

高速公路交通状态预测的多源数据融合研究^{*}

姚午开, 韩子雯, 高志波

(长沙理工大学 交通运输工程学院, 湖南 长沙 410114)

摘要: 针对目前高速公路交通状态预测大量使用单一数据源的不足, 通过对收费站抽样刷卡数据与浮动车 GPS 数据的预处理, 得到体现同一时间和同一空间的交通状态多源数据; 采用极限学习机算法对高速公路交通状态进行预测和估计, 并用行程时间指数 TTI 对道路交通状态进行估计; 以广州机场高速公路南线为例, 对比分析结果表明, 采用多源数据融合模型得出的行程时间均方根误差和平均绝对百分比误差均小于单一数据源模型, 且极限学习机在数据准确性与稳定性上均有良好效果, 优化率分别为 80.6%、53.7%。

关键词: 公路交通; 交通状态预测; 数据融合; 极限学习机; 行程时间

中图分类号: U491.2

文献标志码: A

文章编号: 1671-2668(2019)03-0016-04

高速公路交通状态预测的准确性与时效性对交通信息发布与交通诱导实施方案制订有相当大的影响, 特别是在一些较忙碌的路段。由于用于检测交通流数据的检测器的特性、功效各不相同, 甚至在同一路段存在多组检测精度、检测广度和检测内容不相同的检测器, 且这些检测器所反映的数据都只是交通流的一个侧面, 需将不同检测器所采集的交通流状态数据进行融合得到更真实、全面的交通状态数据用于交通状态预测。Jun Xia 等提出了一种基于贝叶斯网络的信息融合模型, 将融合过程分为信息感知层、特征聚类层和决策融合层, 基于融合算法融合各类交通信息。Bachmann C. 等将浮动车数据与检测器数据融合预测高速公路行车速度。Tang-Hsien Chang 等使用 ETC 系统和传统车辆检测器数据融合技术建立卡尔曼滤波短期预测模型和傅立叶变换长期预测模型进行交通状态预测, 预测误差不大于 10%。史岩等使用神经网络与回归分析法, 将浮动车与微波检测器得到的速度进行融合, 得到更全面的速度数据。蒋肖利用 SVM 模型与 RBF 模型对交通量、占有率和平均车速等交通参数进行融合, 实现短时交通流预测。

在交通流密度、行车速度与行程时间等交通参数中, 出行者最关心并据此进行规划决策的指标是行程时间。但行程时间受交通流量、驾驶员状态、天气环境等因素的影响, 需使用多源数据更全面地反映交通流状态。该文以行程时间为主要预测对象,

利用数据中的时间和空间匹配特性, 选用相同的路段地图数据、浮动车 GPS 数据和收费站抽样刷卡数据, 通过极限学习机对交通流状态数据进行融合。

1 数据融合模型构建

1.1 模型提出

利用机器学习算法进行高速公路交通状态预测是一研究热点。机器学习效率高, 且具有模拟人类思维推理的特点, 在数据量不断增大的今天, 可加大数据的利用效率与深入挖掘数据内部含义。经过多种方法验证, 结合交通数据的基础结构和特点, 选择 BP 神经网络与极限学习机 (ELM) 模型对多源数据进行融合。

1.2 极限学习机模型

ELM 模型中包含输入层、隐含层和输出层 3 层神经元, 其中输入层的每个神经元为一个输入变量 (模型中有 n 个神经元), 隐含层的每个神经元为一个运算隐藏变量 (模型中有 l 个神经元), 输出层的每个神经元为一个输出变量 (模型中有 m 个神经元)。3 层神经元中还存在隐含的权重值, 将输入层与隐含层间的权重值设为 w , 隐含层中神经元的阈值设为 b , 隐含层与输出层间的权重值设为 β , 得:

$$w = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ w_{l1} & w_{l2} & \cdots & w_{ln} \end{bmatrix}_{l \times n}$$

^{*} 基金项目: 国家自然科学基金项目 (51678076; 51408065)

$$\beta = \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{m1} & \beta_{m2} & \cdots & \beta_{mm} \end{pmatrix}_{l \times m}, b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_l \end{pmatrix}_{l \times 1}$$

式中: w_{ij} 与 β_{jk} 分别为输入层与隐含层、输出层与隐含层之间的权重值。

设 Q 个样本的训练输入矩阵 X 及输出矩阵 Y 分别为:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1Q} \\ x_{21} & x_{22} & \cdots & x_{2Q} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nQ} \end{pmatrix}_{n \times Q}$$

$$Y = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1Q} \\ y_{21} & y_{22} & \cdots & y_{2Q} \\ \vdots & \vdots & \vdots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mQ} \end{pmatrix}_{m \times Q}$$

