图

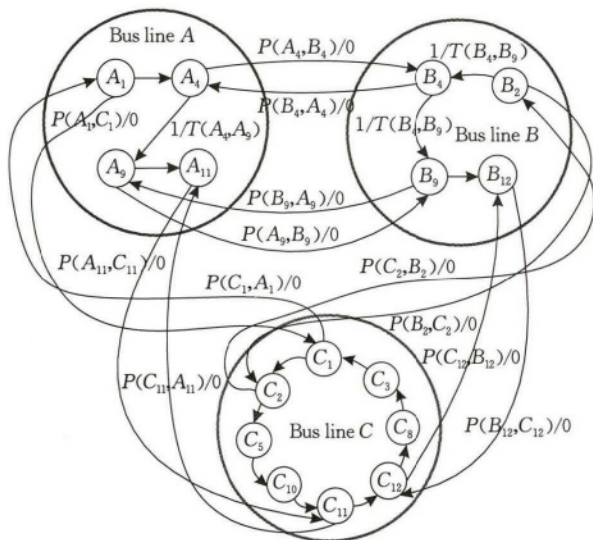


图 8 与图 7 公交网对应的概率时空图

#### 4.2 公交车运行时间预测

为了准确预测公交车在路段上的运行时间, 我

们把公交车的历史运行数据看作时间序列, 对它进行分析, 识别出其所适合的模型, 从而建立公交车运行时间的预测方程.

采用时间序列分析方法是考虑到公交车的历史运行数据的规模并不是很大, 而时间序列分析方法可以在有限样本数据的情况下建立起相当精确的数学模型, 其预测精度和可操作性都较好.

具体而言, 设已有  $x$  天的公交车工作日运行历史数据, 为了预测数据沿一公交线路  $A$  从一个路段  $e_i$  被携带到相邻下一个路段  $e_{i+1}$  的延迟. 首先, 将公交车每天的运行时间 (即 6:00~22:00) 按照 20 min 为一区间 (period) 划分为 48 个区间, 这样公交车每天行驶通过路段  $e_i$  所需的运行时间可以看作按区间划分的 48 组时间序列, 记为  $\{T_t^{\text{period}}, t \in [1, x]\}$ . 接着, 对给定一个具体的时间区间  $y$  下的公交车运行时间序列, 简记为  $\{T_t, t \in [1, x]\}$ , 进行平稳性检验, 然后, 通过模型识别、参数估计, 建立  $T_t (t > x)$  的预测方程, 即可预测第  $t$  天 ( $t > x$ ) 的公交车在这个给定的时间区间内的运行时间.

考虑到公交车运行时间的分布规律未知, 而游程检验法<sup>[22]</sup>无需假设被检验的随机序列的分布规律, 因此我们采用该方法对时间区间  $y$  下的公交车运行时间序列 (历史) 数据进行分析, 发现按照时间区间划分的公交车运行时间数据是平稳的.

在模型识别阶段, 首先需要对公交车运行时间序列进行自相关分析. 将公交车运行时间序列  $\{T_t\}$  分为两组间隔为  $k$  天的序列, 一共有  $x-k$  对, 即

$$(T_1, T_{k+1}), (T_2, T_{k+2}), \dots, (T_{x-k}, T_x).$$

显然,  $\{T_1, \dots, T_{x-k}\}$  和  $\{T_{k+1}, \dots, T_x\}$  都具有均值  $\bar{T}$ .

接着, 观察自相关函数  $\rho_k$  的值, 自相关函数的定义如式 (1) 所示:

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad (1)$$

其中,

$$\begin{aligned} \gamma_k &= E[(T_t - \bar{T})(T_{t+k} - \bar{T})] \\ &= \frac{1}{x-k} \sum_{t=1}^{x-k} (T_t - \bar{T})(T_{t+k} - \bar{T}), \\ k &= 0, 1, \dots, x-1. \end{aligned}$$

自相关函数反映了在  $T_{t+1}, T_{t+2}, \dots, T_{t+k-1}$  影响下, 间隔为  $k$  天的运行时间  $T_t$  和  $T_{t+k}$  的相关关系. 若自相关函数  $\rho_k$  在  $k > q$  时全为零, 表明该公交车运行时间序列只有  $q$  步相关性, 该  $q$  值就是滑动平均 (MA) 模型的阶数.



其次,我们对公交车运行时间进行偏相关分析,偏相关分析用于反映在剔除了  $T_{t+1}, T_{t+2}, \dots, T_{t+k-1}$  天运行时间的情况下,间隔为  $k$  的运行时间  $T_t$  和  $T_{t+k}$  的关系. 偏相关系数  $\varphi_{kk}$  通过 Yule-Walker 方程求解,该方程式的矩阵表示为

$$\begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \varphi_{k1} \\ \varphi_{k2} \\ \vdots \\ \varphi_{kk} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix} \quad (2)$$

将  $k=1, 2, 3, \dots$  代入后可以解出

$$\varphi_{11} = \rho_1, \varphi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}, \varphi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}, \dots$$

若  $\varphi_{kk}$  在  $k > p$  时都为零,表明该公交车运行时间序列的偏相关函数是  $p$  步截尾,此  $p$  值是自回归模型 (AR) 的阶数.

由于时间序列的随机性,对  $\rho_k$  和  $\varphi_{kk}$  的估计可能有误差,  $k > p$  或  $k > q$  时可能不全为零,而是在零附近波动,因此  $\rho_k$  和  $\varphi_{kk}$  的截尾性需要用统计方法去检验. 若  $\rho_k$  和  $\varphi_{kk}$  均不截尾,但收敛于零的速度较快,那么仍可以认为  $\{T_t\}$  符合  $\text{ARMA}(p, q)$  模型,其中  $p, q$  的值可以根据相关系数的值是否超过置信边界来确定.

