

文章编号: 2095-2163(2022)08-0132-04

中图分类号: TP311

文献标志码: A

基于 GA-BP 神经网络的旅游客流预测应用研究

林思涛, 林 钊, 陈家琦, 吴凤舞, 张学健

(福州外语外贸学院 大数据学院, 福州 350202)

摘要: 随着人们日益增长的物质文化需要, 旅游业蓬勃发展。各大旅游景点人满为患的现象也随之出现, 为有关部门的管理带来了极大的困扰。在大数据技术正日趋普及的背景下, 产生了大量与旅游相关的数据。文章尝试优化 BP 神经网络模型, 预测一段时间内的客流量变化, 为有关部门制定管理策略和推出应急措施提供参考。经实证结果分析, 利用遗传算法优化 BP 神经网络, 可对客流量进行更精准的预测。

关键词: 智慧旅游; 旅游需求管理; 百度搜索指数; 遗传算法; BP 神经网络

Research on application of tourist flow forecasting based on GA-BP neural network

LIN Sitao, LIN Zhao, CHEN Jiaqi, WU Fengwu, ZHANG Xuejian

(Big Data Institute, Fuzhou University of International Studies and Trade, Fuzhou 350202, China)

[Abstract] With the growing cultural and material needs of the people, tourism is booming. The phenomenon of overcrowding in major tourist attractions has also appeared, which has caused great trouble to the management of relevant departments. With the increasing popularity of big data technology, a large number of tourism related data have been generated. This paper attempts to optimize the BP neural network model to predict the change of passengers flow in a period of time, so as to provide reference for the management department to formulate management strategies and emergency measures. Through the analysis of empirical results, using genetic algorithm to optimize BP neural network can predict the passengers flow more accurately.

[Key words] smart tourism; tourism demand management; Baidu search index; genetic algorithm; BP neural network

0 引言

随着国内经济飞快发展和人民生活水平的提高, 人们对物质文化的需求也在不断增长。愈来愈多地, 人们在闲暇时间选择外出旅游, 接触新鲜事物、提高生活品质, 旅游业因此蓬勃发展。与此同时, 客流量的持续攀升也造成了交通堵塞、景区拥挤等问题, 而传统旅游管理方式已无法应对这一现状, 给城市和景区的管理带来了极大的困扰。若能够及时预测未来一段时间内的游客量变化, 提供更准确的旅游景点客流量预测, 相关部门就可以提前制定安全防范措施和客流引导, 避免拥堵事件的发生, 更加合理地利用旅游资源。

传统的旅游需求预测模型, 主要是针对定量与定性方面的研究, 其共同特点是先建立时序数据的主观模型, 再进行预测。因此, 传统模型缺乏对数据样本的学习过程, 不具有自学习和泛化能力, 对客

流量的预测误差较大。随着人工神经网络(ANN)的快速发展, 越来越多的学者开始将 BP 神经网络(BPNN)^[1]及支持向量回归(SVR)^[2]等算法应用于旅游需求预测领域。

1 方法简介

1.1 网络搜索指数

如今, 大数据时代的到来, 人们可以通过互联网了解更多的旅游相关知识, 搜索引擎中对旅游业的优化也在逐渐完善。针对网络搜索数据在生活中的应用, 尤其是在经济、社会管理领域以及市场营销预测中, 网络搜索数据的应用越发广泛。2013年, H7N9 禽流感的大爆发, 对国内的家禽养殖经济造成了巨大的经济影响。谭小林等人^[3]利用网络搜索指数, 对 H7N9 爆发趋势进行相关性分析, 利用对网络搜索关键词的跟踪, 可以预测 H7N9 禽流感爆发的进程。袁恒^[4]研究了利用网络搜索指数对市场的预测, 探讨了