隐含层神经元的激活函数为 $g(x)$, 则网络的输出为:

$$T = (t_1, t_2, \dots, t_Q)_{m \times Q}$$

$$t_j = \begin{bmatrix} t_{1j} \\ t_{2j} \\ \vdots \\ t_{mj} \end{bmatrix}_{m \times 1} = \begin{bmatrix} \sum_{i=1}^l \beta_{i1} g(w_i x_j + b_j) \\ \sum_{i=1}^l \beta_{i2} g(w_i x_j + b_j) \\ \vdots \\ \sum_{i=1}^l \beta_{im} g(w_i x_j + b_j) \end{bmatrix}_{m \times 1}$$

$$(j = 1, 2, \dots, Q) \quad (1)$$

式中: $w_i = [w_{i1}, w_{i2}, \dots, w_{in}]$, $x_j = [x_{1j}, x_{2j}, \dots, x_{nj}]^T$ 。

式(1)可表示为:

$$H\beta = T' \quad (2)$$

使用最小二乘法对隐含层与输出层间的权重值矩阵 β 进行求解,得:

$$\min ||H\beta - T'|| \quad (3)$$

$$\hat{\beta} = H^+ T' \quad (4)$$

式中: H^+ 为隐含层输出矩阵 H 的 Moore-Penrose 广义逆矩阵。

通过以上公式,得到图 1 所示极限学习机的网络结构。

按照 ELM 模型的输入层特点,将浮动车 GPS 数据与收费站抽样刷卡数据作为神经网络输入层,神经元为 3 个,并将最初的隐含层神经元设为 10 个。将通过输入训练数据得到的结果与真实数据进行

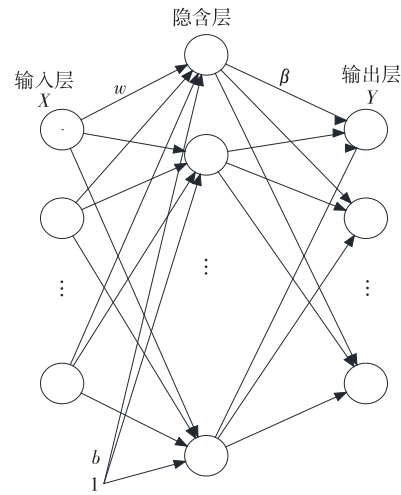


图 1 极限学习机的网络结构

行比较,得到预测结果的误差,若误差超出界限,则加大隐含层神经元数目。利用 Sigmoid 函数激活隐含层神经元,函数公式如下:

$$g(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

确定输入数据源后,对浮动车 GPS 数据、收费站抽样刷卡数据进行融合,以融合后的行程时间作为输出参数,通过实际数据进行验证,得到完整有效的面向高速公路交通状态预测和估计的数据融合方法,使交通状态预测结果更全面。融合流程见图 2。

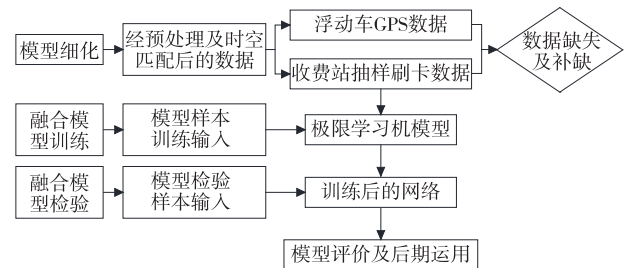


图 2 交通数据融合流程

2 数据获取及预处理

2.1 原始数据获取

以广州市机场高速公路南线(三元里—机场南路段)的地图数据、浮动车 GPS 数据、收费站抽样刷卡数据为研究对象,数据内容见表 1。

2.2 数据预处理

不同来源数据有不同的格式与时间划分,即使是同一来源的数据也会受到各种现实条件(天气环境、检测器好坏等)的影响出现错误、缺失等现象,为提升数据融合精度,对数据进行预处理。

表1 原始数据

数据来源	数据内容
地图数据	提供可以与浮动车 GPS 数据进行地图匹配的地图数据
浮动车 GPS 数据	包含 10 d 的浮动车 GPS 数据,每条 GPS 数据包含车辆行驶方向、瞬时速度、经纬度位置等信息。采样间隔为 15~120 s
收费站抽样刷卡数据	包含 10 d 的收费站数据,每条数据包含车辆进出收费站的编号 ID 和时间。每 5 min 时间段内随机抽取 10 个数据

