

文章编号:1005-0523(2015)04-0045-07

基于BP神经网络分析事故对路段通行能力的影响

凌璐,李枫

(同济大学道路与交通工程教育部重点实验室,上海 201804)

摘要:以发生交通事故的人、机、环境等多方面采集的调查数据为基础,解析突发性交通状态局部演变过程,并总结事发点通行能力的7类主要影响因素和3类评价指标。采用SPSS主成分分析与BP神经网络相结合的方法,对调查数据进行降维,并在保证数据丢失最小原则下,将七类影响因素提取为五类,简化神经网络拓扑结构,提高建模质量,通过BP神经网络模型调整离散输入量,以预测的评估值与实际评估值的误差最小为学习训练目标,求解最佳的连接强度权值与偏置值,得出事故因素与通行能力的定量关系。

关键词:交通事故;主成分分析;BP神经网络;通行能力

中图分类号:U491.31

文献标志码:A

城市道路具有交通流密度大、连续性强等特点,交通事故的出现可能降低路段所有车道的通行能力,并涉及到整个路网的运行状态,此时无论用“瓶颈”断面处的最大通行能力,还是用未受到影响处的最大通行能力来评价路段的服务水平,都不能真实地反映整个路段车流的运行状况。

目前国内外已经有通过交通流理论建模以及基于实际数据来评价事故对道路通行能力的影响,Smith等发现城市高速公路容量减少主要是交通事故造成的,并且证明了将不确定性事故率作为随机变量是最好的建模方法^[1]。Frantzeskakis等也在交通事故率、碰撞暴露、事故类型对公路通行能力的影响做了相关研究^[2]。张海军等通过微观仿真将事发点通行能力的影响因素归纳为事件性质、阻塞行车道宽度、事件发生位置、现场行车秩序与上游交通量等五类来分析交通流系统的平衡状态^[3];杨莹等从实际事故调查视频中获取数据,分析道路使用的基本条件,分析了路段车辆排队长度与事故横断面实际通行能力、事故持续时间、路段上游车流量间的关系,建立车队消散模式的理论模型^[4];Shi等则提出了在不同流量组合条件下,交通事故持续时间对道路通行能力的影响^[5];易富军等以遗传算法优化的RBF神经网络模型,提出了公路隧道群交通事故微观预测方法,对以任意断面为中心的单元长度内的交通事故率进行合理预测^[6]。

国内外基于事故对通行能力的影响有了一定的研究,但具体对影响通行能力的因素选定,缺乏数据分析的理论支持,同时过于依赖交通流基本理论而忽略了交通本身存在的不稳定性与随机性。研究突发性交通事故对城市道路通行能力的影响存在很多不确定因素,本文从突发性交通状态局部演变过程出发,对事发点通行能力的影响因素进行总结,通过VISSIM微观交通仿真软件得出各影响因素与道路通行能力的评价值,采用SPSS主成分分析方法与BP神经网络相结合的方法,求得变量因素与输出值的连接强度权值 W 与偏置值 B 。

1 基于VISSIM交通微观仿真的数据准备

1.1 假设

按影响程度事故类型可分为占用1个车道的交通事故;或者是占用2个车道的交通事故。事故类型不

收稿日期:2015-01-29

作者简介:凌璐(1992—),女,硕士研究生,研究方向为交通规划与管理。

通讯作者:李枫(1966—),男,副教授,研究方向为轨道交通安全,铁路物流规划,城市道路交通。

仅对路段通行能力的影响明显,而且在一定程度上增加了交通冲突的可能性。

除了车道与交叉口信号设置以及转向车的作用因素外,不考虑路容路况以及其他交通设施条件的影响。在VISSIM仿真环境下,当车流速度小于 $5 \text{ km}\cdot\text{h}^{-1}$ 时即为排队开始。

