

面向行程时间预测准确度评价的数据融合方法

李慧兵^{1,2}, 杨晓光¹

(1. 同济大学 交通运输工程学院, 上海 201804; 2. 上海济祥智能交通有限公司, 上海 200092)

摘要: 提出 BP 神经网络融合模型. 该模型由三部分组成: 初始数据产生模块、BP 神经网络数据融合模块、融合结果分析模块. 选择四个参数作为该模型的输入变量, 其中路段交通流密度和交通量由线圈数据提供, 而行程时间估计值与浮动车样本量由浮动车数据提供, 并且给出选择这四个参数的依据与原因. 最后选择杭州市的一条主干道作为目标路段, 采集该路段上的 406 组数据对该模型进行验证, 试验结果表明模型对准确度评价的相对误差仅为 4.86%.

关键词: 行程时间估计值; 准确度评价; BP 神经网络; 浮动车数据; 线圈数据

中图分类号: TB114.2

文献标志码: A

Data Fusion Method for Accuracy Evaluation of Travel Time Forecast

LI Huiying^{1,2}, YANG Xiaoguang¹

(1. College of Transportation Engineering, Tongji University, Shanghai 201804, China; 2. Shanghai Jixiang Intelligent Transportation Technology Co. Ltd., Shanghai 200092, China)

Abstract: A BP neural network model was brought forward, which was composed by the initial data generated module, the BP network-based data fusion module and the result analysis module. Four variables such as link average density, traffic volume, link average travel time based on floating car data (FCD) and floating car sampling size were taken as input variables. Link average density and traffic volume could be obtained by the data of loop detectors, while link average travel time and floating car sampling size could be acquired with FCD. Then, the reasons to choose those four variables were given with the support of a statistical analysis. At last, an arterial road in Hangzhou was chosen as an object link, 406 groups of data were utilized to verify the model. The results show that the mean absolute error (MAE) of the proposed model is only 4.86%.

Key words: estimated travel time; accuracy evaluation; BP

neural network; floating car data(FCD); loop detector data

路段平均行程时间作为反映道路拥挤程度的关键交通参数之一, 历来受到交通管理与控制部门的高度重视. 路段平均行程时间的精确度、完整性与时效性, 直接影响动态交通管理措施的应用效果.

不同数据源需要不同的模型算法来估计和预测路段平均行程时间值. 在我国大中型城市路网上, 浮动车数据是一种具有数据精度高、覆盖范围广、成本低等优点的数据类型, 所以对于路段行程时间的估计浮动车数据目前应用最广^[1], 因此基于浮动车数据的路段行程时间准确度评价就成为一个急需解决的问题. 一些国内外学者对路段行程时间准确度与浮动车样本量之间的关系进行研究. 张存保等^[2]、周舒杰等^[3]在路网的层面上对路段行程时间准确度与最小浮动车样本量之间的关系进行研究. 但是他们提出的模型均为数学分析模型, 而且这些模型没有经过实地验证, 更重要的是, 它们都是以路网为基础建立起来的, 而不是基于单条路段. Qi^[4]试图寻找单条路段行程时间准确度与浮动车样本量之间的关系, 但是其给出的关系是定性的, 而不是定量、精确的数学关系. 上述研究的局限性在于, 它们均认为浮动车样本量是影响路段行程时间准确度的唯一因素, 而实际上并不是这样. Hellinga 等^[5-6]指出, 由于浮动车的时空分布具有随机性, 因此浮动车采样在大部分情况下是有偏差的, 所以不管浮动车样本量有多大, 浮动车样本的路段平均行程时间都有可能不会接近全体车辆的路段平均行程时间. 与此同时, 很多国内学者试图对行程时间准确度进行直接评价, 在南京、杭州等城市的一些 ITS 重大项目中^[7-9], 研究人员提出一系列路段行程时间准确度的评价方法, 如: 浮动车法、试验车法、牌照识别法等. 这些方

收稿日期: 2011-12-06

基金项目: 国家“八六三”高技术研究发展计划(SS2012AA112306)