2.2.1 删除错误数据

使用阈值法对错误数据进行处理。数据处理前进行实地调研,根据调查结果确定合理的区间范围,不在这个区间范围内的数据则删除。

2.2.1.1 高速公路收费站数据

高速公路收费站数据是数据误差的“重灾区”,常存在许多异常数据,需予以删除。

(1) 车辆进出高速公路为同一收费站。在许多情况下,某一车辆进出高速公路为同一收费站,可能是出行者在高速公路上违章掉头或出行者在服务区交换收费卡。对车辆进出高速公路为同一收费站的数据全部删除。

(2) 高速公路行程时间异常。在未发生拥堵的情况下,可根据车辆通过高速公路的长度和运行速度得出其行程时间区间,行程时间未在该区间内的数据会严重干扰数据融合与交通状态预测结果,需全部删除。出现行程时间异常的原因可能是系统故障或车辆在服务区内休息等。

2.2.1.2 高速公路浮动车 GPS 数据

每条 GPS 数据都包含车辆当前时间的经纬度、瞬时速度、行驶方向、车辆类型等信息,因 GPS 本身硬件原因或接收数据出现问题会导致种种问题数据。最容易产生误差的数据是时刻变化的经纬度与瞬时速度,这两类数据也是数据融合与交通状态预测较重要的数据,需对其进行处理。

(1) 经纬度异常数据的删除。车辆轨迹是连续不断的,若轨迹大致准确,只有极少数点不在轨迹上,则可根据正常情况下两相邻经纬度的聚类效应[见式(6)]进行判断,删除异常数据;若轨迹大部分偏离,则将整条 GPS 数据全部删除。

$$0 \leq l \leq f v_m t \quad (6)$$

式中: l 为高速公路上邻近两个数据采集点的距离; f 为修正参数; v_m 为最大可行驶速度的修正值; t 为采样时间。

(2) 瞬时速度异常数据的删除。根据相关法

规,瞬时速度的取值范围见式(7)。瞬时速度异常的数据均删除。

$$0 \leq v_d \leq f_v \cdot v_m \quad (7)$$

式中: v_d 为当前瞬时速度取值; f_v 为瞬时速度最大值的修正参数,通常取 1.3~1.5; v_m 为法规所定的最高行驶速度(km/h),当前中国高速公路的最高行驶速度为 120 km/h。

2.2.2 补充缺失数据

数据采集过程中,检测器故障、天气影响数据传输、车流量过大使车辆缓慢前行,甚至是错误数据的删除都可能导致数据缺失情况,一旦数据缺口扩大,会使数据反映的交通状况与实际道路交通状况存在较大差异,需对缺失数据进行处理。高速公路交通状态随着时间连续变化,可根据该特征,采用相邻时段数据的平均值 $\bar{y}(t)$ 对缺失数据进行修复。计算公式如下:

$$\bar{y}(t) = [y(t-1) + y(t+1)]/2 \quad (8)$$

$$\bar{y}(t) = [y(t-n) + y(t-n+1) + \dots + y(t-1)]/n \quad (9)$$

式中: n 为数据个数。

若拥有 $t+1$ 时段的数据,则按式(8)计算较简便;当数据在线产生,无法获得 $t+1$ 时段的数据时,则按式(9)进行处理,该方法既能用于离线数据处理,也可处理在线数据。

2.2.3 空间匹配

选取浮动车 GPS 数据和收费站抽样刷卡数据作为数据融合基础数据,融合过程中两类数据需按空间属性一一对应,融合的数据必须是同一高速路段的数据,否则融合数据没有任何作用。选择广州三元里(平沙收费站)至机场南(机场收费站)路段的这两类数据。浮动车 GPS 的经纬度数据需地图数据来匹配,提取该路段的收费站抽样刷卡数据与 GPS 相对应即可。

2.2.4 时间匹配

不同种类检测器所反馈的数据间隔不相同,需

统一所采集交通数据的时间周期,并将同一时刻的数据进行匹配。浮动车 GPS 数据的传输间隔可自由设定,常见间隔为 10、20、30、40 s 等,收费站抽样刷卡数据则每 5 min 传输一次,为达到时间上的统一,将两者统一间隔为 10 min。