假设事故点的持续时间均能立即处理,不会造成路网瘫痪。

1.2 数据准备

由实际事故资料可知,影响通行能力评估指标的因素包括事故类型,路段单向车道数,车道宽度,道路实际流量,道路通行能力,事故持续时间,交通流速度变化情况等,以最大排队长度,平均每车延误时间以及停止车辆数为通行能力的影响评估指标。为了更好的得出交通流离散输入的特点,在已有的数据基础上,运用交通仿真模拟软件(VISSIM)对驾驶行为进行模拟,得出大量的事故交通行为仿真数据:

为了使模型更具真实性,在原有的数据基础上,运用VISSIM模拟实际道路路况与驾驶行为心理,得到多组影响因素与评估指标的值。根据调查结果,仿真时将路段假上的车流分为:小汽车:公交车=4:1,设车流的VISSIM仿真的前提数据如表1所示。

在以上前提下,将车道基础数据按照1,2,3,4条递增,且每种情况车道宽度均分为3,3.5 m,事故占用车道均为1,2车道。将最大排队长度、平均每车延误、停止车辆数作为数据输出,可得到13组仿真数据如表2所示。

表1 VISSIM仿真模拟参数

Tab.1 VISSIM simulation parameters

车辆类型	运行参数	
	项目	参数值(范围)/($\text{m}\cdot\text{s}^{-2}$)
公交车	最大加速度	1.1(1.3,2.0)
	期望加速度	1(1.5,2.0)
	最大减速度	5(-8.1,-6.1)
	期望减速度	2(-1,-0.7)
中客车	最大加速度	1.5(2,7.5)
	期望加速度	2.5(2,2.5)
	最大减速度	5.5(-6.0,-5.0)
	期望减速度	1.4(-1.5,-1.3)
小汽车	最大加速度	2.2(3.5,250)
	期望加速度	2(3.5,250)
	最大减速度	4(-8.1,-6.1)
	期望减速度	1(-3,-2.5)
自行车	最大加速度	2.5(1.8,3.5)
	期望加速度	2.5(2.0,3.5)
	最大减速度	7.5(-8.5,-6.5)
	期望减速度	2.95(-3.0,-2.9)

表2 仿真数据

Tab.2 The simulation data

车道数	自变量				因变量		
	实际流量/ ($\text{pcu}\cdot\text{h}^{-1}$)	道路通行能力/ ($\text{pcu}\cdot\text{h}^{-1}$)	速度变化差/ ($\text{km}\cdot\text{h}^{-1}$)	持续时间/s	最大排队长 度/m	平均每车 延误时间/s	停止车辆数/ pcu
1	800~1 200	1 630, 1 760, 1 980	8.55~12.13	1 200, 1 080, 900, 600, 400	133~193	93.1~188.4	112~151
2	1 660~1 760	3 260, 3 520, 3 960	7.5~9.7	1 200, 1 080, 900, 600, 400	80~157	64.4~87.1	57~113
3	2 300~2 700	4 488, 4 890, 5 740	6.0~8.2	1 200, 1 080, 900, 600, 400	77~137	62.2~76.1	54~96
4	3 000~3 500	5 216, 5 632, 6 336	2.8~5.9	1 200, 1 080, 900, 600, 400	64~89	55.8~67.1	32~49

2 自变量信度分析与因子分析

为了减少因素选取时,主观性评判的影响,将VISSIM仿真得到的7类影响因素:车道宽度、事故占用车道数、实际流量、道路通行能力、速度变化差、持续时间、最大排队长度、平均每车延误时间以及停止车辆数进行信度分析,并将这7类影响因素通过主成分分析法进行降维处理,科学化、合理化提取影响因素,最后简化为5类作为输入量,提高BP神经网络建模的精度。

2.1 信度分析

利用SPSS19.0统计软件对交通事故数据采集信度与合理性的检验,分析结论的有效性,并作为改善修正的根据^[7]。根据调查交通流在发生事故的过程中到达和消散均为连续型,用Cronbach's α 系数法对仿真得到的影响参数进行信度检测。通常当信度检:0.60~0.65为不可信;0.65~0.70为最小可接受值;0.70~0.80为相当好;0.80~0.90为非常好^[7]。

