

# 基于BP神经网络的公路客运量预测方法

王 栋

(西安航空学院 车辆工程学院,陕西 西安 710077)

**摘 要:**公路客运量是交通科学管理的基础性数据资料,能够反映出公路运输产出成果,对提高公路交通管理层次及建立畅通、高效的公路交通系统,具有重要意义。为提高公路客运量的预测精度,选择与公路客运量相关的主要社会指标(包括公路客运量、汽车保有量、国民总收入、人均GDP、人口总量、城镇居民人均可支配收入、社会消费品零售总额和城市化率),运用灰色关联分析法进行计算分析,最终确定公路客运量影响因子为汽车保有量、人均GDP、人口总量和城市化率。将所确定的因子作为公路客运量的预测指标,建立基于BP神经网络的公路客运量预测模型,并对模型进行了应用测试。结果表明:BP神经网络模型具有较高的精度,最小相对误差为1.1%,平均相对误差为2.78%。

**关键词:**灰色关联分析;BP神经网络;公路客运量;预测

**中图分类号:**U491.1

**文献标识码:**A

**文章编号:**1673-629X(2017)02-0187-04

**doi:**10.3969/j.issn.1673-629X.2017.02.043

## Prediction Method of Highway Passenger Transportation Volume Based on BP Neural Network

WANG Dong

(Vehicle Engineering Institute, Xi'an Aeronautical University, Xi'an 710077, China)

**Abstract:** Highway passenger transportation volume is basic data of traffic scientific management and can reflect the results of highway transportation, which is of great significance to improve the road traffic management level and establish a smooth and efficient highway traffic system. In order to improve the forecasting accuracy of highway passenger transportation, the gray correlation method is used to compute and analyze. The main predictors are car ownership, per capita GDP, total population number and urbanization. The prediction model of highway passenger transportation is establish based on BP neural network, and then verified with tests. The results show that highway passenger transportation can be predicted accurately by the model based on BP neural network. The minimum relative error is 1.1% and the average relative error is 2.78%.

**Key words:** grey relational analysis; BP neural network; highway passenger transportation; forecasting

### 0 引言

2013年国内汽车产销量均突破2 000万辆,增速已连续五年蝉联全球第一。机动车保有量飞速增长,但随着城市化进程的迅速发展,道路基础设施建设的滞后,道路交通阻塞的问题日益严重。公路客运量是交通科学管理的基础性数据资料,能够反映出公路运输产出成果,对提高公路交通管理层次及建立畅通、高效的公路交通系统,具有重要意义<sup>[1-3]</sup>。能够科学准确地预测公路客运量,掌握公路客运量发展的趋势、特点、规律和数量,可以很好地为制定公路网发展规划、交通发展战略、行业管理和决策等做先期准备<sup>[4-6]</sup>。

国内外对公路客运量的预测方法进行了大量研究。Garrido等建立了多项概率模型并用蒙特卡洛方法对其计算精度进行评价<sup>[7]</sup>;Godfrey等使用指数平滑法进行预测分析<sup>[8]</sup>;Suyani等利用系统动力学构建模型进行预测分析<sup>[9]</sup>。李明伟等提出了混沌加速遗传算法建立了客运量预测模型<sup>[10]</sup>;陈鹏等运用灰色G(11)方法建立了客运量预测模型并用马尔可夫方法对预测值进行修正<sup>[11]</sup>;彭辉等利用构造的OD模型预测未来年OD分布法构造通道OD的基本原理进行了分段公路客运量的推断<sup>[12]</sup>。

针对公路客运量的预测,主要采用的预测方法有弹性系数法、回归分析法、灰色预测法、支持向量机、灰

收稿日期:2015-07-15

修回日期:2015-12-14

网络出版时间:2017-01-10

基金项目:陕西省教育科学“十二五”规划项目(SGH140711,SGH140790);西安航空学院科研基金项目(2014KY1212)

作者简介:王 栋(1987-),男,硕士,讲师,研究方向为道路交通安全和人车路系统动力学等。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20170110.1010.022.html>

