

# 基于 BP 神经网络模型的笔记本电脑销售预测

李向东 刘清华 陈 静

(浙江育英职业技术学院,浙江 杭州 310018)

**摘要:**针对笔记本电脑销售的不确定性,利用 BP 神经网络算法构建模型,采用 MATLAB 对某店铺的 2021 年笔记本电脑的销售数据进行了网络训练,通过 Pearson 相关系数来分析输入层各节点和输出层销售额之间的相关性,确定了访客数 UV、收藏数、加购物车次数、客单价和支付转化率等 5 个输入层节点,销售额为输出层节点,实验证明构建的 BP 神经网络模型预测值和实际值误差小,具有很高的预测准确度,为笔记本电脑销售计划安排及库存管理提供科学依据。

**关键词:**BP 神经网络; Pearson 相关系数; 笔记本销售; 预测

**中图分类号:** TP183

**文献标志码:** A

**文章编号:** 0199(2023)02-0047-04

近年来,受到一些不可控因素的影响,笔记本电脑行业的出货表现和销售额走势表现并不一致。2022 年 Q2 季度全国笔记本电脑整体销量约 160 万件,同比增长约 7%,整体销售额约 86 亿元,却同比下滑 1%左右。随着人们购物习惯的改变,淘宝、京东等电商平台销量大增,尤其是 618、双 11 等电商购物节的优惠力度更是刺激了笔记本电脑的市场需求。2022 年 6 月份,京东平台笔记本电脑月销量超 90 万件,环比增长 150%,销售额超 48 亿元,环比增长 170%<sup>[1]</sup>。笔记本电脑产品的迭代更新很快,因此,为了控制库存风险和制定销售计划,进行笔记本电脑销售预测就极具意义,为库存管理和销售计划的安排提供科学决策。

目前,针对销售预测的方法有很多研究。王辉等提出了一种基于 Stacking 集成策略销售预测方法,在多个机器学习的模型融合下减少了预测误差<sup>[2]</sup>。滕宇宇等利用回归法对销售收入可以进行很好的预测<sup>[3]</sup>。张志新等利用人工神经网络在网络直播中预测销量<sup>[4]</sup>。杨庆斗等分别构建了普通线性回归、Lasso 回归,以及 BP 神经网络预测模型对长城汽车销售量进行预测,结果发现 BP 神经网络预测误差最小<sup>[5]</sup>。周博军等运用 BP 神经网络模型对我国体育彩票销售金额进行预测,预测的精准度较高<sup>[6]</sup>。综合来看,鉴于 BP 神经网络具有很强的非线

性映射能力和柔性的网络结构,相比其它方法,在销售预测中的应用相对较多,准确度较高。本文采用 BP 神经网络的分析方法,利用联想笔记本电脑在某电商店铺的历史销售情况进行预测,为其提供销售决策依据。

## 1 BP 神经网络

BP(back propagation)神经网络是一种按照误差逆向传播算法训练的多层前馈神经网络,在各种预测应用中使用比较广泛的算法,是从对人脑神经网络进行模拟而建立起来的数学模型。人脑神经网络是由神经元互联而成的,能接收并处理信息,而信息处理主要是由神经元之间的相互作用,即通过神经元之间的连接权值来处理并实现的。如图 1 所示,BP 神经网络一般分为输入层、隐含层和输出层,其中输入层为输入的训练数据,隐含层一般具有一层或多层,单层神经网络只能用于表示线性分离函数,一般用于简单问题,多个隐含层可以用于拟合非线性函数,由输出层向外界输出信息处理结果。这三层中的每一层只影响下一层的神经元状态,若预期预测结果得不到期望输出,网络则进行反向传播。主要思路是输入数据,利用反向传播算法不断调整参数,根据最终调

收稿日期: 2023-03-19

基金项目: 浙江省 2021 年度高校国内访问工程师校企合作项目:基于数据挖掘的联想笔记本销售预测研究(FG2021242)

作者简介: 李向东(1982- ),男,浙江育英职业技术学院信息技术学院,副教授,硕士,研究方向:数据挖掘;

刘清华(1984- ),女,浙江育英职业技术学院信息技术学院,副教授,硕士,研究方向:大数据及信息管理;

