

基于灰色关联与RBF神经网络的港口吞吐量预测

赵 晗, 孙德山

辽宁师范大学数学学院, 辽宁 大连
Email: 1094008976@qq.com

收稿日期: 2020年10月7日; 录用日期: 2020年10月19日; 发布日期: 2020年10月26日

摘 要

在中国的运输体系中, 港口扮演着重要的角色。通过建立改进的神经网络算法, 运用灰色关联分析, 从多个宏观经济指标中提取相关性较强的指标因素来确定RBF神经网络的输入指标。实证结果表明, 该模型不仅很好地反映了港口吞吐量与第一产业总产值、第二产业总产值和客运量之间的关系, 同时有效提高了预测精度, 为决策者提供决策参考。

关键词

灰色关联分析, RBF神经网络, 港口吞吐量

Port Throughput Prediction Based on Grey Correlation and RBF Neural Network

Han Zhao, Deshan Sun

School of Mathematics, Liaoning Normal University, Dalian Liaoning
Email: 1094008976@qq.com

Received: Oct. 7th, 2020; accepted: Oct. 19th, 2020; published: Oct. 26th, 2020

Abstract

In China's transportation system, ports play an important role. By establishing the improved neural network algorithm and using the grey correlation analysis, the input index of RBF neural network was determined by extracting the index factors with strong correlation from several macroeconomic indicators. The empirical results show that the model not only well reflects the rela-

relationship between port throughput, gross output value of the primary industry, gross output value of the secondary industry and passenger volume, but also effectively improves the prediction accuracy and provides reference for decision makers.

Keywords

Grey Correlation Analysis, RBF Neural Network, Port Throughput

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

港口吞吐量不仅是衡量港口发展的重要数据,同时也是腹地综合经济实力的重要指标。港口吞吐量预测是一个复杂的问题,其准确性受到不确定的社会经济因素的影响。合理的预测港口吞吐量,发现港口的发展优势以及存在的问题,可以为港口提供发展规划和合理建设的新思路。目前,国内外文献对港口吞吐量预测研究较多,主要有回归分析法、神经网络法、指数平滑法和时间序列分析法等。Liupeng Jiang 等人(2019)引入博弈论对多元贝叶斯回归模型进行了改进,精确模拟了重叠腹地的港口货流状况[1]; Intihar M 等人(2017)采用动态因子分析和自回归移动平均模型(ARIMAX)对集装箱吞吐量建立了预测模型[2]; Feng-Ming Tsai and Linda J.W. Huang (2015)利用人工神经网络同时预测亚洲主要港口之间的集装箱流量[3]; Gang Xie 等人(2013)提出了三种基于最小二乘支持向量回归模型的港口集装箱吞吐量预测混合方法[4]。以往的研究大多为根据经验选取影响指标,本文采用灰色关联分析从候选变量中提取出相关性较高的影响因素,将其作为神经网络的输入变量,建立 RBF 神经网络模型。最后,以上海港口吞吐量及其影响因素的数据为例,比较三种方法预测的实际值,验证了基于灰色关联度的 RBF 神经网络模型的性能有很大提高。

2. 灰色关联分析

灰色关联度是两个系统或两个指标之间关联性大小的度量[5]。其具体计算过程如下:

第一步:确定参考指标列和子序列。

根据评价目的选取参考指标列 $X_0 = [X_0(1), X_0(2), \dots, X_0(n)]$, 其他因素序列为子序列 $X_i = [X_i(1), X_i(2), \dots, X_i(n)], i = 1, 2, \dots, l$, 其中 n 为数列的长度, l 为收集到的指标个数。

第二步:数据无量纲化处理。

由于收集到各指标量纲不同,不便于比较分析,这里采用均值化方法,公式如下:

$$X_i = \frac{X'_i(k)}{\frac{1}{n} \sum_{k=1}^n X'_i(k)}, i = 1, 2, \dots, l \quad (1)$$

第三步:计算灰色关联系数。

$$\delta_i(k) = \frac{\min_i \min_k |X_0(k) - X_i(k)| + \rho \cdot \max_i \max_k |X_0(k) - X_i(k)|}{|X_0(k) - X_i(k)| + \rho \cdot \max_i \max_k |X_0(k) - X_i(k)|}, k = 1, 2, \dots, n \quad (2)$$

其中 ρ 为分辨系数,取值范围为(0,1)。若 ρ 越小,关联系数间差异越大,区分能力越强。

第四步:计算灰色关联度。

计算各指标与参考指标的关联系数的均值, 即关联序列。

$$r_{0i} = \frac{1}{n} \sum_{k=1}^n \delta_i(k)$$

(3)

