

基于灰色遗传 BP 神经网络的铁路货运量预测

李建国¹, 向万里², 王久梗³

(1. 中国铁路兰州局集团有限公司, 兰州 730000; 2. 兰州交通大学 交通运输学院, 兰州 730070;
3. 兰州交通大学 甘肃省物流与信息技术研究院, 兰州 730070)

摘要:铁路货运量受限于社会建设、投资和政策调控等多方面的影响。为了给货运系统提供更好的调配支撑,基于铁路货运量影响因素,加强 BP 预测性能,提出灰色遗传 BP 神经网络的铁路货运量预测模型(GR_GA_BP),对铁路货运量进行灰色关联分析,找出与铁路货运量关联程度较高的指标作为 GR_GA_BP 预测模型的输入对货运量进行预测。通过 MAPE、RMSE 和 MAE 对比 GR_GA_BP 模型与其他模型,实验结果表明,GR_GA_BP 模型优于其他模型,可以作为新方法预测和研究铁路货运能力。

关键词:铁路货运量;评价指标;灰色关联分析;遗传算法;BP 神经网络

中图分类号:U294 **文献标志码:**A **文章编号:**1671-1807(2022)01-0119-06

铁路货运量是铁路运输能力的表现,其趋势发展综合体现了铁路建设、投资和政策调控的有效性,预测货运量可以为其更好地发展提供科学依据。

在指标建立和构建灰色关联模型分析影响因素方面,吴华稳^[1]从铁路行业的安全出发,颜保凡等^[2]从铁路货运量的内部影响因素出发建立指标体系,赵怀鑫等^[3]用灰熵法分析了公路的货运量和周转量的关联因素,唐力等^[4]借助灰色关联分析评价了铁路的物流服务,钱隼驰等^[5]对灰色关联分析中的分辨系数进行了定量研究。

在预测方面,周程等^[6]用改进的 PSO-BP 预测了货运量,王龙^[7]通过研究公路的 OD 推算模型预测了铁路的货运量,幸运燕等^[8]借助 BP 神经网络模型预测了航班延误时旅客的行为,刘笑佟等^[9]结合不确定理论预测了货运量,张健等^[10]采用事变模型平均方法预测了航空客运量。

1 铁路货运量及其影响因素评价指标体系

本文的铁路货运量影响因素主要从铁路货物运输的主要品类所关联的行业进行分析,从而构建了铁路货运量影响因素评价指标体系,见表 1。

表 1 铁路货运量及其影响因素评价指标体系

一级指标	二级指标	三级指标
铁路货运量及其影响因素评价指标	经济发展	国内生产总值 X_1 /亿元
		第一产业增加值 X_2 /亿元
		第二产业增加值 X_3 /亿元
		社会固定资产投资 X_4 /亿元
		人口总数 X_5 /万人
		进出口总额 X_6 /亿元
	运输业	快递量 X_7 /万件
		公路货运量 X_8 /万 t
		水运货运量 X_9 /万 t
		民用航空货运量 X_{10} /万 t
	部分运输品类产量	全国原煤产量 X_{11} /亿 t
		全国钢材产量 X_{12} /万 t
		全国粮食产量 X_{13} /万 t
		全国生铁产量 X_{14} /万 t
		全国十种有色金属产量 X_{15} /万 t
	铁路发展	铁路营业里程 X_{16} /万 km
		铁路货运量 Y /万 t
		铁路货物周转量 $X_{17}/(\text{亿 t}\cdot\text{km})$
		国家铁路货车拥有量 X_{18} /辆
	生活水平	城镇单位就业人员平均工资 X_{19} /元
		社会消费品零售总额 X_{20} /亿元
		居民消费水平 X_{21} /元
		城镇居民消费水平 X_{22} /元

