

文章编号: 1005-0523(2020)05-0060-06

基于改进 BPNN 的高速公路交通事故持续时间预测

许宏科¹, 赵威¹, 杨孟², 林杉¹, 刘冬伟¹

(1. 长安大学电子与控制工程学院, 陕西 西安 710064; 2. 招商局重庆交通科研设计院有限公司, 重庆 40067)

摘要:针对高速公路交通事故持续时间难以预测的问题,构建了因子分析和BP神经网络(BPNN)相结合的交通事故持续时间预测模型。以包茂高速西延段近两年交通事故数据为基础,采用因子分析法提取影响事故持续时间的公共因子,将提取的公共因子作为输入,利用BP神经网络实现交通事故持续时间的预测。结果表明:与经典回归算法和支持向量机相比,采用因子分析和BP神经网络相结合的方法将预测准确率提高了7.8%,同时解决了传统BP神经网络迭代速度慢与数据处理效率低的问题。

关键词:公路交通;交通事故;持续时间;神经网络;因子分析

中图分类号:U491.3

文献标志码:A

本文引用格式:许宏科,赵威,杨孟,等.基于改进BPNN的高速公路交通事故持续时间预测[J].华东交通大学学报,2020,37(5):60-65.

Citation format:XU H K,ZHAO W,YANG M,et al. Prediction of expressway traffic accident duration based on the improved BPNN[J]. Journal of East China Jiaotong University,2020,37(5):60-65.

DOI:10.16749/j.cnki.jecjtu.2020.05.009

截至2018年底,我国的高速公路总里程已达14.26万km^[1]。随着高速公路网基本形成以及机动车保有量增加,由此带来的交通事故频发问题成为大众关注的焦点。目前,已有许多有效的模型与方法应用于交通事故持续时间预测的研究,早期的主要算法有概率分布^[2],回归分析^[3],时间序列^[4],模糊模型^[5]等。Wang等^[6]针对不同类型的事故类型,采用偏最小二乘回归(PLSR)建立了事故持续时间与影响因素之间的模型,并通过一个不区分事故类型的模型作为对比;Junhua等^[7]提出基于生存分析和对数逻辑分布的加速失效时间模型,可以接受不同分布的缺失数据并准确预测事故持续时间;马阿瑾^[8]总结性地阐述了交通事故时空影响分析的相关理论基础,分别针对不同阶段建立报警方式分类法、决策树法、交通波理论和排队论,将各个阶段的预测时间总和作为交通事故持续时间;陈建军等^[9]总结了预测事故持续时间的过程,详细地介绍了数据收集、数据处理、模型建立步骤,提出了基于条件概率的生存分析事故持续时间预测方法,最终通过得到的累积结束概率图,来反映事故持续时间的预测值和对应的预测概率。传统的方法为交通事故持续时间预测提供了研究思路,但由于模型自身特点导致预测的范围和精度有限。随着交通流理论和人工智能新技术的发展,越来越多的交通事故持续时间预测模型被提出,如决策树模型^[10],贝叶斯网络模型^[11-12],支持向量机^[13-14],人工神经网络^[15-16]。Ahmad等^[17]为了识别和量化影响因素的影响,提出了一种基于“综合数据库”历史数据研究事故持续时间的方法,建立了参数化加速故障时间生存模型,总结出事故持续时间的因素包括事故特征(严重程度、类型、伤害、医疗需求等)、基础设施特征(道路肩的可用性)、一天中的时间和交通特征;赵蕾^[18]构建以C4.5算法为核心的决策树模型来预测事故延迟时间,构建基于动态空间占有率的车流波模型来预测交通恢复时间,在不同交通环境及交通状态下可对交通事故延误进行估计。综上所述,目前的交通事故持续时间预测方法大都侧重于追求预测精度,极大地增加了模型的复杂度,而实时的持续时间预测才能对事故处理具有指导作用。

收稿日期:2020-04-17

基金项目:国家自然科学基金项目(71971029);陕西省重点研发项目(2019GY-002)

作者简介:许宏科(1963—),男,教授,博士生导师,研究方向为交通控制与管理、智能运输系统(ITS)。E-mail:xuhongke@chd.edu.cn。

通信作者:林杉(1983—),男,博士,研究方向为交通信息工程及控制。E-mail:linshan@chd.edu.cn。

针对现有交通事故持续时间预测方法的局限,提出因子分析和 BP 神经网络相结合的方法,通过因子分析法简化预测指标个数,得到公共因子,从而用较少的公共因子代替较多的原始数据。进一步将公共因子作为 BP 神经网络结构的输入层参数,简化了预测模型的结构,提高了模型的迭代速度和运算效率。真实的交通事故数据测试结果表明,与经典的交通事故持续时间预测模型相比,本文提出的预测模型预测准确率更高,是一种准确、合理的交通事故持续时间预测模型。