第一作者: 李慧兵(1983—), 男, 博士生, 主要研究方向为交通数据挖掘与融合. E-mail: lihuibing_1983@yahoo.com.cn

通讯作者: 杨晓光(1959—), 男, 教授, 博士生导师, 工学博士, 主要研究方向为智能交通系统. E-mail: yxgits@vip.sina.com

法有两个共同的局限性,分别如下:

(1) 这些方法只能对行程时间准确度进行离线评价,而不能进行实时、在线评价,而实时、在线评价是非常重要的,因为实时准确度较高的行程时间可以被保存到系统数据库中以作后期的数据融合和挖掘之用,对实时准确度较低的行程时间来说,可以建立相应的修正模型对其进行修正与改进.

(2) 由于成本和费用的原因,这些方法所能评价的路段样本量和评价持续时间都有限^[7-9],因此会影响系统评价结果的可信度.

为了克服前人研究的局限性,本文提出一个 BP 神经网络的融合模型,该模型不仅考虑了浮动车样本量这一影响行程时间准确度的因素,还同时考虑了线圈数据提供的交通状态信息.该模型可以对浮动车行程时间准确度进行实时、在线的评价,同时可以对任何路段、任何时段内的行程时间准确度进行长期、有效的评价.

基于 BP 神经网络的融合模型包括三个部分:初始数据产生模块、BP 神经网络数据融合模块和融合结果分析模块.初始数据产生模块包括线圈检测模块和 GPS 浮动车检测模块.模型的输入参数为路段交通流密度、交通量、行程时间估计值和浮动车样本量. BP 网络结构将依据实地采集到的数据样本情况确定.融合结果分析模块主要是分析融合后的行程时间准确度评价值与实测行程时间准确度的接近程度,以此判断融合结果的优劣,并作为模型改善的依据(见图 1).

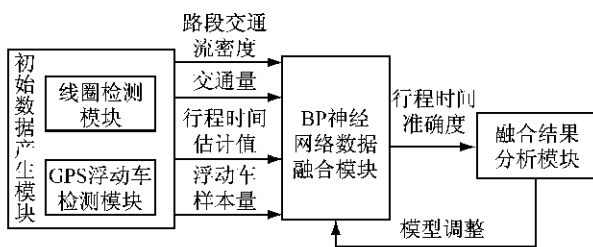


图 1 融合模型结构模块图

Fig.1 Configuration of the fusion model

1 BP 神经网络模型

随着多传感器系统应用领域的不断扩大,数据融合技术得到了较大的发展.为了使多传感器系统能够自适应、并行、高效地融合信息,数据融合算法越来越多地将神经网络应用其中.神经网络作为对

人脑功能的抽象和模拟,是探索人类智能奥秘的强有力工具.神经网络以其泛化能力强、稳定性高、容错性好等优势,在数据融合的应用中日益受到重视.另外,神经网络具有较强的学习能力和适应能力,能够综合各方面的信息,在进行数据融合过程中具有较大的优势^[10].

神经网络融合多传感器信息应具有如下性能^[10]:

(1) 神经网络的信息统一存储在网络的连接权值和连接结构上,使得多传感器信息的表示具有统一的形式,便于管理和建立知识库.

(2) 神经网络可增加信息处理的容错性,当某个传感器出现故障或检测失效时,神经网络的容错功能可以使融合系统正常工作,并输出可靠的信息.

(3) 神经网络的自学习和自组织功能,使融合系统能适应工作环境的不断变化和信息的不确定性.

(4) 神经网络的并行结构和并行处理机制,使得信息处理速度快,能够满足信息融合的实时处理要求.

本文将采用基于 BP 神经网络的融合模型.神经网络的模型很多,其中反向传播网络(back propagation network, BPN)是被采用最多的方法之一. BPN 是一种多层前馈神经网络,其神经元的变换函数是 Sigmoid 型函数,它可以实现从输入到输出的任意非线性映射.由于权值的调整采用反向传播(BP)的学习算法,因此也常称其为 BP 网络^[11].

