

基于 PSO-LSSVM 的高速公路短时行程时间预测

刘伟铭, 雷焕宇, 翟聪, 李松松

(华南理工大学 土木与交通学院, 广东 广州 530640)

摘要: 高速公路行程时间预测对指导公众出行具有重要意义。鉴于高速公路数据采集难, 文中以高速公路收费数据为研究对象, 建立最小二乘支持向量机模型对高速公路行程时间进行预测, 并引入粒子群算法对最小二乘支持向量机的参数进行优化; 利用广东某高速公路收费数据进行验证, 结果显示该模型与以往研究相比其相对误差较小, 用于高速公路行程时间预测具有一定的可靠性和实用性。

关键词: 公路交通; 高速公路; 行程时间; 最小二乘支持向量机(LSSVM); 粒子群优化(PSO)

中图分类号: U491.1

文献标志码: A

文章编号: 1671-2668(2017)03-0036-04

行程时间不仅反映了道路运行状况与交通拥堵水平, 也是公众选择出行路线的合理依据。随着车辆数量的增加, 道路交通环境越发复杂, 行程时间预测难度加大, 准确预测道路行程时间成为智能交通信息服务系统的研究难点, 也吸引了大批学者进行研究。以往主要通过浮动车数据和固定检测器数据对高速公路行程时间进行预测, 这些方法在采集数据过程中对硬件要求高、成本大、维护难, 导致数据质量不高或数据不能完整反映实际道路运行状况, 需建立复杂的数据处理和修正算法模型。而目前高速公路是封闭的, 路段间距长, 收费数据中包含完整的车辆行驶信息, 合理利用收费数据可有效获得当前时段高速公路各路段的交通情况。因此, 越来越多的学者尝试利用高速公路收费数据对高速公路行程时间进行分析和预测, 如文献[7]对基于高速公路联网收费数据的行程时间预测进行了可靠性分析, 文献[8]~[10]分别利用卡尔曼滤波模型和 BP 神经网络模型对行程时间进行预测, 预测误差在 10% 以内, 具有一定实用价值。但以往方法需建立复杂的预测模型, 同时所设计算法收敛速度较慢, 准确率不高, 无法满足实际要求。为此, 该文使用最小二乘支持向量机(LSSVM)对高速公路行程时间进行预测, 通过对支持向量机模型的约束条件进行改进使其更符合实际要求, 同时利用粒子群优化算法(PSO)对模型中的参数进行优化, 降低模型对核心参数的灵敏度。

1 高速公路原始收费数据及其预处理

主要对高速公路收费数据进行分析。表 1 为高速公路原始收费数据中的部分信息, 其中包含高速

公路站点出入车辆及其他附加系统信息。该文利用起点、终点、出发时间、到达时间等主要信息进行行程时间预测。

表 1 高速公路原始收费数据中的基本信息

数据类别	基本记录
入口处记录数据	入口记录号、入口站编码、入口站名、入口车道编码、入口车型、入口车种、入口车牌、入口日期和时间
出口处记录数据	出口记录号、出口站编码、出口车道编码、出口站名、出口车型、出口车种、出口车牌、出口日期和时间

原始数据是含有噪声的有质量问题的数据, 不能直接使用, 需进行数据预处理, 得到基本可用的有效数据。处理步骤:

(1) 剔除错误数据。错误数据大部分是由软件异常导致的。根据以下规则删除错误数据: 1) 出入口记录号是否唯一; 2) 出口时间是否大于入口时间; 3) 出口时间与入口时间之差是否小于 24 h。删除不满足上述 3 个条件之一的数据, 得到基本有效的数据。

(2) 筛选有效数据。由第一步得到的基本有效数据含有部分特殊情况下的数据, 如由个别司机的驾驶特性及突发交通事故所产生的数据, 这部分数据不能反映普遍车辆的行驶特征, 对行程时间预测会产生较大干扰。文献[14]指出单位时间内个体行程时间的分布服从正态分布, 服从标准正态分布时, 样本数据在两倍标准差范围内的概率为 95.4%。据此采用如下方法进行数据筛选: 1) 计算旅行时间的