随后,我们建立如下的自回归滑动平均模型 (ARMA) 预测方程:

$$T_t = \varphi_1 T_{t-1} + \varphi_2 T_{t-2} + \cdots + \varphi_p T_{t-p} + \alpha_t - \theta_1 \alpha_{t-1} - \cdots - \theta_q \alpha_{t-q}, \quad t > x \quad (3)$$

其中,  $p$  和  $q$  分别是在模型识别阶段获得的自回归和滑动平均模型的阶数,  $\alpha_t \sim N(0, \delta^2)$  ( $i=1, 2, \dots$ ) 为白噪声序列,  $\varphi_i$  ( $i=1, 2, \dots, p$ ),  $\theta_j$  ( $j=1, 2, \dots, q$ ) 分别是自回归系数和滑动平均系数. 我们采用矩估计法<sup>[22]</sup>估计方程系数  $\varphi_i$  ( $i=1, 2, \dots, p$ ),  $\theta_j$  ( $j=1, 2, \dots, q$ ).

最后,对建立的预测方程进行检验,保证预测方程的估计值和测量值的残差序列  $\alpha_t = T_t - \hat{T}_t$  ( $t=1, 2, \dots, x$ ) 为随机干扰误差,即白噪声序列. 使用 Ljung-Box Q 统计量进行检验.

这样,利用上述预测方程即可获得公交车在第  $t$  天 ( $t > x$ ) 的时间区间  $y$  内通过路段  $e_i$  所需的运行时间.

对公交车双休日的历史运行数据也按上述方法

做相同的处理,也可预测双休日内公交车的运行时间.

#### 4.3 预测方法的验证

本节用实际的公交车 GPS 轨迹数据来验证 4.2 节提出的公交车运行时间预测方法的可行性. 我们将 2013 年 3 月连续 4 周 (6:00~22:00) 的北京公交车 GPS 轨迹数据按工作日和双休日分为两组. 令北四环主路北辰西桥-安慧桥路段为路段  $e$ , 我们从公交车 GPS 轨迹信息中提取通过路段  $e$  的时间,得到相应的公交车运行时间序列,然后采用时间序列分析方法,用前三周工作日早高峰 [8:00~8:20] 的通行时间数据预测最后一周公交车通过路段  $e$  的时间.

首先,对公交车运行时间序列进行自相关分析和偏相关分析.

从图 9 可以看到,自相关系数在拖尾 2 阶时,超出了置信边界,且自相关值在拖尾 2 阶之后逐渐趋近于 0. 从图 10 中可以看到,在拖尾 2 阶后偏相关值均小于显著边界. 因此,路段  $e$  上的公交车运行时间序列应遵循  $\text{ARMA}(2, 2)$  模型.

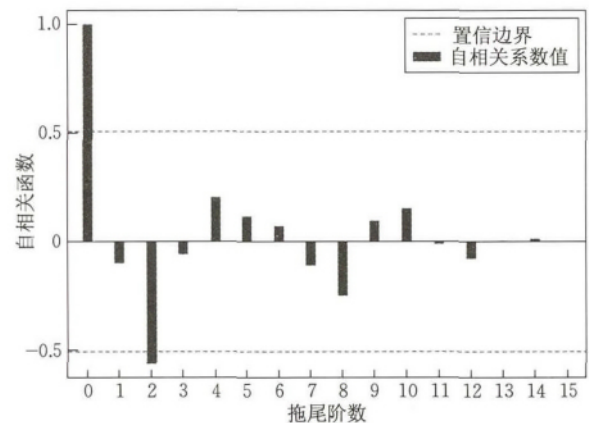


图 9 自相关分析

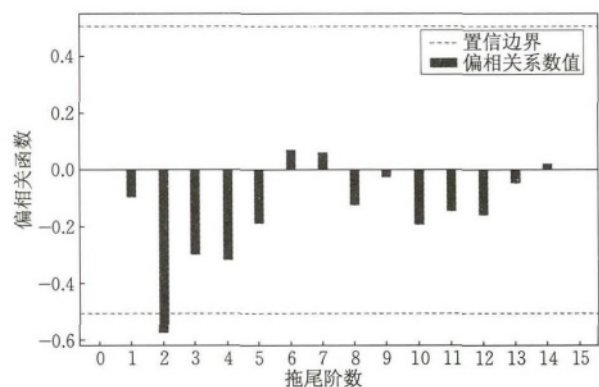


图 10 偏相关分析

然后,我们根据矩估计法确定了预测方程的参数,并预测了 5 步 (即第 4 周 5 个工作日早高峰的运



行时间),结果如图 11 所示.图 11 所示的预测的运行时间与真实的通行时间之间的残差序列通过了 Ljung-Box Q 检验.