Cronbach's α 信度系数为

$$\alpha = \frac{K}{K-1} \left(1 - \frac{\sum_{i=1}^K \sigma_{y_i}^2}{\sigma_x^2} \right) \quad (1)$$

其中: K 为量表中评估项目的总数; $\sigma_{y_i}^2$ 为第*i*个项目的表内方差; σ_x^2 为全部项目总和的方差。

根据公式(1)得到Cronbach's的 α 系数如表3所示,信度系数为0.754,标准化后为0.789,于0.70~0.80之间,因此总体上各参数的内在信度是比较理想的。

2.2 因子分析

因子分析的目的是寻求变量的基本结构,简化观测系统,减少变量维度,首先需要通过KMO和Bartlett来检验数据的有效性。

表3 可靠性统计量

Tab.3 Reliability statistics

Cronbach's Alpha	Alpha	项数
0.754	0.789	7

2.2.1 因子分析的前提条件

因子分析要求变量之间应存在较强的相关关系,当巴特利特球度检验的统计量值比较大,且其相对应的相伴概率值小于用户指定的显著性水平时,认为适合作因子分析;反之,不适合作因子分析^[8]。由表4的KMO和Bartlett的检验可以看出,本次调查数据的KMO值为0.709,大于0.7,巴特利特球度检验值为778.372,显著性为 $P < 0.001$,小于0.05,则表示非常显著,说明变量有较高的相关性,因此本次调查数据适合做因子分析。

表4 KMO和Bartlett的检验

Tab.4 KMO and Bartlett's test

取样足够度的 Kaiser-Meyer-Olkin 度量	Bartlett 的球形度检验		
	近似卡方	自由度	显著性双侧检验
0.709	778.372	21	0.001

2.2.2 基于主成分分析的因子抽取

主成分分析的基本思想是将原来众多的具有一定相关性的指标,重新组合成一组新的互不相干的综合指标来代替原来指标,用较少的主成分得到较多的信息量^[9]。

主成分分析通常是将原始指标做线性组合,即将*P*个原始指标向量 X_1, X_2, \dots, X_p 做线性组合,得到综合指标向量

$$\begin{cases} F_1 = a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p \\ F_2 = a_{12}X_1 + a_{22}X_2 + \dots + a_{p2}X_p \\ \dots \\ F_p = a_{1p}X_1 + a_{2p}X_2 + \dots + a_{pp}X_p \end{cases} \quad (2)$$

其中: F_i 与 F_j ($i \neq j, i, j = 1, \dots, p$)互不相关,即 $Cov(F_i, F_j) = 0$ 。其中, F_1 称为第1主成分,在 F_1 所含信息不足条件下寻找第2主成分 F_2 ,第3主成分 F_3, \dots ,第*p*个主成分; a 为各个主因子的相关系数。

在提取综合变量时,一般要满足大于70%的贡献率,经过统计分析,选取基本特征值为0.8,旋转后,各主成分的特征值及累计贡献率如表5所示,提取方法为主成分法。其中,第1主成分主要与车道数、道路实际流量以及道路通行能力关系紧密;第2主成分主要与事故类型(事故所占用车道数)以及持续时间关系紧

密;2个主成分的累计方差贡献率达到77.676%见表6,则表示因子分析的结果能较好的代替原始变量。

表5 成分矩阵

Tab.5 Ingredient matrix

因素	贡献比例	
	1	2
车道数	0.969	-0.022
车道宽度	-0.152	-0.824
事故类型	0.595	0.824
实际流量	0.956	-0.110
道路通行能力	0.954	-0.020
平均速度	-0.950	0.074
持续时间	-0.053	0.803

表6 解释的总方差

Tab.6 Total variance of the explanation

成分	初始特征值			提取平方和载入			旋转平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	4.04	57.798	57.798	4.046	5.798	57.798	4.042	57.748	57.748
2	1.391	19.878	77.676	1.391	19.878	77.676	1.395	19.928	77.676
3	0.89	11.705	89.381						
4	0.556	7.943	97.325						
5	0.107	1.522	98.847						
6	0.075	1.073	99.92						
7	0.006	0.08	100						

