

# 基于优化BP神经网络的销售预测算法研究

孙艳文, 詹天明

(南京审计大学 信息工程学院, 江苏 南京 211200)

**摘要:** 已知国内房屋售价具有一定的不完整的规律性, 其会因季节变换、人群流动、国家相关政策等一系列因素而呈现一定的规律。与此同时, 该规律性并没有确定的单一因子可以直接影响, 故其售价与全部因素之间的关系也是非线性的。针对这一问题, 利用神经网络输入量的非线性、冗余性和可不完备性, 对一段时期内的房屋售价进行预测是一种合理的预测方法。基于BP神经网络传输阈值的不确定性, 利用时间序列方法对因子数据进行平行预测, 再利用遗传算法和BP神经网络对所得结果进行二次优化, 以达到接近实际的精准预测的目的。经过使用某房地产企业的历史销售数据进行反复仿真实验, 其结果表明所提出的优化算法模型预测精度逼近于实际销售结果, 达到了精准预测的目标。

**关键词:** 时间序列; 优化BP神经网络; 遗传算法; 销售预测; BP神经网络

中图分类号: TP391.1

文献标识码: A

文章编号: 1673-629X(2022)01-0035-05

doi: 10.3969/j.issn.1673-629X.2022.01.007

## Research on Sales Forecasting Based on Improved BP Neural Network

SUN Yan-wen, ZHAN Tian-ming

(School of Information Engineering, Nanjing Audit University, Nanjing 211200, China)

**Abstract:** It is known that the price of houses in China has some incomplete regularity, which will show a certain regularity due to a series of factors such as seasonal change, crowd flow, and relevant national policies. At the same time, there is no definite single factor that can directly affect the regularity, so the relationship between the price and all factors is nonlinear. In order to solve this problem, it is a reasonable forecasting method to forecast the house price in a period of time by using the nonlinear, miscellaneous and incomplete input of neural network. Based on the uncertainty of the transmission threshold of BP neural network, the time series method is used to predict the factor data in parallel, and then the genetic algorithm and BP neural network are used to optimize the results for the purpose of accurate prediction. Through repeated simulation experiments with the historical sales data of a real estate enterprise, it is showed that the prediction accuracy of the proposed optimization algorithm model is close to the actual sales results, and achieves the goal of accurate prediction.

**Key words:** time series; optimized BP neural network; genetic algorithm; sales forecast; BP neural network

## 0 引言

销售预测是销售实施单位根据已有历史销售数据进行规律分析和数据挖掘, 综合考虑多种影响因子, 设计科学合理的预测销售模型, 对未来销售情况的一种预测<sup>[1-2]</sup>。销售预测的准确、可靠性一直是企业作为市场需求规划的必要保证, 销售预测作为营销的重要一环, 是进行行情分析和市场研究的重要依据, 销售预测是否精准可直接影响工程预算、资金回笼、经营决策甚至是企业的未来发展走向。

常见的销售预测模型有移动平均模型、时间序列预测模型、灰色预测模型和回归分析模型等<sup>[3]</sup>。移动平均方法<sup>[4]</sup>相对简单, 通过对一定数组的平均值进行

计算, 得出未来的数值趋势; 缺点是数值的间隔选定有一定难度, 只能了解变化的短期未来趋势。时间序列模型<sup>[5]</sup>运用矩阵及统计学的方法, 可以预测销售产品将会进入高销量区还是低销量区的未来短期走势。灰色预测模型<sup>[6]</sup>所需建模信息少, 其模型是依靠灰色系统生成, 特点是能从混乱的数据找到数据之间的规律, 它大多数情况用于小样本预测, 该环节存在不确定性。回归分析模型<sup>[7]</sup>对未来的预测较为精准, 目前有专业的统计软件 SPSS/SAS 支持, 但因为相对专业和复杂, 需要的自变量及因变量数据要多而且精确。

利用时间序列预测的稳定性<sup>[8]</sup>, 使用遗传算法对BP神经网络进行优化, 既能保持数据挖掘环节搜索的

收稿日期: 2021-02-02

修回日期: 2021-06-03

基金项目: 江苏省自然科学基金面上项目(BK20191409)

作者简介: 孙艳文(1998-), 女, 硕士生, 研究方向为目标检测、图像处理; 詹天明, 博士, 副教授, 研究方向为图像处理、模式识别。

全局性和并行化,又能兼顾预测过程的灵活性和可扩展性,最终还能得到精确度较高的预测结果。经过实验验证,基于优化 BP 神经网络的销售预测算法不仅提高了销售预测的预测准确度,而且大大缩短了收敛时间,简化了网络结构,数据预测的误差进一步缩小,对于精准营销具有较高的理论研究意义和实际应用价值。