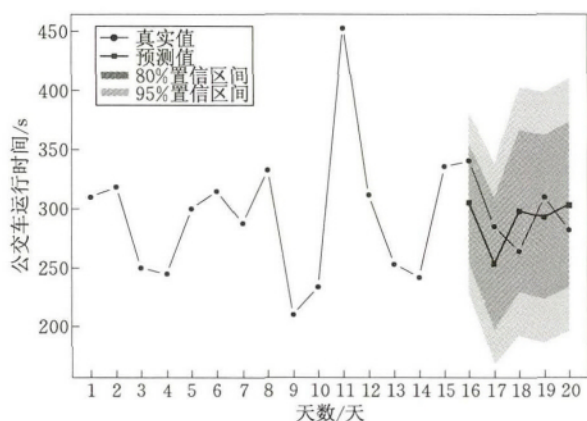


图 11 预测的工作日高峰时段的通行时间

接着,为了预测一天不同时段公交车通过路段  $e$  所需的时间,我们为每个时间区间都建立了预测方程(共 48 个),并让每个预测方程执行一步预测,得到工作日各个时段(6:00~22:00)公交车运行时间的预测值(如图 12 所示)和双休日各个时段运行时间的预测值(如图 13 所示).

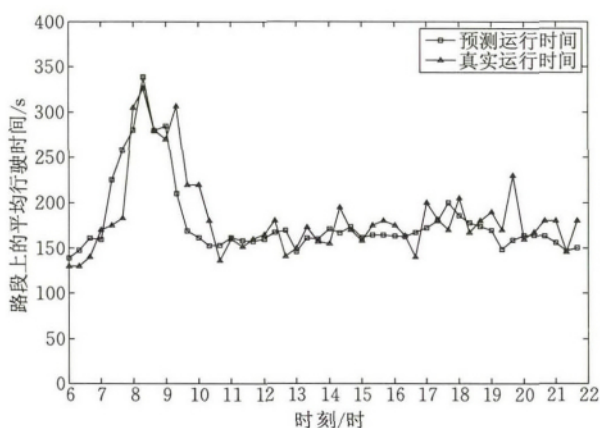


图 12 工作日采样路段公交车通行时间预测

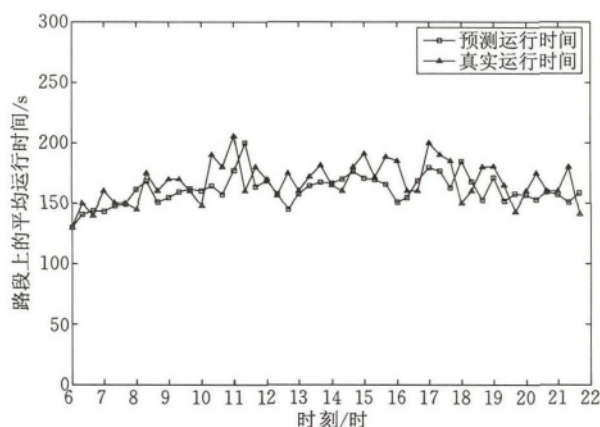


图 13 双休日采样路段公交车通行时间预测

公交车运行时间预测精度估计采用平均绝对误差  $MAE$  (Mean Absolute Error)、平均绝对百分误差  $MAPE$  (Mean Absolute Percentage Error) 和均方根误差  $RMSE$  (Root-Mean-Square Error). 它们的计算公式如式(4)~(6)所示:

$$MAE = \frac{1}{N} \sum_{i=1}^N |t_i - \hat{t}_i| \quad (4)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|t_i - \hat{t}_i|}{t_i} \times 100\% \quad (5)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (t_i - \hat{t}_i)^2} \quad (6)$$

其中,  $N$  是将一天从早晨 6 点到晚上 22 点以 20 分钟为单位划分所得的数目,  $t_i$  是第  $i$  个时段内运行时间的真实值,  $\hat{t}_i$  是第  $i$  个时段内运行时间的预测值.

从图 11 的数据可知,工作日高峰时段公交车通过路段  $e$  的预测运行时间与真实的通行时间之间的平均绝对百分误差为 9.5%. 而表 2 列出了对工作日与双休日预测的精度估计值,其中,双休日的预测平均绝对百分误差为 8.02%,工作日的平均绝对百分误差为 10.42%,该精度优于基于实时 GPS 的方法<sup>[16]</sup>,与文献<sup>[18]</sup>有类似的预测精度,但算法复杂度远小于文献<sup>[18]</sup>.

表 2 公交车运行时间预测精度估计

	平均绝对误差/s	平均绝对百分误差/%	均方根误差/s
工作日	19.81	10.42	28.50
双休日	13.56	8.02	17.12

#### 4.4 车辆相遇概率估计

通过对不同公交线路上的公交车历史相遇情况的统计(见 3.2 节),我们发现公交线路在路段上的相遇每周呈现相似的变化趋势,因此我们采用拟合历史平均相遇概率来反映公交车相遇概率的变化趋势.

为了估计  $t$  时刻公交线路  $A$  与公交线路  $B$  在路段  $e$  的相遇概率  $P$ ,我们用式(7)计算时间区间  $t_i (i \in n, n$  为时间区间的数目)的平均相遇概率  $P_i(A_e, B_e)$ :

$$P_i(A_e, B_e) = \frac{\sum_{d \in D} f_{d,t_i}(A_e, B_e)}{Q} \quad (7)$$

其中,  $f_{d,t_i}(A_e, B_e)$  是在第  $d$  天时间区间  $t_i$  内,公交线路  $A$  的公交车和  $B$  的公交车在路段  $e$  上的相遇次数,  $D$  代表进行数据统计的所有日期,  $Q$  代表在  $D$  内,公交线路  $A$  的公交车通过路段  $e$  的总次数. 这

样我们可以获得历史上不同时间区间(即  $t_1, t_2, \dots, t_n$ )的平均相遇概率  $P_{t_1}, P_{t_2}, \dots, P_{t_n}$ .