3 基于BP神经网络建模

BP(back propagation)神经网络模型由输入层、隐含层和输出层组成,是一种按误差逆传播算法训练的多层前馈网络,而且能学习和存贮大量的输入-输出模式映射关系。BP神经网络的工作过程分成两阶段:一个阶段是学习期,各计算单元状态不变,根据计算结果和模型识别修改各连接权值;另一阶段是工作期,各连接权值固定,计算各单元的状态变化,以求达到稳定状态。

3.1 调用函数

将5种主成分影响参数以及3个输出评估指标数据分成两批,85%用于训练,另外15%的仿真数据用于预测验证其绝对误差值。采用各输入量与权重值的乘积和作为激活函数,在强化学习训练自适应的过程中先后选用不同的激活函数模型,反复调整学习率以及权值,S形函数与双极S形函数都是可导的(导函数是连续函数),因此适合用在BP神经网络中。根据路段交通流的特性,以及调用的各函数训练结果的精确性,并选择训练效果最好的激活函数。最终确定隐含层、输出层激活函数均选用正切S型 $\tan sig$ 函数^[10]

$$\tan sig(x) = g(x) = \frac{1}{1 + \exp(-2\beta x)} \quad (3)$$

3.2 模型建立

将最大排队长度,平均每车延误时间以及停止车辆数当作输出神经元;将事故类型,路段单向车道数,

车道宽度,道路实际流量,道路通行能力,事故持续时间,交通流速度变化作为Input传入的 N 个信号,经权值 W 处理后,用线性求和器得到“综合印象”,再由激活函数 g 对此综合印象作出非线性反应 $\zeta^{[10]}$ 。其中学习速率 $\eta=0.05$;样本模式为 ζ^j ,经程序调试确定系统分为三层,其中中间隐层有 15 个节点,如图 1 所示。

借助 MATLAB 软件编程,通过模型的不断学习调整,将预测的评估指标与目标值相比较,直到模拟数据的方差小于目标值 10^{-3} ,满足学习精度要求,即此模型下所选的参数值符合要求。运用 BP 神经网络模型输出训练的结果方差如图 2 所示,经过训练 4 450 次后达到最优,误差值为 0.000 999 88。

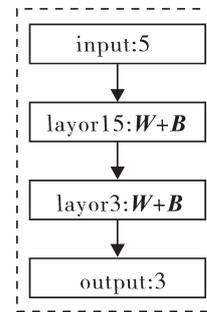


图1 BP神经网络结构图
Fig.1 The BP neural network structure

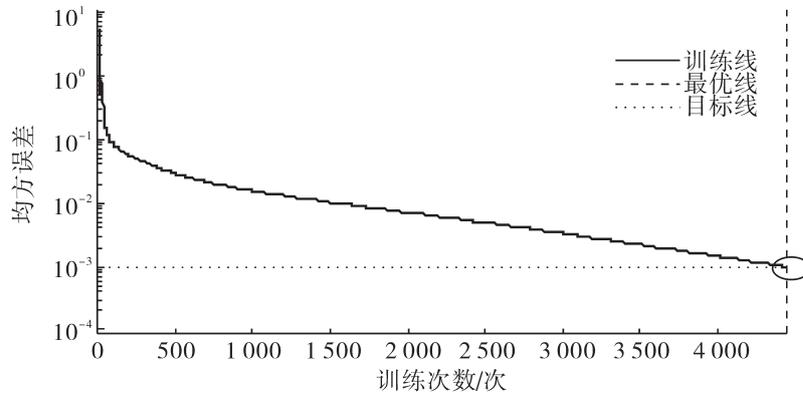


图2 BP回归神经网络模型输出训练结果方差

Fig.2 Output variance of BP regression neural network model

对输出的结果用 15% 的检测数据进行对比分析,可以看出实际值与预测值的误差很小,即所选隐层参数符合精度要求,其中得出 4 组模拟值预测的绝对误差为

$$P = \begin{bmatrix} 0.0428 & 0.0284 & 0.0245 & 0.0432 \\ 0.0339 & 0.0323 & 0.0212 & 0.0312 \\ 0.0312 & 0.0223 & 0.0196 & 0.0275 \end{bmatrix} \quad (4)$$

