

Forecasting Domestic Tourism Revenue in Nanning Based on a Grey-Neural Network Combined Model

Qiubo Liu Min Zhao Yan Liu*

School of Artificial Intelligence and Software, Nanning University, Nanning, Guangxi, 530200, China

Abstract

With the continuous expansion of the scale of China's tourism industry, domestic tourism revenue has shown a significant growth trend. Against this backdrop, exploring the influencing mechanism and evolution law of domestic tourism revenue has important theoretical value and practical significance. This paper takes Nanning City as the research object, selects data from 2006 to 2019, and builds a grey-neural network combined prediction model based on the R statistical software platform. The research aims to systematically identify the key influencing factors of domestic tourism revenue in Nanning City, reveal its internal mechanism of action, and on this basis, conduct scientific predictions, thereby providing empirical evidence and decision-making references for local governments to formulate tourism industry development policies.

Keywords

Domestic tourism revenue; Grey prediction; Neural network; Combination model

基于灰色-神经网络组合模型的南宁市国内旅游收入预测

刘秋波 赵敏 刘雁*

南宁学院人工智能与软件学院, 中国·广西 南宁 530200

摘要

随着我国旅游产业规模持续扩大, 国内旅游收入呈现显著增长态势。在此背景下, 探究国内旅游收入的影响机制及其演变规律具有重要的理论价值和现实意义。本文以南宁市为研究对象, 选取2006—2019年数据, 基于R统计软件平台构建灰色-神经网络组合预测模型, 研究旨在系统识别南宁市国内旅游收入的关键影响因素, 揭示其内在作用机制, 并在此基础上进行科学预测, 从而为地方政府制定旅游产业发展政策提供实证依据和决策参考。

关键词

国内旅游收入; 灰色预测; 神经网络; 组合模型

1 引言

随着国民经济的快速增长, 旅游业已经成为规模最大、发展速度最快的产业之一。近年来, 南宁市文化和旅游产业快速发展, 产业发展水平全区领先。据《2022年南宁市旅游统计调查分析报告》, 2022年全市共接待国内游客11626.75万人次, 实现国内旅游收入1278.59亿元, 旅游人数、旅游收入居全区首位, 国内旅游收入占全市国内生产总值

(Gross Domestic Product, GDP)的24.5%。

目前, 学术界关于国内旅游收入的研究已经有不少成果。许芳惠^[1]采用逐步回归、主成分回归和岭回归方法对比分析湖南省长沙市的国内旅游收入影响最大的因素为住宿餐饮业法人单位数和居民旅游消费价格指数(CPI)。吴顺^[2]利用2000—2014年内蒙古旅游数据, 采用计量经济学的多元线性回归分析, 发现国内旅游人数、星级宾馆个数以及城镇居民家庭人均可支配收入对内蒙古国内旅游业有影响。李姣^[3]的研究结果表明国内旅游收入的主要因素是国内旅客人数和旅行社数量。吕孝敬和许菁^[4]采用探索性数据分析方法得出海南省居民消费价格指数对海南省的国内旅游收入具有明显的影响。刘志茜等^[5]利用1994—2016年我国国内旅游收入数据与5个相关变量分别建立逐步回归和灰色预测模型, 并对未来5年的国内旅游收入进行预测。俞彤和叶月李^[6]运用主成分分析和回归分析对广东省2005—2014年国内旅游收入影响因素进行了分析。张翊^[7]运用灰色模型对山东省未来旅游总收入进行预测, 并运用灰色关联

【基金项目】南宁学院校级科研项目(项目编号: 2023XJ24); 南宁学院首批虚拟教研室建设试点项目(项目编号: 2023XNJYS08)。

【作者简介】刘秋波(2003-), 女, 中国广西灵山人, 在读本科生, 从事智能科学与技术研究。

【通讯作者】刘雁(1995-), 女, 中国广西灵山人, 硕士, 讲师, 从事应用统计、经济社会统计研究。

分析和主成分分析定量分析影响山东省旅游业的主要因素。He and Jie^[8]利用粗糙集理论方法,以1994—2008年样本为基础,选取解释变量建立了国内旅游收入计量模型。

上述对国内旅游收入的影响分析及预测主要集中在国内学者,大多基于计量经济学模型,或采用回归模型,侧重点各有不同,灰色-神经网络组合模型在国内旅游收入预测的研究较少。灰色系统理论是一种在数据有限或信息不完整的系统下进行建模、分析和预测的数学理论^[9]。灰色系统理论包括一系列模型,其中GM(1,1)是最常用的灰色预测模型。GM(1,1)模型使用不完全的或者少量的信息预测未来事物的发展规律,是一种适合只有一个变量的一阶预测方法,因其无须计算均值、标准差等特征值,预测精度较高,常用于短期和中长期预测。神经网络的优点在于可以处理非线性问题,能够适应不同的数据类型和模型复杂度,同时具有较好的泛化能力^[10]。为了结合两者优势,研究在单一GM(1,1)模型基础上构建灰色-神经网络组合模型进行预测。