阈值下限。高速公路限速 120 km/h,假设最大速度为限速值的 120%,以最快速度行驶,可求出最小行程时间,删除低于该值的数据。2) 计算剩余样本的均值和方差。按式(1)计算某路段当天某时间段 t 内所有车辆 N 的旅行时间样本均值,按式(2)计算旅行时间样本标准差,样本方差开平方即为样本标准差。标准差具有样本相同的单位量纲,能描述样本数据的离散程度。3) 删除落在区间 $[\bar{\tau}^{(t)} - 2S, \bar{\tau}^{(t)} + 2S]$ 外的数据,转入步骤 2,直到数据全部落入两倍标准差范围内,然后进入下一步。4) 计算筛选后的样本均值 $T^{(t)} = \bar{\tau}^{(t)}$ 。

$$\bar{\tau}^{(t)} = \frac{1}{N} \sum_{n=1}^N \tau_n^{(t)} \quad (1)$$

$$S = \sqrt{\frac{\sum_{n=1}^N (\bar{\tau}^{(t)} - \tau_n^{(t)})^2}{N-1}} \quad (2)$$

(3) 插补缺失数据。采用线性插值法补充缺失数据,左右数据段不全时使用左右插值法,前后数据不全时使用上下插值法。

2 最小二乘支持向量机模型的建立

最小二乘支持向量机模型将传统支持向量机中不等式约束改成等式约束,并将误差平方和损失函数作为训练集的经验损失,将解二次规划问题转化为求解线性方程组问题,从而提高求解速度和收敛精度。

对于给定的训练样本集 $D = (x_i, y_i) (i=1, 2, \dots, n)$,在向量机回归算法中,通过非线性映射函数 $\varphi(x)$ 将样本输入空间 R^n 映射到高维特征空间 H ,并采用下式估计未知的非线性函数:

$$f(x) = \omega^T \varphi(x) + b \quad (3)$$

式中: ω 为加权系数向量, $\omega = [\omega_1, \omega_2, \dots, \omega_n]^T$; $\varphi(\cdot)$ 为映射函数, $\varphi(\cdot) = [\varphi_1(\cdot), \varphi_2(\cdot), \dots, \varphi_n(\cdot)]^T$; ω, b 为待定的未知参数, $\omega \in H, b \in R$ 。

根据结构风险最小化原理,综合考虑复杂度和拟合误差,将不等式约束表示为等式约束的优化问题,目标函数为:

$$\min_{\omega, b, e} J(\omega, e) = \frac{1}{2} \omega^T \omega + \frac{\gamma}{2} \sum_{i=1}^n e_i^2 \quad (4)$$

$$\text{s.t. } e_i = y_i - \omega^T \varphi(x_i) - b \quad (i=1, 2, \dots, n)$$

式中: γ 为惩罚项参数; e_i 为第 i 个数据实际输出 y_i 和预测输出 $f(x_i)$ 之间的误差, $e_i \in R$ 。

式(4)中,第一项表示模型泛化能力,第二项表示模型精确性。

建立拉格朗日等式进行求解,将约束问题变为无约束优化问题:

$$L(\omega, b, e, a) = J(\omega, e) - \sum_{i=1}^N a_i (\omega^T \varphi(x_i) + b + e_i - y_i) \quad (5)$$

式中: $a_i (i=1, 2, \dots, n)$ 为拉格朗日乘子。

根据 KKT(求解凸二次规划最优解的充分必要条件)条件对式(5)进行求导,得:

$$\begin{cases} \frac{\partial L}{\partial \omega} = 0 \rightarrow \sum_{i=1}^N a_i \varphi(x_i) - \omega = 0 \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^N a_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 \rightarrow a_i - \gamma e_i = 0 \\ \frac{\partial L}{\partial a_i} = 0 \rightarrow \omega^T \varphi(x_i) + b + e_i - y_i = 0 \end{cases} \quad (6)$$