## 1 优化 BP 神经网络的算法设计基础

### 1.1 时间序列模型

时间序列分析模型<sup>[9]</sup>是根据已有的系统观测所得时间序列数据,通过特定方法进行参数评估和曲线拟合来建立数学模型的一种方法。把与时间有因变量关系的数据序列称之为动态数据<sup>[10]</sup>,采用线性插值法对曲线进行拟合,参数估计一般使用非线性最小二乘法或指数平滑法<sup>[11]</sup>。时间序列模型对已有数据序列依赖性较高,并假设从已有数据分析所得的变化模式会延续至后续阶段,这种必然式的规律预测不能综合考虑房地产及其他行业的多种复杂影响因素的综合作用结果,但其时间跨度可以任意调节,常用单位有年、季度、月或日,统计跨度灵活。同时,预测所依托的前期数据允许具有不规则性,甚至可以不考虑事务间的因果关系,具有较好的适用性。

指数平滑法<sup>[12]</sup>是平稳测试方法中的一种,由移动平滑法演变而来,通过对已有历史数据进行加权平均进而得到预测结果,对离预测区间紧邻的数据赋予较大的权值系数,对较远的后期预测,其权值则按指数规律递减,所得曲线相对平滑,因此称之为指数平滑方法。根据所选函数的幂级分为一级指数平滑、二级指数平滑等。

二次指数平滑法(second exponential smoothing method)<sup>[13]</sup>是在一次指数平滑的基础上再进行一次指数平滑,因而称之为二次指数平滑。数据经过二次指数平滑处理之后得到一条二次平滑曲线,各数据点经过拟合处理之后,避免了时间序列曲线上出现突兀式的瞬间抖动或数据点的大尺度偏离现象<sup>[14]</sup>,对特殊点能加以自动调整,依据该算法计算所得的数据能预测时间序列的短期变化规律。

### 1.2 BP 神经网络

BP(back-propagation,反向传播)神经网络<sup>[15]</sup>是传统神经网络中的一种经典算法模型,其网络结构由输入层、隐层和输出层构成<sup>[16]</sup>,信息的正向传递与误差的反向传播是其核心,并进行  $N$  次迭代。BP 神经网络算法通过将循环模拟中系统产生的误差值返回到输出结果中用于调整神经元的权重,从而得到可以模拟原始问题的人工神经网络系统。简单地讲就是不停地

修正网络中各层节点的权值和阈值,直到获得符合预期目标的输出结果。

对于 BP 神经网络算法的步骤,可以简单归纳如下:

- (1) 给定 BP 神经网络的输入向量  $X$  和输出向量  $Y$ ,初始化权值和阈值;
- (2) 经过计算,得出 BP 神经网络的实际输出;
- (3) 计算 BP 神经网络实际输出与输出向量  $Y$  之间的误差;
- (4) 误差反向传播,更新权值;
- (5) 权值进行学习,使得误差最小。

重复步骤(1)~步骤(5),直至误差最小,如图 1 所示。

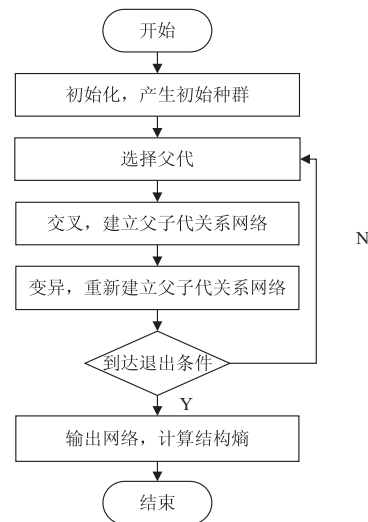


图 1 BP 神经网络算法过程

BP 神经网络是一种具有非线性映射与自学习能力的全局算法,具有非常高效的泛化能力。对于任意的非线性映射关系,BP 神经网络均可以逼近,然而其是一种基于梯度下降的算法,具有以下缺陷:

- (1) 误差存在于误差曲面上,然而有些误差曲面比较平坦,此时误差感受不到权值的变化,导致下降速度变慢,训练时间较长;
- (2) 算法存在多个极小值点,将导致算法陷入极小值点而无法得到全局最优解;
- (3) 算法在学习数据的过程,由于其权值的变化不定,将导致学习过程变得不稳定,出现振荡。