2 灰色关联模型构建

本文在研究铁路货运量时统计了可能影响其

收稿日期:2021-04-24
基金项目:国家自然科学基金(61563028);中国国家铁路集团有限公司科技研究开发计划(N2020X015);中国铁路兰州局集团有限公司科技项目(2021013-2,2020010-2,2020011-2)。
作者简介:李建国(1965—),男,甘肃通渭人,中国铁路兰州局集团有限公司,高级工程师,研究方向为铁路运输、现代物流;向万里(1978—),男,湖南张家界人,兰州交通大学运输学院,系主任,教授,博士,研究方向为铁路货物运输;王久梗(1963—),男,江苏连云港人,兰州交通大学甘肃省物流与信息技术研究院,院长,教授,硕士,研究方向为铁路运输、现代物流。

发展的相关指标,借助灰色关联度模型铁路货运量与不同量纲的指标关联程度之间的关联程度。选择铁路的货运量这一定性指标作为参考序列,然后分析铁路货运量的发展与其他指标关联的程度,探究各个指标内部存在的互相牵制的关系,从而为货运量提供可靠的数据支撑。

2.1 模型

灰色关联模型是通过分析影响系统行为的一系列指标和表现系统行为的参考序列之间的相似程度,来评价各个指标之间的关联程度的一种方法。该模型首先对各个指标进行无量纲化处理,再通过几何关系得到灰色关联度。

根据评价指标体系中的数据建立参考序列和比较序列:

$$\begin{cases} Y = Y(k) \mid k = 1, 2, \cdots, n \\ X_i = X_i(k) \mid k = 1, 2, \cdots, n, i = 1, 2, \cdots, m \end{cases}$$

(1)

2.1.1 灰色绝对关联度

采用初始化零点法进行数据的无量纲化:

$$\begin{cases} Y(k) = \begin{cases} 0, k = 1 \\ Y(k) - Y(1), \text{否则} \end{cases} \\ X_i(k) = \begin{cases} 0, k = 1 \\ X_i(k) - X_i(1), \text{否则} \end{cases}, i = 1, 2, \cdots, m \end{cases}$$

(2)

绝对关联度计算:

$$\begin{cases} |R_0| = \left| \sum_{k=2}^{n-1} Y(k) + \frac{1}{2} Y(n) \right| \\ |R_i| = \left| \sum_{k=2}^{n-1} X_i(k) + \frac{1}{2} X_i(n) \right|, \\ i = 1, 2, \cdots, m \\ |R_i - R_0| = \left| \sum_{k=2}^{n-1} [X_i(k) - Y(k)] + \frac{1}{2} [X_i(n) - Y(n)] \right| \\ \xi_{0i} = \frac{1 + |R_0| + |R_i|}{1 + |R_0| + |R_i| + |R_i - R_0|}, \\ i = 1, 2, \cdots, m \end{cases}$$

(3)

2.1.2 灰色相对关联度

进行数据的无量纲化处理

$$\begin{cases} Y'(k) = \begin{cases} 1, k = 1 \\ \frac{Y'(k)}{Y'(1)}, \text{否则} \end{cases} \\ X'_i(k) = \begin{cases} 0, k = 1 \\ \frac{X'_i(k)}{X'_i(1)}, \text{否则} \end{cases}, i = 1, 2, \cdots, m \end{cases}$$

(4)

相对关联度值计算:

$$\begin{cases} |R'_0| = \left| \sum_{k=2}^{n-1} Y'(k) + \frac{1}{2} Y'(n) \right| \\ |R'_i| = \left| \sum_{k=2}^{n-1} X'_i(k) + \frac{1}{2} X'_i(n) \right|, \\ i = 1, 2, \cdots, m \\ |R'_i - R'_0| = \left| \sum_{k=2}^{n-1} [X'_i(k) - Y'(k)] + \frac{1}{2} [X'_i(n) - Y'(n)] \right| \\ \gamma_{0i} = \frac{1 + |R'_0| + |R'_i|}{1 + |R'_0| + |R'_i| + |R'_i - R'_0|}, \\ i = 1, 2, \cdots, m \end{cases}$$