求出灰色关联度后, 对指标进行排序, 便可直观得出哪些指标与参考数据关系密切, 一般选取关联度大于 0.8 的指标。最后, 重复以上步骤, 将提取出的影响指标依次作为参考数列, 求出各影响因素指标之间的灰色关联度, 得出灰色关联矩阵, 保留关联度较小的指标, 剔除关联度高的指标数据, 避免影响指标间的共线性[6]。

3. RBF 神经网络基本原理

径向基神经网络(RBF Neural Network)是由 J. Moody 和 C. Darken 于 20 世纪 80 年代末提出的, 是一种模拟人脑神经元细胞局部调节的网络结构。它通过特有的网络拓扑结构解决研究问题, 具有网络结构简单、训练速度快且稳定和全局逼近最优等优点。神经网络通常由三部分组成: 输入层, 非线性传递的隐藏神经层和输出层[7]。

RBF 神经网络模型的隐含层常采用高斯函数:

$$R_i(x) = \exp\left(-\frac{\|x - c_i\|^2}{2\sigma_i^2}\right), i = 1, 2, \dots, n$$

(4)

RBF 神经网络输出公式为:

$$g(x) = \sum_{i=1}^n w_i R_i(\|x - c_i\|), i = 1, 2, \dots, n$$

(5)

其中, c_i 为径向基函数的中心, $\|x - c_i\|$ 代表输入变量 x 和中心点 c_i 之间的距离; w_i 代表网络连接权重; σ_i 表示基函数围绕中心点 c_i 的神经元宽度; i 表示隐含层结点数个数; $g(x)$ 表示实际输出。

4. 实证分析

4.1. 数据选取与说明

选取上海市 1969~2018 年的年度港口货物吞吐量作为参考指标列。由于港口吞吐量的影响因素广泛且复杂, 综合相关文献, 初步选取 10 项影响指标, 见表 1。即选取 1969~2018 年港口货物吞吐量影响因素的年度数据作为灰色关联分析的子序列。数据均来源于《中国统计年鉴》。

Table 1. Factors influencing port cargo throughput
表 1. 港口货物吞吐量影响因素

影响因素	指标
经济发展的整体水平	GDP (X_1)
经济发展潜力	固定资产投资总额(X_2)
经济结构	第一、二、三产业总产值(X_3, X_4, X_5)
城市需求能力	居民消费水平(X_6)
经济发展活力	社会消费品零售总额(X_7)
经济外贸水平	外贸进出口总额(X_8)
货运需求	货运量(X_9)
客运需求	客运量(X_{10})

4.2. 数据处理与模型建立

4.2.1. 数据的处理与分析

本文使用的软件为 Matlab, 首先对数据建立灰色关联分析模型。由于这些影响因素之间的相关性大小不一, 为选取具有代表性且相关性较小的影响指标, 首先求解出以上指标与港口货物吞吐量的灰色关联度。为消除各原始数据量纲差异的影响, 将参考指标列数据和子序列数据进行均值化处理, 如式(1)。再根据式(2)、(3)计算出灰色关联系数和灰色关联度。其中, 分辨系数的选取并不改变相关性的强弱, 这里选取 $\rho = 0.4$ 。具体计算结果及排序见表 2。

Table 2. Grey correlation degree and sorting
表 2. 灰色关联度及排序

指标	灰色关联度
客运量(X_{10})	0.8827
货运量(X_9)	0.8817
第一产业总产值(X_3)	0.8656
第二产业总产值(X_4)	0.8346
固定资产投资总额(X_2)	0.8122
GDP (X_1)	0.7837
社会消费品零售总额(X_7)	0.7690
第三产业总产值(X_5)	0.7490
外贸进出口总额(X_8)	0.7333
居民消费水平(X_6)	0.6368

选取与上海市港口货物吞吐量相关度在 0.8 之上的指标, 有客运量、货运量、第一产业总产值、第二产业总产值和固定资产投资总额, 且关联度 $r_{10} > r_9 > r_3 > r_4 > r_2$ 。其余指标在 0.6 至 0.8 之间, 说明这些指标对港口货物吞吐量有一定的影响, 但不是主要影响, 因此在预测港口货物吞吐量时可以忽略这些指标。为了避免以上指标对港口货物吞吐量影响的重叠性, 消除指标间的相互共线性, 求出这 5 个指标的相关系数矩阵, 见表 3。

Table 3. Correlation coefficient matrix
表 3. 相关系数矩阵

	X_3	X_4	X_2	X_{10}	X_9
X_3	1.0000	0.7329	0.7502	0.7313	0.7944
X_4	0.7465	1.0000	0.9020	0.6320	0.6615
X_2	0.7502	0.8951	1.0000	0.6451	0.6451
X_{10}	0.8053	0.7071	0.7306	1.0000	0.8216
X_9	0.8427	0.7138	0.7153	0.8073	1.0000