3 模型结果对比

对模型进行数据训练,选择均方根误差 $RMSE$ 、平均绝对百分误差 $MAPE$ 和行程时间指数 TTI 对处理结果进行评价,验证模型的训练是否满足需求。其中 TTI 是估计城市路网拥堵程度的量化指标, $TTI = \text{交通拥堵情况下行程时间} / \text{正常情况下行程时间}$ 。 $TTI \geq 1$,随着 TTI 的增大,交通拥堵状况不断加剧。

将周一 16:00—16:20 的行程时间作为预测对象,选用之前一周的行程时间作为训练数据,将 756 组模型训练数据分为存在交通事故与不存在交通事故两类分别进行训练,将周一 16:00—16:20 的实际数据作为验证数据。数据处理结果的误差见表 2,行程时间指数 TTI 计算结果见图 3。

表 2 不同模型数据处理结果的误差

模型	对应数据	$RMSE$	$MAPE/\%$
收费站抽样刷卡	收费站数据	1.294 1	0.195 3
浮动车 GPS	GPS 数据	1.011 3	0.125 3
BP 神经网络融合模型	融合数据	0.845 1	0.095 1
极限学习机融合模型	融合数据	0.598 8	0.037 8

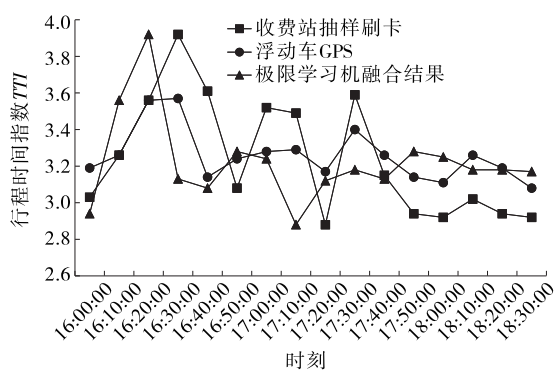


图 3 不同模型的 TTI 计算结果

由表 1 可知:与单一数据源即收费站抽样刷卡数据、浮动车 GPS 数据相比,极限学习机与传统 BP 神经网络两种融合模型的行程时间误差均有所下降。BP 神经网络融合模型的稳定性(均方根误差)分别提高 34.7%、16.4%,极限学习机融合模型分别提高 53.7%、40.8%;BP 神经网络融合模型的准确

性(平均绝对百分比误差)分别提高 51%、24.1%,极限学习机融合模型分别提高 80.6%、69.8%。无论是稳定性还是准确性,极限学习机模型都比 BP 神经网络模型更精确,极限学习机融合后的时间更接近于真实值,融合数据质量更高。

由图 3 可知:按单一源数据与极限学习机融合数据计算的 TTI 变化趋势基本一致,但融合数据计算结果变化与最大值均比其他两种结果更高。说明融合数据比单一数据的 TTI 更全面,通过融合数据对高速公路交通状态进行估计更精确、可靠。

4 结语

该文引入极限学习机方法,提出以极限学习机为基础的数据融合模型,其融合后数据的稳定性与准确性均高于单一数据源和 BP 神经网络模型,采用该模型进行数据融合可克服目前单源数据不够精确的不足,与实际交通状况更贴近,并可运用行程时间指数 TTI 对道路交通状态进行估计。该模型对交通流激增、交通事故突发等情况具有更好的适应性与预测效果。今后的研究中,需进一步考虑天气环境、交通流量等现实因素的影响,提高交通事件检测系统的整体性能。

参考文献:

- [1] 罗向龙,张生瑞,牛力瑶.基于检测器优化选择的短时交通流预测[J].计算机工程与应用,2017,53(8).
- [2] Jun Xia, Yuqiang Feng, Luning Liu, et al. An information fusion model of innovation alliances based on the Bayesian network[J]. Tsinghua Science and Technology, 2018, 23(3).
- [3] Bachmann C, Roorda M J, Abdulhai B, et al. Fusing a bluetooth traffic monitoring system with loop detector data for improved freeway traffic speed estimation[J]. Journal of Intelligent Transportation Systems, 2013, 17(2).
- [4] Tang-Hsien Chang, Albert Y Chen, Yu-Ting Hsu, et al. Freeway travel time prediction based on seamless spatio-temporal data fusion: case study of the freeway in Taiwan[J]. Transportation Research Procedia, 2016, 17.
- [5] 史岩,董宏辉,张瑜,等.多源检测器的交通数据融合研究[J].道路交通与安全,2015,15(3).
- [6] 蒋肖.基于混沌理论和数据融合的短时交通流预测[D].重庆:重庆邮电大学,2016.

