

基于机器学习组合模型的电商商品销量预测^①



韩亚娟, 高欣

(上海大学 管理学院, 上海 200444)

通信作者: 高欣, E-mail: JohnsonGao29@163.com

摘要: 如何准确高效地预测销量是企业一直以来关注的重要问题. 传统的时间序列预测方法虽然在研究和实践中占主导地位, 但是存在一定的局限性. 随着大数据的发展, 电商企业能获取前所未有的数据量和数据特征, 仅利用过去的行为和趋势很难准确地对销量进行预测. 本文提出一种基于随机森林、GBDT、XGBoost 算法的成本厌恶偏向性组合预测模型, 并利用每个商品的成本数据实现对样本的精细化赋权, 进而输出预测结果. 结果表明, 组合预测模型能更精确预测销量, 对电商企业降低商品管理成本有重要意义.

关键词: 销售预测; 机器学习; 组合模型; 特征构建; 样本赋权

引用格式: 韩亚娟, 高欣. 基于机器学习组合模型的电商商品销量预测. 计算机系统应用, 2022, 31(1): 315-321. <http://www.c-s-a.org.cn/1003-3254/8345.html>

E-commerce Commodity Sales Forecast Based on Machine Learning Combination Model

HAN Ya-Juan, GAO Xin

(School of Management, Shanghai University, Shanghai 200444, China)

Abstract: How to accurately and efficiently forecast sales data is an important issue for enterprises. Although the traditional time series prediction method is dominant in research and practice, it has some limitations. With the development of big data, e-commerce enterprises can obtain unprecedented data volume and data characteristics, and it is difficult to accurately predict sales only by using past behaviors and trends. This study proposes a risk aversion-biased combination forecasting model based on the random forest, GBDT, and XGBoost algorithm and used the cost data of each commodity to realize the accurate sample weighting and to output the forecasting results. The experimental results show that the combination forecasting model can predict sales more accurately, which is of great significance for e-commerce enterprises to reduce the cost of commodity management.

Key words: sales forecast; machine learning; combination model; feature construction; samples weighting

销量预测是企业供应链的各个层次中必不可少的环节^[1]. 近年来, 电子商务以及电商物流在互联网时代背景下蓬勃发展, 使交付周期缩短、客户期望增加. 电商企业为了赢得持续竞争优势, 在有限资源下, 销量预测变得更加重要^[2]. 与此同时, 电商企业也能从消费者行为中获得大量数据, 数据成为企业未来发展的核心竞争力, 海量的数据对于企业来说是其独特的优势和资源^[3]. 如何有效地利用数据、精确地对销量进行预

测, 成为了电商企业关注的焦点. 随着大数据和机器学习的快速发展, 合适的算法技术, 将为企业建立预测方案提供关键的技术支持.

许多学者专家在销量预测方面做了系统的研究. 在传统的销量预测方法中, 大多采用了如线性模型、指数平滑等统计方法. 如陈日进^[4]在销售预测中比较了指数平滑与时间序列分解法, 指出指数平滑法受季节性、周期性制约. Papacharalampous 等人^[5]评估了

① 收稿时间: 2021-03-20; 修改时间: 2021-04-29; 采用时间: 2021-06-08; csa 在线出版时间: 2021-12-17

各种统计方法在时间序列的可预测性,并在性能上做了比较,体现出这些方法在需求基本稳定市场中的良好表现。但是这种情况并不适用于现在的预测,因为需求还依赖于除时间外的其他因素,而这些因素并不能有效地用过去的需求量来表示。例如, Uber 或滴滴出行等按需乘车服务不能仅依靠时间序列来估计其需求,它们必须考虑其他因素,如天气条件(湿度、温度等)、一天中的时间段或一周中的某一天^[6]。而传统的统计预测方法通过外推历史趋势和季节波动来预测未来,所以它们通常很难利用对需求产生重大影响特征进行预测^[1]。

因此,包括机器学习和神经网络在内的人工智能的新方法因其增强预测性能和建模非线性模式的能力而受到关注^[7]。由于神经网络在机器视觉、自然语言处理等方面良好表现,人们开始将其运用于销量预测。如 WaveNet^[8]、长短期记忆神经网络(long short-term memory, LSTM)^[9]等。虽然与传统的预测方法相比,神经网络有更好的预测表现,但是其预测结果的可解释性较差。所以预测的结果所能带给企业的实际价值是有限的。与之相对的是,一些学者在研究中展现了随机森林(random forest, RF)杰出的可解释性水平、良好的精度和适当的计算时间^[10,11]。随机森林也被视为预测性分析约定俗成的工具,因为它让管理者了解模型背后的原因,并了解其如何影响最终的结果。此外,运用迭代与梯度提升思想的梯度提升树(gradient boosting decision tree, GBDT)算法在生产与服务性需求预测中表现出比一般模型更好的性能与稳定性^[12]。在基于梯度提升的基础上,极限梯度提升(extreme gradient boosting, XGBoost)算法在工业实践中有着优秀的准确性,在销售预测的研究中验证了其良好的精度^[13]。