(5)

2.1.3 灰色综合关联度

ξ_{0i} 与 γ_{0i} 分别为Y和 X_i 的绝对关联度和相对关联度, $\mu \in [0, 1]$,Y和 X_i 的灰色综合关联度 α_{0i} 为

$$\alpha_{0i} = \mu \xi_{0i} + (1 - \mu) \gamma_{0i}$$

(6)

2.2 实例测算

根据式(1)2000—2019年中国铁路货运量及其影响因素的数据建立参考序列和比较序列的矩阵,所用到的数据来源于国家统计局的年度数据。根据式(2)、式(3)求解得到灰色绝对关联度值,根据式(4)、式(5)求解得到灰色相对关联度值,结果见表2。

表 2 灰色绝对关联度和灰色相对关联度

指标	绝对 关联度	相对 关联度	指标	绝对 关联度	相对 关联度
$Y \& X_1$	0.710	0.704	$Y \& X_{12}$	0.703	0.667
$Y \& X_2$	0.672	0.644	$Y \& X_{13}$	0.534	0.837
$Y \& X_3$	0.998	0.718	$Y \& X_{14}$	0.635	0.729
$Y \& X_4$	0.771	0.600	$Y \& X_{15}$	0.509	0.720
$Y \& X_5$	0.525	0.794	$Y \& X_{16}$	0.500	0.883
$Y \& X_6$	0.993	0.693	$Y \& X_{17}$	0.557	0.677
$Y \& X_7$	0.564	0.531	$Y \& X_{18}$	0.864	0.902
$Y \& X_8$	0.555	0.901	$Y \& X_{19}$	0.606	0.713
$Y \& X_9$	0.761	0.779	$Y \& X_{20}$	0.990	0.697
$Y \& X_{10}$	0.501	0.880	$Y \& X_{21}$	0.530	0.775
$Y \& X_{11}$	0.500	0.916	$Y \& X_{22}$	0.538	0.847

由表2的关联度值得到图1,不难看出Y与 X_1 、 X_3 、 X_6 、 X_9 、 X_{18} 、 X_{20} 的关联度值较为接近,与其他指标的差异较大,这主要是由量纲过大的差距引起的。所以,分别取 $\mu = 0.4, 0.5, 0.6$,再根据式(6)得到综合关联度(图2),取3个条件下关联度值均在0.700以上的7个指标: X_3 、 X_6 、 X_9 、 X_{12} 、 X_{17} 、 X_{18} 、 X_{20} ,数据见表3。