## 2 优化 BP 神经网络的具体算法设计

### 2.1 遗传算法优化设计

遗传算法是根据自然界生物种群优胜劣汰、适者生存的遗传学原理及进化规律而提出的一种算法模型<sup>[17-18]</sup>。算法通过模拟自然界生物进化的过程来逐步探索筛选出研究问题的最优解。算法的一般实现流程如下:

- (1)通过算法对每条染色体评估个体的适应度;
  - (2)以高适应度和高选择率为基准原则,从染色体队列中找出父母双亲;
  - (3)对父母双亲进行染色体的抽取,用以交叉生成子代;
  - (4)将子代进行变异操作;
  - (5)不断去重复步骤(2)~步骤(4),直到有新的种群生成;
  - (6)结束循环,返回结果。
- 算法流程如图 2 所示。

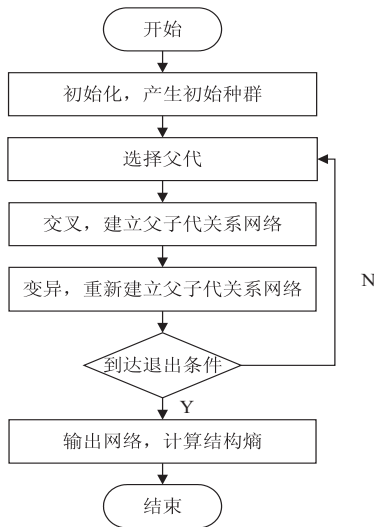


图 2 遗传算法流程

遗传算法的缺点是无法直接处理空间参数,为解决这一问题学者们引入了优化参数进行调优,通过适应度函数形成编码串联群体,对其进行交叉结合和基因变异,最终优胜劣汰地筛选出优秀个体,根据适应度阈值不断迭代,直至满足条件为止。

优化 BP 神经网络由 BP 神经网络的网络结构、算法优化和算法预测三个部分组合而成。模拟生物个体的长度信息由网络结构进行表征,在实际算法设计中通常表示为拟合函数的输入或输出参数;通过探索和调整特定权值达到优化 BP 神经网络算法的目的,最终迭代筛选出具有最后适应度值的种群个体;网络预测通过遗传算法获取最优个体的阈值和权值用以训练网络,预测出函数的输出。

对 BP 神经网络进行一次实验,实验所用模拟数据来源于《中国房地产统计年鉴》。实验抽取一家名为碧桂园的房地产公司作为实验对象,并对其进行仿真实验。根据该公司提供的销售额,统计出了 1 000 条数据,在数据清洗后取前 975 条利用建立的 BP 神经网络进行训练学习,再利用后 25 条已经完备的数据进行缺失销售额的实验,得到实验结果与原数据进行对比,得出了结果——利用 BP 神经网络进行实验后的预测误差为 29.550 1%,如图 3 所示。