预测值与目标值的相关系数 $R=0.999$,适应度如图 3 所示。

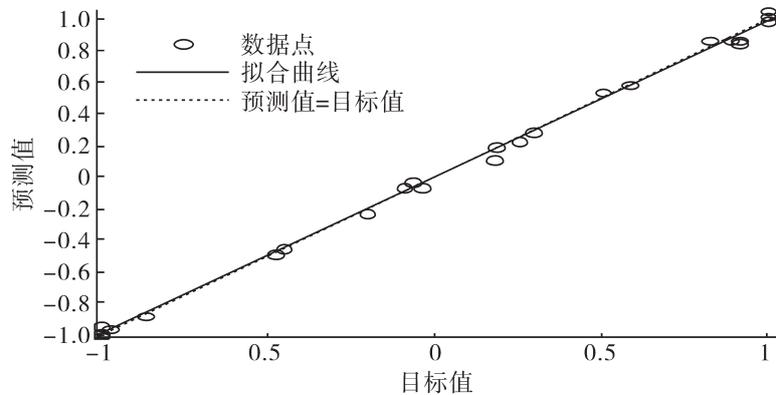


图3 预测值与目标值的适应度

Fig.3 Fitness of predicted values and the target value

通过训练、验证预测值,在误差最小的情况下,可以得到偏置值 B 以及权值矩阵 W 。即事故类型,路段

单向车道数,道路实际流量,道路通行能力和事故持续时间这5个相关参数与3个通行能力评估值所对应的连接权重与偏置值。

$$B_1 = \begin{bmatrix} -2.3259 \\ 2.3935 \\ -1.6583 \\ -1.3215 \\ 1.3523 \\ 0.6536 \\ -0.4697 \\ -0.3259 \\ 0.2479 \\ 0.6690 \\ 1.4022 \\ -1.3074 \\ 1.9225 \\ 2.0558 \\ 2.4894 \end{bmatrix}, B_2 = \begin{bmatrix} -0.2209 \\ 0.4537 \\ -0.3180 \end{bmatrix} \quad (5)$$

$$W_1 = \begin{bmatrix} 0.0849 & -1.3468 & -0.4765 & -1.5593 & -1.2565 \\ 0.0050 & 0.5471 & 1.4180 & 0.7419 & 1.4350 \\ 0.2402 & 1.0334 & 0.7033 & -1.5305 & -1.3058 \\ 1.6417 & 0.9091 & -0.9567 & -0.5113 & 1.1067 \\ -1.7205 & -0.2029 & -0.6135 & -0.9411 & -0.8098 \\ -1.5277 & 1.0906 & 0.9228 & 0.4086 & 1.2123 \\ 1.2711 & -1.5196 & -0.2380 & 0.5089 & 1.1995 \\ -0.8334 & -1.0770 & 1.0591 & 1.4947 & 1.0827 \\ 0.0189 & 0.8425 & -1.7317 & -0.3081 & 1.6740 \\ 1.0629 & -0.4225 & 0.8293 & 1.3428 & 1.3595 \\ 1.2857 & -0.5315 & 0.3806 & 0.7366 & 1.4833 \\ -0.7059 & 1.6137 & 0.5954 & -0.5285 & 1.4512 \\ 0.3752 & 0.8808 & -0.8269 & 1.1323 & -1.5216 \\ 1.7792 & -0.0647 & -0.2559 & 0.9808 & -1.2525 \\ 1.1491 & -0.9307 & -1.5933 & -0.1439 & -0.8553 \end{bmatrix} \quad (6)$$

在隐含层有15个节点的情况下,权值 W 与偏置值 B 矩阵能使得拟合的相关系数均 R 值达到0.999,与斜率为1的直线非常接近,说明基于BP神经网络的模型建立使得预测值与实测值吻合良好,能够准确地预测出五类因素对路段通行能力的影响。并且各实际值与预测值的绝对误差不超过5%,可见所建立的神经网络模型具有很高的预测精度,能够充分描述输入、输出变量之间的复杂映射关系。