1 高速公路交通事故基本特征分析

本文涉及的高速公路交通事故主要是指造成高速公路交通流状态异常的交通事件,不考虑高速公路定期维修、基础建设、计划性改造等非紧急事件的影响。

高速公路交通事故持续时间是指从交通事故发生到恢复交通整个过程的时间,主要由事故发生,事故响应、事故清除和交通恢复时间 4 个阶段组成,如图 1 所示。前 3 个阶段的总时间表示事故延迟时间,主要受事故检测能力和事故清除效率的影响。交通恢复阶段是不可控的变量,不仅受到事故发生期间道路交通状况的影响,还与事故的形态等级、天气等因素有关。本文从时间信息、路段信息、事故信息和环境角度考虑影响因素,主要包括交通事故总量及其类型分布、时间、地点、环境、车辆类型、事故原因、特征及造成损失等影响因素。

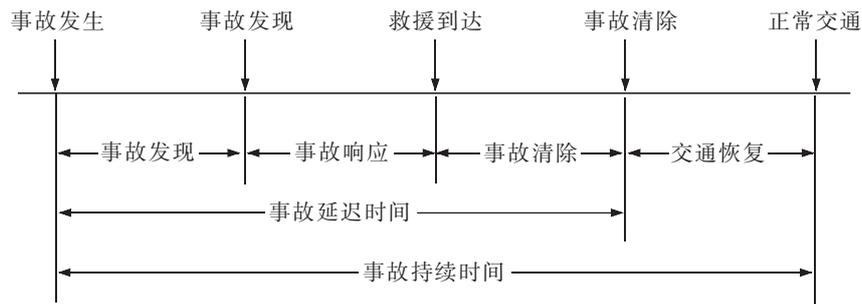


图 1 高速公路交通事故持续时间分布图

Fig.1 Distribution chart of expressway traffic accident duration

2 交通事故持续时间预测模型

2.1 事故影响因子提取

本文采用因子分析法对交通事故持续时间初始影响因素进行筛选,因子分析法是通过研究多个变量间协方差矩阵的内部依赖关系,找出能综合所有变量主要信息的少数几个不可观测的随机变量。这些随机变量称为因子,各个因子间互不相关,所有的变量都可以表示成为公因子的线性组合。

将高速公路交通事故持续时间的影响因素作为初始变量,通过因子分析法减少变量的数目,用较少的公共因子代替所有变量去分析整个问题,在不影响最终预测结果准确性的同时,简化了网络结构和提高了运算效率。其计算过程可以分为四个步骤,设 n 为交通事故数据样本总数, p 为影响因素个数, m 为公共因子个数, X 为总影响因素矩阵, F 为公共因子矩阵,具体原理及步骤如下:

第一步:计算总影响因素矩阵的协方差矩阵 R ,并计算协方差矩阵的特征根,同时计算特征根对应的特征向量 $\gamma_1, \gamma_2, \dots, \gamma_p$

$$X=(X_{ij})_{n \times p} \tag{1}$$

$$R=(r_{ij})_{n \times p} \tag{2}$$

第二步:计算因子载荷矩阵 A

$$A=(a_1, a_2, \dots, a_m)=(a_{ij})_{p \times m} \tag{3}$$

$$a_{ij}=\sqrt{\lambda_j} \gamma_{ji} \tag{4}$$

第三步:建立因子得分模型

$$X_i = a_{i1}F_1 + a_{i2}F_2 + \dots + a_{im}F_m + \varepsilon_i \quad (5)$$

其中: ε_i 为特殊因子,表示公因子以外的影响因素所导致的变量差异,实际分析时忽略不计。

第四步:通过回归估计等方法,求解因子得分模型,可解出公因子矩阵 F 的表达式

$$F_i = b_{i1}X_1 + b_{i2}X_2 + \dots + b_{ip}X_p \quad (6)$$

其中: $(b_{ij})_{p \times m}$ 为因子得分系数,实现总影响因素矩阵 X 和公共因子矩阵 F 的转换。

2.2 事故持续时间预测

BP神经网络算法也称为误差逆传播算法。它采用梯度下降策略,能够根据预设的参数更新规则,不断调整网络中的参数,以达到最符合期望的输出。基于改进BP神经网络进行交通事故持续时间预测的过程如下:

