

基于趋势预测的合乘收益研究

宋 飞,李 蓉,张思东,张宏科

(北京交通大学电子信息工程学院 下一代互联网互联设备国家工程实验室,北京 100044)

摘 要: 首先提出一种基于车载自组网的城市出租车合乘的方案;其次针对北京市 6499 辆出租车的 GPS 信息进行了较为全面的数据分析,获取 81785 个乘客上车时间及位置,从整体和局部两个方面定量地讨论了出租车合乘带来的益处,运用灰色预测和曲线拟合两种不同的方法建立数学模型预测上车点个数的变化趋势;最终得出在北京市六环内乘客采用合乘方式后,可减少 1/3 以上的上车点以及约 20~30% 的行驶里程,对缓解交通压力及构建绿色城市具有十分重要的意义。

关键词: 车载自组网; 出租车; 合乘方案; 数据分析

中图分类号: TP393 **文献标识码:** A **文章编号:** 0372-2112 (2014)07-1353-07

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2014.07.016

The Research on Taxi Sharing Benefit Based on Tendency Estimation

SONG Fei, LI Rong, ZHANG Si-dong, ZHANG Hong-ke

(National Engineering Laboratory for Next Generation Internet Interconnection Devices, School of Electronics and Information Engineering, Beijing Jiaotong University, Beijing 100044, China)

Abstract: An urban taxi sharing scheme based on VANET is proposed firstly. We elaborates the necessary processing of the data derived from 6499 GPS sets of Beijing taxis in operation and mainly analyzes the data from both macro and micro scopes. The location and timing of 81785 picking up points are obtained as well. The feasibility and benefits of the sharing scheme are discussed. Two different prediction methods, i. e. grey prediction and curve fitting, are employed to create mathematical models to predict the trend of picking up points. The conclusion is that by adopting the taxi sharing method, more than 1/3 of the picking up points and about 20~30% mileage will be reduced, which will certainly contribute towards relieving the traffic pressure and creating an environment - friendly city.

Key words: VANET; taxi; sharing scheme; data analyses

1 引言

出租车在一定程度上满足了人们个性化的交通需求,但近些年来在北京、上海、广州等大城市的早晚交通高峰时期“打车”却成为困扰诸多乘客的难题^[1].早期的调查表明^[2],在取样地点交通准高峰期间,出租车数量占行经车辆总数的 58%,但每辆出租车的平均载客数仅为 0.74 人.北京市近 10 年来出租车总量一直控制在 6.66 万辆以内^[3],打车难的现象在热点地区和特定时段也一直存在.2012 年出台的北京市“十二五”交通规划提出,出租车总量将继续控制在 6.66 万辆以内^[4].由此看来,暂时不能依靠车辆的增加来解决打车难的问

题.2012 年 3 月,北京市针对打车难的问题,提出鼓励乘客合乘出租车^[5],特别是早晚高峰时段,合乘者各付共同路段车费的 60%,并可打印多份发票.目前 96103 和 96101 两部叫车电话都可以提供“合乘订车”业务.同时,出租车拒载将从重处罚.

虽然有了政府的支持,但在实际生活中选择合乘方式的乘客却为数不多.一方面,很多乘客并不了解合乘的相关政策,具体内容有待宣传;另一方面,一些乘客虽然了解合乘政策,但苦于没有方便快捷的方法定位合适的合乘对象并完成拼车.基于以上原因,本文提出可以利用近年来发展较快的车载自组网和智能移动终端来实现快速高效地合乘.通过对实际采集的出租车 GPS

数据进行分析,能够定量地计算乘客周围满足合乘条件的用户数与相关参数的关系,同时通过现有数据预测 4 万辆出租车情况下采用合乘方式后带来的收益情况。

2 车载自组网简介

汽车的普及带来了气候、交通等诸多问题,智能交通系统^[6]正在得到学术界越来越多的关注.由于我国人口众多,道路建设速度远远滞后于汽车数量的增长.车辆价格的下调又让城市的道路拥堵雪上加霜.因此在当前环境下,交通系统的智能化特别是车载自组网(VANET)技术具有非常重要的意义^[7].该技术包括车辆间通信(V2V)和车辆与基础设施间的通信(V2I)^[8],如图 1 所示.其中基础设施一般指固定在路边的通信单元,简称路边单元(RSU).2003 年美国联邦通信委员会为车辆间通信划分了专用的频段并制定相应协议 DSRC^[9],2009 年 IETF 制定了 WAVE 标准(802.11p)^[10],为 VANET 的发展提供了重要的底层协议支持.通过该技术,装有移动车载设备的车辆之间可以根据需要随时建立网络进行通信.随着 GPS 和电子地图在移动车载设备上的普及,该技术将会越来越多地应用于城市交通的方方面面^[11].