基金项目: 国家级大学生创新创业训练计划项目(202113762015)。

作者简介: 林思涛(2001-), 男, 本科生, 主要研究方向: 数据挖掘、人工智能; 林 钊(2000-), 男, 本科生, 主要研究方向: 数据挖掘; 陈家琦(2001-), 男, 本科生, 主要研究方向: 数据挖掘; 吴凤舞(2000-), 女, 本科生, 主要研究方向: 数据挖掘; 张学健(2000-), 男, 本科生, 主要研究方向: 数据挖掘。

收稿日期: 2022-02-17

国内网络搜索数据的市场预测价值。迟依涵^[5]研究了基于网络搜索数据的房地产价格预测。赖凯声等人^[6]研究了网络搜索数据与社会心理学相关性;马碧云^[7]基于网络搜索数据对旅游客流量进行了非线性预测的研究。戚明远^[8]利用网络搜索数据对商品住宅市场进行相关性的研究。王炼等人^[9]从中国电影市场来研究网络搜索数据对票房量的预测。白笑笑^[10]利用网络搜索数据,可以预测冰箱的销量并对冰箱的需求进行了分析。

1.2 GA 算法

遗传算法(Genetic Algorithm, GA)是一种近似优化的算法,起源于对生物体系中的天然演化,通过电子计算机的模拟展开研究。

在 GA 算法中,每一种染色体组型都对应了算法的一种解决对策。通常情形下,可以使用适应度函数(fitness function)判断解决对策的优劣。因此,从每种染色体组型到求解结果的适应度就构成了一种映射。可以将 GA 算法求解的整个过程看作在多元函数中寻求最佳解的过程。

可如此设想,在一个多维曲面中有众多的“顶峰”,每个顶峰对应的便是局部最优解。而其中会有某个“顶峰”的海拔最高,则这个“顶峰”就是全局最优解。那么,GA 算法的主要任务便是尽可能爬到最高的顶峰,而不是限制于某个小山峰。

1.3 BP 神经网络

BP(Back Propagation)神经网络,是一种多层前馈型的神经网络,主要包含输入层、隐含层(中间层)和输出层。BP 神经网络层与层之间采用全互连方式,由数量众多的单神经元通过可调的连接权值实现全连接。中间层(隐含层)可以有一层或多层,而同一层间的各神经元却不能相互连接。

BP 神经网络的传播方式分为前向传播和反向传播,其信号分为函数信号和误差信号。在前向传播过程中,函数信号从输入层经过中间层(隐含层)流向输出层,成为一个输出信号。若输出情况与期望的情况相差过大,则进入反向传播。此时,在网络的一个输出神经元中产生误差信号,一层接一层地反向传播调整权值和阈值,直到误差都在允许范围为止。BP 神经网络的典型拓扑结构如图 1 所示。

BP 神经网络可用于分类、聚类、预测等。目前,大部分的神经网络都是以 BP 神经网络为基本架构,并针对现实应用续以后期的优化完善而得。因此 BP 神经网络已经成为目前使用得最广泛的神经网络模型之一。

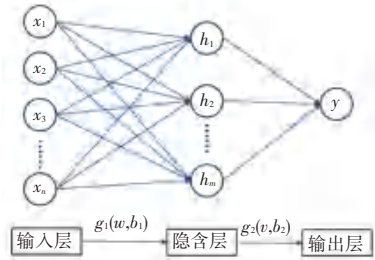


图 1 BP 神经网络拓扑结构

Fig. 1 BP neural network topology

1.4 GA-BP 神经网络预测模型

基于 GA 算法优化的 BP 神经网络模型,是由 BP 神经网络系统架构的确定、GA 算法优化以及 BP 神经网络预测三部分组成。其中,BP 神经网络系统架构的确定,主要是以拟合函数中输入、输出的参数个数来确定其模型架构。根据模型的参数个数,可以判断出 GA 算法进一步优化的参数个数,从而判定出 GA 算法中个体的编码长度,由 GA 算法优化的参数就是 BP 神经网络的初始阈值和权值。因此,如果已确定 BP 神经网络模型的系统结构,就可知阈值和权值的个数。群体中所有个体均包括了某个网络系统中的所有权值和阈值,通过统计适应度函数确定了个体适应度值,用 GA 算法通过选择、变异和交叉训练,寻找出适应度值最高的个体。BP 神经网络模型采用 GA 算法获得的最佳个体,对网络系统完成初始权值和阈值的赋值,网络系统经训练后将输出预测样本。GA 算法优化 BP 神经网络模型权值和阈值的流程如图 2 所示。