定义核函数 $k(x, y_j) = \varphi^T(x) \varphi(y_j)$ 代替非线性映射函数,从方程组消去后,得:

$$\begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & k(x_1, x_2) + \frac{1}{\gamma} & \cdots & k(x_1, x_2) \\ \vdots & \vdots & \cdots & \vdots \\ 1 & k(x_1, x_2) & \cdots & k(x_1, x_2) + \frac{1}{\gamma} \end{bmatrix} \times \begin{bmatrix} b \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix} \quad (7)$$

由式(7)即可求出 a, b ,得到如下最小二乘支持向量机模型:

$$f(x) = \sum_{i=1}^N a_i k(x, x_i) + b \quad (8)$$

式中: $k(x, x_i)$ 为核函数。

用核函数代替映射函数,是为了降低计算的复杂度。常用核函数主要有多项式核函数、高斯核函数、感知器核函数。不同的核函数,对模型计算复杂度的影响不一样。高斯核函数是正定核,且调节参数少,故确定使用高斯核函数。LSSVM 的惩罚项 γ 和核参数的选择直接影响模型的效果,为此,引入粒子群优化算法 PSO 进行参数优选。

3 粒子群优化算法

PSO 是一种基于仿生学原理的进化计算技术,具有相当快的逼近最优解的速度,可有效对系统参

数进行优化。粒子群算法的本质是利用当前位置、全局最优值和个体最优值来指导粒子下一步迭代位置,其个体充分利用自身经验和群体经验调整自身状态是粒子群算法具有优异特性的关键。粒子的位置和速度根据如下方程进行更新:

$$\begin{aligned} v_{iD}^{k+1} &= \omega v_{iD}^k + c_1 \xi (p_{iD}^k - x_{iD}^k) + c_2 \eta (p_{gD}^k - x_{iD}^k) \\ x_{iD}^{k+1} &= x_{iD}^k + v_{iD}^{k+1} \end{aligned} \quad (9)$$

式中: ω 为动力常量,其大小决定前一速度对当前速度的影响; c_1 和 c_2 为学习因子或加速系数, c_1 调节粒子飞向自身最好位置方向的步长, c_2 调节粒子飞向全局最好位置的步长; $\xi, \eta \in U(0, 1)$ 。

适应度函数对粒子解的选择至关重要,这里选择 K-CV 意义下的准确率作为 PSO 的适应度函数值。K-CV 将原始数据分为 K 组,将每个子集数据分别进行一次验证集,其余 $K-1$ 组子集数据作为训练集,这样可得到 K 个模型,用这 K 个模型最终验证集下准确率平均数作为 K-CV 意义下的性能指标。一般 $K \geq 2$,这里取 $K=3$ 。粒子群优化算法优化最小二乘支持向量机参数的步骤:

(1) 初始化。设置粒子群的种群规模 M 、进化迭代次数 N 、最大速度 v_{\max} 、动力常量 ω 、学习因子 c_1 和 c_2 、各粒子的初始位置 y_0^i 和初始速度 v_0^i 。

(2) 通过适应度函数计算每个粒子的适应值。

(3) 更新最优位置。对每个可行的粒子,比较每个粒子的当前适应值和其历史最好适应值,选取最优适应度的粒子位置作为当前最优位置;然后比较每个粒子的当前适应值和全局经历过的最好适应值,选择最好适应值的粒子位置作为全局最好位置。

(4) 更新粒子位置及速度。根据式(9)计算粒子的新速度 v_{i+1}^i ,并根据事先设定的范围对各粒子进行速度限制。

(5) 终止迭代。若满足终止条件(达到种群进化代数或目标值达到设定阈值范围),则输出所求最优解,否则返回步骤 2。

4 基于粒子群最小二乘支持向量机的旅行时间预测

选取某年广州—东莞高速公路路段的收费数据,对 3 月 2 日—4 月 11 日的数据进行预处理,以每 15 min 为一个时间点,计算该时间段内的平均行程时间,统计每天 7:00—19:00 共产生的 48 组数据。前 20 d 共 960 组数据用于模型训练,然后对第 21 d 的旅行时间进行预测。按图 1 所示流程进行试验,试验在 MATLAB R2014a 上进行。

