

# 基于 BP-AdaBoost 的电商短期销量预测模型<sup>①</sup>



王丽红

(烟台汽车工程职业学院 经济管理系, 烟台 265500)

通讯作者: 王丽红, E-mail: wanglh1123@163.com

**摘要:** 电子商务是伴随互联网技术快速兴起的一种规模大、潜力大的新型商业模式, 对产品进行短期销量预测能够帮助电商企业对市场变化采取更加迅速的响应和措施. 本文通过电商销量历史数据和门户商品链接点击量建立了一种应用于电子商务会计系统的短期销量预测模型. 借助 AdaBoost 思想集合多个传统的 BP 神经网络的预测结果, 使其具备更高的预测准确率, 根据电商短期销量变化的特点规划时间窗口的时序设计, 建立考虑周末效应的以日为本的销量预测模型. 实验证明, 该预测模型的预测误差可以控制在 20% 以内.

**关键词:** AdaBoost; BP 神经网络; 电子商务; 销量预测; 短期销量; 时间序列

引用格式: 王丽红. 基于 BP-AdaBoost 的电商短期销量预测模型. 计算机系统应用, 2021, 30(2): 260-264. <http://www.c-s-a.org.cn/1003-3254/7790.html>

## Forecast Model of Short-Term Sales in E-Commerce Based on BP-AdaBoost

WANG Li-Hong

(Department of Economic Management, Yantai Automobile Engineering Professional College, Yantai 265500, China)

**Abstract:** E-commerce is a new business mode on a large scale and with great potential that is flourishing along with the emerging Internet technology. Forecasting short-term sales of products can help e-commerce companies respond more quickly to market changes. This study establishes a forecast model of short-term sales applied to the e-commerce accounting system based on historical data on e-commerce sales and clicks on portal products. With the adoption of AdaBoost idea, the forecast results of multiple traditional BP neural networks are assembled, leading to a higher accuracy. According to the characteristics of the short-term sales in e-commerce, we plan the timing design of time window and establish a forecast model of sales in the unit of day considering the weekend effect. Experiments show that the forecast error of this model can be controlled within 20%.

**Key words:** AdaBoost; BP neural network; e-commerce; sales forecast; short-term sales; time series

近年来, 电子商务的发展势头强劲, 一度对传统的实体经济产生了非常大的冲击. 仅 2019 年上半年, 销售额已经达到了惊人的 195209.7 亿元<sup>[1]</sup>, 且 2019 年天猫双十一购物节活动成交额突破 100 亿, 仅用了 1 分 36 秒, 这在以往的传统零售行业内几乎是无法想象的<sup>[2]</sup>. 随着电商行业的兴起, 国内涌现出天猫、淘宝、京东等大型电商平台, 并产生了 B2B、B2C、C2C 等许多

新模式, 这些足以看出电商行业巨大的发展潜力<sup>[3]</sup>.

由于电子商务具有数据化、集成化、互动化的特性, 使得电子商务会计系统与其他传统会计系统相比具有更大的数据规模, 同样也蕴藏了更多待挖掘的数据信息<sup>[4,5]</sup>. 各电商企业在产品销售过程中会产生大量的销售数据, 如何处理这些销售数据, 并从中发掘销量变化的规律, 指导电商企业制订策略, 对电商企业管理

① 收稿时间: 2020-06-23; 修改时间: 2020-07-21, 2020-07-29; 采用时间: 2020-07-31; csa 在线出版时间: 2021-01-27

具有重要的意义<sup>[6]</sup>。同时,销量预测也是新型电子商务会计系统中一个重要的组成模块<sup>[7]</sup>。

目前,各个行业的销量预测方法大多是依据历史销量对未来销量的预测。付敏娟等人考虑了指数平滑模型较大的预测误差和神经网络的过拟合问题,提出了用二次指数平滑序列改进BP神经网络的指数平滑神经网络太阳能热水器销售预测模型<sup>[8]</sup>。圣文顺等用遗传算法改进传统的BP神经网络,并用时间序列对模型进一步校正,大大降低了预测误差<sup>[9]</sup>。王建伟利用产品销售共性提取产品聚类簇,再使用时间序列模型得出预测结果并通过隐马尔科夫预测模型给出预测结果的概率分布<sup>[10]</sup>。上述销量预测方法均采用常用的时间序列模型,说明基于历史数据的销量预测具有更高的可行性和可靠性。由于BP神经网络存在的过拟合、收敛慢、误差大等问题,采用其他方法对神经网络进行改进也是较为可行的方法。