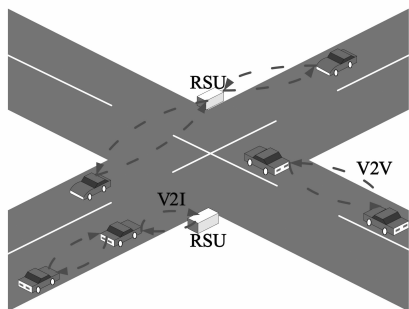


图1 车载自组网示意图

持有智能移动终端(如手机,平板电脑,笔记本等)的乘客,可以依托 VANET 技术和出租车、路边单元通过无线方式进行通信.出租车可将当前的状态(有无乘客、行驶路线、目的地等)广播给周围的乘客或路边单元.如果有乘客与其路线相同,便可向出租车发出合乘申请.需要出租车的乘客也可将自己的起点和终点广播给周围乘客、路边单元和出租车,当有满足条件的合乘对象或出租车出现时及时将相关信息进行显示,为进一步沟通提供基础.

合乘能够在降低用户乘车费用、提高搭乘成功率的同时减少二氧化碳排放,提高出租行业整体运输能力.因此有必要对合乘收益进行定量计算.

3 数据源介绍

我们选取了 2010 年 5 月 1 日全天、北京市实际运

营的 28000 辆出租车 GPS 记录进行了分析(每辆车的信息采样周期均为 60 秒).为了更好地进行数据处理,对原始数据进行数据清洗:除去北京六环以外、速度超限、GPS 定位无效、当天出现频率小于 200 次、一直处于静止或空车状态的车辆,最终得到 6499 辆出租车的有效数据.数据内容格式如表 1 所示:其中, ID 表示车辆编号、T 表示从 1900 年 1 月 1 日起累计的时间(单位:秒)、LAT 和 LONG 分别表示纬度和经度(单位:度)、HEAD 表示方向(单位:度,以正东为 0 度,取值 0~360 之间)、V 表示车速(单位:米/秒)、HEAVY 表示空重车信息(0 为空车,1 为重车).

由于目标是对乘客合乘情况进行分析,需要研究 Origin Destination(OD)的时空分布.设由空车变为重车的位置为上车点,重车变为空车的位置为下车点.定位之后发现部分上车点与对应下车点间的距离很近,时间间隔也非常短.为减小分析误差,将上下车时间差小于 10 分钟,直线距离小于 500 米的点全部去除,最后筛选出 81785 个上车点及对应的下车点,后续分析均基于上述数据完成.

表 1 数据格式

ID	T	LAT	LONG	HEAD	V	HEAVY
1	1241107200	39.80288	116.39645	357	1131	0

4 数据分析

文中的数据分析使用 MATLAB 软件,结合地图的分析选取地理信息系统软件 MAPINFO^[12].基本分析流程为:首先对数据的整体分布情况进行研究(包括地域分布与时间分布);其次对是否存在出租车合乘的条件进行分析;最后假定在实现合乘的条件下,定量讨论该方案所带来的益处.



图2 上车点地域分布

4.1 宏观位置分布

将提取的 81785 个上车点加载在北京市地图上(如图 2 所示),数据中被记录的出租车主要活动范围是四环以内.分析不同区域的点密度可以得到:在西单、王

府井、北京站、北京西站、西直门、北京南站、中关村、东直门、工体、建国门、国贸、大望路周边地区上车点比较集中.上述地区均为繁华的商业区或交通枢纽,人流量非常大,打车难的现象尤其严重.结合北京市行政区的范围(为了便于细化,仍沿用崇文区和宣武区的划分),分别统计各区域内的上车点数,如图 3 所示.朝阳区的比例最高,海淀次之.