BPN 采用广义的 Delta 规则进行网络训练和学习,是一种有教师的学习,其学习算法如下^[12](见图 2):

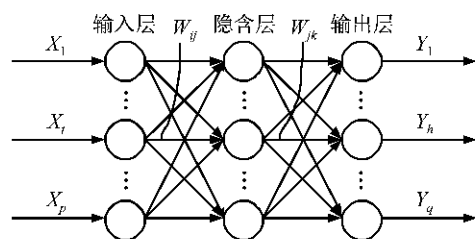


图 2 BP 网络模型

Fig.2 BP network model

步骤 1 权重 W_{ij} , W_{jk} 和神经元阈值 θ_j , θ_k 取随机小值.

步骤 2 输入一个学习样本 X_i 和已知的输出值(期望输出) T_k .

步骤 3 计算神经元的输出.

隐含层

$$Z_j = f\left(\sum_{i=1}^n W_{ij} X_i + \theta_j\right)$$

输出层

$$Y_k = m\left(\sum_{j=1}^n W_{jk} Z_j + \theta_k\right)$$

式中: f, m 是 Sigmoid 型函数.

步骤 4 计算神经元的误差.

输出层

$$\delta_k = (Y_k - T_k) Y_k (1 - Y_k)$$

隐含层

$$\delta_j = Z_j (1 - Z_j) \sum_{k=1}^q \delta_k W_{jk}$$

步骤 5 调整 W_{ij}, W_{jk} 和 θ_j, θ_k .

$$W_{ij} = W_{ij} + \alpha \delta_j X_i, \quad \theta_j = \theta_j + \beta \delta_j$$

$$W_{jk} = W_{jk} + \alpha \delta_k Z_j, \quad \theta_k = \theta_k + \beta \delta_k$$

式中: i, j, k 分别指输入层、隐含层与输出层的节点编号; α 和 β 分别是权重函数和神经元阈值函数的修正系数.

步骤 6 输入下一个学习样本, 返回步骤 3, 直至全部学习样本训练完毕, 令计算误差和

$$E = \frac{1}{2} \sum_{p=1}^P \sum_{k=1}^q (T_{pk} - Y_{pk})^2$$

若 E 小于预先设定的某一精度 ϵ , 则训练结束, 否则重新将学习样本再次学习, 直至 $E < \epsilon$.

当 BPN 训练结束以后, 就可以利用它进行信息处理和优化.

2 模型参数选择

本文选择行程时间准确度作为模型的输出参数, 选择路段交通流密度、交通量、基于浮动车数据的行程时间估计值以及浮动车样本量作为模型的输入参数, 下面对参数选择的依据和原因进行阐述.

2.1 模型输出参数选择

选择行程时间准确度作为模型的输出参数. 准确度是指计算值或测量值与真值相接近的程度^[13], 而相对误差可以表征这种接近程度, 因此本文用行程时间相对误差来表征行程时间准确度. 基于浮动车的实测行程时间准确度被当作输出参数来对 BP 神经网络进行训练, 其计算公式如下:

$$A_n = \left| \frac{t-r}{t} \times 100\% \right| \quad (1)$$

式中: r 为基于浮动车数据的行程时间估计值, t 为实测行程时间. 其中实测行程时间是指在检测间隔内

通过目标路段全体车辆的平均行程时间, 在本文中, 实测行程时间是利用车辆牌照法得到的.

2.2 模型输入参数

从式(1)可看出, 影响行程时间准确度的因素有两个: ① 基于浮动车数据的行程时间估计值; ② 实测行程时间. 模型输入参数中的交通量和路段交通流密度是通过线圈检测器获取的, 它们与行程时间实测值关系密切. 而行程时间估计值和浮动车样本量是通过浮动车数据获得, 其中行程时间估计值除了代表其本身之外, 还与行程时间实测值关系密切; 浮动车样本量既与行程时间估计值直接相关, 更与行程时间实测值有密切的关系. 图 3 是模型输入参数的选择依据图, 输入参数的具体选择原因将在下文阐述.