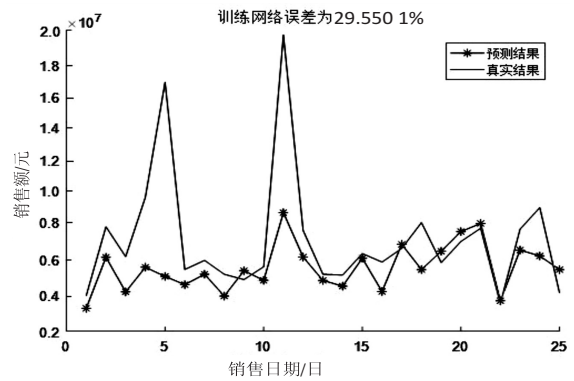


图 3 BP 神经网络预测结果

对上述数据集利用遗传算法优化后的 BP 神经网络模型进行仿真实验,如图 4 所示,其预测结果误差为 21.472 8%。

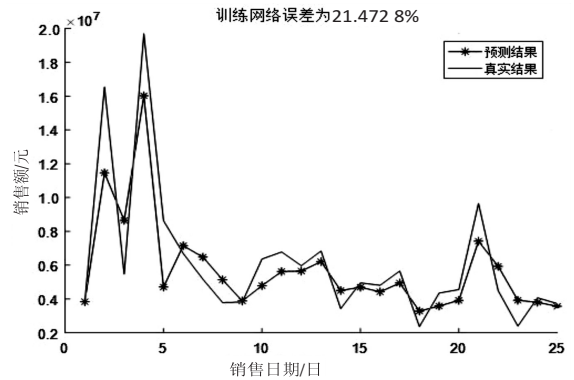


图 4 GA-BP 神经网络预测结果

### 2.2 对 BP 神经网络的再优化

就之前利用 BP 神经网络和 GA-BP 神经网络得到的结果来看,预测误差仍然较大。数据集的大小不一,对于一些数据量比较小的数据集,上述两种神经网络都没有足够的训练量,这样会导致阈值调整不充分、过快收敛等等问题,从而使最后的预测精度大打折扣。面对这一问题,该文提出了一种应用方法,即利用时间序列的结果对 GA-BP 神经网络进行优化校正(简称 TC-GABP 神经网络模型)。由于时间序列预测模型对于事物发展之间的不规则性和因果关系和对预测精度和时长呈反比都不考虑的特点,基于样本容量较小的情况,对原 GA-BP 神经网络的阈值和结果进行一些优化。

时间序列预测模型的实现流程如下:

- (1)人工收集历史资料,按照传统分类方法进行整理并根据时间序列绘成统计图;
- (2)对时间序列中每一时期的数值进行诱发的原因进行分析;
- (3)对长期趋势、季节变动以及不规则变动进行求值操作,并找到最近似的数学模式来代表;
- (4)计算未来的时间序列预测值。利用步骤(3)中得到的模型以及预测不规则变动值通过加法或乘法

模式进行计算。

利用 BP 神经网络的单次传递值和时间序列模型的结果所产生的误差进行一元线性回归,并通过计算误差的函数拟合关系进行预测。最后通过找出两者的误差规律,对 BP 神经网络的传递值进行校正,择优选择遗传算法,进入 GA-BP 神经网络进行结果计算得出校正后的预测值。

在面对样本数据量小的情况下,利用每个个体独立讨论又会导致工程量过大,那么就有必要利用线性关系来处理这种非线性问题,但是最后误差一定较大。为此设定一个元线性回归函数为  $y_r = f(x_{\text{误差}})$  作为校正函数,再使用 BP 神经网络的输出误差  $E$  减去  $y_r$  的值,得到校正过后的值,  $E$  值。该值在 GA-BP 的三层网络架构中影响各个神经元的阈值和各传递权值,并在校正 GA-BP 神经网络缺少训练样本的情况下,达到调整传递值的目的。

$y_r$  修正 GA-BP 神经网络的预测值,得到  $E$ 、 $E_c$  以及  $E$  的修正值,其中  $E_c$  为误差平方和。根据数学方法,对 GA-BP 神经网络的阈值与权值进行再一次带有优化因子的优化,步骤如下:

(1) 输入  $\alpha_n = \sum_{i=1}^d v_{in} x_i$ ,  $\alpha_n$  为隐含层的第  $n$  个神经元接收到的;

(2) 给定  $\beta_j = \sum_{n=1}^q \omega_{nj} b_n$ ,  $\beta_j$  为输出层神经元  $j$  的接收参数,  $b_n$  为隐含层第  $n$  个神经元的输出结果;

(3) 计算 GA-BP 神经网络隐含层与输出层的连接权值  $\Delta v_{in} = -\eta \frac{\partial E_c}{\partial v_{in}}$ , 隐含层的阈值  $\Delta \gamma_n = -\eta \frac{\partial E_c}{\partial \gamma_n}$ 。

其中,对训练例  $(x_k, y_k)$ , 假定神经网络的输出为:  $\hat{y}_k = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_m)$ , 即  $\hat{y}_k = (\beta_j - \theta_j)$ , 其中  $\theta_j$  的更新与  $E_c$  相关,为  $\Delta \theta_j = -\eta \frac{\partial E_c}{\partial \theta_j} = -\eta \frac{\partial E_c}{\partial \hat{y}_j} \cdot \frac{\partial \hat{y}_j}{\partial \theta_j}$ , 网络在