色马尔可夫预测法、时间序列法等。其中时间序列法以时间为自变量,依据客运量的历史数据构建模型,但模型较为简单,无法表征不同因素的内在联系,此外也不能研究分析多个因子的关联性。而回归分析法在构建模型的过程中,会预设一些假定条件,因此预测的精度较差。因为重点考虑单个因素,所以灰色预测法与弹性系数法都无法整体把握客运量的变化。文中从定性的角度出发分析公路客运量相关的社会指标,采用灰色关联分析法计算客运量与相关的指标定量间的关联度,找出影响公路客运量的因素。运用 BP 神经网络构建预测模型,并对其预测精度进行了分析测试。

# 1 基于灰色关联分析法的公路客运量影响因子的确定

影响公路客运量的因素有很多,在构建客运量预

测模型时,要充分考虑到各个可能的因素,模型的因素指标越多,那么最后的预测精度越高。而在运用过程中,如果输入的因素指标过多,会导致模型过于复杂,一方面会增大计算负担,另一方面无法有效解决实际问题。此外,不同的指标之间存在一定的相关性,指标过多会使计算变得困难。所以,构建模型时尽量减少输入因素指标的个数,同时要保证不同因素指标之间有较好的独立性。

综合参考国内外的研究现状,根据实际需求,选取了以下相关因素指标:城镇居民人均可支配收入、国民总收入、人均 GDP、汽车保有量、人口总量、社会消费品零售总额和城市化率。查阅《中国统计年鉴》获取 1994—2001 年公路客运量以及相关因素的统计数据,见表 1。

表 1 统计数据

年份	公路客运量/ $10^4$ 人	汽车保有量/ $10^4$	国民总收入/ $10^8$ 元	人均 GDP/元	人口总量/ $10^4$ 人	城镇居民人均可支配收入/元	社会消费品零售总额/ $10^8$ 元	城市化率/%
1994	953 940	941.95	48 108.46	4 044.00	119 850	3 496.2	18 622.9	28.51
1995	1 040 810	1 040.00	59 810.53	5 045.73	121 121	4 283.0	23 613.8	29.04
1996	1 122 110	1 100.08	70 142.49	5 845.89	122 389	4 838.9	28 360.2	30.48
1997	1 204 583	1 219.09	78 060.85	6 420.18	123 626	5 160.3	31 252.9	31.91
1998	1 257 332	1 319.3	83 024.28	6 796.03	124 761	5 425.1	33 378.1	33.35
1999	1 269 004	1 452.94	88 479.15	7 158.50	125 786	5 854.0	35 647.9	34.78
2000	1 347 392	1 608.91	98 000.45	7 857.68	126 743	6 280.0	39 105.7	36.22
2001	1 402 798	1 802.04	108 068.2	8 621.71	127 627	6 859.6	43 055.4	37.66
2002	1 475 257	2 053.17	119 095.7	9 398.05	128 453	7 702.8	48 135.9	39.09
2003	1 464 335	2 382.93	134 977.0	10 541.97	129 227	8 472.2	52 516.3	40.53
2004	1 624 526	2 693.71	159 453.6	12 335.58	129 988	9 421.6	59 501.0	41.76
2005	1 697 381	3 159.66	183 617.4	14 185.36	130 756	10 493.0	68 352.6	42.99
2006	1 860 487	3 697.35	215 904.4	16 499.70	131 448	11 759.5	79 145.2	44.34
2007	2 050 680	4 358.36	266 422.0	20 169.46	132 129	13 785.8	93 571.6	45.89
2008	2 682 114	5 099.61	316 030.3	23 707.71	132 802	15 780.8	114 830.1	46.99
2009	2 779 081	6 280.61	340 320.0	25 607.53	133 450	17 174.7	132 678.4	48.34
2010	3 052 738	7 801.83	399 759.5	30 015.05	134 091	19 109.4	156 998.4	49.95
2011	3 286 220	9 356.32	468 562.4	35 197.79	134 735	21 809.8	183 918.6	51.27