从表 3 中数据可知, X_2 与 X_4 相关度在 0.89 之上, 有较强的相关性, 保留任意一个即可, 且由 $r_4 > r_2$, 去掉相关性较小的指标 X_2 , 保留 X_4 ; 其次在指标 X_9 与 X_{10} 中关联度也超过 0.8, 由 $r_{10} > r_9$, 故剔除 X_9 ; X_2 、 X_4 、 X_9 、 X_{10} 与 X_3 的关联度在 0.8 以下, 说明 X_3 对其他指标具有弱相关性, 保留 X_3 。综上分析, 最后确定 X_3 、 X_4 与 X_{10} 作为模型训练指标。

4.2.2. 模型确立与评估

根据灰色关联分析结果,选取三个主要影响因素,建立 RBF 神经网络模型并进行预测。将 1969~2018 年的第一产业总产值(X_3)、第二产业总产值(X_4)和客运量(X_{10})作为 RBF 神经网络的输入端,1969~2018 年的港口货物吞吐量为输出端。本文共选取 50 组数据,其中,1969~2015 年 47 组数据作为训练样本进行模型的建立,2016~2018 年 3 组数据作为测试样本,验证模型的可行性。由于数据间量纲的影响,以及考虑到 RBF 神经网络的收敛速度,将样本数据归一化,公式为 $\hat{x} = \frac{x - x_{\max}}{x_{\max} - x_{\min}}$ 。其中, $x_{\max}$ 为 x 的最大值, $x_{\min}$ 为 x 的最小值。

由前述可知,本文建立的模型中因变量和自变量的数目分别为 1 和 3,即模型的输入端有 3 个节点,输出端为 1 个节点。RBF 神经网络的扩散因子 spread 也会影响模型的精度,spread 值越大,函数拟合效果越平滑。但若 spread 值过大,模型的逼近误差也会变大,若 spread 值过小,会出现过拟合现象,都会影响模型的性能。本文采用网格法来确定 spread 和隐含层节点 n 的最优值。参数优化方法如下[7]:

- 1) 根据经验公式, $n = \sqrt{m+1} + \alpha$, α 取 1 至 10 之间整数。由此初步确定 n 的取值范围为 3 至 12 之间。
- 2) spread 选定默认值 1.0, 动态改变隐含层节点 n , 以预测误差最小为目标找到隐含层节点 n 优选区间为 [10,11], 见图 1(左)。
- 3) 取隐含层节点 $n = 10$, 动态改变 spread, 以预测误差最小为目标找到 spread 的优选区间 [1.8,1.9], 见图 1(右)。

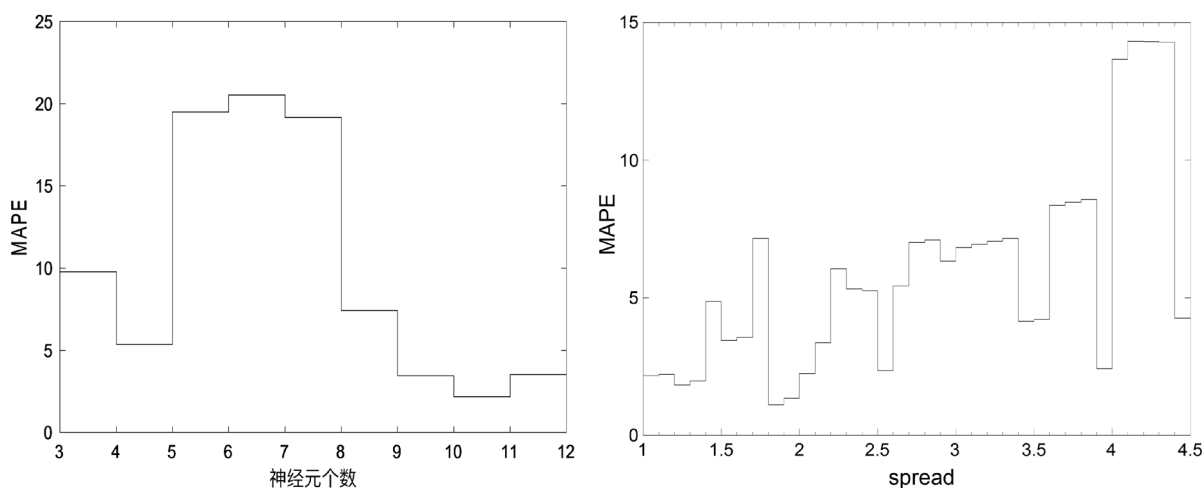


Figure 1. Search the optimal interval for the spread and hidden layer nodes

图 1. 搜索 spread 和隐含层节点的最优区间

4) 搜索全局最优区间。为避免搜索过程存在遗漏,适当的减少搜索步长并外延优选区间。最终得到 RBF 神经网络最优参数为 $\text{spread} = 1.8$, $n = 10$ 。此时模型收敛结果稳定且预测误差最小,模型拟合精度达到 99%。

