

基于神经网络的电商消费者复购行为预测

张靖轩, 张微微

上海工程技术大学管理学院, 上海

收稿日期: 2021年9月18日; 录用日期: 2021年10月11日; 发布日期: 2021年10月21日

摘要

随着互联网零售行业持续高涨, 越来越多的消费者选择在网络购物, 尤其是中国消费者。利用消费者在网络留下的行为数据进行重复购买行为的预测对企业实现精准营销有着重要的意义。本文采用电子商务平台上的历史行为数据对消费者复购行为进行预测有助于提升用户体验和营销效果。提出一种基于神经网络的消费者复购行为预测模型, 实现用户属性、商品属性及用户行为特征的自动抽取与选择, 并以此对消费者复购行为进行预测。在阿里巴巴移动电商平台数据集的实验结果表明, 基于神经网络的预测模型F1值比基准模型平均提升了7%~11%。

关键词

电子商务, 复购行为, 神经网络, 预测模型

Prediction of Repurchase Behavior of E-Commerce Consumers Based on Neural Network

Jingxuan Zhang, Weiwei Zhang

Management Institute, Shanghai University of Engineering Science, Shanghai

Received: Sep. 18th, 2021; accepted: Oct. 11th, 2021; published: Oct. 21st, 2021

Abstract

As the Internet retail industry continues to rise, more and more consumers choose to shop online, especially Chinese consumers. Using consumer behavior data left on the Internet to predict repeat purchase behavior is of great significance for companies to achieve precision marketing. This article uses historical behavior data on the e-commerce platform to predict consumer repurchase behavior, which will help improve user's experience and marketing effects. A neural network-based

consumer repurchase behavior prediction model is proposed to realize the automatic extraction and selection of user's attributes, commodity attributes and user's behavior characteristics, and to predict consumer repurchase behaviors. Experimental results on the data set of Alibaba's mobile e-commerce platform show that the F1 value of the neural network-based prediction model is improved by an average of 7%~11% compared with the benchmark model.

Keywords

E-Commerce, Repurchase Behavior, Neural Network, Prediction Model

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着大数据时代的到来, 可用信息数据爆发式增长。2007 年至 2011 年, 数字信息的容量迅速增长了 9 倍。2020 年, 全球信息量已经达到 35 万亿千兆字节[1]。这些数据具有稀疏性、高维度的特点, 本身并没有意义, 但是经过清洗, 通过降维或特征选择等方式减少冗余、复杂的变量, 再结合机器学习或数据挖掘算法便能得出需要的信息, 并转化为知识。大数据高速增长阶段, 社交媒体分析、生物医学成像、金融风险、监视视频分析和零售等领域都在不同程度地学习并应用大数据, 其中电子商务中的应用最为显著。通过分析消费者在电商平台留下积累的历史行为数据, 如点击浏览、收藏物品、加入购物车等操作, 有助于较好地预测消费者的购物偏好、习惯以及购买意愿。

消费者复购行为预测是国内外学者关注的热点研究。祝歆(2017)通过融合 Logistic 回归与 SVM 模型预测了消费者的购买行为[1]。Chunyan HOU (2018)使用随机森林(RF)和梯度提升决策树(GB)做特征转换, 再结合 Logistic 回归预测了消费者的购买行为[2]。Liu 融合了 Logistic 回归、随机森林以及感知机, 对网络消费者复购行为进行了预测, 并取得了较为理想的结果[3]。王克利运用随机森林将网络消费者历史行为数据分组训练, 减少了数据不平衡造成的影响, 增强了训练模型的泛化能力[4]。

随着电商平台中行为数据的不断累积, 特征向量的构造和选择都需要耗费大量的时间和精力。另外数据逐渐复杂化, 适用于线性数据分析的传统机器学习方法已经不足以精确预测消费者的重复购买行为。因此, 本文采用 back propagation 神经网络(简称 BP 神经网络)模型来预测电商消费者的复购行为。

