

基于加权时变泊松模型的出租车载客点推荐模型

商建东¹, 李盼乐^{1*}, 刘润杰¹, 李润川^{2,3}

(1. 郑州大学 智慧城市研究院, 郑州 450000; 2. 郑州大学 产业技术研究院, 郑州 450000;

3. 郑州大学 互联网医疗与健康服务河南省协同创新中心, 郑州 450000)

(* 通信作者电子邮箱 13137052075@163.com)

摘 要: 针对出租车空载率高、司机寻客难的问题, 提出泊松-卡尔曼组合预测模型(PKCPM)。首先, 采用加权非齐次泊松模型, 针对出租车历史数据进行建模, 得到目标时刻的估计值; 其次, 基于当天的实时数据, 将临近时刻乘客需求的平均值作为目标时刻预测值; 最后, 将预测值和估计值作为卡尔曼滤波模型的输入参数, 实现对目标时刻出租车乘客需求的预测, 同时引入误差反向传播机制, 减小下一次预测误差。基于郑州市出租车轨迹数据集, 对组合模型与非齐次泊松模型(NHPM)、加权非齐次泊松模型(WNHPM)、支持向量机(SVM)等三种模型进行对比, 实验结果显示 PKCPM 的误差比 WNHPM、SVM 分别降低了 8.85 个百分点、14.9 个百分点。该模型能对不同时段内、不同空间网格的乘客需求进行预测, 为出租车寻找乘客提供可靠的依据。

关键词: 空载率; 卡尔曼滤波预测模型; 加权时变泊松模型; 临近时刻乘客需求

中图分类号: TP181 **文献标志码:** A

Recommendation model of taxi passenger-finding locations based on weighted non-homogeneous Poisson model

SHANG Jiandong¹, LI Panle^{1*}, LIU Runjie¹, LI Runchuan^{2,3}

(1. Smart City Research Institute, Zhengzhou University, Zhengzhou Henan 450000, China;

2. Research Institute of Industrial Technology, Zhengzhou University, Zhengzhou Henan 450000, China;

3. Cooperative Innovation Center of Internet Healthcare, Zhengzhou University, Zhengzhou Henan 450000, China)

Abstract: To solve the problem of high taxi empty-loading ratio of taxi and difficulty in finding passengers, a new model called Poisson-Kalman combined prediction Model (PKCPM) was proposed. Firstly, weighted Non-Homogeneous Poisson Model (NHPM) was used to get the estimated value of the target time based on taxi historical data. Secondly, the mean value of the passenger demand in the near time, was taken as the predicted value, based on the real-time data. Finally, the predicted value and the estimated value were used as the inputs of Kalman filtering model to predict the target variance, meanwhile, the error backpropagation mechanism was introduced to reduce the next prediction error. The experimental results on the taxi trajectory dataset in Zhengzhou show that compared with NHPM, Weighted NHPM (WNHPM) and Support Vector Machine (SVM), PKCPM achieves a better optimization effect, and the error of PKCPM is reduced by about 8.85 percentage points and 14.9 percentage points respectively compared with WNHPM and SVM. PKCPM can predict passenger demand within different time and spacial grid, and provides a reliable solution to taxi driver for finding passengers.

Key words: empty-loading ratio; Kalman filtering prediction model; weighted Non-Homogeneous Poisson Model (NHPM); passenger demand in the near time

0 引言

随着城市人口和交通流的不断增加, 交通问题已成为现代化大城市的一个焦点问题, 例如, 在北京上下班高峰期期间, 公交车和地铁将会过载, 乘客需要等待至少 30 min 才能等到出租车^[1], 然而在其他时段, 绝大多数出租车要么空载, 要么停在某处等待乘客。据不完全统计, 绝大多数城市的出租车空载率在 50% 以上, 部分城市甚至高达 80% 以上^[2], 这导致了极大的交通资源浪费, 如何有效地降低出租车空载率, 进而提高出租车的利用率是智能交通的研究热点之一。

出租空载率高不仅造成市民打车难, 而且加剧了城市拥

堵、资源浪费和环境污染。在车载通信系统中, 车载全球定位系统(Global Positioning System, GPS) 会以一定的频率向管理中心发送当前的经纬度、时间、速度、方向以及载客状态等时空轨迹信息, 各地的出租车管理中心积累了大量的出租车 GPS 轨迹数据。通过对出租车数据进行时空分布挖掘, 可以动态感知不同区域内的上客分布规律, 将有助于出租车司机寻找最佳客源, 提升司机的经济收入, 进而减少车辆的空载。

本文基于海量出租车轨迹数据, 深入挖掘出租车上客时空分布特征, 结合非齐次泊松分布模型和指数平滑法对不同区域、不同时段内的乘客需求进行预测, 为出租车寻找乘客提供了可靠的依据。