4 结论

在实地调查的基础上,设定VISSIM交通微观仿真的基础环境,得到不同参数值影响通行能力的评估指标;运用SPSS主成分分析法进行降维处理,将7类影响参数提取为5类作为输入量;采用BP算法对神经网络进行训练,最后得到影响因素与输出评估值的连接权值矩阵 W 与偏置值矩阵 B 。在实际应用中只要调查到路段5类基础数据,通过激活函数、权值矩阵以及偏置值矩阵就能得出对应的通行能力评估值,为量化分析交通事故对道路通行能力的影响提供了手段。采用BP神经网络模型建立动态交通流模型,调整离散输入量,输出非离散量权值,相对于传统数学理论建模的方法能更好的描述突发性交通事故的运行状态,并可以通过学习和训练优化影响通行能力的权值。由于VISSIM的仿真模拟系统本身存在的缺陷,以及SPSS分析法初始数据获取的主观性,会对实际输入的参数带来一定影响,后期工作将围绕基础数据的有效性获取展开进一步研究。

参考文献:

- [1] SMITH B L, QIN L, VENKATANARAYANA R. Characterization of freeway capacity reduction resulting from traffic accidents[J]. Journal of Transportation Engineering, 2003,129(4):362-368.
- [2] FRANTZESKAKIS J M, IORDANIS D I. Volume-to-capacity ratio and traffic accidents on interurban four-lane highways in Greece[J]. Transportation Research Record, 1987,1112:29-38.
- [3] 张海军,张珏,杨晓光. 异常事件下高速道路交通状态的分析与仿真[J]. 交通运输工程学报, 2008,8(2):116-121.
- [4] 杨莹,李婷,茹贞,等. 交通事故对城市路面通行能力的影响[J]. 运城学院学报, 2013,31(5):6-11,20.
- [5] SHI J Q, CHENG L. Simulation and analysis of highway traffic accident based on VISSIM[J]. Applied Mechanics and Materials, 2013,253:1682-1685.
- [6] 易富君,韩直,邓卫,等. 遗传算法优化的RBF神经网络模型在公路隧道群交通事故微观预测中的应用[J]. 交通运输工程与信息学报, 2012,10(1):64-72.
- [7] 时立文. SPSS19.0统计分析[M]. 北京:清华大学出版社, 2012:231-312.
- [8] 陈建宏,刘浪,周智勇,等. 基于主成分分析与神经网络的采矿方法优选[J]. 中南大学学报:自然科学版, 2010,41(5):1967-1972.
- [9] 张慧. VC++环境下的BP神经网络建模和模拟退火优化研究[J]. 电脑与信息技术, 2008,16(2):31-33,53.
- [10] 飞思科技产品研发中心. 基于MATLAB6.5辅助神经网络分析与设计[M]. 北京:电子工业出版社, 2003:97-156.
- [11] 吴中,邢桂光,方昭,等. 基于有序加权平均算子的信号交叉口通行能力优化算法[J]. 华东交通大学学报, 2014(6):34-39.

Analysis of Accident Impact on Road Traffic Capacity Based on the BP Neural Network

Ling Lu, Li Feng

(Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China)

Abstract: Based on survey data of man, vehicles and environment involved in accidents, this study analyzes local evolution of traffic breakdown and summarizes influential factors of accident point capacity which include seven types of influential factors and three kinds of evaluating indicators. Dimensionality reduction of survey data is fulfilled by SPSS principal component analysis in combination with BP neural network. According to the minimum data loss rule, seven types of influential factors are transformed to five principal components and neural network topology is simplified, thus improving the model quality. Adjustment of discrete inputs which use BP neural network model learning and training generates the optimal connection weights- W and bias values- B when the error between predicted value and actual value reaches the minimum. The quantitative relationship between the factors of the accident and the capacity is thereby worked out.

Key words: traffic accidents; principal component analysis; BP neural network; capacity

(责任编辑 姜红贵)