$(x_k, y_k)$  上的  $E_c = \frac{1}{2} \sum_{j=1}^m (\hat{y}_j^k - y_j^k - y_r)$ 。

### 2.3 算法构建

基于优化 BP 神经网络的流程图如图 5 所示。

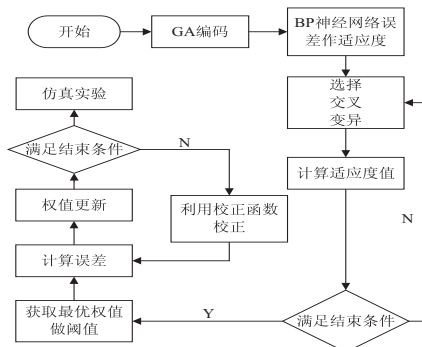


图 5 基于优化 BP 神经网络的流程图

### 3 优化 BP 神经网络在房地产销售额数据集中的应用

利用时间序列预测模型和优化改进后的 TC-GABP 神经网络对数据集进行销售预测。

#### 3.1 时间序列仿真实验

仿真实验使用的是《中国房地产统计年鉴》中碧桂园房产公司数据作为数据集。使用其已经提供的 1 000 条销售额数据,抽取前 975 条进行训练学习,并分别利用一次指数平滑法和二次指数平滑法的时间序列模型对其进行仿真测试,仿真结果如图 6 和图 7 所示。

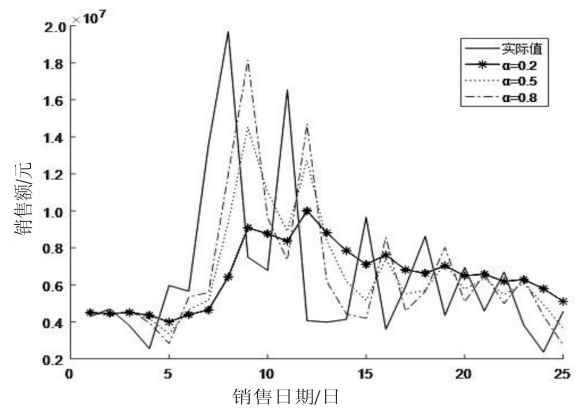


图 6 一次平滑指数预测

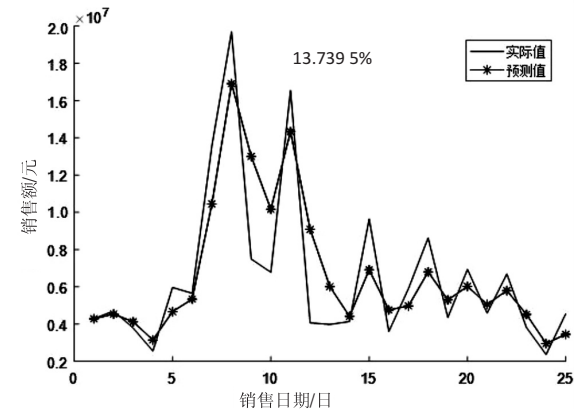


图 7 二次平滑指数预测

从仿真实验结果可以看出,实验结果与理论预测基本相同。但一次指数平滑法的实验结果无论是在  $\alpha$  为何值的情况下都出现了较多的滞后反应。这也衬托出了二次指数平滑法的预测准确度之高(误差率为 25.747 1%),为优化算法提供了可行方案。

#### 3.2 优化后算法仿真实验结果对比

以上文提及的 BP 神经网络算法和遗传算法优化后的 BP 神经网络算法的实验结果误差程度为比较对象。使用已经提供的 1 000 条销售额数据,训练集由数据集中 97.5% 的数据组成,其余数据则作为验证集,来验证误差。TC-GABP 神经网络算法得到的结果如图 8 所示。

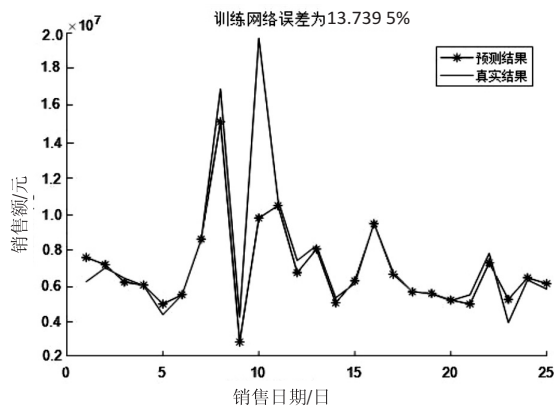


图 8 TC-GABP 神经网络预测

从上述图像中可以直观看出,TC-GABP 神经网络训练后的预测结果更接近真实的结果,比遗传算法改进的 BP 神经网络得到的结果更准确。经过数学计算,它的精度提高了 8%,误差率降低了 13%。表明 TC-GABP 神经网络模型不单单是在预测的准确率上有所提升,其误差率也得到了改善。相较于另外三种模型来说,预测结果更为接近真实值的结果。