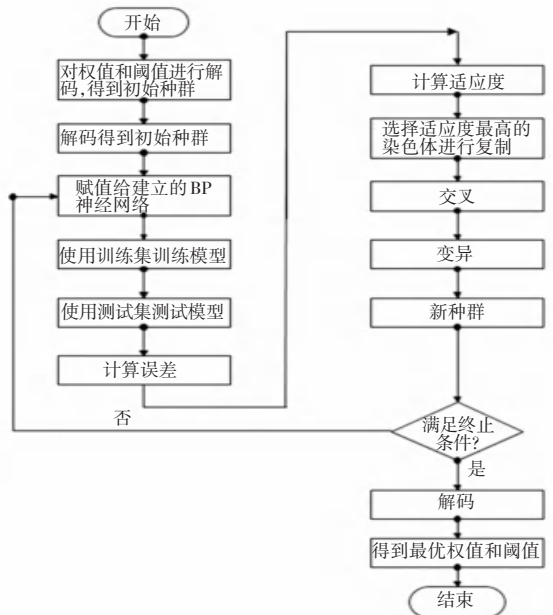


图 2 GA-BP 神经网络流程图

Fig. 2 GA-BP neural network flow chart

2 模型验证

2.1 关键词遴选

首先,本文从“食、住、行、娱、游、购”六个方面选定关键词。通过遴选,选定“厦门美食”、“厦门酒店”、“厦门交通”、“厦门景点”、“厦门地图”、“厦门特产”等基准关键词。其次,在百度搜索指数官网对基准关键词进行查询,同时找出其它相关度较高的关键词。最后,根据皮尔逊相关系数(Pearson)、斯皮尔曼相关系数(Spearman)以及显著性检验 p 值,计算得到的关键词的百度搜索指数与厦门旅游客流量的相关度强弱,选定了“厦门美食”、“厦门大学”、“沙坡尾”、“双子塔”、“厦门岛内”、“演武大桥”6个关键词,见表1。

表2 变量平稳性检验

Tab. 2 Variables stationarity test

变量	差分分数	ADF 检验值	1%临界值	5%临界值	10%临界值	结论
厦门客流量	1	-3.153 71	-3.632 74	-2.948 51	-2.613 02	平稳
厦门岛内	1	-5.066 34	-3.669 92	-2.964 07	-2.621 17	平稳
厦门大学	1	-3.340 09	-3.632 74	-2.948 51	-2.613 02	平稳
沙坡尾	1	-3.845 65	-3.632 74	-2.948 51	-2.613 02	平稳
双子塔	1	-3.175 64	-3.632 74	-2.948 51	-2.613 02	平稳
演武大桥	1	-4.540 92	-3.669 92	-2.964 07	-2.621 17	平稳
厦门美食	1	-3.221 37	-3.632 74	-2.948 51	-2.613 02	平稳

2.2 训练细节

在本文构建的模型中,遗传算法初始种群个数为30,进化代数为50,交叉概率为0.8,变异概率为0.2。BP神经网络部分,由一个输入层、一个隐藏层以及一个输出层构成。其中,隐藏层的节点经过程序计算得出最佳隐含层节点个数为4。模型的学习率为0.01,训练步长为25,学习目标为0.000 1。GA-BP神经网络训练状态如图3所示。由图3可知,该模型经过18次迭代,达到设定学习目标。