本文建立了一种电子商务会计系统中的短期销量预测模型,其以日销量为预测单位,并考虑电商行业内较为常见的“周末效应”改进时间窗的时序设计和预测模型结构。运用AdaBoost思想集合多个传统的BP神经网络的预测结果,通过增加弱预测器数量重点学习预测错误的样本并分配预测器权重,根据预测器权重组合多个BP神经网络(以下简称BPN)的模型输出建立一个准确率更高的预测模型。短期销量预测对于迅速反应电商行情变化,制订短期内稳定销量的应对方案和促销计划具有重要的意义。

## 1 AdaBoost-BPN 预测模型的建立

### 1.1 BPN 弱预测器

BPN是机器学习领域应用较为广泛的一种神经网络算法,常用于模式识别、数据分类、数据挖掘等领域。BPN实际上是一个大型的双向循环的迭代计算模型,正向计算过程是简单的输出结果计算,反向的误差修正计算是BPN的算法核心,通过Widrow-Hoff学习规则对各节点的权值进行反复的修正,直至输出结果误差小于期望阈值<sup>[11-13]</sup>。常见的BPN一般为3层结构,分别为输入层、隐含层和输出层,较为复杂的工程可能用到多隐含层的网络结构,3层BPN的网络结构如图1所示。

### 1.2 AdaBoost 算法

AdaBoost算法也是一种迭代算法,其思想来源于

Hansen提出的同一样本的多个神经网络的组合能够提高网络泛化能力的理论<sup>[14]</sup>。针对于数据预测的算法,AdaBoost算法是按照一定的规则将多个BPN弱预测器集合为一个强预测器。这里所提到的多个BPN预测器集合的规则就是AdaBoost算法的核心规则,与BPN的节点权重修正类似,也是通过计算各BPN弱预测器的权重对BPN弱预测器的结果进行组合<sup>[15,16]</sup>。

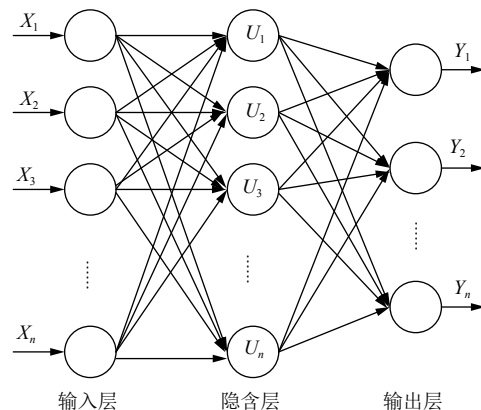


图1 3层BPN网络结构图

AdaBoost算法的具体执行步骤如下:

1) 首先,对每一个输入样本设置权重 $\omega_1(n)=1/N$ , $N$ 为样本总数。利用第1个BPN弱预测器进行第1次训练,并计算错误率 $\varepsilon_i$ 。

$$\varepsilon_i = \sum_{n=1}^N \omega_i(n), h_i(x_n) \neq y_i \quad (1)$$

2) 利用错误率按照式(2)计算BPN弱预测器的权重 $\alpha_i$ , $i$ 为BPN弱预测器编号。

$$\alpha_i = \frac{1}{2} \ln \left( \frac{1-\varepsilon}{\varepsilon} \right) \quad (2)$$

3) 按照式(3)规则确定第2个BPN弱预测器的样本权值,增加预测不准确的样本的权值,降低预测准确的样本的权值,即重点学习预测错误的样本,其中 $h_i(x_n)$ 为第 $i$ 个BPN弱预测器第 $n$ 个样本的预测结果。

$$\begin{cases} \omega_{i+1}(n) = \frac{\omega_i(n)}{Z_i} \times \begin{cases} e^{-\alpha}, & h_i(x_n) = y_i \\ e^{\alpha}, & h_i(x_n) \neq y_i \end{cases} \\ Z_i = SUM(\omega) \end{cases} \quad (3)$$