然后,用多项式拟合两公交线路在时刻  $t$  的相遇概率  $P$ :

$$P(t) = \mu_0 + \mu_1 t + \mu_2 t^2 + \mu_3 t^3 + \dots + \mu_m t^m \quad (8)$$

其中  $m$  代表多项式的次数. 我们利用最小二乘法确定系数  $\mu$ .

这里给出一个拟合 3.2 节历史平均相遇概率的实例,取  $m=4$ ,式(9)为拟合得到的相遇概率多项式,其中的参数值是利用最小二乘法来确定的,拟合曲线如图 14 所示.

$$P(t) = -0.00012t^4 + 0.00713t^3 - 0.15304t^2 + 1.36139t - 3.68266 \quad (9)$$

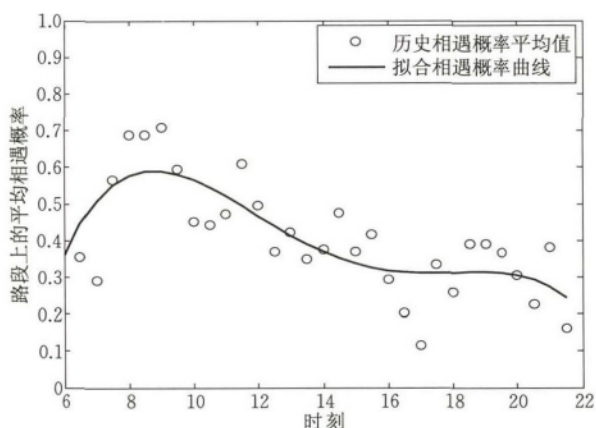


图 14 拟合相遇概率图

## 5 Vela 路由机制

为了基于公交车进行数据传递,公交车间通过心跳消息相互感知. 公交车会周期性地向一跳通信范围内的邻居节点广播自己的心跳消息,心跳消息中包含发送者节点的标识以及位置等信息. 心跳周期为心跳消息发送的时间间隔. 如果公交车连续两个心跳周期未收到相应节点的心跳消息,则认为该节点已经不在自己一跳通信范围内,即不再是自己的邻居节点,此时它会将该节点从心跳消息列表中删除,否则会更新相应节点的时间戳.

某路段上的公交车作为发起数据传递的节点  $V_{source}$ ,为了将数据传递到目的路段,按下面的流程决策路由路径并进行数据传递.

首先,该公交车根据公交路网和已有的公交车运行轨迹数据构造概率时空图. 其次,根据概率时空图中边的权重,以自身为起点,按如下步骤计算到达

目的路段的最优路线.

第 1 步. 根据目的路段的位置,定位能到达目的路段的公交线路集合,从而确定经过目的路段的公交线路节点,这些节点集合记为  $goalSet$ .

第 2 步. 计算  $V_{source}$  到  $goalSet$  中所有节点的前  $k$  个延迟最短路径:对于  $goalSet$  中的每个节点  $V_{dest}$ ,利用 Dijkstra 算法计算在概率时空图中  $V_{source}$  到  $V_{dest}$  的最短路径,并存入前  $k$  路径集合. 将  $V_{source}$  节点进一步扩展,以与  $V_{source}$  相连的节点集为起点,求解它们到  $V_{dest}$  的最短路径,这些路径在各自加上各个节点到  $V_{source}$  这一段后,存入前  $k$  路径集合. 将扩展的节点再次作为起点,并从这些节点继续向外扩展,求解到达  $V_{dest}$  的最短路径,均记录到前  $k$  路径集合中. 该扩展过程一直进行,直到路径集合中有了  $k$  条到达  $V_{dest}$  的路径.

第 3 步. 评估这  $k$  条路径数据传递的可靠性:用路径中路段的数据传递概率的最小值作为该路径可靠性的度量. 将前  $k$  条线路按照其可靠性度量排序,选择可靠性最高的路线,作为数据传递的线路.

然后,  $V_{source}$  节点按照选定的满足延迟要求的、可靠性最高的线路传递数据. 若在指定路段,按照设定路径,数据应该从公交线路  $A$  转移到另一条公交线路  $B$ ,但在实际运行时,公交车在该路段上可能遇不到公交线路  $B$  的车辆. 当  $A$  线路上的车辆检测到数据未在预先设定的数据传递线路中时,它将以自己为源节点,重新计算路径,将数据以新的路径传递到目的路段位置.

在进行数据传递时,公交车采用消息确认机制,即接收到数据包的一方会向发送方发送确认消息,发送方在收到确认消息后才将该数据包删除,否则该数据包会继续保持在发送队列中.

数据传递的路径决策算法的伪代码如下.

算法 1. Bus-based Data Delivery Mechanism.

```
//  $V_{source}$  denotes a bus node that initiates a data delivery request
//  $goalSet$  denotes a set of nodes that pass the destination road segment
1. BEGIN
2. Obtain  $G(V, E)$  according to the road map and bus lines
3. Calculate  $P, T$  and construct the probabilistic spatial-temporal diagram
4.  $shortestPathSet \leftarrow getTopkPath(V_{source}, goalSet)$ 
```

```

5.  path ← sort shortPathSet with P
6.  return path
7.  END
getTopkPath(Vsource, goalSet)
1.  BEGIN
2.  FOR Vdest IN goalSet
3.    Dijkstra(Vsource, Vdest)
4.    queue.push(Vsource)
5.    WHILE queue != NULL DO
6.      curNode = peek the first element from the
                     queue
7.      IF curNode path reaches to Vdest
8.        shortestPathSet add curNode path
9.        IF shortestPathSet.size() = k
10.         RETURN shortestPathSet
11.       END
12.     END
13.     nbrNodes ← curNode's neighbor nodes
14.     Calculate nbrNodes path to Vdest
15.     queue.push(nbrNodes)
16.   END
17. END
18. END