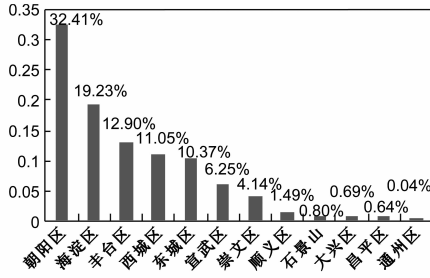


图3 统计各行政区内上车点数所占比例

统计 81785 个上车点中每小时内上车点个数,如图 4 所示.早 9 时至晚 22 时每小时上车点个数维持在 8000 以上,该时段选择乘坐出租车的人数较多(上午 10 时至 11 时上车点最多,有 12078 个).凌晨 3 时至 5 时的上车点最少,均只有 1100 多个.

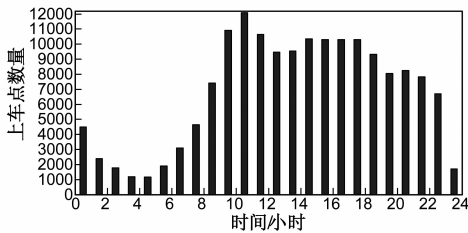


图4 上车点时间分布

4.2 可行性分析

我们设定在两种情况下,乘客具备合乘条件:

- (1) 两个上车点的时间和距离接近且下车点距离接近,乘客可以同时上车并先后下车;
- (2) 一辆车起点到终点的行驶轨迹包含另一辆车起点到终点的行驶轨迹,乘客可以进行沿路搭乘.

为了更贴近于实际情况,现仅对第一种情况进行分析,并做出如下假设:

- (a) 任意两辆车内的乘客只要满足条件均可进行合乘,且每辆车内均只有 1 名乘客;
- (b) 只考虑两辆车上的乘客进行合乘的情况;
- (c) 两辆车上的乘客可以进行合乘的条件为:

$$\Delta t_{on} \leq 10\text{min} \& d_{on} \leq 1\text{km} \& d_{off} \leq 1\text{km},$$

其中 Δt_{on} 为两上车点的时间差绝对值, d_{on} 为两上车点间的直线距离, d_{off} 为两下车点间的直线距离.

值得注意的是:条件的设置主要目标是保证合乘切实可行,在该条件下得到的所有结果均为理论的下

限值,实际情况要优于理论下限值.两点间的直线距离 d_{on} 和 d_{off} 由式 (1) 得到^[13],其中 R 为地球半径, (x_1, y_1) 、 (x_2, y_2) 分别为两点的经纬度坐标, x 为纬度, y 为经度.

$$d = 2R \arcsin$$

$$\left(\frac{\sqrt{2[1 - \cos(x_2 - x_1) + \cos x_1 \cos x_2 - \cos x_1 \cos x_2 \cos(y_2 - y_1)]}}{2} \right) \quad (1)$$

假设上车点的个数为 N_{on} ,某个上车点周围存在与其满足合乘条件的其他上车点数为 X .对每个上车点进行遍历便可得到对应的 X ,统计满足不同 X 的 N_{on} ,如表 2 所示(X 为 9 ~ 14 时, N_{on} 的值均较小,因此未在表中列出).令

$$P_{on} = \frac{N_{on}(X \geq 1)}{N_{on}(X \geq 0)}$$

为满足周围存在可合乘对象的上车点数与总上车点数之比,可计算得到表 2 对应的 P_{on} 为 25.78%,即约有 1/4 的上车点周围存在可以合乘的对象,且可以看出周围有 1 ~ 2 个可合乘对象的情况居多.

表 2 不同 X 对应的上车点数

X	0	1	2	3	4	5	6	7	8
N_{on}	60717	14382	4196	1402	585	238	106	77	37

以上分析中 Δt_{on} 、 d_{on} 、 d_{off} 均是固定值,为了讨论 d_{on} 、 d_{off} 变化时 P_{on} 的变化情况,找出上车点个数最多的 10 分钟(满足 $\Delta t_{on} \leq 10\text{min}$),将此期间的上车点提取出来(共 1028 个), d_{on} 和 d_{off} 均以 0.1 为间隔从 0 变化到 2,从而得到对应的 P_{on} ,如图 5 所示.可以看出 P_{on} 与 d_{on} 或 d_{off} 均近似成线性递增关系.当 d_{on} 和 d_{off} 均为 2km 时 P_{on} 为 0.7,70% 上车点周围存在可合乘的对象.当 $P_{on} \geq 0.5$ 时, d_{on} 和 d_{off} 均大于 1.5km.所以将 d_{on} 和 d_{off} 设置成大于 1.5km 时超过一半的出租车满足合乘条件.