陈 静(1982- ),女,浙江育英职业技术学院信息技术学院,讲师,硕士,研究方向:数据分析。

整的参数建立模型。选择合适的层数以及隐藏层节点数,在很大程度上都会影响神经网络的性能。

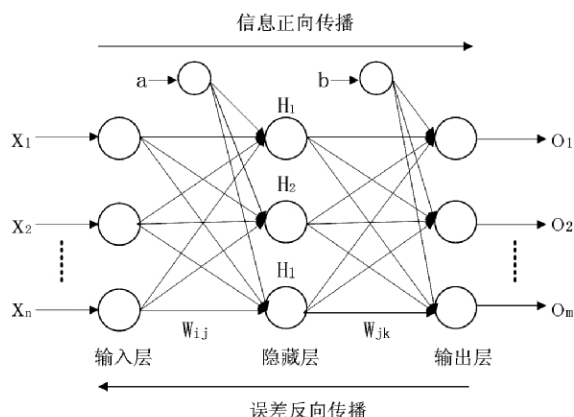


图1 BP神经网络的拓扑结构

## 2 BP神经网络训练步骤

### 2.1 网络初始化

如图1所示,网络输入层的输入为  $x_1 \cdots x_n$ , 隐含层节点为  $H_1 \cdots H_l$ , 输入层到隐含层的权重  $w_{ij}$ , 输出层输出为  $O_1 \cdots O_m$ , 隐含层到输出层的权重为  $w_{jk}$ , 输入层到隐含层的偏置为  $a_j$ , 隐含层到输出层的偏置为  $b_k$ , 其中激励函数为  $g(x)$ , 取 Sigmoid 函数如公式(1)所示。

$$g(x) = \frac{1}{1+e^{-x}} \quad (1)$$

### 2.2 隐含层输出

隐含层 H 的输出如公式(2)所示。其中  $x$  表示输入变量,  $w_{ij}, a_j$  表示输入层和隐含层间的连接权值及隐含层阈值,  $H$  是隐含层节点数,  $g$  为隐含层激活函数。

$$H_j = g\left(\sum_{i=1}^n w_{ij} x_i + a_j\right), j=1,2,3,\dots,l \quad (2)$$

### 2.3 输出层的输出

输出层 O 的输出如公式(3)所示, 其中  $H_j w_{jk}$  和  $b_k$  分别是连接权值和阈值。

$$O_k = \sum_{j=1}^l H_j w_{jk} + b_k, k=1,2,3,\dots,m \quad (3)$$

### 2.4 误差的计算

误差如公式(4)所示, 其中  $Y_k$  是期望输出值,  $O_k$  是实际输出值。

$$e_k = Y_k - O_k \quad (4)$$

### 2.5 权值更新

通过预测误差值  $e$  对网络连接权值更新,  $\eta$  为学习率, 如公式(5)、(6)、(7)、(8)所示。

$$w_{ij} = w_{ij} + \eta H_j (1 - H_j) x_i \sum_{k=1}^m w_{jk} e_k, i=1,2,3,\dots,n; j=1,2,3,\dots,l \quad (5)$$

$$w_{jk} = w_{jk} + \eta H_j e_k, j=1,2,3,\dots;l; k=1,2,3,\dots,m \quad (6)$$

$$a_j = a_j + \eta H_j (1 - H_j) \sum_{k=1}^m w_{jk} e_k, j=1,2,3,\dots,l \quad (7)$$

$$b_k = b_k + \eta e_k, k=1,2,3,\dots,m \quad (8)$$

最后判断迭代是否可以结束, 如果没有结束, 返回第(2)步执行, 直到算法结束。

## 3 建立某网店的联想笔记本销售预测模型

### 3.1 输入层和输出层数据采集

收集了某网店 2021 年全年的销售数据进行整理, 选择了 UV、收藏数、加购物车次数、客单价、支付转化率、销售额共 6 个特征变量, 各特征变量的意义如下。

UV: 访问网店的人数(独立访客), 同一 IP 计算一次, UV 数越大, 说明访问的人数增多, 可能购买的人数将增多, 一般和销售额成正向关系。

收藏数: 商品收藏的用户数, 商品的收藏数越大, 潜在购买人数越多, 一般和销售额成正向关系。

客单价: 店铺每一个顾客平均购买商品的金额, 即平均交易金额。客单价的计算公式是: 客单价 = 销售额 ÷ 成交顾客数, 一般和销售额成正向关系。

支付转化率: 所选时间内访客转化为支付买家的比例, 支付转化率 = 支付买家数 ÷ 访客数, 一般和销售额成正向关系。

销售额: 单位时间内销售的金额。

### 3.2 特征变量相关性分析

项目采用 Pearson 相关系数来分析输入层各节点和输出层销售额之间的相关性。Pearson 相关系数是一种反映两个变量线性相关程度的统计量, 两个变量的线性相关程度用相关系数  $r$  表示, 计算如公式(9)所示, 其中公式(10)表示  $X$  的离均差平方和, 公式(11)表示  $Y$  的离均差平方和。公式(12)表示表示  $X$  与  $Y$  的离均差平方和。