2. BP 神经网络复购行为预测模型

BP 神经网络是一种受生物大脑机制启发构建的智能化机器学习模型, 可以通过反馈机制处理输入值得到适应的输出值。众多神经元相互连接, 当一个神经元处于“兴奋”状态时, 便会对相连接的神经元传送一种化学物质, 使得神经元的电位发生改变; 当某一个神经元的电位超过设置的“阈值”时, 那么这个神经元就会被激活, 也就是会处于“兴奋”状态, 从而向其它相连接的神经元传送化学物质。BP 神经网络由输入层, 隐藏层和输出层三层组成, 如图 1 所示。不同层之间实现了完全连接, 可以相互交流信息, 结构图中箭头代表是实验过程中数据的流向。不同层的神经元的作用也不相同, 输入层负责外接信息的接收, 隐藏层一般用于各类信息的储存, 输出层负责接收并处理来自输入层的各类信息。

由于人工神经网络自适应性强, 处理复杂非线性问题具有较强的优势, 已广泛应用于机器人, 对象识别, 语音和手写识别等各个领域。

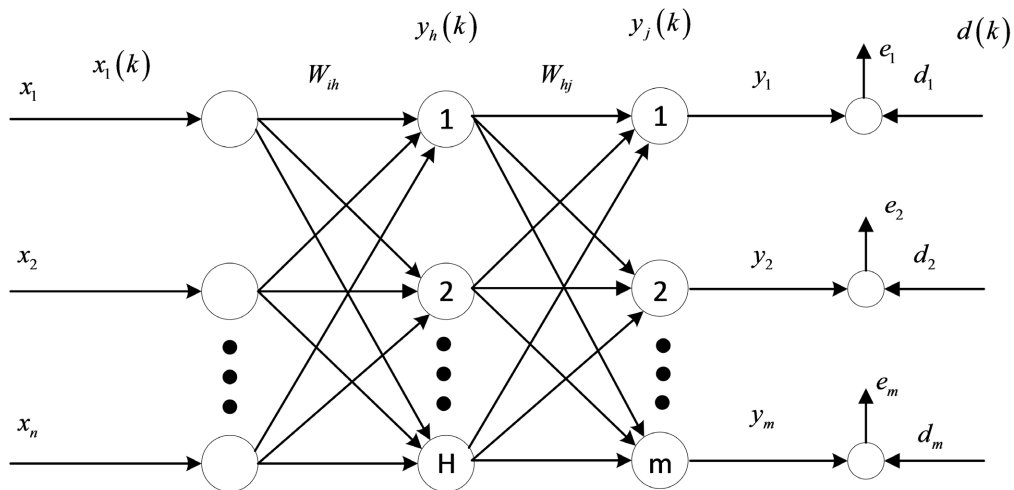


Figure 1. Schematic diagram of artificial neural network

图 1. 人工神经网络示意图

2.1. 预测模型框架

图 1 为融合用户特征、商品特征和用户行为特征的预测电商消费者复购行为的神经网络模型总体架构。数据经过清洗后，剔除受重大促销活动影响产生的不规律数据，并采用抽样的方式使复购样本与未复购样本达到均衡，输入层接收各种影响消费者复购行为的特征，如浏览量、收藏量、加购量、转化率等，经过特征选择和优化后，传入隐层。隐层提取消费者复购行为数据中的重要特征，并根据此完成消费者复购行为的预测。最后由输出层输出结果。

2.2. 电商消费者复购预测实验设计

整理数据发现，消费者对商品进行一次购买后，只有极少数会进行重复购买，导致复购行为数据与未复购行为数据极度不平衡，需要对数据进行抽样处理。由于复购行为数据对研究结果有着重要意义，全部保留。未复购行为数据随机保留，使其数量为复购行为数据的 2 倍。

已有相关研究的特征模型主要分为用户特征和商品特征两类。针对电商消费者是否会对购买过的商品进行重复购买行为这一问题，本文剔除了无关的用户地理位置及行为发生时间这 2 个因素，将留下的信息整理为用户特征、商品特征以及用户与商品交互行为特征，并通过统计分析构建消费者复购行为预测特征，具体如表 1 所示。最后，将整理好的特征集输入到构建的 BP 神经网络预测模型中，输入层将自动提取消费者复购行为的重要特征并训练分析，过程较为简便。见表 1。