收稿日期: 2017-09-28; 修回日期: 2017-12-04。

作者简介: 商建东(1968—), 男, 河南郑州人, 教授, 博士, 主要研究方向: 数据挖掘、高性能计算; 李盼乐(1992—), 男, 河南洛阳人, 硕士研究生, 主要研究方向: 轨迹挖掘; 刘润杰(1972—), 男, 河南安阳人, 副教授, 博士, 主要研究方向: 通信网络特性、混沌分形方法; 李润川(1991—), 男, 河南商丘人, 博士研究生, 主要研究方向: 智慧医疗、机器学习。

1 相关工作

近些年,关于出租车轨迹的研究越来越多,尤其是关于如何提升出租车利用率方面的研究。出租车利用率与许多因素有关,例如人口密度、出租车的策略以及出租车司机的经验等。

从出租车运营的行为方式角度出发,文献[3]使用 L_1 -Norm支持向量机(Support Vector Machine, SVM)对最能反映出租车寻找乘客行为的特征模式进行提取,基于历史数据,根据当前出租车所处的地点、时间以及司机的驾驶模式预测最优寻客策略。由于数据质量对特征提取的结果影响较大,因此在实际环境下,很难获取到与出租车寻客模式相关的特征。

基于历史数据,对城市历史出租车载客热点进行分析,预测当前最佳载客点是相关研究热点之一。文献[4]通过研究出租车上客下车的时空分布特征,提出了出租车上客下车时空分布的线密度探测模型。该模型通过对出租车上客下车事件的探测和分析,获取城市出租车上客下车的时空分布规律。文献[5]采用时间序列距离度量和时间序列自相关性结合的聚类方法,对北京市每个地块的出行量进行分析,从而得出出租车载客的时空分布特征。文献[6]基于历史GPS轨迹,采用改进的基于密度的聚类(Density-Based Spatial Clustering of Applications with Noise, DBSCAN)算法提取上客和未上客的热点区域,并将其可视化。文献[7]使用了K-Means算法对出租车上客地点进行聚类,得到上客点密集区域,为出租车寻找乘客提供了可靠的依据。文献[6-7]主要是通过聚类的手段对历史数据进行分析,寻找载客密集区域,从而预测当前时刻最佳载客地点,指导出租车司机有目的的巡游。但是,聚类方法对于参数较为敏感,很难得到较好的聚类结果;此外聚类算法时间复杂度较高,很难实现在线运行。文献[8-9]为出租车司机和乘客开发了一个完整的推荐系统,该系统对上客概率和等待时间构建了概率模型,通过该模型可以预测出租车成功找到乘客的概率以及乘客需要等待的时间。与聚类方法相比,文献[8-9]方法具有快速、准确的优点。

尽管以上方法可以为出租车推荐上客区域,但主要集中于定性地预测乘客聚集区域。与他们的研究相比,文献[10]基于历史出租车数据,采用加权时变泊松模型对不同时段的乘客需求进行预测,从定量的角度呈现乘客需求分布。加权时变泊松模型使用非齐次泊松模型对乘客需求进行预测,然后使用平滑因子法对同星期属性的历史乘客需求进行加权,很好地解决了突发性问题。

加权非齐次泊松模型仅仅考虑同星期属性的历史乘客需求的影响,忽略了临近时刻乘客需求的影响,使得预测结果并不是很好。本文认为当前乘客需求不仅与同星期属性的历史乘客需求有关,也与临近时刻的乘客需求相关。为此,本文在加权时变泊松模型的基础上,结合卡尔曼预测方法,融合临近时刻乘客需求因素,实现对不同区域内、不同时间段乘客需求的预测,为出租车寻找乘客提供可靠的依据。

2 预测模型

为了准确预测 t 时刻在 Z 地点要乘坐出租车的乘客数目,本文结合卡尔曼滤波预测模型和加权非齐次泊松模型(Weighted Non-Homogeneous Poisson Model, WNHPM),并融

合临近时刻乘客需求因素,实现对乘客需求的预测。

2.1 非齐次泊松模型

非齐次泊松模型(Non-Homogeneous Poisson Model, NHPM)是一种常见的统计可靠性模型,能够较为准确地描述自然界中随机现象发生次数随时间变化的规律,具有非常强的实用背景,例如:电话交换台在某段时间内到来的呼叫次数、服务台某段时间内到来的顾客数量等。由于城市出租车乘客数量具有随机性、有限次、独立性等特点,因此特定区域内到达乘客的数量可以用非齐次泊松分布来描述。

设 $N(r, k)$ 表示第 k 时段区域 r 内乘客数量,则 k 时段内有 n 个乘客的概率为:

$$P\{N(r, k) = n\} = \frac{(\lambda(t) \cdot t)^n}{n!} e^{-\lambda(t) \cdot t} \quad (1)$$