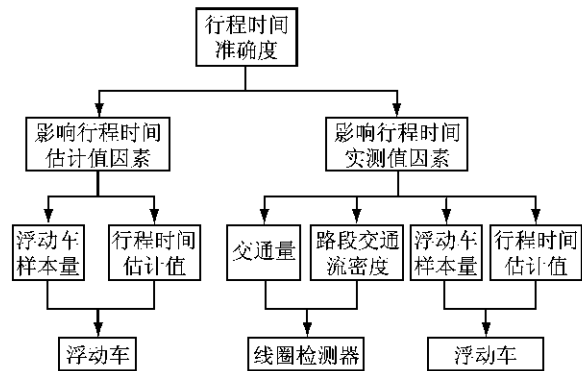


图3 模型输入参数选择依据图

Fig.3 Selection criteria of input parameters

3 初始数据产生模块

本文以杭州市主干道环城东路上南向北方向的一条自然路段(环城北路与体育场路之间的路段)作为研究目标路段, 该路段长度为 530 m, 路段中间没有支路. 如图 4 所示, 目标路段是从上游交叉口出口断面到下游交叉口停车线这一区段. 路段上线圈检测器布设方法为典型的 SCATS(Sydney coordinated adaptive traffic system)线圈布设方法, 这些检测器能检测到每 5 min 内通过检测器断面的交通量. 浮动车数据由实时运行在目标路段上的浮动车所提供, 包括车辆经纬度、车辆编号、车辆瞬时速度等信息, 数据发送间隔为 30 s.

本模块分为两个部分: GPS 浮动车检测模块和线圈检测模块. GPS 浮动车检测模块可以获得路段行程时间估计值和浮动车样本量等参数. 线圈检测模块能直接获取准确的交通量数据, 并且可以通过间接计算的方法获取路段交通流密度.

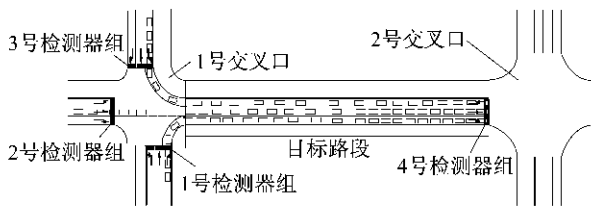


图4 目标路段示意图

Fig.4 Object link diagram

3.1 GPS 浮动车检测模块

浮动车检测技术是基于GPS的动态交通数据采集技术。GPS是一个能够实现全方位、实时定位与导航功能的卫星系统,已经在许多领域得到了成功的应用。在动态交通数据采集方面,基于GPS的动态交通数据采集技术可以采集车辆的瞬时速度、行程时间、行程速度等数据。具体方法是在车辆上配备GPS接收装置,以一定的采样间隔记录车辆的三维位置坐标和时间数据,这些数据传入计算机后与地理信息系统(GIS)的电子地图相结合,经过重叠分析计算出车辆的瞬时车速及其通过特定路段的行程时间和速度指标。浮动车检测技术的精度主要受浮动车数量、车辆类型以及驾驶员的驾车习惯等因素的影响。本模块可以生成行程时间估计值和浮动车样本量等参数。

(1) 行程时间估计值

基于浮动车数据的行程时间估计值可以近似代表实测行程时间值^[14],因此选择行程时间估计值作为模型的一个输入变量。

(2) 浮动车样本量

张存保等^[2]、周舒杰等^[3]以及Qi^[4]均指出,在一般情况下,随着浮动车样本量的增大,利用浮动车直接法计算得到的行程时间估计值会接近于穿越路段全体车辆的路段平均行程时间(实测路段行程时间),因此浮动车样本量与行程时间估计值、实测行程时间之间的关系都很密切,尤其是后者。