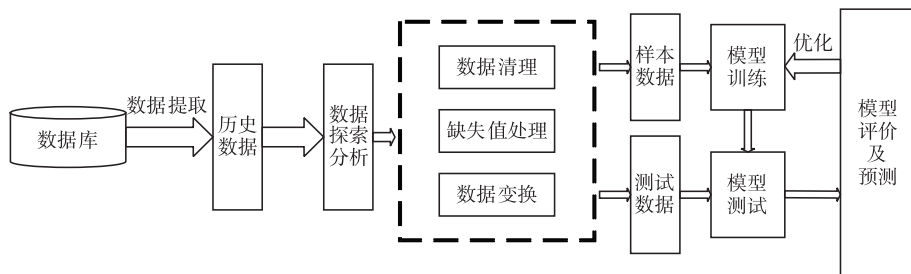


图 1 最小二乘支持向量机模型建立流程

为了使最小二乘支持向量机模型的性能最优,采用粒子群算法优化模型参数。粒子群算法参数如下:学习因子 $c_1=1.5$ 、 $c_2=1.7$,种群进化代数 $N=100$,种群数量 $M=20$,动力常量 $\omega=1$;采用 K-CV 方法对训练样本进行交叉验证,将对训练集进行 K-CV 意义下的准确率作为 PSO 中的适应度函数值。图 2 为参数优化过程中适应度变化曲线。从图 2 可见,随着种群迭代次数的增加,适应度函数值不断减小,迭代至第 58 代时,适应度函数值最低,此时正则化参数 $\gamma=100$,核函数参数 $\sigma=0.01$ 。

通过 PSO 算法对 LSSVM 预测模型的参数进行优化,运用优化后参数构建 PSO-LSSVM 预测

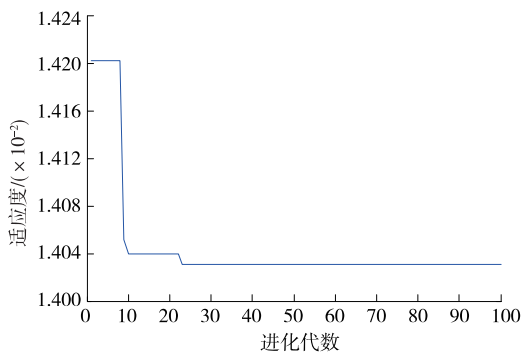


图 2 PSO 适应度变化曲线

模型对后面行程时间进行预测,预测结果与真实数据的比较见图 3。为了验证模型的有效性,选用 BP

神经网络进行性能比较。图 4 为 BP 神经网络模型预测结果与真实数据对比。从中可见, PSO—LSSVM 模型和 BP 神经网络在行程时间预测趋势上具有同步性,但 PSO—LSSVM 模型在局部时间点的精度比 BP 神经网络模型略高,其整体效果更好。但两模型在某些特殊奇异点还存在一定误差。

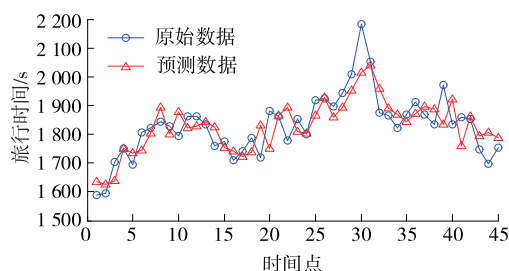


图 3 PSO—LSSVM 模型预测结果与真实数据对比

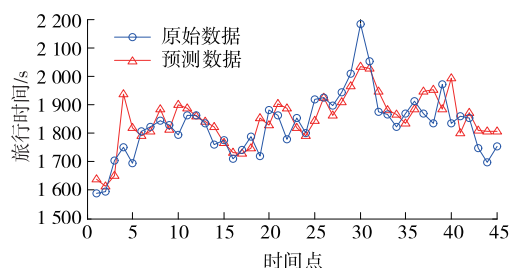


图 4 BP 网络模型预测结果与真实数据对比

图 5、图 6 分别为 PSO—LSSVM 行程时间预测模型的相对误差和绝对误差。从中可见,经过粒子群优化后的 LSSVM 模型的行程时间预测绝对误差在 150 s 内,其中大部分落在 60 s 以内,满足交通诱导的要求。