本文以南宁市2006—2019年的国内旅游收入数据为例,选取了第三产业增加值、客运量、国内旅游人数、公路里程、私人汽车拥有量、城镇居民家庭人均可支配收入等可能影响国内旅游收入的6个指标,采用R语言统计分析平台,基于自适应最小绝对收缩和选择算子(Lasso)方法完成变量筛选后,针对小样本数据特征构建混合预测模型体系。即运用灰色系统理论中的GM(1,1)模型对筛选所得关键变量进行时序外推预测,该模型通过累加生成算子和微分方程构建,可有效解决小样本数据建模问题,获得2020—2022年期间各变量的预测值序列。基于人工神经网络(ANN)构建非线性回归模型,利用其非线性映射能力及鲁棒性特征,对历史观测数据进行网络训练与参数优化。最终将灰色预测结果输入经训练收敛的神经网络架构中,实现多方法融合的智能预测。

2 数据来源与相关分析

2.1 数据来源与指标选取

本文数据来源于广西统计年鉴、南宁统计年鉴和南宁市统计局官方网站,通过整理而得。考虑到数据的真实性以及可获得性,选取2006年至2019年南宁市国内旅游收入(y)为研究对象,并根据定性分析选取第三产业增加值(x_1)等6个自变量。变量说明见表1。

表1 变量指标说明

变量名	指标说明	单位
y	国内旅游收入	(亿元)
x_1	第三产业增加值	(亿元)
x_2	客运量	(万人次)
x_3	国内旅游人数	(万人次)
x_4	公路里程	(公里)
x_5	私人汽车拥有量	(辆)
x_6	城镇居民家庭人均可支配收入	(元)

2.2 相关分析

相关系数用来描述定量变量之间的关系,初步判断因变量与自变量之间是否具有线性关系。表2是各变量间的Pearson相关系数矩阵。

表2 变量 Pearson 相关系数矩阵

	y	x_1	x_2	x_3	x_4	x_5	x_6
y	1.000	0.983	-0.397	0.996	0.840	0.992	0.946
x_1	0.983	1.000	-0.350	0.991	0.912	0.992	0.980
x_2	-0.397	-0.350	1.000	-0.392	-0.206	-0.381	-0.345
x_3	0.996	0.991	-0.392	1.000	0.878	0.997	0.968
x_4	0.840	0.912	-0.206	0.878	1.000	0.894	0.963
x_5	0.992	0.992	-0.381	0.997	0.894	1.000	0.975
x_6	0.946	0.980	-0.345	0.968	0.963	0.975	1.000

由表2可知,客运量(x_2)与国内旅游收入(y)的线性关系不显著,且呈负相关。其余变量均与国内旅游收入呈高度的正相关关系。同时,各自变量间存在严重的多重共线性,需要对变量进行进一步筛选。图1展示了各变量间相关性的强弱程度,颜色越深则相关性越强。

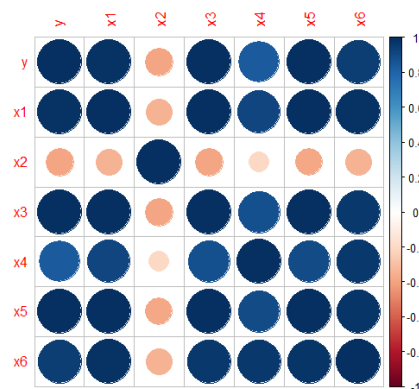


图1 相关性分析

3. 建模过程

3.1 Lasso 变量选择

在R studio统计软件中加载lars包,编制相应的程序后运行得到如下结果,如表3所示。

表3 自适应Lasso变量选择模型系数表

x_1	x_2	x_3	x_4	x_5	x_6
0.14881	0.00000	0.12529	-0.03631	0.00000	-0.01156

由表3可知,客运量(x_2)和私人汽车拥有量(x_5)的系数为0,即这几个变量在Lasso变量选择过程中被剔除了。这是因为私人汽车拥有量(x_5)与第三产业增加值(x_1)存在明显的共线性,Lasso回归方法在变量选择过程中剔除了变量;而客运量(x_2)与国内旅游收入(y)的相关性太小,以致可以忽略。

综上,通过Lasso回归方法识别出影响国内旅游收入(y)的关键因素为:第三产业增加值(x_1)、国内旅游人数(x_3)、

公路里程 (x_4) 和城镇居民家庭人均可支配收入 (x_6)。

3.2 模型构建

对 Lasso 变量选择方法筛选出的因素建立灰色 - 神经网络组合预测模型, 其参数设置为误差精度 10^{-7} , 学习次数 10000 次。第三产业增加值 (x_1)、国内旅游人数 (x_3)、公路里程 (x_4) 和城镇居民家庭人均可支配收入 (x_6) 的 2020 年、2021 年及 2022 年数值通过 R 语言建立 GM(1,1) 灰色预测模型得出, 预测精度等级如表 4 所示。

表 4 灰色预测模型相关因素精度表

变量	x_1	x_3	x_4	x_6
平均相对误差	6.54%	2.44%	2.77%	5.20%
相对精度	93.46%	97.56%	97.23%	94.80%
后验差比值 (C 值)	0.058 (< 0.35)	0.021	0.155	0.088
预测精度等级	好	好	好	好