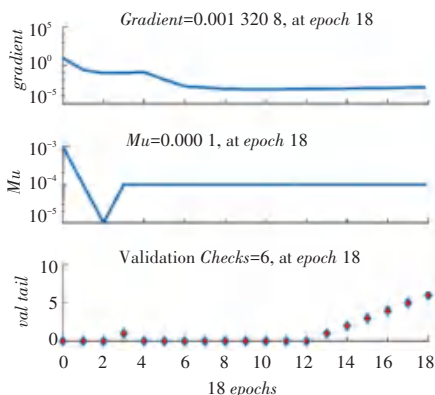


图3 GA-BP神经网络训练状态

Fig. 3 Training state of GA-BP neural network

表1 百度搜索关键词指数与厦门旅游流量的相关性检验
Tab. 1 Keywords correlation test between Baidu search index and tourist traffic in Xiamen

关键词	pearson 相关性	spearman 相关系数	p 值	结论
厦门美食	0.596	0.392	0.000	中度相关
厦门大学	0.472	0.405	0.003	中度相关
沙坡尾	0.653	0.458	0.000	强相关
双子塔	0.659	0.624	0.000	强相关
厦门岛内	0.588	0.564	0.000	中度相关
演武大桥	0.529	0.486	0.000	中度相关

由于文章利用的因变量与自变量均为时序变量,因此在模型建立前,为保证所有变量的平稳性,需要对旅游客流量和各关键词百度搜索指数进行平稳性检验。通过使用ADF检验法对因变量与自变量进行平稳性检验,运行得出各变量均是一阶单整,结果详见表2。

2.3 结果分析

本文从百度指数官网和厦门市文旅局获取了2017年1月到2020年12月的百度搜索指数和旅游客流量数据,由于2020年新冠肺炎疫情爆发,旅游客流量骤减,故不考虑2020年的异常数据。本文模型建立的训练集取自2017年1月到2019年6月,共计30个月的月度数据,模型的预测目标设置为2019年7月至12月,共6个月内厦门市客流量。预测结果如图4所示,预测结果参数见表3。

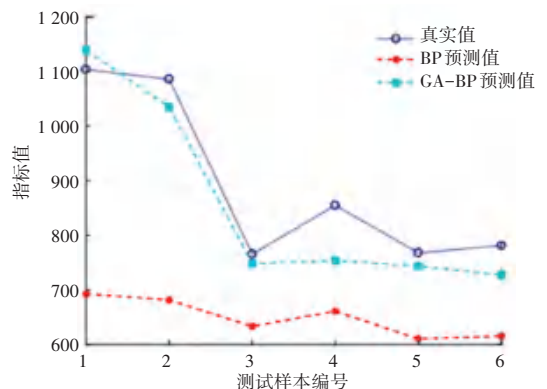


图4 BP神经网络优化效果图

Fig. 4 Optimization effect of BP neural network

表 3 厦门市客流量预测结果

Tab. 3 Forecast results of passengers flow in Xiamen

时间	实际值	BP 模型预测值	BP 模型预测误差	GA-BP 模型预测值	GA-BP 模型预测误差
2019 年 7 月	1 103.44	693.34	-410.10	1 139.22	35.78
2019 年 8 月	1 085.75	681.95	-403.80	1 035.00	-50.75
2019 年 9 月	765.53	633.63	-131.90	748.72	-16.81
2019 年 10 月	854.67	661.47	-193.20	754.07	-100.60
2019 年 11 月	767.94	610.36	-157.58	744.05	-23.89
2019 年 12 月	781.10	615.36	-165.74	727.63	-53.47

从上述结果总体来看,利用 GA-BP 神经网络模型得到的预测结果误差均比 BP 神经网络模型的预测结果误差小。另外,本文采用平均绝对误差、均方误差、均方误差根、平均绝对百分比误差等 4 种模型预测误差的检验标准,分别对 2 个模型进行检验对比,结果见表 4。