其中: $\lambda(t)$ 表示在 k 时段内乘客到达的速率函数。由于区域 r 乘客的到达受很多因素的影响,例如交通状况、个人因素等,所以 $\lambda(t)$ 是随着时间不断变化的。但在时间足够短时,即 $\|k\| < \delta$ (δ 是一个任意给的数), $\|k\|$ 表示第 k 个时间段的长度,乘客的到达可以看成是线性的,为了简化模型可以将 $\lambda(t)$ 看作是与时间相关的线性函数,不妨定义为:

$$\lambda(t) = a + bt \quad (2)$$

可以使用最大似然估计法^[8]对非齐次泊松分布的到达率函数进行求解。

非齐次泊松分布模型仅仅预测了时间相关的平均乘客需求,受交通拥堵、天气等意外因素的影响,每天相同时间段的乘客需求都略有不同。文献[10]基于历史出租车数据,采用平滑指数法对历史同一时刻的乘客需求进行加权,形成WNHPM。指数平滑法依据时间顺序对不同时段内的乘客到达率进行加权,使用逐步衰减的不等权办法,对最近的天赋予较大的权重,其目的是增加临近时间的相关度。权值 w 定义为:

$$w = (\alpha, \alpha(1-\alpha), \alpha(1-\alpha)^2, \dots, (1-\alpha)^{d_i-1}) \quad (3)$$

其中: α ($0 < \alpha < 1$) 是平滑因子,由用户自行设定; d_i 表示样本总个数。

2.2 泊松-卡尔曼组合预测模型

WNHPM 利用平滑指数对不同时段的乘客数量进行平滑处理,弱化了序列中极值的影响,很好地解决突发性问题;但是加权齐次泊松分布仅仅考虑了历史数据的影响,忽略了临近时刻乘客需求的影响,而临近时刻乘客需求对于模型预测精度是非常重要的。

城市中出租车乘客需求在不同时段内有着不同的分布规律,例如早高峰时段、晚高峰时段,并且白天和夜间也有着明显的区别^[11]。通常城市内产生的出租车乘客需求会分布在城市的不同地方,这些出行需求会按照一定的速率被城市出租车进一步转化为城市交通量,在交通路网上交流通流,最后,出租车乘客需求随着交通流转移和疏散。但是由于城市交通资源的限制,在整个连续性的出租车-乘客系统中,乘客不可能在一瞬间被完全转移,短时间内乘客的需求的变化是比较均匀的,并且受临近时刻需求量的影响较大。因此准确理解当天内临近时刻的乘客需求,并将这些信息应用到预测模型中是提高模型精确度的关键步骤。

由于城市环境的复杂性,出租车乘客需求预测往往很难使用单一的预测模型来实现精确预测。不同的假设可以建立不同的预测模型,不同预测模型对于输入数据有着不同的敏

感度。例如,受城市复杂性的影响,即使在短时间乘客到达率函数也很难满足线性性,因此,基于线性到达率函数的加权非齐次泊松分布模型的准确性必然会有有一定的损失。

卡尔曼滤波^[12-13]是一种非平稳的线性递归预测算法,以最小均方误差为估计准则,利用历史估计值和当前时刻的预测来更新当前对参数变量的估计,最终求得目标时刻的预测值,具有非线性特性。由于估计过程中考虑了系统的模型误差,因此模型本身对误差具有一定的鲁棒性,广泛应用于智能控制领域和城市交通状态预测。本文综合卡尔曼滤波预测模型和非齐次泊松模型的优势,提出泊松-卡尔曼组合预测模型,利用历史数据和临近时刻数据两类数据实现不同时段、不同区域内的乘客需求预测。

对于选定区域,按照经纬度划分将其划分为 $100\text{ m} \times 100\text{ m}$ ^[14] 的网格,按照网格的排列顺序对其进行编号,例如,编号 (r, c) 代表处于第 r 行 c 列的网格区域。除了对区域进行划分外,本文还对时间进行了划分,对每个时间段分别进行建模,目的是控制预测的时间间隔,提高预测精度。对于时间 t ,可以通过式(4)获取所处的时间段。

$$k = t_{\text{hour}} \times \frac{60}{\Delta t} + \left\lceil \frac{t_{\text{min}}}{\Delta t} \right\rceil \quad (4)$$

其中: Δt 表示时间间隔,即所要预测的时间跨度,由用户自定义。

因为相同星期属性的相同时刻有着较强的相关性^[15],所以,本文基于目标时刻同星期属性的历史数据,构建加权时变泊松分布模型,用于探测的历史数据中乘客需求规律。对于临近时刻数据,采用均值法实现对目标时刻的预测,则目标时刻的乘客需求组合模型可表示为:

$$\hat{D}_{\text{PKCPM}}(h_t) = \alpha D_N(h_t) + D_{\text{WNHPM}}(h_t) + K(h_t) * (D_N(d_t, h_t) - D_{\text{WNHPM}}(d_t, h_t)) \quad (5)$$