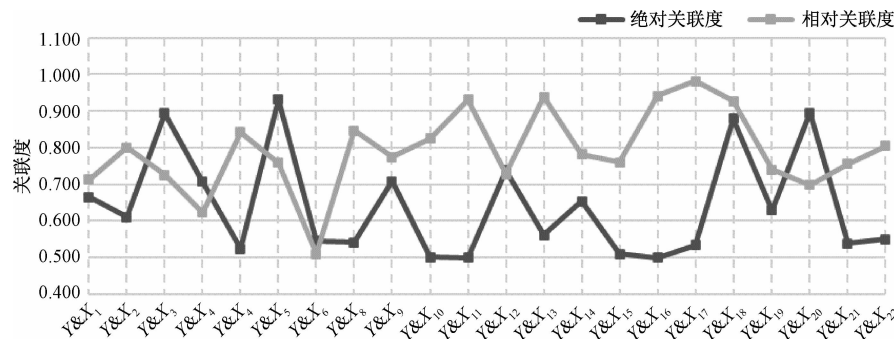


图 1 灰色相对、绝对关联度值对比

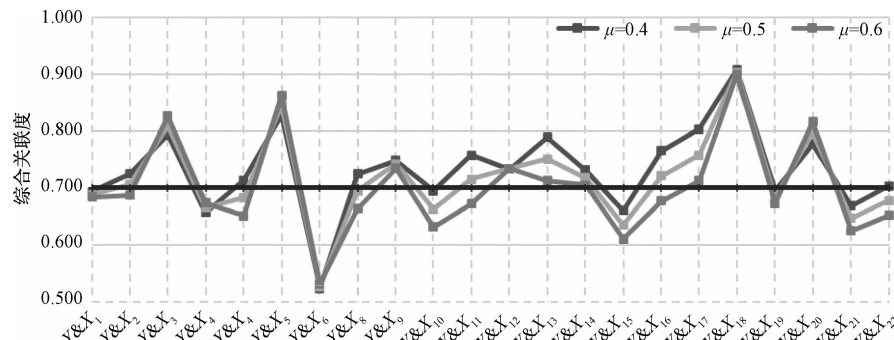


图 2 $\mu=0.4,0.5,0.6$ 时的综合关联度值

表 3 2000—2019 年货运量及相关影响指标数据

年份	铁路货运量 Y/ 万 t	国内生产 总值 X_1 /亿元	第二产业 增加值 X_3 /亿元	进出口总额 X_6 / 亿元	水运货运量 X_9 / 万 t	国家铁路货车 拥有量 X_{18} /辆	社会消费品零售 总额 X_{20} /亿元
2000	178 581	99 066.1	45 663.7	39 273.3	122 391	439 943	39 105.7
2001	193 189	109 276.2	49 659.4	42 183.6	132 675	453 620	43 055.4
2002	204 956	120 480.4	54 104.1	51 378.2	141 832	459 017	48 135.9
2003	224 248	137 422.0	62 695.8	70 483.5	158 070	510 327	52 516.3
2004	249 017	161 840.2	74 285.0	95 539.1	187 394	526 894	59 501.0
2005	269 296	187 318.9	88 082.2	116 921.8	219 648	541 824	68 352.6
2006	288 224	219 438.5	104 359.2	140 974.7	248 703	564 899	79 145.2
2007	314 237	270 092.3	126 630.5	166 924.1	281 199	577 521	93 571.6
2008	330 354	319 244.6	149 952.9	179 921.5	294 510	591 793	114 830.1
2009	333 348	348 517.7	160 168.8	150 648.1	318 996	601 412	133 048.2
2010	364 271	412 119.3	191 626.5	201 722.3	378 949	622 284	158 008.0
2011	393 263	487 940.2	227 035.1	236 402.0	425 968	651 175	187 205.8
2012	390 438	538 580.0	244 639.1	244 160.2	458 705	670 801	214 432.7
2013	396 697	592 963.2	261 951.6	258 168.9	559 785	715 492	242 842.8
2014	381 334	643 563.1	277 282.8	264 241.8	598 283	716 578	271 896.1
2015	335 801	688 858.2	281 338.9	245 502.9	613 567	768 516	300 930.8
2016	333 186	746 395.1	295 427.8	243 386.5	638 238	764 783	332 316.3
2017	368 865	832 035.9	331 580.5	278 099.2	667 846	808 736	366 261.6
2018	402 631	919 281.1	364 835.2	305 010.1	702 684	839 213	380 986.9
2019	431 773	990 865.1	386 165.3	315 504.8	747 225	878 000	411 649.0

3 基于 GR_GA_BP 的货运量预测方法

针对铁路货运量的预测选择具有较强的非线性映射能力的 BP 神经网络,基于灰色关联模型的综合关联度下,考虑到 BP 神经网络容易陷入局部最优和收敛速度慢的缺点,结合遗传算法优化 BP

神经网络的全局搜索能力和收敛速度,构建基于遗传算法下的 GR_GA_BP 模型,加强 BP 神经网络对铁路货运量的预测能力。GR_GA_BP 预测流程如图 3 所示。