4) 按照上述规律一直进行迭代计算直至第 $I$ 个BPN弱预测器,得到 $I$ 个BPN弱预测器的权重集合 $\{\alpha_1, \alpha_2, \dots, \alpha_I\}$ ,按照式(4)计算所有BPN弱预测器的集合预测结果,其中 $sign$ 代表符号函数。

$$H(x) = \text{sign} \left( \sum_{i=1}^I \alpha_i h_i(x) \right) \quad (4)$$

### 1.3 AdaBoost-BPN 预测模型工作原理

BPN 弱预测器均采用 2 层隐含层的 BPN, BPN 的隐含层节点数量采用式 (5) 确定。

$$N_h = \sqrt{N_i + N_o} + R \quad (5)$$

其中,  $N_h$  为隐含层节点数量;  $N_i$  为输入层节点数量, 由预测样本数据的维数决定;  $N_o$  为输出层节点数量, 由预测结果数据维数决定,  $R$  为  $[0,10]$  之间的随机数。神经网络训练过程中, 不是隐含层节点数量越多训练效果越好, 因此具体的隐含层节点数量需要进行大量的训练试验来确定。BPN 弱预测器的输入为历史销售数据, 输出结果为销量预测数据。

AdaBoost-BPN 预测模型的实质是将 BPN 弱预测器融入 AdaBoost 的算法的过程中, AdaBoost 算法通过不断增加 BPN 弱预测器数量强化训练那些预测结果不准确的样本, 直至预测误差达到设置的误差阈值。AdaBoost 算法的输入为 BPN 弱预测器模型和样本数据, 输出为多个 BPN 弱预测器的预测模型和权重, 并组合得到一个强预测器。AdaBoost-BPN 预测模型流程图如图 2 所示。

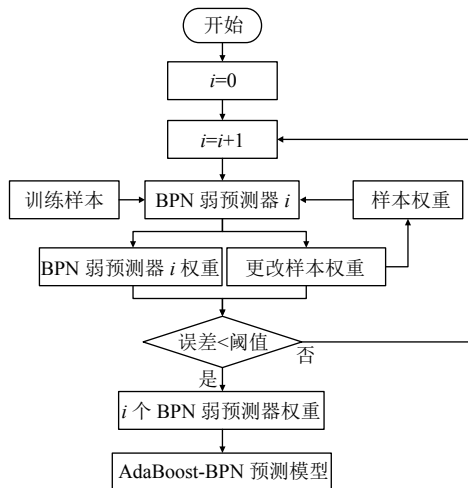


图 2 AdaBoost-BPN 预测模型流程图

## 2 基于时间序列的短期销量预测模型

目前常见的销量预测模型大多以月销量或者周销量作为基本预测单位<sup>[17,18]</sup>, 随着电商行业规模的不断扩大, 入驻电商领域的商家越来越多, 甚至以个人为主导的 C2C 商业模式也展现出较大的活力和发展潜力, 电

商行业的竞争愈发激烈<sup>[19]</sup>。在这种情况下, 以往的销量预测方法已经无法保证预测的及时性, 因此本文建立的 AdaBoost-BPN 短期销量预测模型以日销量作为预测基本单位。

基于时序数据的预测任务通常用时间窗口切割时间轴来获得训练样本, 本文建立的时间轴设计了两个固定跨度的时间窗口, 观测时间窗和预测时间窗<sup>[20,21]</sup>。本文以某电商平台商家 LMAZ-CG1520 型号牛仔裤的历史销量和链接点击量数据作为模型的训练和测试数据。该组数据具有较为明显的分布特征, 同一周内工作日的销量和周末的销量会一般存在较大的差别, 周末两天的日销量会明显高于工作日 5 天的日销量, 本文称之为电商的“周末效应”。除了特殊节假日和平台促销活动期间, 其他时间的销量基本遵循以一周为周期的短期循环。为了使模型的训练能够全面采集一周内的销量变化特征, 同时考虑“周末效应”的影响, 本文采用观测时间窗跨度为 5, 预测时间窗跨度为 3, 当时间轴总跨度为  $n$  时, 即以  $n$  天的历史销量数据作为训练数据集, 将得到  $n-5+1$  组训练数据, 时间轴的划分如图 3 所示。