表 5 南宁市国内旅游收入真实值与预测值对比

年份	真实值	预测值	误差值	相对误差	年份	真实值	预测值	误差值	相对误差
2006	97.5476	97.634	-0.087	-0.09%	2015	729.93	732.084	-2.154	-0.30%
2007	114.2201	115.376	-1.156	-1.01%	2016	903.24	902.175	1.065	0.12%
2008	141.21	139.097	2.113	1.50%	2017	1109.8	1110.387	-0.587	-0.05%
2009	179	181.300	-2.300	-1.29%	2018	1368.42	1367.549	0.871	0.06%
2010	234.78	233.615	1.165	0.50%	2019	1699.02	1699.831	-0.811	-0.05%
2011	307.05	306.909	0.141	0.05%	2020	1215.49	1088.477	127.013	10.45%
2012	397.13	397.066	0.064	0.02%	2021	1529.14	1352.283	176.857	11.57%
2013	469.64	470.379	-0.739	-0.16%	2022	1278.59	1699.831	-421.241	-32.95%
2014	598.73	596.805	1.925	0.32%					

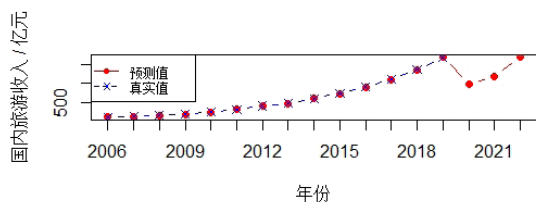


图 2 南宁市国内旅游收入真实值与预测值对比图

4 结论

本文在对相关研究进行梳理和总结的基础上, 构建出灰色 - 神经网络组合模型的方法, 对 2006 至 2019 年南宁市国内旅游收入进行分析, 利用 Lasso 变量选择方法剔除具有共线性和相关程度较低的变量, 并通过灰色预测模型对自变量进行预测从而确定神经元个数, 最后通过神经网络模型对国内旅游收入进行预测, 通过将 2006—2019 年的真实值和预测值进行比较, 相对误差均在 1% 以内, 预测效果较好, 证明 GM(1,1)-BP 神经网络组合模型能够用于对国内旅游收入进行预测, 并为相关部门提供较有意义的参考依据。

参考文献

[1] 许芳惠. 网红经济下湖南省长沙市国内旅游收入影响因素分析

由表 4 可知, x_1 、 x_3 、 x_4 、 x_6 这 4 个变量的预测精度等级都是“好”可以用于构建神经网络预测模型, 即神经元 (变量) 个数为 4 个。

为避免量纲的影响, 将灰色预测得到的数据及历史数据进行标准化, 代入国内旅游收入所建立的神经网络预测模型, 得到 2020 至 2022 年国内旅游收入预测值。相关数据如表 5 所示。

由表 5 可知, 模型得到的 2020 年—2022 年的南宁市国内旅游收入预测值。南宁市国内旅游收入真实值和预测值在 2006 年至 2019 年的相对误差绝对值均在 5% 以内, 说明建立的神经网络模型拟合效果较好。

为了更直观地观察南宁市国内旅游收入真实值与预测值的变化趋势, 绘制预测对比图如图 2 所示。南宁市国内旅游收入预测值平稳上升, 2020 年出现明显下降, 随后显著回升。2022 年真实值略有下降, 与预测值有较大偏差, 可能与疫情反复以及经济环境有关。

[J].西部旅游,2023(03):68-70.

[2] 吴顺. 内蒙古旅游业发展影响因素分析[J]. 环渤海经济瞭望, 2018 (08):90-91.

[3] 李姣. 我国国内旅游收入影响因素的实证分析[J]. 重庆工商大学学报(自然科学版), 2019, 36(05):105-114.

[4] 吕孝敬, 许菁. 海南省居民消费价格指数对国内旅游收入的影响——基于探索性数据分析方法[J]. 商业经济, 2018(11):62-63.

[5] 刘志茜, 刘莹, 丁瑾华, 孙翔, 朱家明. 基于多元回归和 GM(1,1) 的国内旅游发展加权组合预测[J]. 喀什大学学报, 2018, 39(03):17-24.

[6] 俞彤, 叶月李. 广东省国内旅游收入影响因素实证研究[J]. 武汉商学院学报, 2016, 30(04):19-22.

[7] 张翊. 基于灰色系统理论的山东省旅游影响因素探究[J]. 旅游纵览, 2023(15):31-33+70.

[8] He Xiaoya, Jie Zhiben. Research on econometric model for domestic tourism income based on rough set[J]. Communications in Computer and Information Science, 2011, 232(2):259-266.

[9] 朱强. 基于灰色系统理论的经济建模方法[J]. 统计与决策, 2010 (11):27-28.

[10] 张良均, 云伟标, 王路, 刘晓勇. R 语言数据分析与挖掘实战[M]. 北京:机械工业出版社, 2015.9.