为检测基于灰色关联分析的 RBF 神经网络模型的可靠性,对相同数据建立递归神经网络模型和 BP 神经网络模型,具体结果见表 4。使用 MAPE 来对比四种神经网络预测的平均绝对百分比误差。公式为:

$$\text{MAPE} = \frac{\sum \frac{|y_t - y'_t|}{y_t}}{n} \times 100 \quad (6)$$

其中, y_t 为实际值, y'_t 为预测值, n 为样本数量。

Table 4. Comparison of prediction results and errors
表 4. 预测结果及误差比较

年份	实际值	RBF 神经网络		灰色关联 - RBF 神经网络		灰色关联 - 递归神经网络		灰色关联 - BP 神经网络	
		预测值	误差	预测值	误差	预测值	误差	预测值	误差
2016	64482	66416	2.99%	66329	2.86%	65721	1.8%	63816	1.0%
2017	70542	66509	5.71%	70671	0.18%	69037	2.3%	67504	4.3%
2018	68392	65673	3.97%	68211	0.26%	70545	3.8%	67810	0.8%
MAPE		4.23%		1.10%		2.4%		2.0%	

由表 4 可知, 运用灰色关联分析优化后的模型相比直接使用神经网络模型, 减少了相关性较小的输入变量, 模型计算时间明显减少, 预测误差也显著下降。基于灰色关联分析的 BP 神经网络单年误差预测结果不稳定, 最大误差达到 4.3%, 最小 0.8%。基于灰色关联分析的递归神经网络误差预测结果较稳定, 但整体误差略大。基于灰色关联分析的 RBF 神经网络预测单年误差最小可达到 0.18%, 平均绝对百分比误差达到 1.1%, 预测结果较稳定。通过上述实证分析, 基于灰色关联分析的 RBF 神经网络适合具有非线性指标的港口货物吞吐量预测建模。

5. 总结

本文从影响港口货物吞吐量的多个指标来分析和预测港口货物吞吐量, 将灰色关联分析与 RBF 神经网络模型结合用于港口吞吐量预测。通过比较几种常用数学模型的结果, 该模型的预测精度良好, 具有一定的实用性和可操作性, 可应用于港口建设项目, 通过调整港口货物吞吐量的影响因素, 提供重要的港口规划和建设。同时, 该模型具有较强的泛化能力, 对文献[8]中算例, 应用此方法建模结果均有提升。因此, 对于具有多种复杂影响因素的问题预测时, 该模型是可行且有效的, 但在神经网络参数优化方法上还有待进一步改进。

基金项目

辽宁省自然科学基金指导计划项目(编号: 2019-ZD-0471)。

参考文献

[1] Jiang, L., Wang, J., Jiang, H., *et al.* (2019) Prediction Model of Port Throughput Based on Game Theory and Multimedia Bayesian Regression. *Multimedia Tools and Applications*, **78**, 4397-4416. <https://doi.org/10.1007/s11042-018-5766-2>

[2] Intihar, M., Kramberger, T. and Dragan, D. (2019) Container Throughput Forecasting Using Dynamic Factor Analysis and ARIMAX Model. *Promet-Traffic & Transportation*, **29**, 529-542. <https://doi.org/10.7307/ptt.v29i5.2334>

[3] Tsai, F. and Huang, L. (2017) Using Artificial Neural Networks to Predict Container Flows between the Major Ports of Asia. *International Journal of Production Research*, **55**, 5001-5010. <https://doi.org/10.1080/00207543.2015.1112046>

[4] Xie, G., Wang, S., Zhao, Y. and Lai, K.K. (2013) Hybrid Approaches Based on LSSVR Model for Container Throughput Forecasting: A Comparative Study. *Applied Soft Computing*, **13**, 2232-2241. <https://doi.org/10.1016/j.asoc.2013.02.002>

[5] 张军芳, 张利民, 安达. 基于灰色关联度和 BP 神经网络的全国人口预测研究[J]. *农村经济与科技*, 2019, 30(22): 257-258.

[6] 刘枚莲, 朱美华, 黄键. 港口吞吐量预测影响因素筛选方法研究[J]. *水运工程*, 2011(3): 76-80.

[7] 池可. 服装销售预测方法及 RBF 神经网络模型研究[D]: [硕士学位论文]. 苏州: 苏州大学, 2009.

[8] 魏爽. RBF 神经网络在气象时间序列预测中的应用[D]: [硕士学位论文]. 南京: 南京信息工程大学, 2012.