Table 1. Predictive characteristics of consumer repurchase behavior

表 1. 消费者复购行为预测特征

特征类型	特征简要描述	特征表示
消费者特征	消费者点击数	x1
	消费者收藏数	x2
	消费者加购数	x3
	消费者购买数	x4

Continued

商品特征	商品被点击数	x5
	商品被收藏数	x6
	商品被加购数	x7
	商品被购买数	x8
用户与商品交互特征	用户对某一商品的点击总次数	x9
	用户对某一商品的收藏总次数	x10
	用户对某一商品的加购总次数	x11

3. 实验

本文的数据来自阿里巴巴天池大赛中移动电商平台, 共 2,084,859 条消费者历史购买行为数据, 内容包括用户信息、商品信息、消费者行为类型及行为发生时间等 6 个字段。实验将根据该数据对 12 月 19 日用户是否将对已购买的商品再次购买进行预测。对数据进行清洗时, 删除商品位置、行为发生时间以及用户位置及三个对用户复购行为影响较小的字段, 增加购买点击转化率、购买加购转化率、购买收藏转化率三个用户与商品交互行为特征。

3.1. 评估指标

结合数据案例实际类别与文中构建模型预测类别, 结果可以划分为真正例(TP)、假正例(FP)、真负例(TN)、假负例(FN)四种情况, 具体如表 2 所示。混淆矩阵可具体表现为表示为精确率 Precision、召回率 Recall 和 F1 值[5]。本文通过查准率、查全率以及 F1 值来评估 BP 神经网络预测模型性能, 具体公式如下:

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (3)$$

Table 2. Confusion matrix of user's repurchase behavior

表 2. 用户复购行为混淆矩阵

实际结果	预测结果	
	正例	负例
正例	TP	FN
负例	FP	TN

3.2. 实验结果与分析

为了验证构建的 BP 神经网络模型的预测性能, 本文将与 5 种传统机器学习模型进行比较。参数设置具体如表 3 所示。

Table 3. Algorithm parameter settings of each model
表 3. 各模型算法参数设置

模型类别	主要参数	取值
支持向量机	核函数	高斯核函数
	错误项的惩罚系数	1.0
	阈值	0.01
	最大迭代次数	1000
随机森林	决策树棵树	200
	特征数	11
	最大深度	10
	结点最小样本数	15
XGBoost	学习率	0.2
	最小权重	1
	树深度	8
	决策树棵树	100
BP 神经网络	学习率	0.01
	卷积核个数	32
	卷积核大小	5 × 5
	池化核大小	2 × 2
	激活函数	ReLU

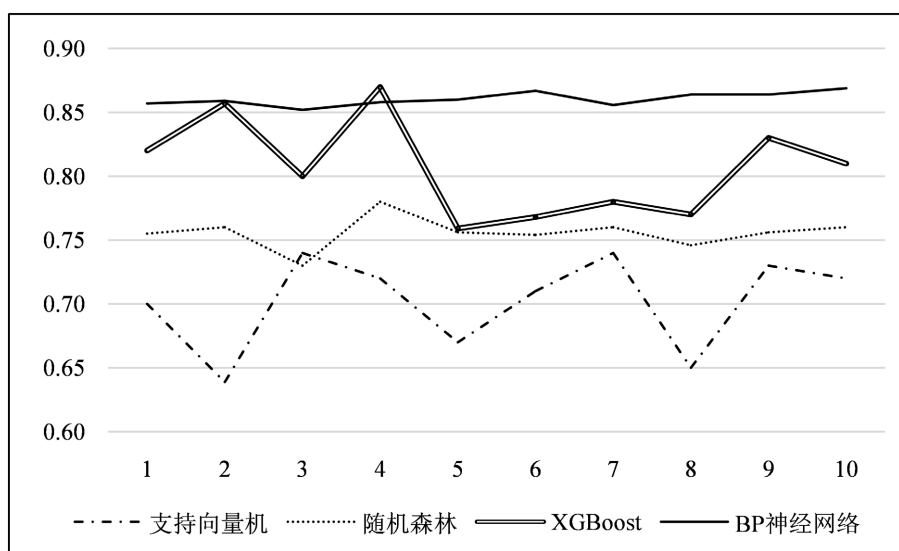
另外, 文章采用 K 折交叉验证来保证实验结果的准确和客观, 实质上就是将实验重复地进行 K 次。首先将原始数据随机分成 K 个数据集, 每次实验都从这 K 个数据集中选择一个不同的集合作为测试集, 剩余的 $K - 1$ 个作为训练集进行实验, 最后将得到的 K 个实验结果取平均值。该方法可以避免欠拟合和过拟合, 更好地说明模型结果。本文中使 K 为 10, 即每个实验进行 10 次, 取查准率、召回率以及 F1 值的平均值作为预测模型最终结果, 求得结果如表 4 所示。