其中: h_t 表示目标时刻 t 所属的时间段; $\hat{D}_{\text{PKCPM}}(h_t)$ 为该时段内泊松-卡尔曼组合预测模型(Poisson-Kalman Combined Prediction Model, PKCPM)的预测值; $D_N(h_t)$ 表示临近的 N 个时段乘客需求的平均值,即

$$D_N(h_t) = \frac{1}{N} \sum_{i=1}^N D(h(t) - i) \quad (6)$$

K_{h_t} 为增益函数,根据前一时刻的临近时刻预测值 $D_N(h_t - 1)$ 与历史估计值 $D_{\text{WNHPM}}(h_t)$ 的误差来推测目标时刻的误差比,可以通过调节 m 值来控制对目标时刻的影响,其定义为:

$$K_{h_t} = \frac{D_N(h_t - 1) - D_{\text{WNHPM}}(h_t)}{D_N(h_t - 1) - D_{\text{WNHPM}}(h_t) + m} \quad (7)$$

算法 出租车乘客需求预测算法。

输入 出租车历史数据集 x_h , 当天数据集 x_r 。

输出 乘客需求预测值。

- 1) 初始化参数 α, m 。
- 2) While $|D_h| < N_h$ do
- 3) If $x \in x_h \wedge x$ 与目标时刻具有相同的星期属性
- 4) 将 x 加入 D_h
- 5) End
- 6) For $d_i \in D_h$
- 7) 求解到达率函数 $\hat{\lambda}_i$
- 8) End

- 9) 加权时变模型计算历史估计值 D_{WNHPM}
- 10) 基于当天数据计算临近时刻的预测值 D_N
- 11) 计算增益函数 K
- 12) 利用组合模型计算预测值 \hat{D}_{PKCPM}
- 13) End

3 实验与分析

3.1 数据集

实验数据来源于郑州市 2017 年 1 月 20 日—2 月 28 日的出租车 GPS 轨迹数据,共 35 d 数据 110 GB,该数据集大约由 30 个字段组成,包含经纬度、乘客状态、速度、方向等。本次数据分析仅提取了原始数据集中经纬度、状态、时间、车牌号 4 个字段作为有效字段。

本文使用 KS-Tests 测试上客量是否满足泊松分布,为了避免稀疏性的影响,本文以 10 min 为时间单位,按照上客量对不同的研究区域进行排序,分别选取 top1000、top2000、top3000 进行 KS-Tests 测试。

图 1 显示了在置信度为 95% 时,不同时段通过 KS-Tests 的比例。从图 1 可以看出,在各个时间段内通过 KS-Tests 的比例 top1000 大于 top2000, top2000 大于 top3000,这主要是因为随着研究区域的扩大,稀疏区域随之增多,而稀疏区域由于上客量较少,因此浮动较大,很难找到规律。考虑到本文的研究主要集中在繁忙区域,因此可以判定郑州市上客量满足泊松分布。

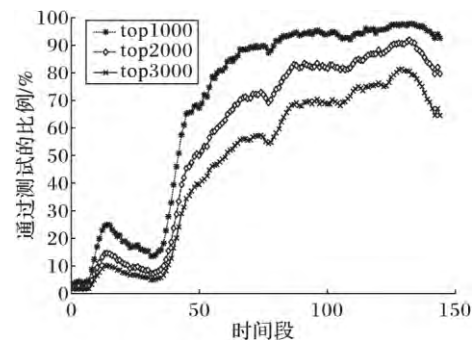


图 1 95% 置信度下通过 KS-Tests 的比例

Fig. 1 Proportion of successful KS-Tests at the 95% confidence level

泊松-卡尔曼组合预测模型主要利用了两部分数据:一部分为历史同时刻的乘客出行需求数据;另一部分是目标时刻同一天的前 N 个时刻的乘客出行需求数据。本文选取 3 个临近时刻对当前时刻进行预测,即 $N = 3$ 。此外,选择其中 21 d 数据作为训练集来训练模型,7 d 数据则作为测试集进行模型有效性的测试,如表 1 所示。

表 1 数据选择
Tab. 1 Data selection

预测日期	历史数据日期
2 月 20 日 周一	02 月 13 日 02 月 06 日 01 月 23 日
2 月 21 日 周二	02 月 14 日 02 月 07 日 01 月 24 日
2 月 22 日 周三	02 月 15 日 02 月 08 日 01 月 25 日
2 月 23 日 周四	02 月 16 日 02 月 09 日 01 月 26 日
2 月 24 日 周五	02 月 17 日 02 月 10 日 02 月 03 日
2 月 25 日 周六	02 月 18 日 02 月 11 日 02 月 04 日
2 月 26 日 周日	02 月 19 日 02 月 12 日 02 月 05 日