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} = \frac{l_{XY}}{l_{XX} l_{YY}} \quad (9)$$

$$l_{XX} = (X - \bar{X})^2 \sum X^2 - \frac{(\sum X)^2}{n} \quad (10)$$

$$l_{YY} = (Y - \bar{Y})^2 \sum Y^2 - \frac{(\sum Y)^2}{n} \quad (11)$$

$$l_{XY} = \sum (X - \bar{X})(Y - \bar{Y}) = \sum XY - \frac{(\sum X)(\sum Y)}{n} \quad (12)$$

通过 MATLAB 编程计算, 得到访客数 UV 等 5 个变量和销售额的相关系数  $r$  值, 如表 1 所示。相关系数  $r$  简单的分类如表 2 所示。

表 1 变量和销售额的相关系数

变量	访客数 UV	收藏次数	加购物车次数	客单价	支付转化率
相关系数 $r$	0.9362	0.9836	0.9550	0.5725	0.7307

表 2 相关系数  $r$  的相关性分类

相关系数 $r$	0.8-1.0	0.6-0.8	0.4-0.6	0.2-0.4	0.0-0.2
相关性	极强相关	强相关	中等程度相关	弱相关	极弱相关或无相关

从表 1 和表 2 中对比可知, UV、收藏数、加购物车次数三个变量和销售额相关性极强, 支付转化率和销售额相关性强, 客单价和销售额中等程度相关, 因此, 选择 UV、收藏数、加购物车次数、客单价、支付转化率等 5 个变量作为 BP 神经网络的输入层节点数据, 将销售额作为输出层数据。

### 3.3 数据归一化处理

销售预测模型将每周的 UV、收藏数、加购物车次数、客单价、支付转化率 5 个变量作为输入变量, 销售额作为输出变量, 为避免各个输入变量的数量级差异过大, 影响求解算法的效果, 需要对数据进行归一化处理, 采用公式(13)处理。

$$x_1 = \frac{x_0 - x_{min}}{x_{max} - x_{min}} \quad (13)$$

公式中的  $x_0$  表示原始值,  $x_{min}$  表示所有变量中的最大值,  $x_{max}$  表示所有变量中的最小值,  $x_1$  表示归一化后的变量值, 从公式可知, 计算后所有变量的取值范围为  $[0, 1]$ 。某网店 2021 年 52 个周销售数据归一化后部分数据如表 3 所示。

### 3.4 BP 神经网络隐藏层神经元确立

本项目为单隐藏层, 隐藏层神经元节点数在很大程度上将影响神经网络的性能, 但节点数选择没有统一的标准, 当节点数过少时, 训练只能通过经验得出的公式来确定, 通常采用公式(14)确定。其中  $n$  和  $l$  分别为输入层、输出层节点数,  $a$  为 0 到 10 之间的调节常数, 通过计算, 得到隐藏层神经元节点个数  $m$  的值为 3-13 之间。

表 3 网店销售部分核心数据归一化表

统计周	UV	收藏次数	加购数	客单价	转化率	销售额
1	0.5888	0.8153	0.6051	0.7652	1.0000	0.7849
2	0.2659	0.2097	0.2333	0.4879	0.1654	0.1709
3	0.1105	0.0799	0.1205	0.1621	0.3459	0.0693
4	0.1729	0.1431	0.1564	0.8510	0.1429	0.1079
5	0.2097	0.1780	0.2077	0.0000	0.3684	0.0773
6	0.1117	0.1281	0.1231	0.8172	0.4962	0.1378
...	...	...	...	...	...	...
51	0.1301	0.0982	0.1308	0.6551	0.3158	0.1062
52	0.2780	0.2579	0.2821	0.5641	0.3609	0.1924

$$m = \sqrt{n+l} + a \quad (14)$$

为选择最优的节点数, 依次按照不同的节点数进行模型训练, 得到如表 4 所示结果, 确定最佳隐含层的神经元个数为 4, 此时 MSE(均方误差)最小为 0.0008。