文中采用灰色关联分析法对公路客运量以及相关因素指标进行关联分析,找出指标中与客运量有较强关联度的指标,具体计算步骤如下:

(1)将原始数据展开进行无量纲化处理,令

$$Y = \left\{ \frac{y_1}{y_1}, \frac{y_2}{y_1}, \dots, \frac{y_i}{y_1}, \dots, \frac{y_m}{y_1} \right\} \quad (1)$$

$$P_j = \left\{ \frac{x_{j1}}{x_{j1}}, \frac{x_{j2}}{x_{j1}}, \dots, \frac{x_{ji}}{x_{j1}}, \dots, \frac{x_{jm}}{x_{j1}} \right\} \quad (2)$$

其中,  $Y$  表示公路客运量量纲为 1 的参考序列;  $y_i$

表示第  $i$  组公路客运量实际值;  $P_j$  表示第  $j$  个因素量纲为 1 的比较序列;  $x_{ji}$  表示第  $j$  个因素在第  $i$  组的值,  $1 \leq i \leq n, 1 \leq j \leq 7$ , 其中  $i$  和  $j$  均为正整数。

(2)计算参考序列与比较序列之间的差值

$$\Delta_{ji} = |Y_i - P_{ji}| \quad (3)$$

$$Y_i = \frac{y_i}{y_1}$$

$$P_{ji} = \frac{x_{ji}}{x_{j1}}$$

其中,  $\Delta_{ji}$  表示  $Y_i$  与  $P_{ji}$  之差的绝对值;  $Y_i$  表示参考序列中的第  $i$  组值;  $P_{ji}$  表示比较序列中第  $j$  个因素在第  $i$  组的值。

(3) 计算关联系数。

$$\varepsilon_{ji} = \frac{\Delta_{\min} + \rho\Delta_{\max}}{\Delta_{ji} + \rho\Delta_{\max}} \tag{4}$$

其中,  $\rho$  表示分辨系数(取值为 0.5);  $\varepsilon_{ji}$  表示第  $j$  个因子在第  $i$  组的值与第  $i$  组公路客运量的关联系数;

$\Delta_{\min}$  和  $\Delta_{\max}$  分别表示取值中的最小值与最大值。

(4) 计算灰色关联度。

$$\varepsilon_j = \frac{1}{m} \sum_{i=1}^m \varepsilon_{ji} \tag{5}$$

其中,  $\varepsilon_j$  表示第  $j$  个因子与公路客运量  $y$  之间的灰色关联度。

运用灰色关联度计算客运量与不同因子指标之间的关联度,结果如表 2 所示。

表 2 关联度计算

指标	汽车保有量	国民总收入	人均 GDP	人口总量	城镇居民人均可支配收入	社会消费品零售总额	城市化率
关联度	0.84	0.73	0.86	0.89	0.75	0.69	0.93

如表 2 所示,与公路客运量有较强关联性的指标有人均 GDP、城市化率、汽车保有量以及人口总量。这 4 个指标与客运量的关联度均大于 0.80,因此采用它们作为预测模型的最终预测指标。

2 基于 BP 神经网络公路客运量预测模型的建立

2.1 BP 神经网络的原理

神经网络是模仿高级神经网络和功能构建起来的一种前馈型信息处理系统,含有输入层、隐含层以及输出层。其中,BP 神经网络是运用非线性可微分函数展开权值训练的一种高级神经网络,分为多层,可以采用输出层的误差反馈估计直接前导层误差,进而能计算出前一层可能的误差<sup>[13-14]</sup>。

根据研究结果,BP 神经网络可以按照给定的精度要求逼近任意指定函数<sup>[15-16]</sup>。

2.2 BP 神经网络模型的构建

文中分 4 步构建 BP 神经网络模型:

(1) 将模型输入层设为:

$$X_k = (x_1, x_2, \cdots, x_n) \tag{6}$$

其中,  $x_1$  为公路客运量第 1 相关预测指标;  $x_2$  为第 2 相关预测指标;  $x_n$  为第  $n$  相关预测指标。

(2) 将模型输入层对应的输出层设为:

$$Y = y \tag{7}$$

(3) 隐含层各个单元的输入为:

$$S_j = \sum_{i=1}^k w_{ij}x_i - \theta_j, j = 1, 2, \cdots, p \tag{8}$$

其中,  $w_{ij}$  为输入层至隐层的连接权重;  $\theta_j$  为隐层单元的阈值;  $p$  为隐层单元的个数。

模型转移函数为 sigmoid 函数:  $f(x) = 1/(1 + e^{-x})$ 。则隐含层单元的输出为:

$$b_j = 1/[1 + \exp(\theta_j - \sum_{i=1}^k w_{ij}x_i)] \tag{9}$$

(4) 输出层单元的输入为:

$$L_i = \sum_{j=1}^n v_{ji}b_j - \gamma_i \tag{10}$$

输出层单元的输出为:

$$Y_i = 1/[1 + \exp(\gamma_i - \sum_{j=1}^n v_{ji}b_j)] \tag{11}$$

其中,  $\gamma_i$  表示输出层单元阈值;  $v_{ji}$  表示隐层到输出层的连接权重。

步骤(1)到(4)为模型的正向计算传播。而在误差反向传播过程中,则要训练设置好的 BP 神经网络,通过调整阈值  $\gamma_i$  以及连接权重  $w_{ij}$ 、 $v_{ji}$  不断缩小误差至要求精度范围。其中式(11)为最终构建的客运量预测模型。

2.3 模型的设计及调试

随机抽取 13 组数据作为样本展开模型训练(1994 年-2000 年和 2006 年-2011 年),在模型训练之前,先对选取的数据进行归一化处理。

将模型设计为三层,经反复多次的调试训练,最终将隐含层的神经元个数设置为 9。采用的隐含层传递函数为 logsig;输出层传递函数选择为 tansig;trainlm 代表模型的训练函数;取 learnlm 为学习函数;mse 则代表网络性能函数。其他参数的设定如表 3 所示。

表 3 训练参数的选择

训练次数设定	训练目标	学习速率
3 000	0.001	0.01

通过 8 次训练,模型的误差达到了预期范围,最终 mse=0.000 976 817。此外,模型能保持较高的收敛速度。

模型的训练收敛曲线图与误差曲线图分别如图 1、图 2 所示。

图 2 表明,文中预测模型能较好地控制误差,误差始终控制在 0.08 以内。

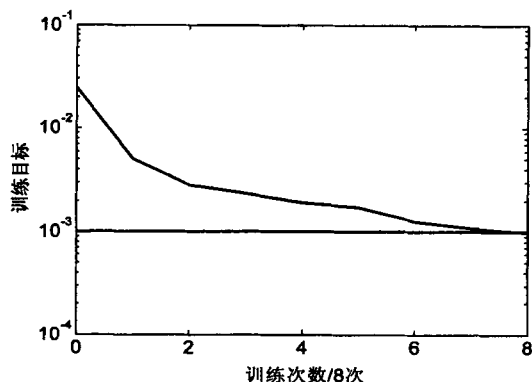


图1 收敛曲线

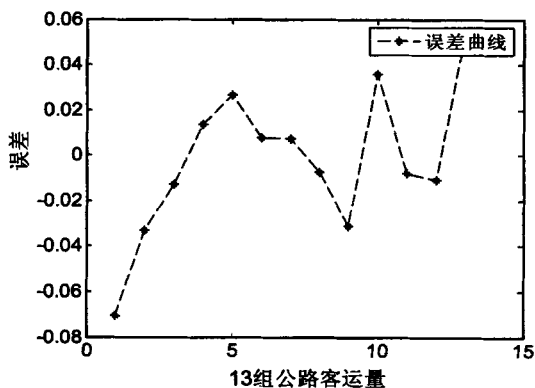


图2 误差曲线

### 3 模型的测试

选取训练样本外的剩余 5 组数据(2001 年-2005 年)进行测试,用以测试模型应用的误差范围,结果如表 4 所示。

表 4 实际值与预测值对比

年份	公路客运量/ $10^4$ 人	预测值/ $10^4$ 人	误差/ $10^4$ 人	相对误差/%
2001	1 402 798	1 418 500	15 702	1.1
2002	1 475 257	1 525 000	49 743	3.3
2003	1 464 335	1 502 200	37 865	2.5
2004	1 624 526	1 687 700	63 174	3.8
2005	1 697 381	1 753 000	55 619	3.2