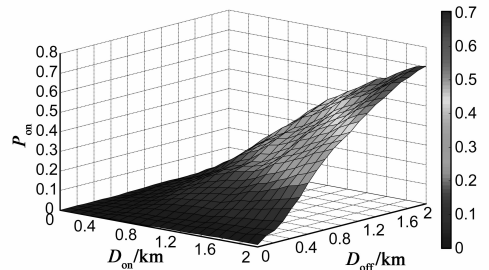


图5 合乘条件变化时可合乘的上车点比例

通过上述分析可知,实际中满足合乘条件的乘客很多,随着前提条件范围的扩大,符合合乘条件的乘客比例会大幅度增加.

4.3 收益分析

设上车点与下车点间的折线距离(垂直和水平方

向距离和)为车辆的行驶里程,计算出原始数据 81785 个上车点对应的行驶里程总数为 697606.2723km,对所有上车点和下车点的数据进行遍历,在满足合乘条件

$$\Delta t_{on} \leq 10\text{min} \& d_{on} \leq 1\text{km} \& d_{off} \leq 1\text{km}$$

的两个上车点中,将时间较大的上车点删除,得到合乘后的上车点数为 70964,行驶里程数为 632657.1438km. 通过合乘减少了 13.23% 的上车点和 9.31% 的行驶里程. 这些减少的上车点可以为更多的乘客提供服务.

以上分析的数据只是北京市出租车的一部分,并不能准确反映实际合乘后带来的宏观收益,为了分析北京市出租车的整体情况,下面对现有数据进行预测分析,得到随着车辆数的增加,相关参数的变化趋势.

5 数据预测

目前,北京市约有 6.66 万辆出租车,本文研究的 GPS 数据分布在北京市六环以内,假设六环内的出租车有 4 万辆,下面分别采用灰色预测和非线性曲线拟合两种方法预测这 4 万辆出租车采用合乘方案后的收益情况.

5.1 预测方法讨论

在已有的 6499 辆出租车中,分别随机抽取 1000、2000、3000、4000、5000、6000 辆,其中 2000 辆包含之前的 1000 辆出租车,以此类推. 令车辆数为 N_c , 计算不同 N_c 对应的上车点数 N_{on} 和行驶里程数 D (单位 km), 以及采用合乘方式后的上车点数 N_{onh} 和行驶里程数 D_h (单位 km), 并计算其减少的百分比 $P_{non} = (N_{on} - N_{onh}) / N_{on}$ 和 $P_d = (D - D_h) / D$, 结果如表 3 所示, 大体可知所有数均随 N_c 的增大而增大. 具体关系如图 6、图 7 和图 8 所示, N_{on} 、 N_{onh} 、 D 、 D_h 、 P_{non} 、 P_d 均随 N_c 呈近似线性递增关系, 但是曲线的斜率随着 N_c 的增大而略微减小. 本文基于以下考虑对数据进行预测:

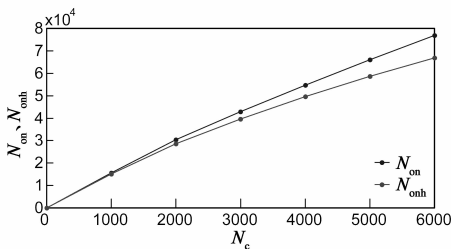


图6 N_{on} 和 N_{onh} 的原始数据图

(a)只考虑两个上车点合并成一个上车点的情况,因此 P_{non} 的最大值为 50% ;

(b) N_{on} 和 D 不仅与 N_c 有关还与乘客的数量有关,而乘客的数量不会无限增多,因此推测随着 N_c 的增长, N_{on} 、 N_{onh} 、 D 、 D_h 增长速率放缓,逐渐趋于稳态,这与图 6 和图 7 的曲线趋势相吻合.

(c)由图 6 和图 7 可以看出上车点个数和行驶里程数有相同的变化趋势,因此仅对上车点进行预测分析.