(下转第 43 页)

综合比较,新增旅游观光轨道交通在工程建设上虽然投资相对较大,但其建设期对生态环境破坏小,且不会对核心景区正常游赏组织造成干扰;运行期具有安全、舒适、低碳环保、噪声低、无污染等优势,更有利于风景区可持续发展并满足“全球低碳生态景区”环境保护要求,能有效改善核心景区的交通能力,满足不同类型游客出行和游赏的需要。构建多层次立体化旅游交通模式符合黄果树风景区可持续发展需求。

4.3 基于旅游观光轨道交通的交通组织模式

结合旅游景点与客源分布、自然地形地貌等因素,采用“枢纽+集散中心+景点”的交通组织方案,构建多层次立体化旅游交通模式,将旅游观光轨道交通作为风景区内部交通出行的骨干运输方式。

(1) 通过黄果树新城枢纽停车换乘功能,组织对外交通方式中的私家车、旅游巴士及出租车等换乘出行。

(2) 依靠大容量旅游观光轨道交通方式衔接枢纽和集散中心,同时加强集散中心和景区交通接驳,充分发挥枢纽截流功能和集散中心客流组织功能。

(3) 以旅游观光轨道交通站点为集散中心,依托景区内道路和慢行通道,规划内部公交系统(环保观光大巴)和慢行系统(自行车和步行走廊)与轨道交通衔接,实现旅游观光轨道交通游客的安全、快捷集散。

(4) 在旅游旺季,结合景区的游客容量限制,通过旅游观光轨道交通售票方式灵活地引导和管控客流,同时加大环保观光大巴的集散运力,确保游客的安全疏散和游览。

5 结语

山地带状多组团旅游景区由于其地形环境的特

殊性,在游客量日益增长和生态环境保护的双重压力下,如何处理好风景区的可持续发展成了诸多景区的难题。该文以黄果树风景区为例,结合山地带状多组团旅游景区的特征,提出树立“保护风景名胜,多层次组织方式”的交通发展理念,构建基于旅游集散中心模式的以旅游观光轨道交通为骨干、旅游公路和慢行系统为补充的多层次立体化交通模式,为今后山地带状多组团旅游景区的交通改善与优化提供借鉴和参考。

参考文献:

- [1] 陈科,张殿业,姜克锦,等.基于旅游交通出行链的山地旅游交通模式研究[J].交通运输研究,2009(增刊1).
- [2] 张翊.旅游景区交通需求管理规划研究:以凤凰古城景区为例[J].公路与汽运,2014(3).
- [3] 冯佩雨,李文权.旅游景区综合交通规划的探析:以长白山为例[J].交通运输工程与信息学报,2014,12(3).
- [4] 冯德显.山地旅游资源特征及景区开发研究[J].人文地理,2006(6).
- [5] 全球低碳生态景区推荐参考指标[EB/OL].http://www.sohu.com/a/201564525_163959,2017-10-31.
- [6] 赵磊.基于集散中心模式的景区旅游交通体系规划探讨:以巽寮湾为例[A].中国城市交通规划年会暨学术研讨会论文集[C].2014.
- [7] 熊刚,张云龙,孙超,等.深圳市大鹏新区旅游交通发展模式研究[A].中国城市交通规划年会暨学术研讨会论文集[C].2015.
- [8] 铁道第三勘察设计院.安顺市旅游扶贫小火车示范段工程可能性研究报告[R].天津:铁道第三勘察设计院,2016.
- [9] 张铁映.城市不同交通方式能源消耗比较研究[D].北京:北京交通大学,2010.

收稿日期:2018-11-09

(上接第19页)

- [7] 王翔,陈小鸿,杨祥妹.基于K最近邻算法的高速公路短时行程时间预测[J].中国公路学报,2015,28(1).
- [8] Ding A L, Zhao X M, Jiao L. Traffic flow time series prediction based on statistics learning theory[A]. Proceedings of The IEEE 5th International Conference on Intelligent Transportation Systems[C].2002.
- [9] 李楠,路小波.基于小波融合的车牌模糊图像复原算法[J].公路交通科技,2011,28(3).

- [10] Huang Guang Bin. Extreme learning machine: theory and applications[J].Neuro-computing,2006,70.
- [11] 李彬,李贻斌.基于ELM学习算法的混沌时间序列预测[J].天津大学学报,2011,44(8).
- [12] 蔡磊,程国建,潘华贤.极限学习机在岩性识别中的应用[J].计算机工程与设计,2010,31(9).
- [13] 张扬,何承,张伟.上海市道路交通状态指数简介及应用案例[J].交通与运输,2016(3).

收稿日期:2018-11-07