3.2 模型训练误差

为了对比并分析预测模型的预测性能,对预测模型的准确性进行评价,本文采用均衡平均绝对误差百分比 (Symmetric Mean Absolute Percent Error, sMAPE)^[15] 评价指标,定义如下:

$$sMAPE = \frac{1}{t} \sum_{i=1}^t \frac{|R_i - X_i|}{\delta_i} \quad (8)$$

$$\delta_i = \begin{cases} R_i + X_i, & R_i > 0 \text{ 或 } X_i > 0 \\ 1, & R_i = 0 \text{ 且 } X_i = 0 \end{cases} \quad (9)$$

图2展示了一周(周一至周日)的模型平均误差随时间的变化。白天误差随时间逐渐上升,上升幅度大约为0.2个百分点,这主要是由于白天市民出行率较高,且出行目的多样化,因此造成平均误差有不断上升趋势。到了晚上22:00以后,大多数人在家,整体出行率较低,因此误差有下降趋势。此外,不同日期之间的误差比较接近,没有太大差异,这表明了PKCPM具有较高的稳定性。整体来看,PKCPM训练误差较小(小于1.3%),表明该模型具有较好的精确性。

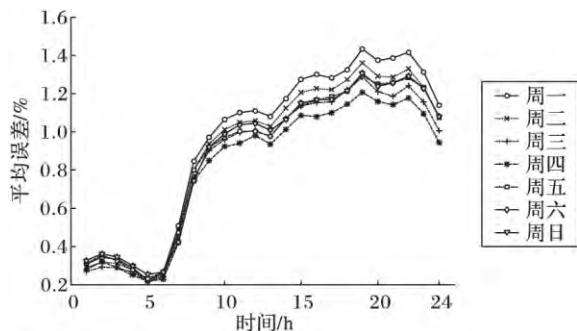


图2 PKCPM训练平均误差对比

Fig. 2 Comparisons of mean error on PKCPM

3.3 不同平滑因子的比较

为了获取最佳的预测结果,对PKCPM中的平滑因子进行调优,在不同平滑因子 α 下的预测误差是不同的,结果如表2所示。

表2 不同平滑因子下误差对比
Tab. 2 Error comparison with different α

α	最大误差/%	平均误差/%	α	最大误差/%	平均误差/%
0.1	41.37	10.15	0.6	31.61	8.76
0.2	37.48	9.88	0.7	30.72	8.51
0.3	34.08	9.59	0.8	30.26	8.26
0.4	33.28	9.31	0.9	30.37	8.33
0.5	32.49	9.03	1.0	30.81	8.76

从表2可以看出,不同的平滑因子 α 所对应的最大、平均误差略有变化,但是并不和 α 呈明显的线性关系,因此选择合适的 α 对于获取较低的误差比较重要。可以看出,当设置 $\alpha = 0.8$ 时,最大误差和最小误差均达到最小,即最大误差为30.26%,平均误差为8.26%。

3.4 模型效果对比与分析

为了验证PKCPM的执行效率,分别针对非齐次泊松模型(NHPM)、加权非齐次泊松模型(WNHPM)、支持向量机(SVM)以及泊松-卡尔曼组合预测模型(PKCPM)的运行时间和准确率作比较。仿真实验是普通台式机上的,其配置为i5-4460 CPU 8 GB内存。表3为当 $\alpha = 0.8$ 时,四种模型

在同一天内不同班次下的误差百分比对比。

表3 不同模型的sMAPE对比
Tab. 3 Comparisons of sMAPE on different models

模型	误差百分比					
	00:00—8:00		8:00—16:00		16:00—24:00	
	最大	平均	最大	平均	最大	平均
NHPM	34.28	5.40	38.37	12.01	39.58	12.40
WNHPM	33.17	4.57	36.41	11.14	37.96	12.40
PKCPM	27.32	4.36	29.32	9.78	29.11	10.65
SVM	28.27	6.34	34.48	19.56	38.18	25.55

从表3可以看出无论最大误差还是平均误差,PKCPM都最优。由于SVM训练过程难以掌控,容易出现过度拟合情况,因此,在整体上SVM的训练误差要高于其余三种方法,尤其是在平均误差方面,最大可达到25.55%,高出PKCPM模型14.9个百分点。WNHPM采用了对历史数据加权的方法,因此,相对于传统的NHPM方法,最大误差降低了1.96个百分点,平均误差降低了0.87个百分点。PKCPM组合模型在WNHPM基础上,结合卡尔曼滤波预测模型,充分考虑到临近时刻的对当前时刻的影响,大大降低了模型的误差。与WNHPM相比,最大误差降低了8.85个百分点,平均误差降低了1.75个百分点。

为保证运行时间的可信度,分别进行10次重复实验取得运行平均值,得出的运行结果如图3所示。实验结果表明PKCPM算法运行时间与预测数量呈近似线性关系,相对于SVM算法具有较高的运行效率,与NHPM和WNHPM具有相近的运行效率,但PKCPM比NHPM和WNHPM具有更好的预测效果。