表 3 不同车辆个数对应的上车点及行驶里程情况

N_c	N_{on}	N_{onh}	P_{non}	D/km	D_h/km	P_d
1000	15588	15114	3.04%	134835.83	131863.95	2.20%
2000	30395	28617	5.85%	261779.41	250949.14	4.14%
3000	42938	39624	7.72%	369981.05	350097.85	5.37%
4000	54745	49571	9.45%	476089.33	444317.23	6.67%
5000	66097	58633	11.29%	571454.36	525738.90	8.00%
6000	76930	66863	13.09%	659250.65	598781.14	9.17%

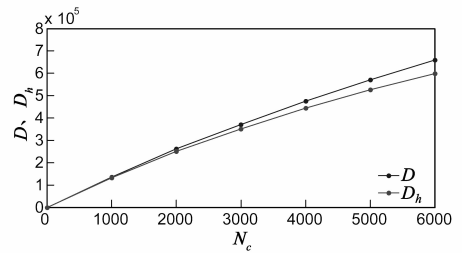


图7 D 和 D_h 的原始数据图

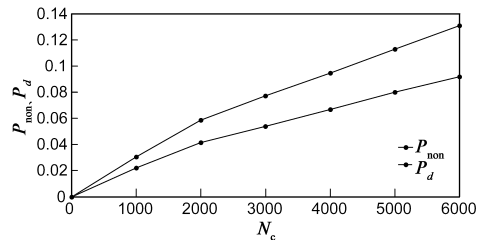


图8 P_{non} 和 P_d 的原始数据图

5.2 灰度预测

上车点的变化具有一定规律,可以采用灰色预测模型 GM(1,1) 来进行短期预测与长期趋势分析. 灰色系统分析方法^[14]是通过鉴别系统各因素之间发展趋势的相似或相异程度(即进行关联度分析),并通过对原始数据的生成处理来寻求系统变化的规律. 生成数据序列具有较强的规律性,可以用它来建立相应的微分方程模型,从而预测事物未来的发展趋势和未来状态.

5.2.1 GM(1,1)模型介绍

GM(1,1)模型的建立过程为:

- (1) 设 $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$ 为原始建模序列;
- (2) 设 $X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\}$ 为 $X^{(0)}$ 的一次累加生成序列,其中

$$x^{(1)}(k) = \sum_{i=0}^k x^{(0)}(i) = x^{(1)}(k-1) + x^{(0)}(k), k = 1, 2, \dots, n;$$
- (3) 设 $Z^{(1)} = \{z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n)\}$ 为 $X^{(1)}$ 的紧邻均值生成序列,其中

$$z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1);$$

(4)GM(1,1)的灰微分方程模型为:

$$x^{(0)}(k) + ax^{(1)}(k) = b \quad (2)$$

式(2)中 a 为发展系数, b 为灰色作用量, 设 $\alpha = (a, b)^T$, $\alpha = (B^T B)^{-1} B^T Y_n$, 其中

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \dots & \dots \\ -z^{(1)}(n) & 1 \end{bmatrix}, Y_n = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \dots \\ x^{(0)}(n) \end{bmatrix};$$

(5)计算出灰微分方程(2)的时间响应序列为:

$$\hat{x}^{(1)}(k+1) = \left(x^{(1)}(0) - \frac{b}{a} \right) e^{-ak} + \frac{b}{a}, k = 1, 2, \dots, n \quad (3)$$

(6)得到预测方程为:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \quad (4)$$

由式(4)得到的便是对原始数据的预测值. 利用 MATLAB 软件完成上述计算步骤, 便可求出已知数据的预测值.

为了检验预测数据是否合格, 需要做残差检验、关联度检验和后验差检验, 具体计算过程为:

(a)残差检验. 计算原始序列 $X^{(0)}$ 和预测序列 $\hat{X}^{(0)}$ 的绝对残差序列 $\Delta^{(0)} = \{\Delta^{(0)}(i), i = 1, 2, \dots, n\}$,

其中 $\Delta^{(0)}(i) = |x^{(0)}(i) - \hat{x}^{(0)}(i)|$, 以及相对残差序列 $\phi = \{\phi_i, i = 1, 2, \dots, n\}$, 其中

$$\phi_i = \left[\frac{\Delta^{(0)}(i)}{x^{(0)}(i)} \right] \%,$$

得到平均相对残差:

$$\bar{\phi} = \frac{1}{n} \sum_{i=1}^n \phi_i \quad (5)$$

按照实际需要设定一个 α , 当 $\bar{\phi} < \alpha$ 且 $\phi_n < \alpha$ 时, 该模型残差合格.