3.2 线圈检测模块

虽然基于浮动车数据的行程时间估计值可以近似代表实测行程时间值^[14],然而浮动车数据只是由少量运行在目标路段上的浮动车所发送,同时由于浮动车存在时空分布不均匀的特性,因此基于浮动车数据的行程时间估计值很难在任何情况下都真实反应实测行程时间。由于SCATS线圈可以检测到通过目标路段的全体车流量,因此本文把线圈检测到的车流量以及通过间接计算得到的路段交通流密度作为模型的两个输入参数,用它们来反映和代表路段实测行程时间。

(1) 交通流密度

对于给定路段来说,路段容纳车辆数 N_a 是与该路段的交通流拥挤状况紧密相关的。首先,对研究路段在某一时段 a 的结束时刻所容纳的车辆数 N_a 进行测算。该路段所包含车辆包括排队车辆和正在运动中车辆两部分。假设中间路段无进出开口,则有平衡方程^[15]

$$N_a = N_{a-1} + Q_{a,进} - Q_{a,出} \quad (2)$$

式中: N_a 为第 a 时段结束时刻路段容纳车辆数, N_{a-1} 为第 $(a-1)$ 时段结束时刻路段容纳车辆数, $Q_{a,进}$ 为第 a 时段进入该路段的总车辆数, $Q_{a,出}$ 为第 a 时段驶出该路段的总车辆数。

考虑到路段长度和车道数,第 a 时段路段交通流密度

$$K_a = \frac{N_a}{Ls} = \frac{N_{a-1} + Q_{a,进} - Q_{a,出}}{Ls} \quad (3)$$

式中: L 为路段总长度, s 为路段车道数。

从式(3)可以看出,必须得到第 $(a-1)$ 时段结束时刻路段容纳车辆数,才能对第 a 时段路段交通流密度 K_a 进行计算。然而, N_{a-1} 是一个未知变量,因此,有必要对式(3)进行如下变换:

$$K_a = \frac{N_a}{Ls} = \frac{N_{a-1} + Q_{a,进} - Q_{a,出}}{Ls} = \frac{N_l^{min} + \sum_{n=l+1}^a (Q_{n,进} - Q_{n,出})}{Ls} \quad (4)$$

式中: N_l^{min} 为第 l 时段结束时刻路段容纳车辆数,同时也是路段上容纳的最小车辆数,一般来说,第 l 时段属于凌晨时分(3:00~4:00 am),因为那时候路段上车辆很少或者几乎没有车辆,这里把 N_l^{min} 设为零。

从图4可以看出,1~3号检测器组检测到的总流量即为进入目标路段的总车辆数 $Q_{进}$,而4号检测器组检测到的流量即为驶出目标路段的总车辆数 $Q_{出}$ 。从式(2)~(4)可以看出,计算第 a 时段路段交通流密度 K_a 的关键在于找到路段上容纳最小车辆数的时段。因为4号检测器组(见图4)检测到的车流量可以反映在该时段通过目标路段的车辆数,因此4号检测器组检测到车流量最小的时段就是该路段上容纳最小车辆数的时段。

利用2010年4月到8月间早高峰8:00~9:00 am内的108组数据(路段交通流密度数据和实测行程时间数据)对路段交通流密度与实测行程时间之间的关系进行拟合分析。如图5所示,路段交通流密度与实测行程时间之间服从比较明显的线性

关系,因此可以用路段交通流密度来反映和代表实测行程时间。

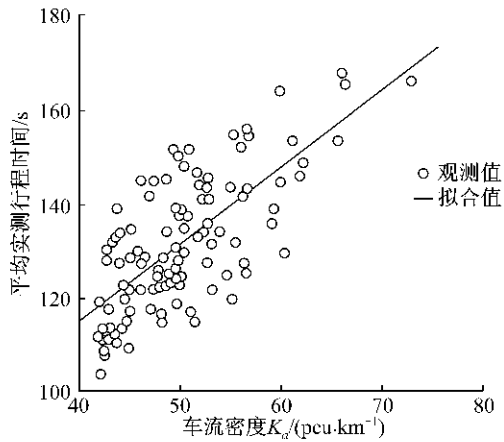


图5 路段交通流密度与实测行程时间的关系图
Fig.5 Relationship between traffic flow density and actual travel time

(2) 交通量

选择4号检测器组检测到的交通量作为模型的一个输入变量,因为该交通量代表检测间隔内通过目标路段全体车辆的交通量,因此可以较好地反映路段交通状态以及路段的实测行程时间。

4 BP神经网络数据融合模块

本模块是通过构造一个BP神经网络,对初始数据进行融合处理以得到与实际数据更为接近的数据.本文用于数据融合的BP神经网络的设计方案如下:

(1) 神经网络的输入数据和输出数据.模型的输入输出参数见图6.