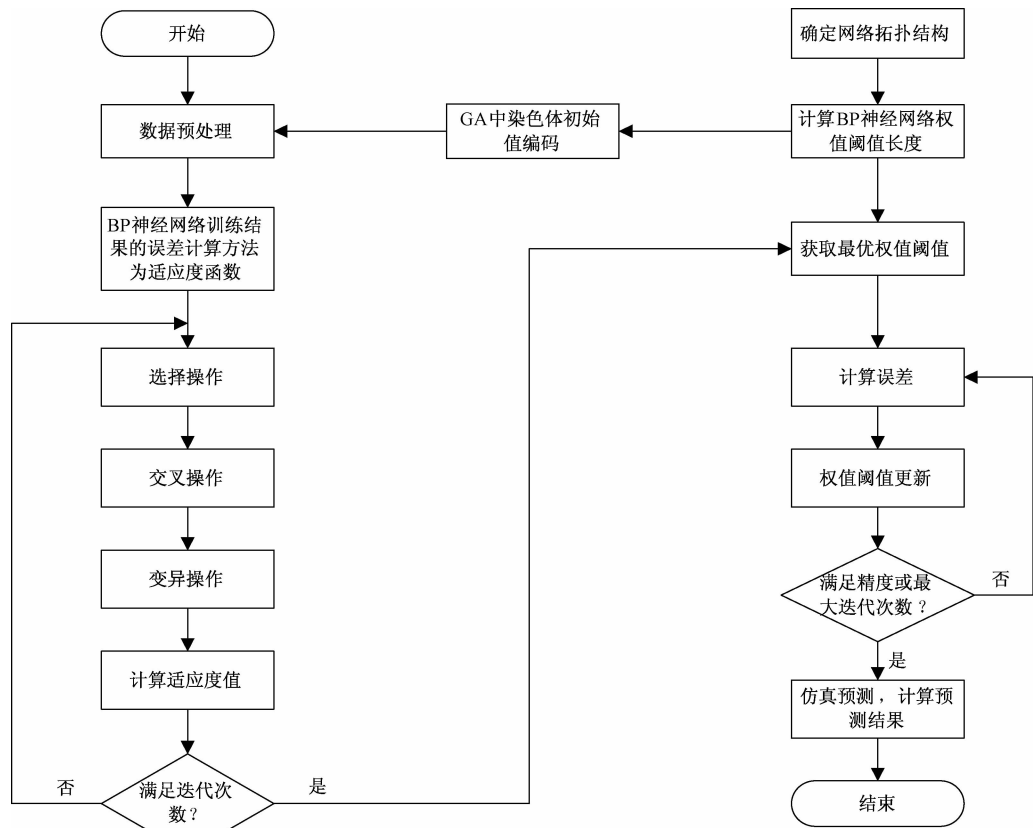


图 3 GR_GA_BP 预测流程

4 基于 GR_GA_BP 的货运量预测实例

4.1 货运量预测与评价指标

以 2000—2019 年全国的铁路货运量数据为例,分别在 BP 模型、GR_BP 模型和 GR_GA_BP 模型下对 2015—2019 年全国铁路货运量预测,将预测结果进行比较。

为了评价 GR_GA_BP 模型预测的结果,使用平均绝对百分比误差 (MAPE)、均方根误差 (RMSE) 和平均绝对误差 (MAE) 来衡量预测性能 [计算方法见式 (7)、式 (8) 和式 (9)]。其中 MAPE 用来衡量模型预测结果的可靠性, RMSE 是用来衡量观测值同真值之间的偏差, MAE 反映预测值误差的实际情况。

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{\hat{X}_i - X_i}{X_i} \right|$$

(7)

$$RMSE = \frac{1}{N} \sqrt{\sum_{i=1}^N (\hat{X}_i - X_i)^2}$$

(8)

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{X}_i - X_i|$$

(9)