(b)关联度检验. 首先计算原始序列 $X^{(0)}$ 和预测序列 $\hat{X}^{(0)}$ 的关联系数:

$$\eta(i) = \frac{\min_l |x^{(0)}(l) - \hat{x}^{(0)}(l)| + P \max_l |x^{(0)}(l) - \hat{x}^{(0)}(l)|}{|x^{(0)}(i) - \hat{x}^{(0)}(i)| + P \max_l |x^{(0)}(l) - \hat{x}^{(0)}(l)|} \quad (6)$$

关联度为:

$$r = \frac{1}{n} \sum_{i=1}^n \eta(i) \quad (7)$$

根据经验, 关联度大于 0.6 符合要求.

(c)后验差检验. 首先计算原始序列的平均值和均方差:

$$\bar{x}^{(0)} = \frac{1}{n} \sum_{i=1}^n x^{(0)}(i) \quad (8)$$

$$S_1 = \left(\frac{\sum_{i=1}^n [x^{(0)}(i) - \bar{x}^{(0)}]^2}{n-1} \right)^{\frac{1}{2}} \quad (9)$$

然后利用类似式(7)和(8)的公式计算残差的均值 $\bar{\Delta}$ 和均方差 S_2 , 得到方差比为:

$$C = S_2/S_1 \quad (10)$$

最小残差概率为:

$$P = P\{e_i < S_0\} \quad (11)$$

其中, $S_0 = 0.6745S_1$, $e_i = |\Delta^{(0)}(i) - \bar{\Delta}|$. 根据 C 和 P 的值判断模型是否合格, 如表 4 所示. 若相对残差、关联度、后验差检验在允许的范围内, 则可用所建模型进行预测, 否则进行残差修正.

表 4 根据 C 和 P 判断模型精度

P	C	模型精度
> 0.95	< 0.35	优
> 0.80	< 0.5	合格
> 0.70	< 0.65	勉强合格
< 0.70	> 0.65	不合格

5.2.2 直接进行灰色预测

将 N_{on} 和 N_{onh} 的原始数据(在表 3 中列出)带入灰色预测模型, 得到的预测值分别为:

$$\hat{N}_{on} = \{0, 24575, 31289, 39837, 50720, 64576, 82218\},$$

$$\hat{N}_{onh} = \{0, 23293, 29155, 36492, 45676, 57171, 71559\}.$$

分别对 \hat{N}_{on} 和 \hat{N}_{onh} 进行模型检验, 得到各检验参数值如表 5 所示, 可以看出所有参数均在合格范围之内, 因此可以预测 N_c 为 1 万以内的数据 N_{on} 和 N_{onh} . 不考虑其他因素的影响, 直接利用 GM(1,1) 模型对 N_c 变化到 4 万时 N_{on} 和 N_{onh} 的值进行预测, 得到的预测数据曲线如图 9 所示. 可以看出 N_{on} 和 N_{onh} 均随 N_c 呈指数规律变化, 通过 N_{on} 和 N_{onh} 可以求出 P_{non} (如图 10 所示). P_{non} 与 N_c 呈近似线性递增关系, P_{non} 的最大值大于 50%, 这些结论与本文之前的考虑是不相符的. 因此直接对数据进行灰色预测得到的结果并不准确.

表 5 模型检验各参数的值

检验对象	平均相对残差	关联度	方差比	最小残差概率
\hat{N}_{on}	0.1205	0.6419	0.1115	1
\hat{N}_{onh}	0.1161	0.6425	0.1179	1