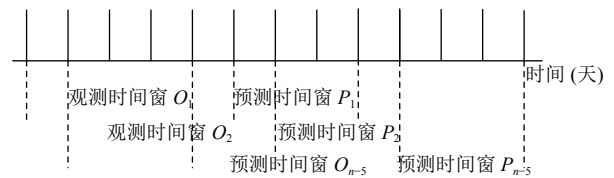


图 3 时间轴划分图

AdaBoost-BPN 短期销量预测模型的观测时间窗跨度为 5 天, 常规的事件窗采样无法判断观测时段内是否存在周末, 因此本文在模型的输入部分增加了 1 个输入层节点, 该输入节点称为“周末因子”。电子商务销量历史数据中包含该条记录的时间标签, 判断该条记录是否为周末的销量较为简单。当“周末因子”为 1 时, 代表此次训练样本中包含周末销量, 能够帮助网络调整权重, 5 组输入数据相邻的两组明显偏高的数据或者边缘 1 组明显偏高的数据即为周末销量; 当“周末因子”为 0 时, 代表此次训练样本均为工作日销量。

电商平台中某商品的销量与该商品链接的点击量有直接关系, 模型输入中应包含该商品的历史销量、历史点击量和“周末因子”。因此, 单个 BPN 弱预测器一共有 11 个输入节点, 3 个输出节点, 共设置两层隐含层, 第一层隐含层采用  $\text{tansig}$  函数作为传递函数, 第二

层隐含层采用 pureline 函数作为传递函数, 经过大量的训练测试, 隐含层节点数分别为 13 和 3 时, 模型的训练效果最佳, 训练次数为 1000 次时, 网络计算精度可以达到 0.01. 加入“周末因子”的 AdaBoost-BPN 短期销量预测模型中单个 BPN 弱预测器的结构如图 4 所示.

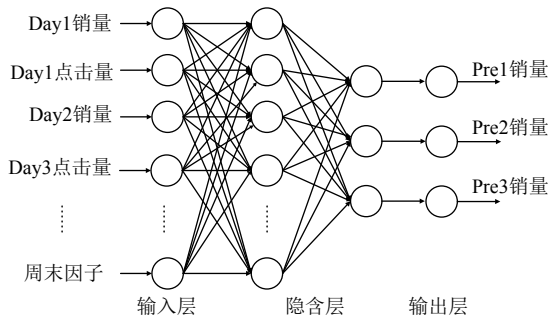
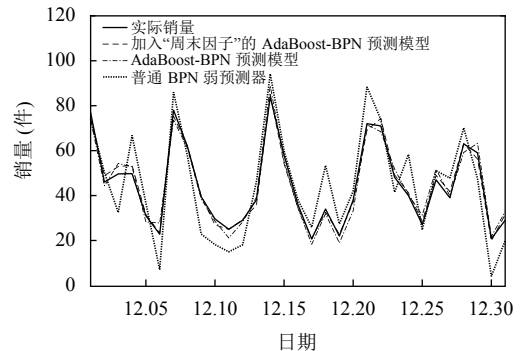