```

## 6 实验评估

本节通过模拟实验评估 Vela 机制的性能和可伸缩性,采用的度量指标为数据传递到达率和数据传递延迟,前者是指在规定的时间内目的车辆在指定路段收到的消息数与发起数据传递的车辆发出的消息数的比率,后者是指数据从源车辆发出到被传递到目的路段所经历的时间。

我们选择 R2R<sup>[14]</sup> 作为对比的路由机制。R2R 是一种基于公交车线路的路由机制,它基于公交车线路相遇频率进行数据传递。考虑到 BLER<sup>[13]</sup> 和 R2R 是同类型机制中的典型代表,而 R2R 的性能又优于 BLER,故模拟实验将把 Vela 与 R2R 进行比较。

### 6.1 实验搭建

我们采用 NS-3<sup>①</sup> 作为模拟实验平台,基于北京市海淀区的真实地图(如图 15 所示)分别以公交车真实轨迹和模拟轨迹为输入设计了两组实验。

在第 1 组实验中,我们选取了海淀区一块 4000 m × 5000 m 的区域,该区域内有双向行驶的 944 路、939 路、881 路 3 条公交线路。在实验开始前,我们提取了该区域在 2013 年 3 月 29 日 10:30 到 12:00 这一

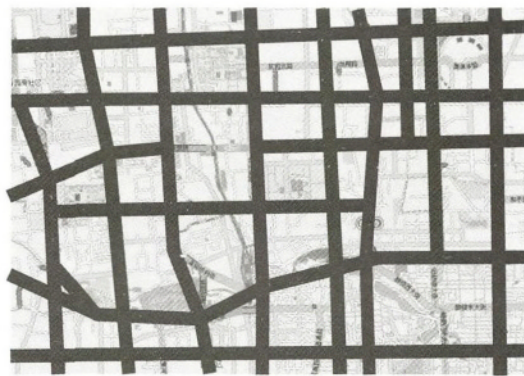


图 15 北京市海淀区的真实地图

个半小时内上述 3 条公交线路上运行的所有 35 辆公交车的轨迹数据。将这些公交车每 20 s 报告一次的轨迹信息用插值方法转化成 NS-3 中记录每秒钟车辆 GPS 位置的 tcl 轨迹文件,然后采用第 4.2 节的时间序列分析方法,预测该区域路段的运行时间,并用 4.4 节的方法估计了车辆间的相遇概率。

实验开始后,3 条公交线路按照真实 GPS 轨迹信息运行,公交车也按照真实的进入时刻进入实验区域,所有公交车都每隔 1 s 产生一条消息,该消息以随机方式选取公交车运行经过的道路为数据传递的目的路段。实验持续时间为 5300 s,在整个实验过程中,车辆发送消息总量为 3500 条。实验观察了数据传递的到达率和延迟随时间的变化。该实验用于评估 Vela 进行数据传递的性能。

在第 2 组实验中,我们选取了海淀区一块 4000 m × 8000 m 的区域,该区域内的道路总长度为 145 km,包含了 58 个十字路口和 86 条双向车道。在实验开始前,采用开源的城市交通模拟工具 SUMO<sup>[23]</sup> 生成了公交车轨迹 tcl 文件以及 15 条公交线路轨迹(如图 16 所示),其中,公交车的运动模型采用了能反映城市场景车辆运动特征的智能驾驶员模型,每条道路上模拟了公交车停靠车站,站台停靠时间为 30 s 到 80 s 之间的随机值。所有车辆根据指定的公交线路行驶。实验开始后,每隔 200 s 为每条公交线路增加 1 辆进入实验区域的公交车,所有公交车每隔 1 s 产生一条消息,该消息以随机方式选取公交车经过的道路为数据传递的目的路段。实验持续时间为 4000 s,车辆发送消息总量为 3000 条。实验观察了公交线路数量的变化对性能指标的影响。该实验用于评估 Vela 数据传递机制的可伸缩性。

① NS-3 tutorial. <http://www.nsnam.org/ns-3-19/documentation/2013.12>



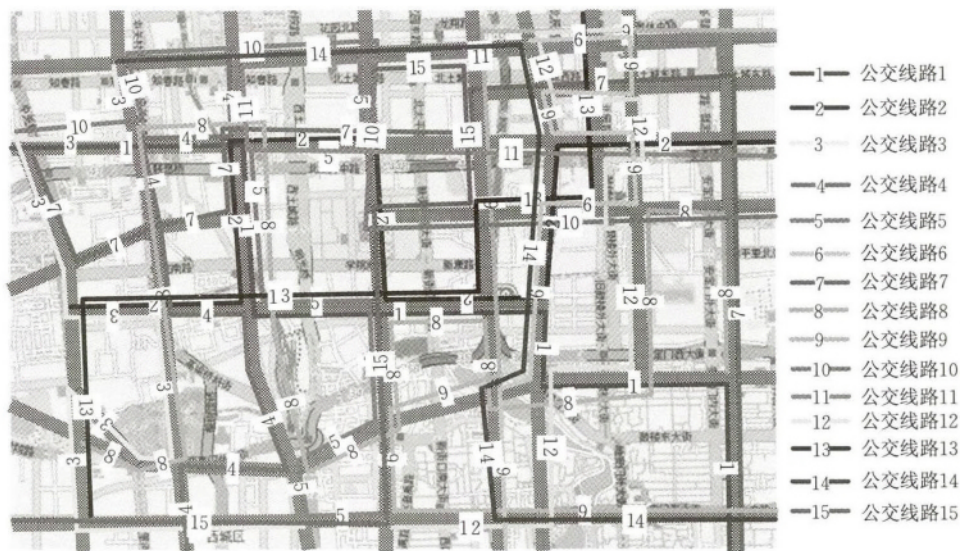


图 16 基于真实地图的 15 条模拟的公交线路

表 3 汇总了第 2 组实验所用到的参数。

表 3 第 2 组实验参数

参数	值
公交车线路数	5~15 条
发车间隔	200 s
站台停靠时间	30~80 s
车速	20~60 km/h
无线通信半径	250 m
数据产生率	1 个/s
心跳消息发送率	1 个/s