表 4 模型精度检验指标

Tab. 4 Model accuracy test index

模型	MAE	MSE	RMSE	MAPE /%
BP 模型	243.72	73 044.86	270.27	25.99
GA-BP 模型	46.89	2 948.34	54.30	5.31

通过检验结果可以看出,GA-BP 神经网络模型各项误差值均远小于 BP 神经网络模型,说明本文建立的 GA-BP 神经网络模型对客流量的预测能力远高于 BP 神经网络。

3 结束语

文章通过爬取 2016 年 1 月到 2019 年 6 月期间,“厦门美食”、“厦门大学”、“沙坡尾”、“双子塔”、“厦门岛内”、“演武大桥”共 6 个关键词的百度搜索指数及旅游客流量数据,建立了 BP 神经网络模型和 GA-BP 神经网络模型,并对厦门旅游客流量进行预测验证。通过模型预测结果与误差指标对比分析,得出遗传算法优化后的 BP 神经网络模型的预测精度和稳定性比标准 BP 神经网络模型都要高,充分说明了遗

传算法可以避免 BP 神经网络陷入局部极小值。因此,GA-BP 神经网络在这一预测方面具有更大的优势。该预测模型可为有关部门及商家提供更精确的旅游客流量预测参考,提高旅游地相关产业的资源配置效率,保证旅游景点各行业的平稳运行。

参考文献

- [1] LI Shaowen, CHEN Tao, WANG Lin, et al. Effective tourist volume forecasting supported by PCA and improved BPNN using Baidu index[J]. Tourism Management, 2018, 68: 116-126.
- [2] PAI Pingfeng, LIN Kuoping, LIN Chishen, et al. Time series forecasting by a seasonal support vector regression model [J]. Expert Systems with Applications, 2010, 37(6): 4261-4265.
- [3] 谭小林,卢洪涛. 网络搜索与 h7n9 禽流感流行趋势的相关性分析[J]. 湖北中医杂志, 2015, 37(10): 15-18.
- [4] 袁恒. 基于网络搜索指数的市场预测模型及应用研究[D]. 重庆:重庆邮电大学, 2016.
- [5] 迟依涵. 基于网络搜索数据的房地产价格预测研究[D]. 西安:西安建筑科技大学, 2020.
- [6] 赖凯声,陈浩,盛明明,等. 网络搜索数据与社会心理学研究 [C]//第十八届全国心理学学术会议摘要集-心理学与社会发展. 天津:中国心理学会, 2015:791-792.
- [7] 马碧云. 基于网络搜索数据的旅游客流量非线性预测研究[D]. 南昌:江西财经大学, 2019.
- [8] 戚明远. 网络搜索数据与商品住宅市场的相关性研究[D]. 广州:华南理工大学, 2019.
- [9] 王炼,贾建民. 基于网络搜索的票房预测模型-来自中国电影市场的证据[J]. 系统工程理论与实践, 2014, 34(12): 3079-3090.
- [10] 白笑笑. 基于网络搜索数据的冰箱销量预测研究[D]. 西安:西安理工大学, 2020.

(上接第 131 页)

4 结束语

对于幼鱼单类别小目标检测,本文通过对 YOLOv5m 删减不同尺度的预测层,得到 YOLOv5m-s 和 YOLOv5-sm 模型。为了验证模型精度,对 YOLOv3m、YOLOv5s、YOLOv5m-s、YOLOv5-sm 和 YOLOv5m 网络进行识别精度评估,结果表明 YOLOv5m-s 模型性能最优,准确度为 0.923,召回率

为 0.886,平均精度为 0.912,实现了幼鱼检测的高精度和高速度,满足实时检测需求。

参考文献

- [1] MARTINEAU P R, MOURRAIN P. Tracking zebrafish larvae in group-status and perspectives [J]. Method, 2016, 62(3): 292-303.
- [2] 陈小娥. YOLO 算法在交通监控目标检测中的研究与应用[J]. 智能计算机与应用, 2020, 10(06): 199-201, 205.