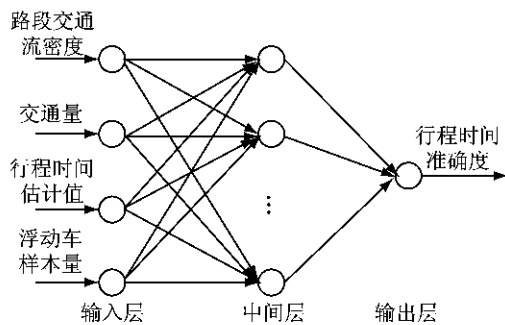


图6 BP神经网络数据融合模块结构图
Fig.6 Architecture of BP network model based on data fusion

(2) 网络层数.神经网络由输入层、输出层和若干隐含层构成.隐含层数量可以为一层,也可以为多

层.隐含层越多,网络结构越复杂,其训练时间也越长.通常情况下,包括一个隐含层的三层结构神经网络可以满足应用要求,因此,本文选择的网络层数为三层.

(3) 神经元节点数量.根据神经网络的输入数据和输出数据,确定神经网络的输入层包含四个节点(分别输入路段交通流密度、交通量、行程时间估计值以及浮动车样本量),输出层仅包含一个节点(输出融合后的行程时间准确度).隐含层节点数为15个.

在确定了BP网络的结构后,利用输入输出样本集对其进行训练,也即对网络的权值和阈值进行学习和调整,以使网络实现给定的输入输出映射关系。

5 融合结果分析模块

该模块负责对融合得到的行程时间进行准确性检验,检验参数的绝对误差 ϵ_h 、平均绝对误差 $\bar{\epsilon}_h$ 和误差离散程度 δ 由式(5)~(7)确定.

$$\epsilon_h = |A_h - e_h| \quad (5)$$

$$\bar{\epsilon}_h = \frac{1}{n} \sum_{k=1}^n \epsilon_k \quad (6)$$

$$\delta = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (\epsilon_k - \bar{\epsilon}_h)^2} \quad (7)$$

式中: A_h 为时段 h 内行程时间准确度实测值(该值可由式(1)计算得到), e_h 为时段 h 内行程时间准确度的评价价值。

6 试验与分析

通过对文中提出的基于浮动车与感应线圈的神经网络融合模型进行试验,从而验证算法的有效性,其中BP融合模型是用Matlab编写.本文以杭州市主干道环城东路上南向北方向的一条自然路段(环城北路与体育场路之间的路段)作为研究目标路段,该路段长度为530m,路段中间没有支路.视频检测器被布设在上下游交叉口处,它们能检测到每辆车分别通过目标路段前后断面的具体时刻.因此,利用视频检测器能获得路段的实测行程时间值。

利用2010年4月到8月间早高峰(8:00~9:00am)内的406组数据对文中提出的神经网络模型进行验证.其中356组数据用来训练,50组数据用来检验.结果证明,利用这50组数据得到的行程时间准确度评价价值的平均绝对误差 $\bar{\epsilon}_h$ 为4.86%,误

差离散程度 δ 为 2.16%。

7 结论

(1) 提出一种面向实时行程时间准确度评价的 BP 神经网络数据融合模型,同时利用 406 组数据对该模型的有效性进行验证。结果表明,行程时间准确度评价值的平均绝对误差 $\bar{\epsilon}_a$ 为 4.86%,误差离散程度 δ 为 2.16%。