## 6.2 实验结果与分析

第 1 组实验观察了 Vela 和 R2R 两种策略随着运行时间的增加(从 3500s 到 5300s 的过程中)性能指标的变化. 图 17 显示了两种策略的数据到达率, 实验在运行到 5300s 时, Vela 机制的数据到达率已达到 91.3%, 而 R2R 的数据到达率仅是 72.3%. 从图 17 可以看到, 两种机制在数据到达率上始终有较大的差距, 这说明基于路段的相遇规律和延迟估计更为准确. 图 18 显示了两种策略的数据延迟, 可以看到 Vela 机制的延迟始终低于 R2R 机制, 这说明 Vela 机制达到了其原定的减少延迟的目标, 而 R2R 机制数据传递延迟较长, 可能的原因来自以下 3 方面: (1) 数据传递以线路间高相遇频率为依据, 这种方式可以增加数据传递的可靠性, 但是由于没有考虑数据传递的延迟, 所以数据可能被携带到距离目的路段更远的线路上; (2) 数据发起车辆可能需要携带数据前行较长时间才与频繁相遇的公交线路进行数据传递, 有可能错失了很多数据传递机会; (3) 与 Vela 将数据直接路由到目的路段不同, R2R 机制将数据传递到目的线路的公交车之后, 目的线

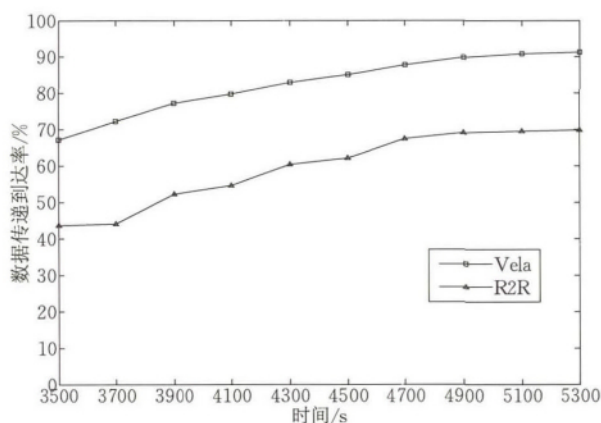


图 17 数据传递到达率

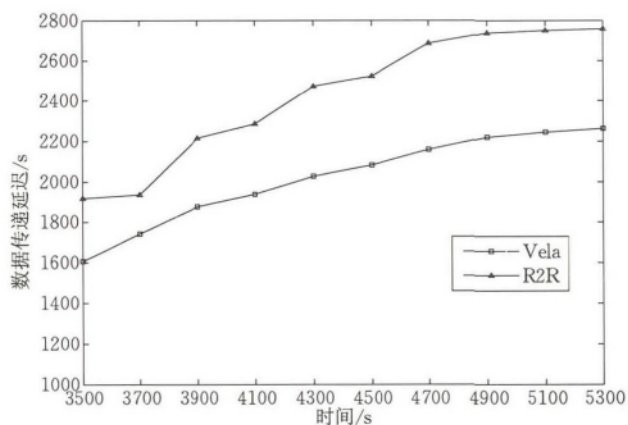


图 18 数据传递延迟

路的公交车可能需要携带数据向其反向行驶车辆传递数据并携带到目的路段, 这无疑会增加数据传递的延迟。

第 2 组实验观察公交车线路数量对两种数据传递机制的影响. 实验中, 公交车线路数量从 5 条增加



到 15 条,实验运行时间为 4000 s.

图 19 显示了在不同公交线路数下的数据传递到达率. 可以看到,随着公交车线路数的增加,Vela 数据传递到达率从 80.5%增长到了 91.1%. 这说明对于 Vela 策略,线路数的增加有效提高了数据传递到达率. 图 20 显示了不同公交线路数下的数据传递延迟. 从图上可以看到随着线路数量的增加,Vela 的数据传递延迟一直在下降. 实验结果表明基于 Vela 的数据传递具有较好的可伸缩性. 这是由于 Vela 是以数据传递的延迟和可靠性为指标从多条备选路径中遴选路径,而公交线路的增加为数据传递增加了更多的备选路径,从而增加了 Vela 选到低延迟、高可达路径的可能性.