从表 4 的实验结果可以看出, SVM 的查准率较低, 仅为 0.665, 表明消费者行为数据呈现线性不可分状态。随机森林准确率、查全率、F1 值均高于支持向量机, 但依然不够理想。相比前两种预测模型, XGBoost 在查准率和查全率都有不错的结果, 因此 F1 值也较高。但从查准率、查全率、F1 值三个角度的综合考虑下, BP 神经网络模型具有最好的电商消费者行为预测性能。

图 2 从以折线图的形式更加直观地比较了四个预测模型的预测性能。如图所示, 支持向量机的预测结果较差, F1 值几乎都处于最下方。XGBoost 的实验结果 F1 值有了较大的提升, 但波动较大, 较不稳定。BP 神经网络的 F1 值较高的同时, 也不随着实验的变化而变化, 表示具有较强的稳定性。

Table 4. Comparison of precision rate, recall rate and F1 value of four user's repurchase behavior prediction models**表 4.** 四种用户复购行为预测模型查准率、召回率和 F1 值对比

预测模型	查准率	查全率	F1 值
支持向量机	0.665	0.723	0.713
随机森林	0.731	0.787	0.758
XGBoost	0.818	0.834	0.826
BP 神经网络	0.828	0.854	0.841

**Figure 2.** F1 change line chart of four models**图 2.** 四种模型 F1 变化折线图

4. 结语

在本文中, 我们提出了一种 BP 神经网络模型对电商消费者复购行为进行预测。与现有的复购行为预测模型相比, 我们的特征工程不仅考虑了用户特征和商品特征, 还考虑了用户与商品的交互行为特征。另外通过与传统的机器学习模型比较, 验证了该模型用于预测电商消费者重复购买行为的有效性。在后续研究中, 我们将从电商消费者重复购买行为的影响机制层面构建一种包含用户个体差异性的预测模型, 进而为增加电商消费者重复购买行为提供有价值的建议。

参考文献

- [1] 祝歆, 刘潇蔓, 陈树广, 李静, 张天宇. 基于机器学习融合算法的网络购买行为预测研究[J]. 统计与信息论坛, 2017, 32(12): 94-100.
- [2] Hou, C., Chen, C. and Wang, J. (2018) Tree-Based Feature Transformation for Purchase Behavior Prediction. *IEICE Transactions on Information and Systems*, **E101-D**, 1441-1444. <https://doi.org/10.1587/transinf.2017EDL8210>
- [3] Liu, W., Wang, J., Sangaiah, A.K. and Jing, J. (2018) Dynamic Metric Embedding Model for Point-of-Interest Prediction. *Future Generation Computer Systems*, **83**, 183-192. <https://doi.org/10.1016/j.future.2017.12.014>

- [4] 王克利. 面向重复购买背景的预测模型研究[D]: [硕士学位论文]. 广州: 华南理工大学, 2018.
- [5] 胡晓丽, 张会兵, 董俊超, 吴冬强. 基于 CNN-LSTM 的用户购买行为预测模型[J]. 计算机应用与软件, 2020, 37(6): 59-64.