式中: $\hat{X}_i$ 是预测的全国铁路货运量数据集; X_i 为实际的全国铁路货运量数据集; N 为数据集的长度。

4.2 预测结果和分析

本实验中,每个模型独立运行 10 次,记录每个模型的误差准则 (MAPE、RMSE、MAE) 的最小、最大、均值和方差值,结果见表 4~表 6。

表 4 不同预测模型之间 MAPE 的比较

模型	最小值	最大值	平均值	标准差
GR_GA_BP	1.883 *	8.172 *	5.367 *	2.283 *
GR_BP	4.342	13.500	7.648	2.811
BP	6.885	37.751	15.886	11.562

表 5 不同预测模型之间 RMSE 的比较

模型	最小值	最大值	平均值	标准差
GR_GA_BP	0.873 *	4.055 *	2.349 *	0.982
GR_BP	1.876	5.018	3.285	0.939 *
BP	3.647	18.181	7.523	5.536

表 6 不同预测模型之间 MAE 的比较

模型	最小值	最大值	平均值	标准差
GR_GA_BP	0.732 *	3.260 *	2.023 *	0.853 *
GR_BP	1.618	4.982	2.853	1.005
BP	2.816	14.851	6.239	4.608

综合表 4、表 5 和表 6,GR_GA_BP 在全国铁路货运量的预测更优,跟模型 GR_BP 和 BP 相比,GR_GA_BP模型的 MAPE、RMSE 和 MAE 最小值、最大值和平均值均优于其他两个模型。10 次预测中 MAPE、RMSE、MAE 的平均值分别为 5.367、2.349、2.023,与 GR_BP 模型相比 MAPE、RMSE 和 MAE 值分别降低了 29.83%、28.48% 和 29.08%;与 BP 模型相比 MAPE、RMSE 和 MAE 值分别降低了 66.22%、68.78% 和 67.57%。综合图 4 和图 5,比较 GR_BP 和 BP 模型发现通过灰色关联模型得到的指标跟铁路货运量的预测更具相关性可以提高 BP 模型预测的准确度,比较和 GR_BP 模型发现可以通过遗传算法优化提高预测能力,图 6 实验结果表明 GR_GA_BP 模型相比 BP 模型和 GR_BP 模型表现最好,预测铁路货运量更加可靠。

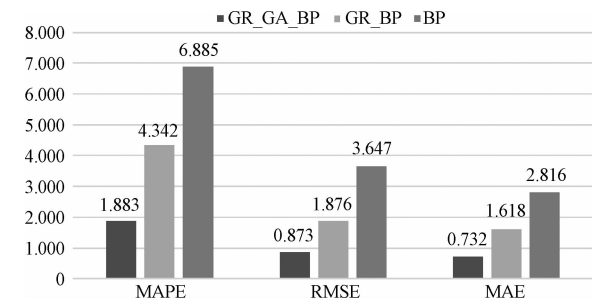


图 4 不同模型下最优预测结果的 MAPE、RMSE 和 MAE 值对比

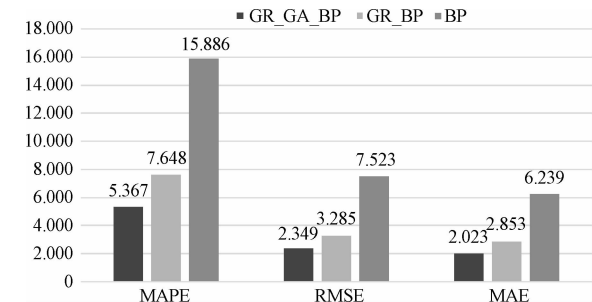


图 5 不同模型下 MAPE、RMSE 和 MAE 的平均值对比

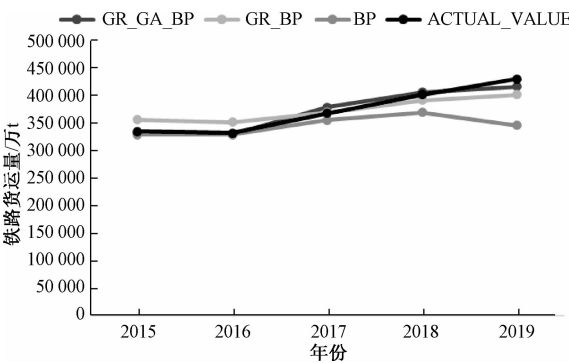


图 6 全国铁路货运量预测结果对比

5 结论

对全国铁路货运量态势的预测有利于相关资源的优化配置和改革,促进全国经济与社会的发展。本文结合仿真测试和全国铁路货运量的面板数据,构建灰色模型及其影响因素评价指标与优化方法,主要研究结论如下:

- 1)利用灰色关联模型对铁路货运量及其影响因素进行关联度分析,有效减少干扰货运量预测的其他非线性指标,降低原始 BP 模型的复杂度,提高预测精度。
- 2)结合遗传算法构建 BP 神经网络,优化 BP 神经网络的权值和阈值,提高了全局搜索能力,对货运量的预测提供了有效优化方法。
- 3)基于 GR_GA_BP 的预测模型,从各项指标(MAPE、RMSE 和 MAE)来看优于经典 BP 模型和 GR_BP 模型,可以作为货运量预测的有效方法。

参考文献

[1] 吴华稳. 铁路行业运输与安全统计指标体系构建[J]. 中国安全科学学报,2018,28(S2):124-128.

[2] 颜保凡,郭垂江,李夏苗. 铁路货运量的内部影响因素及其敏感度分析[J]. 铁道科学与工程学报,2018,15(5):1341-1346.

[3] 赵怀鑫,孙星星,徐倩倩,等. 基于灰熵法的公路货运量和货物周转量关联因素分析[J]. 交通运输工程学报,2018,18(4):160-170.

[4] 唐力,刘启钢,孙文桥. 基于灰色关联分析法的铁路物流服务质量评价[J]. 铁道运输与经济,2019,41(1):7-12.

[5] 钱隼驰,仇蕾. 灰色关联分析中分辨系数取值的定量研究[J]. 统计与决策,2019,35(10):10-14.

[6] 周程,施文,马士华. 基于改进 PSO-BP 非线性补偿的货运量“分解-集成”预测[J]. 统计与决策,2018,34(9):81-85.

[7] 王龙. 新建铁路运量预测的公路 OD 推算模型研究[J]. 铁道运输与经济,2019,41(6):81-87.

[8] 辜运燕,杨建华,王聪会,等. 航站楼延误航班旅客行为状态预测研究[J]. 系统仿真学报:1-7[2021-12-29]. DOI:

10.16182/j.issn1004731x.joss.20-0228.

[9] 刘笑佟,任爽. 基于不确定理论的铁路货运需求预测[J]. 运筹与管理,2020,29(3):135-141.

[10] 张健,孙玉莹,张新雨,等. 基于时变模型平均方法的我国航空客运量预测[J]. 系统工程理论与实践,2020,40(6):1509-1519.

Forecasting of Railway Freight Volume Based on Grey Genetic BP Neural Network

LI Jianguo¹, XIANG Wanli², WANG Jiugeng³

(1. Freight Department, China Railway Lanzhou Group Co., Ltd., Lanzhou 730000, China; 2. School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China; 3. Research Institute of Logistics and Information Technology of Gansu Province, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: Railway freight volume is limited by the influence of social construction, investment and policy regulation. In order to provide better deployment support for the freight system, based on the influence factors of railway freight volume and strengthening BP prediction performance, the gray genetic BP is proposed Neural network railway freight volume prediction model (GR_GA_BP). The grey correlation analysis of railway freight volume is used to find out the index with higher correlation with railway freight volume, which is used as GR_GA_ The input of BP prediction model predicts the freight volume. By comparing the GR_GA_BP model with other models through MAPE, RMSE and MAE, the experimental results show that the GR_GA_BP model is superior to other models and can be used as a new method to predict and study railway freight capacity.

Keywords: railway freight volume; evaluation index; grey relational analysis; genetic algorithm; BP neural network