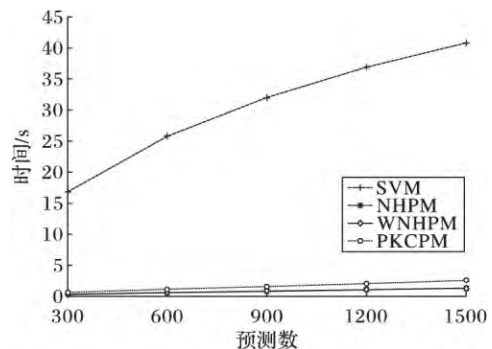


图3 运行时间对比

Fig. 3 Comparisons of running time

4 结语

有效地预测乘客需求不仅可以帮助出租车公司决定合理的车辆投放数量,而且还可以提高出租车利用率,增加出租车司机的利润,减少资源的浪费。本文的PKCPM组合模型利用卡尔曼滤波预测模型将加权时变泊松模型的预测值和临近时刻预测值融合起来,有效地降低了模型的误差,与WNHPM相比,最大误差降低了8.85个百分点,平均误差降低了1.75个百分点,在指导出租车司机巡游方面有着重要的作用。不过该模型仍然存在不足,即乘客需求是在不断变化的,但是对于卡尔曼滤波中的预测值,采用了临近时刻的平均值,过于简单。因此未来的工作将在原有基础上引入马尔可夫模型提高预测值的精确度,从而提高PKCPM的精确度。

参考文献(References)