图 4 单个 BPN 弱预测器结构图

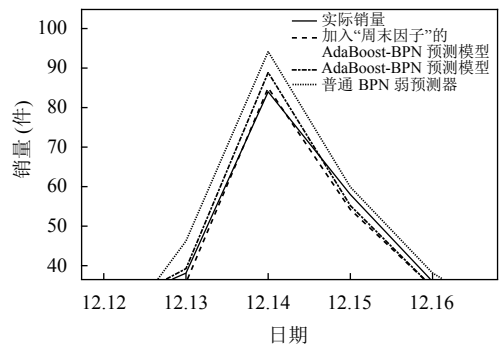
### 3 短期销量预测实验

实验数据取自某电商商家 2019 年第三、四季度 LMAZ-CG1520 型号牛仔裤的历史销量数据和历史点击量数据, 其中有效数据段共 184 条, 采用 7、8、9、10、11 月份的数据作为训练数据集, 数据段共 153 条, 有效训练样本 149 组. 12 月份数据作为测试数据集, 数据段共 31 条. 分别采用普通 BPN 弱预测器、AdaBoost-BPN 预测模型和加入“周末因子”的 AdaBoost-BPN 预测模型进行销量预测实验, 预测结果采用预测误差进行评价. 图 5 为不同模型的销量预测测试结果, 以及预测结果与真实数据的对比. 图 5 中不同预测模型的预测曲线表明, 加入“周末因子”的 AdaBoost-BPN 预测模型的预测曲线与实际销量的重合度最高, 没有加入“周末因子”的 AdaBoost-BPN 预测模型表现稍差, 而只用一个普通的 BPN 弱预测器的预测结果偏差很大.

表 1 为不同模型的平均预测误差对比. 由表 1 结果分析, 采用普通 BPN 弱预测器的误差远远大于 AdaBoost-BPN 预测模型的误差, 说明基于 AdaBoost 思想集合多个 BPN 弱预测器结果能够提升预测的准确率; AdaBoost-BPN 预测模型与普通 BPN 弱预测器相比, 预测效果得到了较为明显的提升, 但是由于没有考虑“周末效应”, 一些特殊的样本数据拉低了整体的预测效果; AdaBoost-BPN 预测模型在添加了“周末因子”后, 平均预测误差明显降低, 可以达到 20% 以内, 预测效果得到了极大的提升.



(a) 12月份销量预测结果



(b) 局部放大图

图 5 销量预测测试结果

表 1 不同模型平均预测误差对比 (%)

预测模型	平均预测误差
普通BPN弱预测器	53.23
AdaBoost-BPN预测模型	26.32
加入“周末因子”的AdaBoost-BPN预测模型	18.89

本文在建立时间轴时, 考虑到电商销量的“周末效应”, 确定的观测时间窗跨度为 5 天, 为了探究选取的观测时间窗跨度对预测效果的影响, 采用加入“周末因子”的 AdaBoost-BPN 预测模型, 将历史时间窗跨度分别设置为 3、4、5、6、7 再次进行预测实验, 实验结果如表 2 所示.

表 2 不同时间窗平均预测误差对比 (%)

观测时间窗跨度	平均预测误差
3	63.35
4	36.32
5	19.32
6	24.68
7	27.59

由表 2 结果分析, 观测时间窗跨度为 3 和 4 时, 样本数据隐藏的信息量不足, 而且预测时间窗为 3 天, 理论上观测时间窗应大于预测时间窗, 较短的观测时间

窗跨度无法准确体现销量变化特征,因此平均预测误差相对较大;观测时间窗跨度为6和7时,本文所建立的“周末因子”形同虚设,模型无法准确反映“周末效应”带来的影响,因此平均预测误差偏大;只有观测时间窗跨度5时,预测模型发挥出最良好的预测能力,同时证明了“周末因子”的加入可以有效提升预测的准确度。

#### 4 结论

在电商行业迅速发展的大背景下,本文针对目前已有的商品销量预测模型预测误差大、无法满足商家对行情变化的迅速应变的问题,提出了一种电子商务会计系统中的短期销量预测模型。该模型以日销量作为预测基本单位,采用近5天的销量数据和点击量数据对未来3天的销量变化情况进行预测。借鉴AdaBoost思想集合多个BPN弱预测器形成强预测器,提高预测准确率。提出合理的日销量预测的时间轴设计方法,根据电商产品销量的变化特性为预测模型添加“周末因子”,使预测模型具备辨识“周末效应”的能力。实验测试结果表明,加入“周末因子”的AdaBoost-BPN预测模型平均预测误差在20%以内,能够有效预测未来短期内的销量变化情况,对电商企业及时掌握产品销售情况,并对短期内存在的销售风险制订及时有效的应对措施具有重要的意义。