### 3.5 MATLAB 模拟仿真

使用 MATLAB 进行网络模型训练, 模型的输入节点数为 5, 输出节点数 1, 隐含层神经元节点数为 4。划分训练集占比 70%, 共 36 个样本数; 验证集占比 15%, 共 8 个样本数; 测试集占比 15%, 共 8 个样本数。采用 levenberg-Marquardt 算法作为训练算法, 使用 MSE 衡量网路性能, 网络允许迭代的次数最大 1000, 实际迭代 10 次, 最大误差为 0.386, 实际误差 0.000211, 网络性能良好, 具有参考性。如图 2 所示, 在数据仿真过程中, 得到在迭代次数为 4 时输出性能最佳, 输出误差为 0.00081633, 超出期望输出, 训练的效果良好。

从图 3 中可以看出训练过程中下降梯度和步长的变化, 呈递减趋势, 迭代到 10 次时梯度下降为 0.00015751。表明随着训练次数的增加, 总体误差随之下降并趋于稳定, 动量因子为 10-5。

图 4 体现模型拟合优度, 确定系数 R 的最大值为 1, 计算得出的值越接近 1, 拟合效果越好。训练集拟合值  $R=0.99835$ , 验证集拟合值  $R=0.97962$ , 测试集拟合值  $R=0.96214$ , 全部数据的拟合值  $R=0.99685$ , 说明拟合的效果很好, 可以用于预测。

表 4 隐含层不同节点数 MSE 的值

节点数	3	4	5	6	7	8	9	10	11	12	13
MSE	0.0009	0.0008	0.0016	0.0015	0.0015	0.0014	0.0019	0.0016	0.0015	0.0029	0.0018

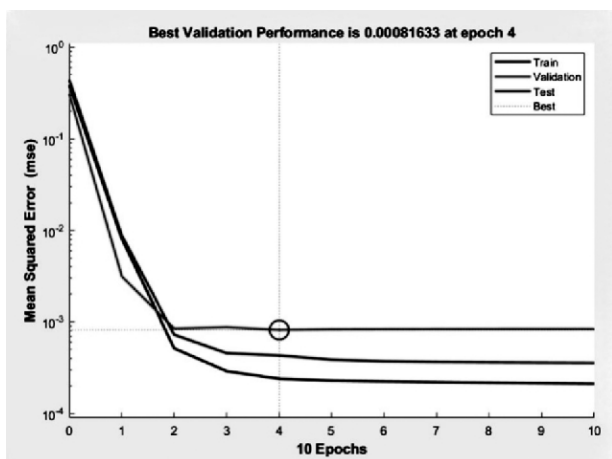


图2 误差性能曲线

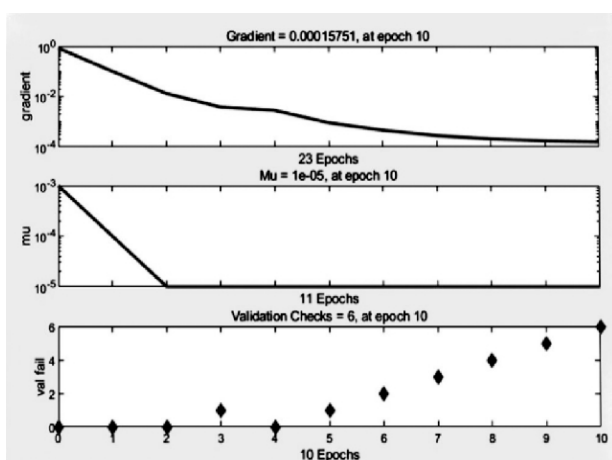


图3 梯度图

#### 4 预测和结果分析

依据项目建立的 BP 神经网络模型进行预测，输入店铺 2022 年前 8 周的销售数据，得到如表 5 所示的预测值。通过和实际销售额进行对比，相对误差都在  $\pm 10\%$  以内，预测效果优良，进一步验证了预测模型的输入变量和输出变量之间的关系信度高。UV、收藏次数、加购数、客单价、转化率对店铺销售额的影响较大。

#### 5 小结

本文构建了一个 BP 神经网络模型，进行某店铺的联想笔记本电脑销售预测，采集了 2021 年 52 个周的输入层和输出层数据，然后利用 Pearson 相关系数来分析输入层各节点和输出层销售额之间的相关性，确定了访客数 UV、收藏数、加购物车次