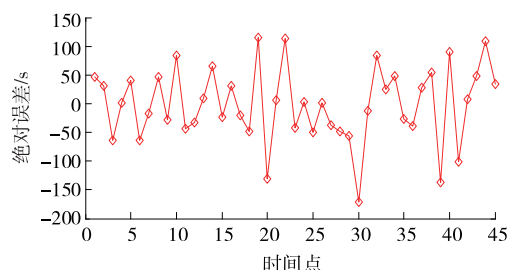


图 5 PSO—LSSVM 预测模型的绝对误差

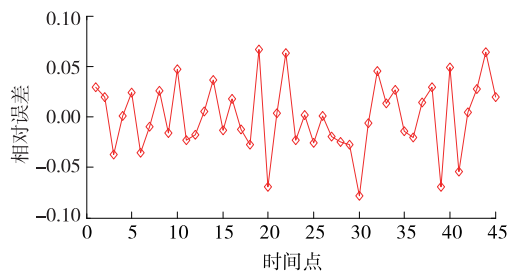


图 6 PSO—LSSVM 预测模型的相对误差

引入平均相对误差、平均绝对误差两个评价指标对 BP 神经网络和 PSO—LSSVM 模型进行比较,结果见表 2。从中可见, PSO—LSSVM 模型的平均相对误差和绝对误差均比 BP 神经网络低,预测精度比神经网络好, PSO—LSSVM 模型的预测效果更理想。

表 2 PSO—LSSVM 模型和 BP 神经网络模型预测效果对比

预测模型	平均相对误差/%	平均绝对误差/s
PSO—LSSVM	3.04	45.575 9
BP 神经网络	7.23	55.341 3

5 结语

该文在分析高速公路收费数据的基础上,通过数据预处理筛选出能反映道路实际交通状况的有效数据,建立高速公路站间短时行程时间预测模型,并利用某高速公路收费数据对该模型进行验证。结果表明, BP 神经网络和 PSO—LSSVM 模型都能对道路旅行时间进行预测, LSSVM 模型的预测效果在某些局部时间点比 BP 模型好,其平均相对误差和平均绝对误差更低。但两者在某些奇异点处的预测误差均较大。粒子群算法早熟收敛问题及如何提高奇异点的预测精度还有待进一步研究。

参考文献:

- [1] Yang J S. Travel time prediction using the GPS test vehicle and Kalman filtering techniques[A]. Proceedings of the American Control Conference[C]. 2005.
- [2] Chien S I, Kuchipudi C M. Dynamic travel time prediction with real-time and historic data[J]. Journal of Transportation Engineering, 2003, 129(6).
- [3] 朱爱华. 基于浮动车数据的路段旅行时间预测研究[D]. 北京: 北京交通大学, 2007.
- [4] Yeon J, Elefteriadou L, Lawphongpanich S. Travel time estimation on a freeway using discrete time Markov chains[J]. Transportation Research Part B: Methodological, 2008, 42(4).
- [5] 杭明升, 杨晓光, 彭国雄. 基于卡尔曼滤波的高速道路行程时间动态预测[J]. 同济大学学报: 自然科学版, 2002, 30(9).
- [6] Wosyka J, Pribyl P. Real-time travel time estimation on highways using loop detector data and license plate recognition[A]. 2012 ELEKTRO[C]. 2012.

(下转第 48 页)

出租车服务创造和谐有序的竞争环境。其实网约车与出租车并不是真正意义上的竞争关系,二者都属于一种出行方式,只不过是运营与管理方式不同,网约车只是传统出租车某些程度上的替代。面对新兴的网约车行业,传统出租车的弊端暴露无遗,如何提升自身的服务与管理才是目前需重点关注的问题,而不是一味地寻求监管部门的保护来压制网约车的发展。也正是由于网约车的高速发展,在一定程度上促进了传统出租车行业的变革,相信经过协调有序的发展,在未来一定可实现二者的融合发展。