(2) 仅选取杭州高峰时段(8:00~9:00 am)的数据对模型进行验证,而未考虑其他时段,因此接下来需要对该模型在不同时段下的适用性进行研究。

参考文献:

- [1] 姜桂艳,常安德,吴超腾.基于 GPS 浮动车的交通信息采集方法[J].吉林大学学报:工学版,2010,40(4):971.
JIANG Guiyan, CHANG Ande, WU Chaoteng. Traffic information collection method based on GPS equipped floating car [J]. Journal of Jilin University: Engineering and Technology Edition, 2010,40(4):971.
- [2] 张存保,杨晓光,严新平.移动交通检测系统中探测车的样本数量[J].中国公路学报,2007,20(1):96.
ZHANG Cunbao, YANG Xiaoguang, YAN Xinping. Probe vehicles sample size for mobile traffic detection system[J]. China Journal of Highway and Transport, 2007,20(1):96.
- [3] 周舒杰,廖孝勇,赵敏,等.面向道路网的浮动车最小覆盖率模型[J].重庆工学院学报:自然科学,2009,23(4):11.
ZHOU Shujie, LIAO Xiaoyong, ZHAO Min, et al. Road network-oriented minimum coverage model for floating vehicle [J]. Journal of Chongqing Institute of Technology: Natural Science, 2009,23(4):11.
- [4] QI Yi. Travel time data quality assessment; development of a design to test new generation traffic monitoring systems[C]// TRB Annual Meeting. Washington D C: Transportation Research Board, 2005:234-235.
- [5] Hellinga B R, FU Liping. Reducing bias in probe-based arterial link travel time estimates[J]. Transportation Research Part C, 2002, 12(5): 257.
- [6] Hellinga B R, FU Liping. Assessing expected accuracy of probe vehicle travel time reports[J]. Journal of Transportation Engineering, 1999, 24(3): 524.
- [7] YANG Xiaoguang, YUN Meiping, LI Huibing, et al. Traffic state intelligent forecasting and management support system [R]. Shanghai: Tongji University, 2008.
- [8] YANG Xiaoguang, YANG Fan, YUN Meiping, et al. Research and application of intelligent transportation system based on mobile data[R]. Shanghai: Tongji University, 2011.
- [9] CHU Hao, YANG Xiaoguang, LI Huibing. Accuracy evaluation of traffic condition distinguishing [C]//The 4th China Annual Conference and International Exhibition on Intelligent Transport Systems. Beijing: China Academic Journal Electronic Publishing House, 2008:909-915.
- [10] 李平勇,游磊,苏长明.嵌入式系统教学体系建设研究[J].教育与教学研究,2011,25(5):95.
LI Pingyong, YOU Lei, SU Changming. On study of construction of embedded system teaching architecture[J]. Education and Teaching Research, 2011, 25(5): 95.
- [11] 熊文华,徐建闽,林思.基于 BP 网络的浮动车与线圈检测数据融合模型[J].计算机仿真,2006,26(9):235.
XIONG Wenhua, XU Jianmin, LIN Si. A fusion model for floating vehicle and loop detector data based on BP network [J]. Computer Simulation, 2006,26(9):235.
- [12] 王伟.人工神经网络原理:入门应用[M].北京:北京航空航天大学出版社,1995.
WANG Wei. Artificial neural network theories: fundamental applications[M]. Beijing: Beihaung University Press, 1995.
- [13] 孙立芳,孙永华.临床检验准确度的探讨与评价[J].中国临床医学,2011,23(2):237.
SUN Lifang, SUN Yonghua. Discussion on the accuracy degree of clinical inspection [J]. Medical Journal of Chinese People's Health, 2011,23(2):237.
- [14] Turner M S, Holdener D J. Probe vehicle sample sizes for real-time information; the Houston experience [C]// Proceedings of Vehicle Navigation and Information Systems (VNIS) Conference. New York: IEEE, 1995: 3-9.
- [15] Cheu R L, Liu Q, Lee D H. Arterial travel time estimation using SCATS detectors [J]. Applications of Advanced Technologies in Transportation, 2002,4(12):32.