输入:交通事故影响因素数据集;输入样本数 n ;误差函数 E ;计算精度值 ε ;最大学习次数 M ;学习率 η 。

输出:交通事故持续时间。

- 1) 原始影响因素的采集和预处理,将影响因素数值化得到影响因素数据集 D ;
- 2) 采用 min-max 标准化方法对影响因素数据集 D 进行标准化,得到影响因素矩阵 X ;
- 3) 利用因子分析对影响因素矩阵 X 进行指标提取,生成影响因素的公共因子矩阵 F ;
- 4) for $i=1$ to M or $E > \varepsilon$, 重复执行(a) (e):
 - (a) 网络初始化过程,给各连接权重和偏置分别赋一个(0,1)之间的随机数;
 - (b) 将公共因子矩阵 F 作为网络的输入,依次分别计算隐藏层和输出层的输出值;
 - (c) 根据误差函数 E 来进行误差的计算,如果误差满足要求,则跳出循环,否则进行步骤(d);
 - (d) 误差反向传播过程,使用梯度下降策略依次对输出层单元和隐藏层单元的权值进行更新;
 - (e) 返回步骤(4),开始下一个循环。
- 5) 迭代结束,得到最优的模型参数,进行交通事故持续时间预测。

在数据标准化过程中,采用 min-max 方法,其表达式如下

$$x_i = \frac{d_i - d_{\min}}{d_{\max} - d_{\min}} \quad i=1, 2, \dots, n \quad (7)$$

式中: x_i 为第 i 个交通事故影响因素标准化后的值; d_i 为第 i 个原始的影响因素的值; d_{\min} 为原始数据中的最小值; d_{\max} 为原始数据中的最大值。

BP神经网络中的误差函数选择均方误差,其表达式如下

$$E = \frac{1}{2} \sum_{k=1}^n (y_k - o_k)^2 \quad (8)$$

式中: y_k 为期望输出的交通事故持续时间; o_k 为预测网络输出的持续时间。

2.3 预测模型评价指标

将预测结果的误差作为模型评价指标,本文选择常用的均方误差(RMSE)、平均绝对误差(MAE)和准确率(ACC)作为评价指标,具体定义如下

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (9)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (10)$$

$$\text{ACC} = \left(1 - \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \right) \times 100\% \quad (11)$$

式中: n 为预测样本数; \hat{y}_i 为样本预测值; y_i 为样本真实值。

3 实例分析

3.1 数据描述

研究数据来源于包茂高速陕西省西安—延安高速公路路网交通时间管理系统,选取2016年1月至2017年11月近两年来的153个交通事故数据,其中前120条数据用于BP神经网络预测模型的训练,后33条数据用于对训练的模型进行测试。每组交通事故数据由交通事故的持续时间和影响交通事故持续时间的12个相关影响因素组成,这些影响因素主要包括事故信息、路段信息、时间信息、环境信息。数据经过预处理后,作为各影响因素的原始数据,变量选取及其含义见表1。

表1 变量选取及其含义
Tab.1 Variable selection and its meaning

| 交通事故属性 | 交通事故影响因素 | 变量赋值 |
|--------|----------------|--|
| 时间信息 | 星期 x_1 | 1=周一,2=周二,3=周三,4=周四,5=周五,6=周六,7=周日 |
| | 节假日 x_2 | 1=是,0=否 |
| | 时间 x_3 | 1=0am~4am,2=4am~8am,3=8am~12am,4=12am~4pm,5=4pm~8pm,6=8pm~12pm |
| | 白昼 x_4 | 1=白天(7am~7pm),2=夜晚(7pm~7am) |
| 路段信息 | 管理所 x_5 | 1=黄陵,2=富县,3=铜川,4=西铜,5=延安 |
| | 桩号 x_6 | 5=K500,6=K600,7=K700,8=K800 |
| | 方向 x_7 | 1=延安方向,2=西安方向,3=铜川方向,4=榆林方向 |
| 事故信息 | 事故起因 x_8 | 1=侧翻,2=剐蹭,3=失火,4=追尾,5=撞击栏杆,6=其他 |
| | 人员伤亡 x_9 | 0=无,1=有 |
| | 涉及车辆数 x_{10} | 1=一辆,2=二辆,3=三辆,4=四辆,5=五辆,6=大于五辆 |
| | 车辆类型 x_{11} | 1=小型客车,2=小型货车,3=大型客车,4=大型货车 |
| 环境信息 | 天气 x_{12} | 1=晴,2=阴,3=雨,4=雪 |
| 生存时间 | 事故持续时间 y | 15 min 为一个单元 |