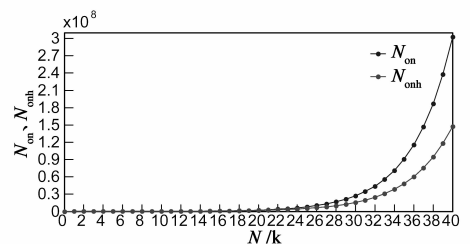


图 9 由 N_{on} 和 N_{onh} 的灰色预测

5.2.3 通过增量变化进行预测

假设原始上车点个数的增长率 K 为每增加 1000

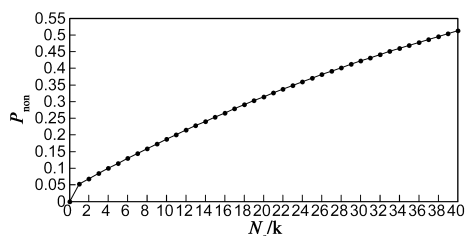


图10 由 N_{on} 和 N_{onh} 的灰色预测得到的 P_{non}

辆出租车时上车点个数的增加数量,即 $K(N_c) = N_{on}(N_c) - N_{on}(N_c - 1000)$, 带入数据得到 $K = \{15588, 14807, 12543, 11807, 11352, 10833\}$, 可以看出随着 N_c 的不断增大, K 呈逐渐下降的趋势, 正因为有 N_c 以外其他因素的影响才会使 N_{on} 的增长速度变慢, 因此可以通过对增量 K 进行预测, 从而计算出相应的 N_{on} . 将 K 带入灰色预测模型进行预测并检验出模型是合格的, 进而得到 N_c 从 0 变化到 4 万时 K 的预测值如图 11 所示, 可以看出 K 随着 N_c 的增加呈负指数趋势下降. 通过 K 便可以得到对应的 N_{on} , 且当 $N_c \leq 6k$ 时, 预测得到的 N_{on} 为 $\{0, 15588, 29797, 42957, 55146, 66434, 76889\}$ 与表 3 中的原始数据对比, 平均相对残差 0.0047, 关联度 0.7215, 方差比 0.0088, 最小残差概率为 1, 预测值与原始值误差很小. 使用同样的方法得到合乘后上车点个数 N_{onh} 的预测值, 由图 12 可以看出, N_{on} 和 N_{onh} 随 N_c 呈现阻滞增长趋势, 当 N_c 不断增大时, N_{on} 和 N_{onh} 变化缓慢并趋于常数. 计算得到的 P_{non} 也有类似的规律, 如图 13 所示, 且 $N_c \leq 6k$ 时 P_{non} 的预测值为 $\{0, 0.0304, 0.0549, 0.0762, 0.0958, 0.1142, 0.1314\}$ 与原始数据很吻合. 当车辆个数 N_c 为 4 万时得到的预测值为 $N_{on} = 198366, N_{onh} = 125879, P_{non} = 0.37$.

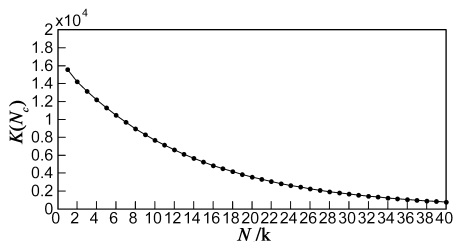


图11 N_{on} 增量 K 的灰色预测

5.3 非线性回归拟合

为了通过对比得到更加准确的结果, 下面采用另一种预测方法——非线性回归拟合进行分析.

根据之前的推测以及增量灰色预测得到的结果, 可以判断需要建立的模型为从 0 开始缓慢递增并逐渐趋于定值的函数, 在对比了多项式、立方根、指数函数、幂函数及 logistic 函数后, 最后选定 logistic 函数:

$$f(x) = \frac{k}{1 + ae^{-bx}}$$

作为预测函数模型^[15]. 根据实际情况, 当车辆个数为 0 时上车点个数也为 0, 因此得到修改后的预测模型为:

$$f(x) = \frac{k}{1 + ae^{-bx}} - \frac{k}{1 + a} \quad (12)$$

其中, x 为出租车数量, $f(x)$ 为上车点数量, k, a 和 b 为常数. 根据表 3 列出的上车点个数, 利用 MATLAB 中的曲线拟合函数 nlinfit 便可以得到 N_{on} 和 N_{onh} 关于 N_c 的函数:

$$f(x, N_{on}) = \frac{1037414}{1 + 0.20e^{-0.11x}} - \frac{1037414}{1 + 0.20} \quad (13)$$

$$f(x, N_{onh}) = \frac{948675}{1 + 0.14e^{-0.16x}} - \frac{948675}{1 + 0.14} \quad (14)$$

根据式(13)和(14)画出曲线如图 12 所示. 可以看出曲线的趋势与通过增量灰色预测方法得到的趋势非常相似. 同时也可以得到 P_{non} (如图 13 所示). 当车辆个数 N_c 为 4 万时, N_{on} 为 170902, 而 N_{onh} 为 114501, P_{non} 为 0.33, 均小于增量预测得到的结果. N_c 为 40k 时取两种方法得到的平均值有 N_{on} 为 184634, N_{onh} 为 120191, P_{non} 为 0.35.