根据表 4,可以计算出基于 BP 神经网络模型构建的预测模型的平均相对误差为 2.78%,最小值为 1.1%,模型的误差较小,能较好地应用于实际。

### 4 结束语

采用了灰色关联分析法分析了人均 GDP 等 8 个相关的社会指标与公路客运量之间的关联性,而通过相关性计算,确定关联性较强的指标有人口总量、城市化率、汽车保有量及人均 GDP。采用相关性较强的 4 个社会指标作为预测模型的输入变量,基于 BP 神经网络模型构建了预测模型。通过训练测试,结果表明

模型误差较小,能较好地应用于实际。由于该模型在训练时数据样本较少,因此在计算时仍然存在一定的相对误差,如果增加数据样本,则可以进一步提高预测模型的预测精度。

#### 参考文献:

- [1] 耿睿,崔德光,徐冰.应用支持向量机的空中交通流量组合预测模型[J].清华大学学报:自然科学版,2008,48(7):1205-1208.
- [2] 孙煦,陆化普,吴娟.基于蚁群优化支持向量机模型的公路客运量预测[J].合肥工业大学学报:自然科学版,2012,35(1):124-129.
- [3] 芮海田,吴群琪,袁华智,等.基于指数平滑法和马尔科夫模型的公路客运量预测方法[J].交通运输工程学报,2013,13(4):87-93.
- [4] 刘芹,张文峰,吴卓葵,等.基于最小二乘支持向量机的城市客运量预测模型[J].交通与计算机,2007,25(5):50-53.
- [5] 李季涛,杨俊锋.基于径向基神经网络的大连站客运量预测[J].大连交通大学学报,2007,28(1):32-34.
- [6] 高王翠,单飞,杨宇翔.基于 SPSS 主成分分析法在公路客运量预测中的应用[J].山东交通科技,2011(2):7-10.
- [7] Garrido R A, Mahmassni H S. Forecasting freight transportation demand with the space-time multinomial probit model[J]. Transportation Research Part B: Methodological, 2000, 34(5):403-418.
- [8] Godfrey G A, Powell W B. An adaptive dynamic programming algorithm for dynamic fleet management, II: multiperiod travel times[J]. Transportation Science, 2002, 36(1):40-54.
- [9] Suyani E, Chou S Y, Chen C H. Air passenger demand forecasting and passenger terminal capacity expansion: a system dynamics framework[J]. Expert Systems with Applications, 2010, 37(3):2324-2339.
- [10] 李明伟,康海贵,周鹏飞.基于 NCAGA-投影寻踪混合优化城市客运量预测[J].系统工程理论与实践,2012,32(4):903-910.
- [11] 陈鹏,孙全欣.基于灰色马尔柯夫过程的铁路客运量预测方法研究[J].铁道运输与经济,2005,27(4):65-67.
- [12] 彭辉,魏金丽,陈宽民.运输通道公路旅客中长距离 OD 模型构造及分段客运量预测[J].中国公路学报,2006,19(2):101-105.
- [13] 王栋.山区高速公路车辆行驶状态监测及预警方法研究[D].西安:长安大学,2013.
- [14] 王栋,邓北川,仇建华,等.山区高速公路直线段车速预测方法[J].交通科学与工程,2014,30(3):81-86.
- [15] Freitas P S A, Rodrigues A J L. Model combination in neural-based forecasting[J]. European Journal of Operational Research, 2006, 173(3):801-814.
- [16] 王栋.基于神经网络算法的公路货运量预测方法[J].北华大学学报:自然科学版,2014,15(3):417-420.