3.2 实验结果及分析

3.2.1 因子分析及数据标准化

通过SPSS对交通事故持续时间影响因素的数据进行因子分析。由Bartlett检验可以看出,独立显著性因子(Sig.)的值为0.001,应该拒绝各变量独立的假设,即认为变量间具有较强的相关性。得到的KMO检验统计量为0.725,说明变量间相关性较强,也证明因子分析可行,Bartlett和KMO的检验结果见表2。按照特征根大于1的默认指标提取了6个公因子,将公因子进行标准化处理,作为BP神经网络预测的输入数据。

表2 Bartlett和KMO的检验结果
Tab.2 Results of Bartlett and KMO

| KMO 检验值 | Bartlett 的球形度检验值 | | |
|---------|------------------|----|-------|
| | 近似卡方 | df | Sig. |
| 0.725 | 332.278 | 28 | 0.001 |

3.2.2 网络的建立

预测模型结构设置为3层,将通过因子分析得到的6个公共因子代替原有12个交通事故持续时间影响因素作为BP神经网络输入层参数,将交通事故持续时间以15 min为一个单位作为输出层参数。

隐藏层神经元的个数越多,预测模型的精确性越高,但同时会增加模型的复杂性,使预测速度和收敛性降低。在满足模型精确度要求的前提下,应该尽可能使模型结构简单化,本文参考隐藏层神经元个数的经验公式 $l = \sqrt{a+b} + c \left(\frac{\pi}{2} - \theta \right)$, 其中: l 为隐藏层的个数; a 为输入层神经元个数; b 为输出层神经元个数; c 为0~10之间的常数。结果表明,当隐藏层神经元个数为13时,改进的BP神经网络预测效果最好,既能保证预测结果的精确性,也能满足收敛性要求。最终确定BP神经网络模型的结构为6-13-1。

3.2.3 仿真训练

设置BP神经网络的学习效率为0.01,神经网络的训练精度为 1×10^{-12} ,最大的学习次数为1000次,基于TensorFlow环境进行仿真。经过训练后的BP神经网络模型对高速公路交通事故持续时间进行预测,预测结果如图2所示。从图2中可以看出,交通事故持续时间预测值与实际值吻合度较高,表明本文所提出的预测模型应用于实际交通事故持续时间预测中具有可行性。

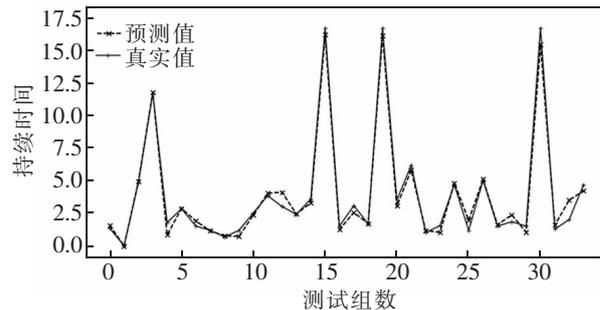


图2 交通事故持续时间预测结果

Fig.2 Prediction results of traffic accident duration

作为对比,本文还建立了经典的回归模型和支持向量机模型,用同样的训练样本对回归模型和支持向量机模型进行训练和测试。将BP神经网络预测模型与回归模型、支持向量机模型各类误差指标进行对比,评价对比结果见表3。从表3可以看出,改进的BP神经网络预测模型在预测精度方面明显优于支持向量机模型和回归模型。其中,基于改进BP神经网络的MAE、ACC值分别为0.85和11.99%,均低于支持向量机和回归模型,RMSE的值与支持向量机相近,但明显低于回归模型,预测准确率比支持向量机提高了7.8%。总体来说,该模型是较好的交通事故持续时间预测方法,具有较高的准确性和较好的实用性。

表3 预测模型结果分析

Tab.3 Analysis of prediction model results

| 预测模型 | 评价指标 | | |
|--------|------|------|--------|
| | RMSE | MAE | ACC |
| BP神经网络 | 0.82 | 0.85 | 11.99% |
| 支持向量机 | 0.81 | 1.13 | 19.79% |
| 回归模型 | 1.57 | 1.62 | 26.97% |