- [1] JING W P, HU L K, SHU L, et al. RPR: recommendation for passengers by roads based on cloud computing and taxis traces data[J]. *Personal and Ubiquitous Computing*, 2016, 20(3): 337–347.
- [2] LEE J. Traveling pattern analysis for the design of location-dependent contents based on the taxi telematics system[EB/OL]. [2017-05-10]. <http://www.dbpia.co.kr/Journal/ArticleDetail/NODE01623693>.
- [3] LI B, ZHANG D Q, SUN L, et al. Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset[C]// *Proceedings of the 2011 IEEE International Conference on Pervasive Computing and Communications Workshops*. Piscataway, NJ: IEEE, 2011: 63–68.
- [4] 唐炳亮, 郑文斌, 王志强, 等. 城市出租车上下客的 GPS 轨迹时空分布探测方法[J]. *地球信息科学学报*, 2015, 17(10): 1179–1186. (TANG L L, ZHENG W B, WANG Z Q, et al. Space time analysis on the pick-up and drop-off of taxi passengers based on GPS big data[J]. *Journal of Geo-Information Science*, 2015, 17(10): 1179–1186.)
- [5] 程静, 刘家骏, 高勇. 基于时间序列聚类方法分析北京出租车出行量的时空特征[J]. *地球信息科学学报*, 2016, 18(9): 1227–1239. (CHENG J, LIU J J, GAO Y. Analyzing the spatio-temporal characteristics of Beijing's OD trip volume based on time series clustering method[J]. *Journal of Geo-Information Science*, 2016, 18(9): 1227–1239.)
- [6] SHEN Y, ZHAO L G, FAN J. Analysis and visualization for hot spot based route recommendation using short-dated taxi GPS traces[J]. *Information*, 2015, 6(2): 134–151.
- [7] LEE J, SHIN I, PARK G L. Analysis of the passenger pick-up pattern for taxi location recommendation[C]// *NCM 2008: Proceedings of the Fourth International Conference on Networked Computing and Advanced Information Management*. Piscataway, NJ: IEEE, 2008: 199–204.
- [8] YUAN J, ZHENG Y, ZHANG L H, et al. T-Finder: a recommender system for finding passengers and vacant taxis[J]. *IEEE Transactions on Knowledge & Data Engineering*, 2013, 25(10): 2390–2403.
- [9] YUAN J, ZHENG Y, ZHANG L, et al. Where to find my next passenger[C]// *UbiComp 2011: Proceedings of the 13th International Conference on Ubiquitous Computing*. New York: ACM, 2011: 109–118.
- [10] 周春姐, 戴鹏飞, 王晓玲, 等. 面向移动用户的乘客需求预测系统[J]. *计算机研究与发展*, 2015, 41(增刊1): 143–148. (ZHOU C J, DAI P F, WANG X L, et al. A system of passenger demand prediction for mobile users[J]. *Journal of Computer Research and Development*, 2015, 41(S1): 143–148.)
- [11] 王芮. 基于 GPS 数据的城市出租车出行需求研究[D]. 济南: 山东大学, 2016: 44–45. (WANG R. Research on urban taxi travel demand based on GPS data[D]. Jinan: Shandong University, 2016: 44–45.)
- [12] 林永杰, 邹难. 基于运营系统的出租车出行需求短时预测模型[J]. *东北大学学报(自然科学版)*, 2016, 37(9): 1235–1240. (LIN Y J, ZOU N. Short-term prediction model of taxi passenger demand based on operation systems[J]. *Journal of North-eastern University (Natural Science)*, 2016, 37(9): 1235–1240.)
- [13] 王涛波, 黄宝军. 基于改进卡尔曼滤波的四维飞行航迹预测模型[J]. *计算机应用*, 2014, 34(6): 1812–1815. (WANG T B, HUANG B J. 4D flight trajectory prediction model based on improved Kalman filter[J]. *Journal of Computer Applications*, 2014, 34(6): 1812–1815.)
- [14] 钱文逸, 蒋新华, 廖律超, 等. 基于时空上下文协同过滤的出租车载客点推荐算法[J]. *计算机应用*, 2015, 35(6): 1659–1662. (QIAN W Y, JIANG X H, LIAO L C, et al. Recommendation algorithm of taxi passenger-finding locations based on spatio-temporal context collaborative filtering[J]. *Journal of Computer Applications*, 2015, 35(6): 1659–1662.)
- [15] MOREIRA-MARTIAS L, GAMA J, FERREIRA M, et al. Predicting taxi-passenger demand using streaming data[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2013, 14(3): 1393–1402.
- [16] HOU Y, LI X, ZHAO Y, et al. Towards efficient vacant taxis cruising guidance[C]// *Proceedings of the 2013 IEEE Global Communications Conference*. Piscataway, NJ: IEEE, 2014: 54–59.
- [17] 张红, 王晓明, 过秀成, 等. 出租车 GPS 轨迹大数据在智能交通中的应用[J]. *兰州理工大学学报*, 2016, 42(1): 109–114. (ZHANG H, WANG X M, GUO X C, et al. Application of taxi GPS big trajectory data in intelligent traffic system[J]. *Journal of Lanzhou University of Technology*, 2016, 42(1): 109–114.)
- [18] 姬波, 叶阳东, 肖煜. 基于信息瓶颈方法的出租车空载聚集区聚类算法[J]. *小型微型计算机系统*, 2013, 34(9): 2139–2143. (JI B, YE Y D, XIAO Y. Empty taxi areas clustering algorithm based on information bottleneck method[J]. *Journal of Chinese Computer Systems*, 2013, 34(9): 2139–2143.)
- [19] 刘大有, 陈慧灵, 齐红, 等. 时空数据挖掘研究进展[J]. *计算机研究与发展*, 2013, 50(2): 225–239. (LIU D Y, CHEN H L, QI H, et al. Advances in spatiotemporal data mining[J]. *Journal of Computer Research and Development*, 2013, 50(2): 225–239.)
- [20] GE Y, XIONG H, ALEXANDER T Z L, et al. An energy-efficient mobile recommender system[C]// *KDD 2010: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2010: 899–908.
- [21] YUAN J, ZHENG Y, XIE X, et al. Driving with knowledge from the physical world[C]// *KDD 2011: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2011: 316–324.
- [22] PHITHAKKITNUKON S, VELOSO M, BENTO C, et al. Taxi-aware map: identifying and predicting vacant taxis in the city[C]// *AMI 2010: Proceedings of the First International Joint Conference on Ambient Intelligence*. Berlin: Springer, 2010: 86–95.
- [23] GUI Z, YU H, TANG Y. Locating traffic hot routes from massive taxi tracks in clusters[J]. *Journal of Information Science & Engineering*, 2016, 32(1): 113–131.

(下转第934页)