### 3.3 结果分析

部分预测结果如表 1 所示。

表 1 不同模型预测结果对比

预测模型	预测结果(误差率)/%
时间序列模型	25.746 1
传统 BP 神经网络	29.554 1
遗传算法改进神经网络	21.562 8
TC-GABP 神经网络	13.739 7

根据表 1 的数据可以看出,TC-GABP 神经网络的预测准确率最佳,达到了 13.739 7%。对于传统的 BP 神经网络,可以看出它的误差率达到了 29.554 1%,接近于 30%。一般来说,接近 30%的误差率的预测模型将被认为是不可靠的。对于时间序列模型和遗传算法改进的神经网络,虽然误差率不是特别高,仿真实验也都表现良好,但是相较于 TC-GABP 神经网络的误差率来说,仍是不太理想。因此,预测模型效果最好的还是 TC-GABP 神经网络,仿真实验也是最接近真实效果的。

## 4 结束语

文章对于销售预测模型的建立,首先选择时间序列模型来减少源数据中的大量误差,再利用优化后的 BP 神经网络进行预测,并具体应用于某房地产销售数据集。实验结果表明,优化后的 BP 神经网络算法不仅大幅提高了预测准确度,缩短了收敛时间,简化了网络结构,也使得数据预测的误差进一步缩小。算法具有较高的理论研究意义和实际应用价值。

### 参考文献:

- [1] 胡雨阳,张承,郑明,等. 基于 LSTM 的销售预测系统的设计与实现[J]. 计算机与网络,2020,46(23):65-67.
- [2] 李钊慧,张康林. 基于 BP 算法和 LSTM 算法的汽车销售预测模型比较研究[J]. 经济研究导刊,2020(20):84-88.
- [3] KIM K I, WANG W, QI Y, et al. Empirical evaluation of work load forecasting techniques for predictive cloud resource scaling[C]//2016 IEEE 9th international conference on cloud computing. San Francisco, CA:IEEE,2016:1-10.
- [4] 陈银光,于守健. 基于改进加权移动平均法的服装销售预测[J]. 智能计算机与应用,2018,8(6):54-56.
- [5] 柳攀. 基于时间序列分析的销售预测方法研究[D]. 大连:大连理工大学,2018.
- [6] LI S, MA X, YANG C. A novel structure-adaptive intelligent grey forecasting model with full-order time power terms and its application[J]. Computers & Industrial Engineering,2018,120:53-67.
- [7] 杨转玲. 基于多元线性回归的盈亏平衡分析[J]. 三明学院学报,2019,36(2):41-45.
- [8] 邹小云,林文学. 基于多目标演化算法和改进概率分类的重尾时间序列预测[J]. 计算机应用与软件,2020,37(12):273-279.
- [9] XIAO Qinkun, XING Li, SONG Gao. Time series prediction using optimal theorem and dynamic Bayesian network[J]. International Journal for Light and Electron Optics,2016,127(23):11063-11069.
- [10] 尚福华,李金成,原野,等. 基于改进 BP 神经网络的地层划分方法[J]. 计算机技术与发展,2020,30(9):148-153.
- [11] 吕丹丹,顾巧祥,邢超. 二次指数平滑法优化马尔科夫预测模型[J]. 中国计量大学学报,2017,28(3):334-339.
- [12] URGAONKAR B B, SHENOY P, CH A, et al. Agile dynamic provisioning of multi-tier internet applications[C]//International conference on autonomic computing. Washington DC:ACM,2010:217-228.
- [13] YANG Q, SONG Y. Data driven confidence intervals for diffusion process using double smoothing empirical likelihood[J]. Journal of Computational and Applied Mathematics,2019,348:282-297.
- [14] 董田田,董学士,张睿,等. 基于指数加权移动平均法的企业销量预测[J]. 青岛大学学报:自然科学版,2020,33(4):50-54.
- [15] 高文,李富星,牛永洁. 基于 BP 神经网络对房价预测的研究[J]. 延安大学学报:自然科学版,2018,37(3):37-40.
- [16] 宋颂,董宝力. 基于改进 BP 神经网络的饮料销售预测模型研究[J]. 成组技术与生产现代化,2018,35(3):9-15.
- [17] MAITY S, ROY A, MAITI M. A rough multi-objective genetic algorithm for uncertain constrained multi-objective solid travelling salesman problem[J]. Granular Computing,2019(1):125-142.
- [18] 胡清准,邱晓晖. 遗传-蚁群算法在智能交通中的应用[J]. 计算机技术与发展,2020,30(4):120-125.