4 结论

1) 构建了因子分析与BP神经网络相结合的预测模型,通过因子分析对原始数据进行降维,将标准化后的公共因子作为BP神经网络的输入层参数,简化了预测模型的结构,提高了模型的迭代速度和数据处理效率。

2) 采用本文提出的神经网络预测方法对包茂高速西延段交通事故持续时间进行预测,预测结果表明,该模型算法结构简单,性能优异,综合预测性能优于典型的支持向量机和回归模型,具有较高的准确率和参考性。

参考文献:

- [1] 中华人民共和国交通运输部. 2018年交通运输行业发展统计公报[R]. 北京:交通运输部综合规划司,2019.
- [2] GOLOB T F, RECKER W W, LEONARD J D. An analysis of the severity and incident duration of truck-involved freeway accidents[J]. *Accident Analysis & Prevention*, 1987, 19(5):375-395.
- [3] 夏正丰. 高速公路交通事故持续时间概率预测模型[J]. *公路与汽运*, 2016(3):52-55.
- [4] 赵小强. 交通事故持续时间预测理论与方法[D]. 北京:清华大学,2010.
- [5] 姚磊,刘渊. 基于ANFIS的交通事件持续时间预测[J]. *计算机工程*, 2014, 40(2):189-192.
- [6] WANG X, CHEN S, ZHENG W. Traffic incident duration prediction based on partial least squares regression[J]. *Procedia - Social and Behavioral Sciences*, 2013, 96:425-432.
- [7] JUNHUA W, HAOZHE C, SHI Q. Estimating freeway incident duration using accelerated failure time modeling[J]. *Safety Science*, 2013, 54:43-50.
- [8] 马阿瑾. 高速公路交通事故持续时间和影响范围研究[D]. 西安:长安大学,2013.
- [9] 陈建军,夏正丰. 基于生存分析的交通事件持续时间预测模型[J]. *公路与汽运*, 2014(6):68-72.
- [10] LIN L, WANG Q, SADEK A W. A combined M5P tree and hazard-based duration model for predicting urban freeway traffic accident durations[J]. *Accident Analysis & Prevention*, 2016, 91:114-126.
- [11] 姬杨蓓蓓. 交通事件持续时间预测方法研究[D]. 上海:同济大学,2008.
- [12] 李东宁,尚云龙,郑中义. 浅议贝叶斯网络应用于海上交通事故致因分析的可行性[J]. *交通建设与管理*, 2009(5):140-144.
- [13] 张卫华,孙浩,穆朝絮. 基于支持向量机的交通安全预测模型及仿真研究[J]. *系统仿真学报*, 2009, 21(19):6266-6270.
- [14] YUAN R, GUANGCHEN B. Determination of optimal SVM parameters by using GA/PSO[J]. *Journal of Computers*, 2010, 5(8):1160-1168.
- [15] 董四辉,卓阜. 道路交通事故BP神经网络预测研究[J]. *中国安全科学学报*, 2010, 20(9):15-20.
- [16] 葛丽娜. 基于灰色神经网络的交通事故预测方法研究[D]. 锦州:辽宁工业大学,2015.
- [17] TAVASSOLI H A, FERREIRA L, WASHINGTON S, et al. Modelling total duration of traffic incidents including incident detection and recovery time[J]. *Accid Anal Prev*, 2014, 71:296-305.
- [18] 赵蕾. 高速公路交通事故持续时间预测[D]. 北京:北京交通大学,2018.

Prediction of Expressway Traffic Accident Duration Based on the Improved BPNN

Xu Hongke¹, Zhao Wei¹, Yang Meng², Lin Shan¹, Liu Dongwei¹

(1. School of Electronics and Control Engineering, Chang'an University, Xi'an 710064, China;

2. China Merchants Chongqing Communications Technology Research & Design Institute Co., Ltd, Chongqing 400067, China)

Abstract: Aiming at the difficulty in predicting the duration of highway traffic accidents, a traffic accident duration prediction model based on factor analysis and back propagation neural network (BPNN) is established. Based on real traffic accident data collected from Baomao Expressway, factor analysis is utilized to extract the common factors influencing the duration of traffic accidents. The extracted public factors are taken as the input, and BP neural network is used to predict the duration of traffic accidents. The experimental results show that compared with the typical regression algorithm and the support vector machine, the improved BPNN method with factor analysis proposed in this paper can not only improve the prediction accuracy by 7.8%, but also solve the problems of low data processing efficiency and slow iteration speed of traditional BP neural network.

Key words: highway transportation; traffic accident; duration; neural network; factor analysis

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>