- ing. Piscataway, NJ: IEEE, 2015: 71 – 80.
- [27] XIANG Z, CHEN Z, GAO X, et al. Solving large-scale TSP using a fast wedging insertion partitioning approach [J]. *Mathematical Problems in Engineering*, 2015, 2015: Article ID 854218.
- [28] JIANG X, CHEN Y, LIU J, et al. Real-time and accurate indoor localization with fusion model of Wi-Fi fingerprint and motion particle filter [J]. *Mathematical Problems in Engineering*, 2015, 2015: Article ID 545792.
- [29] CHEN Z, CHEN Y, HU L, et al. ContextSense: unobtrusive discovery of incremental social context using dynamic bluetooth data [C]// *UbiComp 2014: Proceedings of the 2014 Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. New York: ACM, 2014: 23 – 26.
- [30] ZHANG H, YUAN J, GAO X, et al. Boosting cross-media retrieval via visual-auditory feature analysis and relevance feedback [C]// *MM 2014: Proceedings of the 22nd ACM International Conference on Multimedia*. New York: ACM, 2014: 953 – 956.
- [31] GAO X, HOI S, ZHANG Y, et al. SOML: sparse online metric learning with application to image retrieval [C]// *AAAI 2014: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*. Menlo Park: AAAI Press, 2014: 1206 – 1212.
- [32] CHEN Z, CHEN Y, WANG S, et al. Inferring social contextual behavior from bluetooth traces [C]// *UbiComp 2013: Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*. New York: ACM, 2013: 267 – 270.
- [33] CHEN Z, WANG S, SHEN Z, et al. Online sequential ELM based transfer learning for transportation mode recognition [C]// *Proceedings of the 2013 IEEE Conference on Cybernetics and Intelligent Systems*. Piscataway, NJ: IEEE, 2013: 78 – 83.
- [34] HU L, CHEN Y, WANG S, et al. OKRELM: online kernelized and regularized extreme learning machine for wearable-based activity recognition [J]. *International Journal of Machine Learning and Cybernetics*, 2016: 1 – 14.
- [35] HAGER W W. Updating the inverse of a matrix [J]. *SIAM Review*, 1989, 31(2): 221 – 239.
- [36] WANG W, HOU Z G, CHENG L, et al. Toward patients' motion intention recognition: dynamics modeling and identification of iLeg — an LLRR under motion constraints [J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2016, 46(7): 980 – 992.
- [37] OJETOLA O, GAURA E, BRUSEY J. Data set for fall events and daily activities from inertial sensors [C]// *MMSys 2015: Proceedings of the 6th ACM Multimedia Systems Conference*. New York: ACM, 2015: 243 – 248.
- [38] OJETOLA O. Detection of human falls using wearable sensors [D]. Coventry, UK: Coventry University, 2014: 49 – 66.
- [39] ALTUN K, BARSHAN B. Human activity recognition using inertial/magnetic sensor units [C]// *HBU 2010: Proceedings of the First International Workshop on Human Behavior Understanding, LNCS 6219*. Berlin: Springer, 2010: 38 – 51.

This work is partially supported by the National Natural Science Foundation of China (61702491), the Research Foundation of Hebei University of Economics and Business (2016KYZ05), the Science and Technology Program of State Grid Corporation of China (5442DZ170019), the Science and Technology Innovation Program of China Electric Power Research Institute (5242001600H5).

HU Lisha, born in 1986, Ph. D., lecturer. Her research interests include machine learning, wearable computing.

WANG Suzhen, born in 1964, Ph. D., professor. Her research interests include mobile cloud computing, big data analysis.

CHEN Yiqiang, born in 1973, Ph. D., professor. His research interests include pervasive computing, human-computer interaction.

HU Chunyu, born in 1990, Ph. D. candidate. Her research interests include machine learning, wearable computing.

JIANG Xinlong, born in 1988, Ph. D. candidate. His research interests include machine learning, wearable computing.

CHEN Zhenyu, born in 1985, Ph. D., senior engineer. His research interests include machine learning, wearable computing.

GAO Xingyu, born in 1985, Ph. D., associate professor. His research interests include machine learning, wearable computing, multimedia.

(上接第 927 页)

- [24] ZHANG D, SUN L, LI B, et al. Understanding taxi service strategies from taxi GPS traces [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2015, 16(1): 123 – 135.
- [25] RAO F, ZHANG X, DONG Y, et al. Understanding the mobility pattern of passenger-searching taxis [C]// *Proceedings of the 2014 IEEE 17th International Conference on Intelligent Transportation Systems*. Piscataway, NJ: IEEE, 2014: 290 – 295.
- [26] QIAO S, HAN N, ZHU W, et al. TraPlan: an effective three-in-one trajectory-prediction model in transportation networks [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2015, 16(3): 1188 – 1198.
- [27] QIAO S, SHEN D, WANG X, et al. A self-adaptive parameter selection trajectory prediction approach via hidden Markov models [J]. *IEEE Transactions on Intelligent Transportation*, 2015, 16(1): 284 – 296.
- [28] 乔少杰, 韩楠, 李天瑞, 等. 一种基于空间编码技术的轨迹特征提取方法 [J]. *中国科学: 信息科学*, 2017, 47(11): 1523 – 1537.
- (QIAO S J, HAN N, LI T R, et al. A trajectory feature extraction approach based on spatial coding technique [J]. *Chinese Science: Informationis*, 2017, 47(11): 1523 – 1537.)
- [29] QIAO S, HAN N, WANG J, et al. Predicting long-term trajectories of connected vehicles via the prefix-projection technique [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2017, PP(99): 1 – 11.

SHANG Jiandong, born in 1968, Ph. D., professor. His research interests include data mining, high performance computing.

LI Panle, born in 1992, M. S. candidate. His research interests include trajectory data mining.

LIU Runjie, born in 1972, Ph. D., associate professor. His research interests include characteristics of communication network, chaotic fractal method.

LI Runchuan, born in 1991, Ph. D. candidate. His research interests include intelligent medical treatment, machine learning.