#### 参考文献

- 1 夏平,杨继平. 电商行业财务数据与大数据耦合协同评价. 财会月刊, 2020, (9): 27-34.
- 2 路澳征,张媛媛.“互联网+创新创业”背景下农村电商的模式探索. 福建茶叶, 2020, 42(4): 33-34. [doi: 10.3969/j.issn.1005-2291.2020.04.026]
- 3 申容宇. 探析“互联网+”时代跨境电商发展问题及对策. 智库时代, 2020, (15): 41-42.
- 4 车久菊. B2C电子商务模式下电商企业经营模式变革. 中外企业家, 2020, (11): 88-89.
- 5 刘振,宋寒,代应,等. 新零售”背景下的电商企业渠道决策——基于体验店服务水平与市场需求. 财会月刊, 2020, (8): 115-122.
- 6 李寿喜,汪梦晓. 我国收入会计准则变动对电商企业会计核算的影响. 会计之友, 2018, (5): 142-144. [doi: 10.3969/j.issn.1004-5937.2018.05.033]
- 7 李贞. 互联网环境下电子商务营销渠道优化探索. 财富时代, 2020, (5): 43.
- 8 付敏娟. 基于指数平滑与神经网络模型的太阳能热水器销售预测 [硕士学位论文]. 湘潭: 湘潭大学, 2017.
- 9 圣文顺,赵翰驰,孙艳文. 基于改进遗传算法优化BP神经网络的销售预测模型. 计算机系统应用, 2019, 28(12): 200-204. [doi: 10.15888/j.cnki.csa.007174]
- 10 王建伟. 基于商品聚类的电商销量预测. 计算机系统应用, 2016, 25(10): 162-168. [doi: 10.15888/j.cnki.csa.005423]
- 11 王菲. 一种基于自适应遗传BP算法的神经网络目标分类器设计方法. 自动化技术与应用, 2007, 26(12): 80-82. [doi: 10.3969/j.issn.1003-7241.2007.12.029]
- 12 Hinton GE, Nowlan SJ. The bootstrap widrow-hoff rule as a cluster-formation algorithm. Neural Computation, 1990, 2(3): 355-362. [doi: 10.1162/neco.1990.2.3.355]
- 13 朱兴统. 基于改进粒子群优化BP网络的城市用水量预测. 计算机与现代化, 2012, (8): 21-23, 27. [doi: 10.3969/j.issn.1006-2475.2012.08.006]
- 14 Hansen LK, Salamon P. Neural network ensembles. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, 12(10): 993-1001. [doi: 10.1109/34.58871]
- 15 张晶. 基于AdaBoost回归树的多目标预测算法. 计算机与现代化, 2017, (9): 89-95, 105. [doi: 10.3969/j.issn.1006-2475.2017.09.017]
- 16 闫博,周在金,李国和,等. 基于ARMA和BP\_AdaBoost的组合销售预测模型研究. 计算机与现代化, 2015, (2): 14-18. [doi: 10.3969/j.issn.1006-2475.2015.02.004]
- 17 王雪蓉,万年红. 基于跨境电商可控关联性大数据的出口产品销量动态预测模型. 计算机应用, 2017, 37(4): 1038-1043, 1050. [doi: 10.11772/j.issn.1001-9081.2017.04.1038]
- 18 Ghavamipoor H, Golpayegani SAH. A reinforcement learning based model for adaptive service quality management in E-Commerce websites. Business & Information Systems Engineering, 2020, 62(2): 159-177.
- 19 Zahari A, Jaafar J. Combining hidden markov model and case based reasoning for time series forecasting. Proceedings of the 13th International Conference on Intelligent Software Methodologies, Tools and Techniques. Langkawi, Malaysia. 2015. 237-247.
- 20 孙钰. 基于数据挖掘的电商促销活动效应与销量预测研究 [硕士学位论文]. 上海: 东华大学, 2017.
- 21 黄文明. 基于深度学习的商品销量预测研究 [硕士学位论文]. 南京: 南京理工大学, 2019.