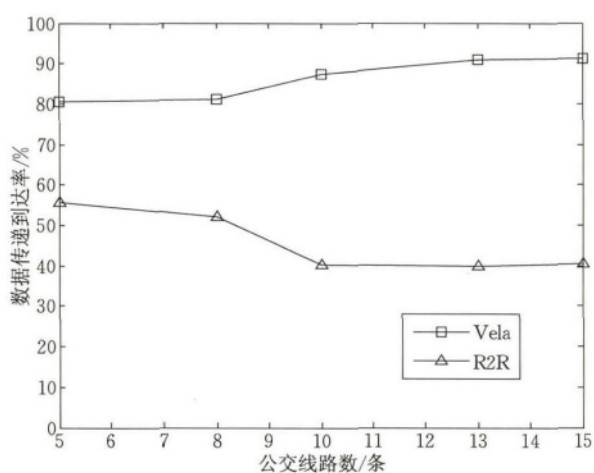


图 19 不同公交线路数下数据传递到达率

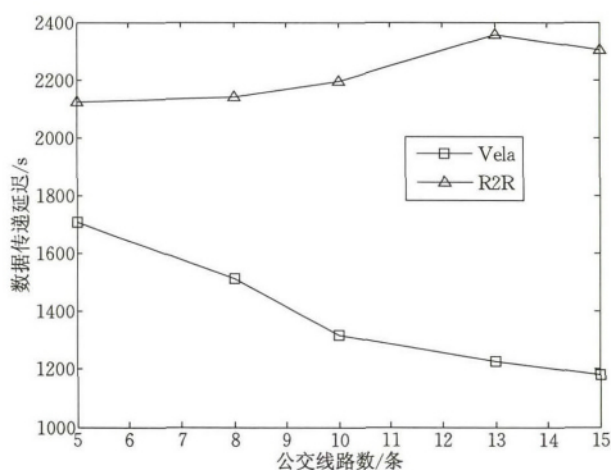


图 20 不同公交线路数下数据传递延迟

在图 19 和图 20 中,公交线路的增加起初对 R2R 的性能影响并不明显,这是因为 R2R 的选路指标并不与数据传递的性能直接相关. 但随着线路数量的不断增加,R2R 的性能在逐渐下降,这是由于面对更多的备选线路,R2R 倾向于选择具有更高相遇概率的路径,而在这样的路径上传递数据,其数

据延迟未必会比备选线路未增加时缩短. 换言之,R2R 机制不具有可伸缩性特征.

## 7 结 论

考虑到公交车网络的特点,本文认为公交车网络数据传递机制的有效性与时空规律的关系密切. 为此,本文对实际的公交车轨迹进行了分析,建立了公交车运行时间预测的 ARMA 模型,从而获得公交车运行的时间特性;同时,基于实际的公交车相遇记录发现公交车之间潜在的空间特性,并用多项式拟合车辆相遇概率曲线,从而获得对公交车相遇概率的估计. 在挖掘公交车轨迹时空规则的基础上,本文建立了公交车数据传递的概率时空模型,并给出了一种实现数据传递的路由机制 Vela,该机制在可容忍的数据传递延迟基础上选择可靠性最高的路径进行传递. 为了验证所提的 Vela 机制的有效性,本文基于真实轨迹和模拟轨迹进行了大量的实验,实验结果表明与已有的 R2R 机制相比,Vela 机制具有较高的数据到达率、较低的数据传递延迟和较好的可伸缩性.

## 参 考 文 献

- [1] Blum J, Eskandarian A, Hoffman L J. Challenges of inter vehicle ad hoc networks. *IEEE Transactions on Intelligent Transportation Systems*, 2004, 5(4): 347-351
- [2] Lochert C, Mauve M, et al. Geographic routing in city scenarios. *ACM SIGMOBILE Mobile Computing and Communications Review*, 2005, 9(1): 69-72
- [3] Zhao J, Cao G. VADD: Vehicle-assisted data delivery in vehicular ad hoc networks. *IEEE Transactions on Vehicular Technology*, 2008, 57(3): 1910-1922
- [4] Jeong J, Guo S, et al. TBD: Trajectory-based data forwarding for light-traffic vehicular networks//*Proceedings of the ICDCS'09*. Montreal, Canada, 2009: 231-238
- [5] Song Chao, Liu Ming, Gong Hai-Gang, Chen Gui-Hai, Wang Xiao-Min. Distributed real-time information based routing protocol in vehicular ad-hoc networks. *Journal of Software*, 2011, 22(3): 466-480(in Chinese)  
(宋超, 刘明, 龚海刚, 陈贵海, 王晓敏. 基于分布式实时信息的车载网络路由协议. *软件学报*, 2011, 22(3): 466-480)
- [6] Zhang L, Jin B, Dubhe A. A reliable and low latency data dissemination mechanism for VANETs. *International Journal of Distributed Sensor Networks*, 2013, (Article ID 581821): 1-16
- [7] Xu F, Guo S, Jeong J, et al. Utilizing shared vehicle trajectories for data forwarding in vehicular networks//*Proceedings*