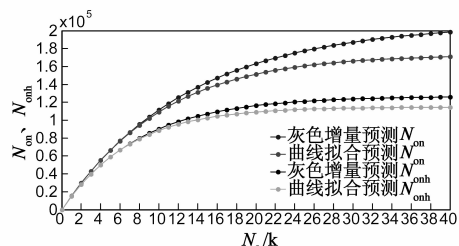


图12 两种预测方法得到的 N_{on} 和 N_{onh}

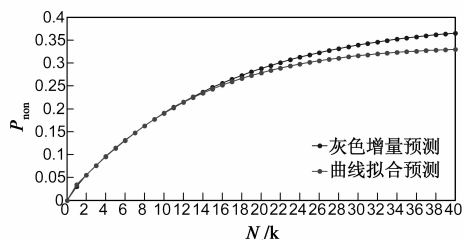


图13 两种预测方法得到的 P_{non}

通过上述预测方法可以得出: 数据采样当天, 北京市六环内上车点约为 17 ~ 20 万个, 出租车采用合乘方案后, 至少可以减少 1/3 的上车点. 根据原始数据的比例, 初步预测能够减少 20% ~ 30% 的行驶里程.

6 结论

本文从实际生活中打车难的现状出发, 提出一种将车载自组网技术应用于出租车合乘的解决方案. 在此基础上, 通过分析实际采集的北京市 6499 辆出租车 GPS 及相关信息, 从整体和局部分别进行研究, 并建立数学模型对市区出租车的合乘情况进行预测. 最终得

出现在现实生活中,乘客合乘出租车的方式是切实可行的,并定量给出了该方案所带来的收益。

上述分析只得到了理论的下限值,现实生活中合乘的方式有多种:不仅可以两人合乘,还可以最多四人合乘;不仅可以临时联系合乘对象,还可以提前约定时间地点实现合乘……。为使这些合乘方案能够更广泛地被人们熟悉和使用,有必要建立更加完备的智能交通系统,搭建更加智能的车载自组网络,设计更加符合现代都市环境的数据分发算法,使得有效信息资源能够更快、更合理、更安全地分发给需要的对象。

参考文献

- [1] 国内城市打车容易度调查[OL]. <http://news.xinhuanet.com>.
- [2] 傅丽叶. 智能出租车方案与多人多车搭乘算法[R]. 第二届北京国际智能交通系统(ITS)技术研讨暨技术与产品展览会, 2002.
- [3] 2012 北京市交通发展年度报告[OL]. <http://ishare.iask.sina.com.cn>.
- [4] 北京市“十二五”时期交通发展建设规划[OL]. <http://wenku.baidu.com>.
- [5] 北京鼓励乘客合乘出租车各付共同路段车费 60% [OL]. <http://news.sina.com.cn>.
- [6] Kianfar R, Augusto B, et al. Design and experimental validation of a cooperative driving system in the grand cooperative driving challenge[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2012, 13(3):994 – 1007.
- [7] Toutouh J, Garcia-Nieto J, Alba E. Intelligent OLSR routing protocol optimization for VANETs[J]. *IEEE Transactions on Vehicular Technology*, 2012, 61(4):1884 – 1894.
- [8] Zhang W X, Chen Y, et al. Multi-Hop connectivity probability in infrastructure-based vehicular Networks[J]. *IEEE Journal on Selected Areas in Communications*, 2012, 30(4):740 – 747.
- [9] Dedicated Short Range Communications home [OL]. <http://www.leearmstrong.com/dsrc>.
- [10] Wireless Access in Vehicular Environments [OL]. http://www.ieee802.org/11/Reports/tgp_update.htm.
- [11] 刘念伯. 车用自组织网络节点移动性研究[D]. 成都:电子科技大学, 2009.
- [12] 牛立军. 基于 Mapinfo 的 WebGIS 图形系统应用研究[D]. 北京:北京交通大学, 2008.
- [13] 卫宇. 考虑地球曲率情况下两点距离问题的求解[J]. *航空兵器*, 2008, (3):7 – 12.
Wei Y, Solution of distance between two points considering curvature of the earth[J]. *Aero Weaponry*, 2008, (3):7 – 12.
- [14] 邓聚龙. 灰色预测与决策[M]. 武汉:华中理工大学出版社, 1988.
- [15] 王兵团. 数学建模基础[M]. 北京:清华大学出版社, 北京交通大学出版社, 2004.

作者简介



宋 飞 男, 1983 年 4 月出生于河北, 博士, 讲师, 硕士生导师. 主要研究方向为新一代互联网体系架构、网络协议优化.

E-mail: fsong@bjtu.edu.cn