4.4 优化价格区间,鼓励公交优先

关于网约车的定价问题,除市场与政府调节外,各市可根据本地实际情况对网约车的价格浮动区间设置上下限,保证乘客利益的同时避免平台公司利用自身市场优势恶性竞争。通过收取交通拥堵费、实行交通管制、鼓励并规范“顺风车”、“拼车”等减少资源浪费,减缓交通拥堵问题。通过大力发展公共交通,鼓励公众公交优先、轨道等大型运输工具优先,减少道路运行压力,自然道路的拥堵、环境等问题也能得到一定程度缓解。

5 结语

网约车等共享经济的出现,对人们的出行与资源共享产生了举足轻重的影响。新政对网约车平台与驾驶员和车辆的严格审查是对乘客、也是对市场负责的表现。政府将权利下放到各省市,各地因地制宜地出台政策来治理交通拥堵、环境、安全等问题,是管理网约车行之有效的办法,对于规范行业具有积极作用。从长远来看,发展网约车可有效利用资源并满足民众多层次出行需求,对缓解出行难问

题具有积极作用。对网约车管理办法进行研究对促进其健康、有序发展具有重要意义。

参考文献:

- [1] 蔡乐渭.网约车新政,小心潜在的负面影响[J].法律与生活,2016(17).
- [2] 七部委令 2016 年第 60 号,网络预约出租汽车经营服务管理暂行办法[S].
- [3] 国办发〔2016〕58 号,国务院办公厅关于进一步深化改革推进出租汽车行业健康发展的指导意见(征求意见稿)[S].
- [4] 刘丽丽.中消协网约车大体验:神州专车第一滴滴第二优步分较低[EB/OL].http://m.ce.cn/cj/gd/201610/25/t20161025_17142621.shtml,2016-10-25.
- [5] 徐燕倩.网约车乱象丛生,不提供完整材料可注册优步司机[EB/OL].http://www.thepaper.cn/newsDetail_forward_1536908,2016-09-30.
- [6] 唐龙.“专车”在国内发展的时间轴[EB/OL].http://www.legaldaily.com.cn/zt/content/2015-10/13/content_6303110.htm?node=78438,2015-10-13.
- [7] 孙兴军.网络约租车的风险及其规避[J].安徽行政学院学报,2016,7(4).
- [8] 孙亚杰.网约车违约问题研究[J].商,2016(8).
- [9] 李少威.恶性事件不断,C2C 网约车四面楚歌[J].南风窗,2016(11).
- [10] 荣朝和.厘清网约车性质,推进出租车监管改革[J].综合运输,2016(1).
- [11] 何鹏.多中心治理视角下的网约车管理对策研究[D].南宁:广西大学,2016.
- [12] 唐培.中国网约车的监管困境及解决对策[J].智能城市,2016(9).

收稿日期:2016-12-30

(上接第 39 页)

- [7] 尹志鹏,刘伟铭,庄岩浩.基于高速公路收费数据的旅行时间可靠性分析[J].公路与汽运,2014(5).
- [8] 赵建东,徐菲菲,张琨,等.融合多源数据预测高速公路站间旅行时间[J].交通运输系统工程与信息,2016,16(1).
- [9] 赵建东,王浩,刘文辉,等.基于收费数据的高速公路站间旅行时间预测[J].同济大学学报:自然科学版,2013,41(12).
- [10] 赵建东,王浩,刘文辉.高速公路旅行时间的自适应插值卡尔曼滤波预测[J].华南理工大学学报:自然科学版,2014,42(2).

- [11] 徐菲菲.基于微波数据的高速公路交通拥挤检测及旅行时间预测[D].北京:北京交通大学,2016.
- [12] 孙宇星,关伟,葛昱,等.基于支持向量机方法的轨道交通乘客旅行时间短时预测方法研究[J].系统工程理论与实践,2014,34(6).
- [13] 黄龙超.基于数据驱动的路段旅行时间估计与预测研究[D].北京:北京交通大学,2015.
- [14] Van Lint J W C. Reliable travel time prediction for freeways [D]. Netherlands: Travel Research School, 2004.

收稿日期:2016-12-23