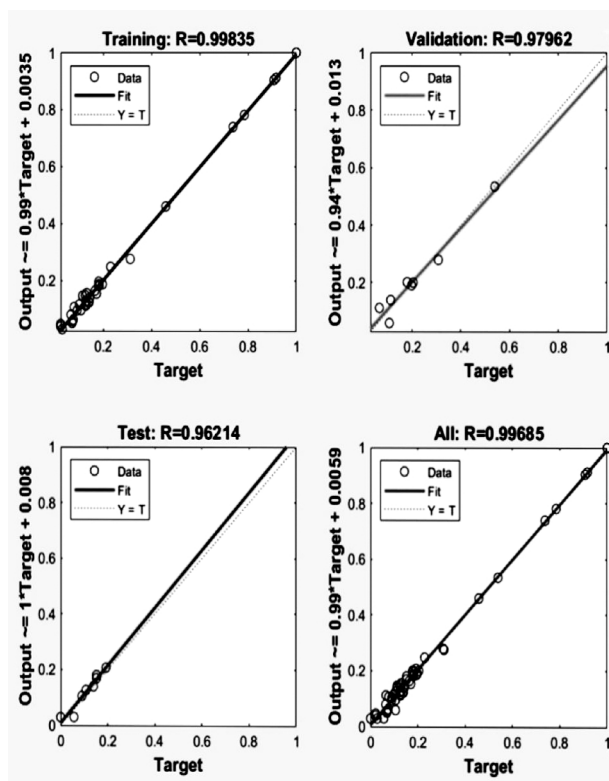


图4 拟合效果

表5 笔记本电脑销售预测结果误差校验

统计周	实际销售额(元)	预测销售额(元)	相对误差
1	2065840	2202540	6.62%
2	769120	774560	0.71%
3	639620	621250	-2.87%
4	659990	640210	-3.00%
5	580520	601250	3.57%
6	764840	695240	-9.10%
7	572340	553210	-3.34%
8	438510	433200	-1.21%

数、客单价和支付转化率 5 个输入变量，销售额作为输出变量，并对数据进行归一化处理，通过经验公式确定了隐含层节点数范围，并通过实验进一步确定个数为 4，利用 MATLAB 进行仿真训练，实验结果表明，模型的训练效果好。最后应用建立的模型对 2022 年前 10 周销售数据进行预测，发现相对误差小，对于商家安排销售计划、营销方案等有重要的参考价值。

(下转第 54 页)

(上接第 50 页)

### 参考文献:

- [1] 鲸参谋电商大数据.Q2 季度笔记本电脑品牌销售排行榜 [EB/OL]. (2022-08-10)[2023-02-11] <https://baijiahao.baidu.com/s?id=1740767060910980942>.
- [2] 王辉,李昌刚.Stacking 集成学习方法在销售预测中的应用[J].计算机应用与软件,2020(8):85-90.
- [3] 滕宁宇,冯润薇,赵智钰,等.回归分析法在销售预测中的应用[J].中国乡镇企业会计,2019(12):107-109.
- [4] 张志新,石文奇,程凡,等.基于 BP 神经网络的网络直播皮革服装销量预测[J].皮革与化工,2021(3):26-30.
- [5] 杨庆斗.基于 BP 神经网络的长城汽车销售量预测研究[D].天津:天津财经大学,2020.
- [6] 周博军,王旺,黄俊达,等.基于 BP 神经网络对中国体育彩票销售金额的预测[J].体育教育学刊,2022(2):46-50.

## Research on Laptop Sales Prediction Based on BP Neural Network Model

*Li Xiangdong Liu Qinghua Chen Jing*

*(Zhejiang Yuying College of Vocational Technology, Hangzhou Zhejiang 310018)*

**Abstract:** Aiming at the uncertainty of laptop sales, BP neural network algorithm was used to build the model, and MATLAB was used to conduct network training on the 2021 laptop sales data of a store. The correlation between the nodes in the input layer and the sales in the output layer was analyzed by Pearson correlation coefficient, and the number of visitors UV, number of favorites, number of shopping cart additions, customer unit price. The experiments proved that the predicted and actual values of the BP neural network model have a small error and high prediction accuracy, which can provide a scientific basis for laptop sales planning and inventory management.

**Key words:** BP neural network; Pearson correlation coefficient; notebook sales; prediction;

(责任编辑:李清平)