- of the INFOCOM. Shanghai, China, 2011: 441-445
- [8] Zhu H, Chang S, Li M, et al. Exploiting temporal dependency for opportunistic forwarding in urban vehicular networks//Proceedings of the INFOCOM. Shanghai, China, 2011: 2192-2200
- [9] Zhu H, Dong M, Chang S, Zhu Y, et al. ZOOM: Scaling the mobility for fast opportunistic forwarding in vehicular networks//Proceedings of the INFOCOM. Turin, Italy, 2013: 2832-2840
- [10] Burgess J, Gallagher B, Jensen D, Levine B. MaxProp: Routing for vehicle-based disruption-tolerant networks//Proceedings of the INFOCOM. Barcelona, Spain, 2006: 1-11
- [11] Liu C, Wu J. Practical routing in a cyclic MobiSpace. IEEE/ACM Transactions on Networking, 2011, 19(2): 369-382
- [12] Chen Li, Li Zi-Jun, Jiang Sou-Xu, Feng Cheng. MGF: Mobile gateway based forwarding for infrastructure-to-vehicle data delivery in vehicular ad hoc networks. Chinese Journal of Computers, 2012, 35(3): 454-463(in Chinese)  
(陈丽, 李治军, 姜守旭, 冯诚. 车载 Ad Hoc 网络中基于移动网关的数据传输. 计算机学报, 2012, 35(3): 454-463)
- [13] Sede M, Li X, Li D, et al. Routing in large-scale buses ad hoc networks//Proceedings of the WCNC. Las Vegas, USA, 2008: 2711-2716
- [14] Li L, Liu Y, Li Z, Sun L. R2R: Data forwarding in large-scale bus-based delay tolerant sensor networks//Proceedings of the IET-WSN. Beijing, China, 2010: 27-31
- [15] Acer U, Giaccone P, Hay D, et al. Timely data Delivery in a realistic bus network. IEEE Transactions on Vehicular Technology, 2012, 61(3): 1251-1265
- [16] Zhu T, Ma F, Ma T, Li C. The prediction of bus arrival time using global positioning system data and dynamic traffic information//Proceedings of the WMNC. Toulouse, France, 2011: 1-5
- [17] Li F, Yu Y, Lin H, Min W. Public bus arrival time prediction based on traffic information management system//Proceedings of the SOLI. Beijing, China, 2011: 336-341
- [18] Sun Yu-Yan, Liu Yan, Zhou Xin-Yun, Sun Li-Min. City bus arrival time prediction based on similarity of road conditions. Journal of Software, 2012, 23(zk1): 87-99(in Chinese)  
(孙玉砚, 刘燕, 周新运, 孙利民. 基于路况相似性的城市公交车到站时间预测机制. 软件学报, 2012, 23(zk1): 87-99)
- [19] Jeong R, Rilett L. Bus arrival time prediction using artificial neural network model//Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems. Washington, USA, 2004: 988-993
- [20] Wu C, Ho J, Lee D. Travel-time prediction with support vector regression. IEEE Transactions on Intelligent Transportation Systems, 2004, 5(4): 276-281
- [21] Zhou P, Zheng Y, Li M. How long to wait? Predicting bus arrival time with mobile phone based participatory sensing//Proceedings of the MobiSys'12. New York, USA, 2012: 379-392
- [22] Zhang Shu-Jing, Qi Li-Xin. Guide to Time Series Analysis. Beijing: Tsinghua University Press, 2003(in Chinese)  
(张树京, 齐立心. 时间序列分析简明教程. 北京: 清华大学出版社, 2003)
- [23] Krajzewicz D, Erdmann J, Behrisch M, Bieker L. Recent development and applications of SUMO—Simulation of urban mobility. International Journal on Advances in Systems and Measurements, 2012, 5(3&4): 128-138



**ZHANG Fu-Sang**, born in 1987, Ph. D. candidate. His research interests include distributed computing and pervasive computing.

**JIN Bei-Hong**, born in 1967, Ph. D., professor, Ph. D. supervisor. Her research interests include distributed computing, mobile and pervasive computing, middleware

and distributed systems.

**WANG Zhao-Yang**, born in 1989, Ph. D. candidate. His research interests include distributed computing and pervasive computing.

**HU Jia-Feng**, born in 1988, Ph. D. candidate. His research interests include distributed computing and pervasive computing.

**ZHANG Li-Feng**, born in 1977, Ph. D. His research interests include distributed computing and pervasive computing.

## Background

As a kind of mobile and self-organizing network, a VANET (Vehicular Ad hoc NETWORK) in combination with the other technologies (such as sensor technologies) has been envisioned to be an underpinning of many valuable applica-

tions, including active safety applications, intelligent transportation management and in-vehicle infotainment, etc. In recent years, the VANET is becoming an important research field, attracting considerable attention from academia and

industries.

In urban scenarios, a bus-based network has distinctive features such as wide coverage, fixed lines and relatively stable schedule so that it can be employed to achieve data delivery in VANET applications. However, a bus-based ad hoc network suffers from the rapidly-changing network topology, network partition and unstable wireless communication quality. These factors directly influence the QoS of data delivery, in addition to the factor that city buildings and other public facilities interfere radio signals. Therefore, how to mine and utilize spatial-temporal rules in a bus-based network to improve data delivery quality is a challenge problem. In other words, it is worth studying on mining the model of bus travel time and patterns of the bus encounter as well as effectively applying to the data routing mechanism. The authors are committed to conduct the research work on the data routing mechanism over bus-based ad hoc networks by taking advantage of trajectory mining results.

This paper puts forward a routing mechanism named Vela which fully exploits the model of bus travel time and patterns

of the bus encounter on road segments. Vela builds a probabilistic spatial-temporal model for data delivery and then gives the routing selection strategy. Compared with existing solutions, the routing mechanism Vela characterizes the spatial-temporal rules of buses in a moderate granularity of road segments, which not only overcomes the shortcoming of low accuracy of the spatial-temporal rules in coarse-grained (i. e. , the bus line level) approaches, but also avoids the low stability of the patterns in fine-grained (i. e. , the bus level) approaches. The research provides some insights for mining real GPS traces to support the data routing mechanism. Meanwhile, the experimental results demonstrate that Vela can provide the best possible QoS levels and scalability.

The research work is funded by the National Natural Science Foundation of China (61472408, 61372182). The former project focuses on the data dissemination issue in the pervasive environments and the latter project concerns the IoT (Internet of Things) technologies and systems (such as the ones over VANETs) under emergency.