在输出预测模型的过程中,学术研究人员和商业实践者经常遇到一个重要问题:是选择合适的建模方法进行预测,还是将这些不同的方法组合成一个单一的预测模型? Lean 等人对候选模型的选择策略与组合策略做了研究,结果发现组合模型的预测效果一般优于个体预测模型,且非线性组合的预测精度会优于线性组合模型^[14]。因此,为进一步优化预测的结果,本文将着眼于电商商品的特点,分析影响电商商品的特征因素,从而构建新的特征集,使用 RF、GBDT、XGBoost 算法建立组合预测模型,对电商商品销量进行预测。

1 相关理论

1.1 RF

RF 是一种基于决策树的回归模型学习算法^[15]。每棵树使用随机选择的特征子集来生长,然后计算在每个树的最后一个节点处获得的预测的平均值,弥补了单个决策树所表现出的偏差低但方差非常大的不足,是一种强大而且不复杂的算法。

1.2 GBDT

GBDT 是一种由迭代思想而来的决策树算法^[16],通过集成基学习器,即 CART 回归树形成强学习器来预测结果。对于一个包含 n 个样本的数据集 $D = \{(x_i, y_i)\} (|D| = n, x_i \in R, y_i \in R)$, 算法具体步骤如下:

(1) 初始化学习器:

$$f_0(x) = \operatorname{argmin}_C \sum_i^n L(y_i, C) \quad (1)$$

其中,初始常数 C 一般设置为样本真实值均值, $L(\cdot)$ 为所选择的损失函数。

(2) 迭代模型,其中迭代次数 $m = 1, 2, \dots, M$:

1) 对于每一个样本 $i = 1, 2, \dots, n$, 计算负梯度,即残差 r_{im} :

$$r_{im} = - \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f(x)=f_{m-1}(x)} \quad (2)$$

2) 将得到的残差更新为样本真实值,并将数据 (x_i, r_{im}) 作为第 m 棵树的训练数据,其对应的叶子节点为 $R_{jm}, j = 1, 2, \dots, J$, J 为回归树的叶子节点数。

3) 对叶子区域 $j = 1, 2, \dots, J$ 计算最优的拟合值:

$$Y_{jm} = \operatorname{argmin}_Y \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + Y) \quad (3)$$

4) 更新强学习器:

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^J Y_{jm} I, x \in R_{jm} \quad (4)$$

其中, I 为指示函数,当 $x \in R_{jm}$ 时值为 1, 否则为 0。

(3) 得到最终的强学习器预测值为:

$$\hat{f}(x) = f_M(x) \quad (5)$$

1.3 XGBoost

XGBoost 是由陈天奇等人提出的集成提升树学习模型^[17]。它高效地实现了 GBDT, 并进行了算法和工程上的许多改进,已经在大量的数据挖掘竞赛中被广泛

地认可,具有高效、灵活且鲁棒性强等优点.对于一个包含 n 个样本的数据集 $D = \{(x_i, y_i)\} (|D| = n, x_i \in R, y_i \in R)$,算法具体步骤如下:

(1) 定义目标函数,由损失函数与正则项两部分组成:

$$L(\varphi) = \sum_i^n l(\hat{y}_i, y_i) + \sum_k^K \Omega(f_k) \quad (6)$$

正则项部分如式(7)所示:

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \omega^2 \quad (7)$$

其中, K 表示共有 K 棵树, f_k 表示第 k 棵树模型, T 表示每棵树的叶子结点数量, ω 表示每棵树的叶子结点的权重值, γ 和 λ 为系数,需要在训练中进行调参.

(2) 模型策略,与GBDT相同,目标函数的求解也是基于迭代思想,对于第 t 次迭代:

$$L^{(t)} = \sum_i^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_k) \quad (8)$$

利用泰勒公式对式(8)进行展开,得到:

$$L^{(t)} = \sum_i^n l\left(g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\right) + \Omega(f_k) \quad (9)$$

其中, $g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)})$, $h_i = \partial_{\hat{y}^{(t-1)}}^2 l(y_i, \hat{y}^{(t-1)})$.

2 基于组合算法的电商商品销售预测模型

2.1 预测模型框架

本文旨在建立一个销售预测模型来预测一段时期的电商销量.为此,将用组合机器学习模型对电商商品的数据进行建模和预测.总体的研究框架如图1所示,主要包括两个部分:数据集处理和组合预测模型构建.

第1部分:数据集处理.对原始数据进行缺失值和异常值的数据预处理以及重新对样本组织与特征构建的过程.

第2部分:组合预测模型构建.对处理过的数据集分别用RF、GBDT、XGBoost模型训练后,组合基础模型的预测结果,并在此基础上利用各个商品的补少补多成本对组合预测的结果赋权,得到最后的预测输出.

2.2 数据集处理

数据集处理的目的是挖掘有效的特征.预测结果的上限由数据和特征所决定,而模型和算法只能不断逼近这个上限.当算法难以突破瓶颈时,优秀的组合特

征通常可以达到良好的预测效果.在本节中,将对数据集进行预处理,并构建新的模型特征集.

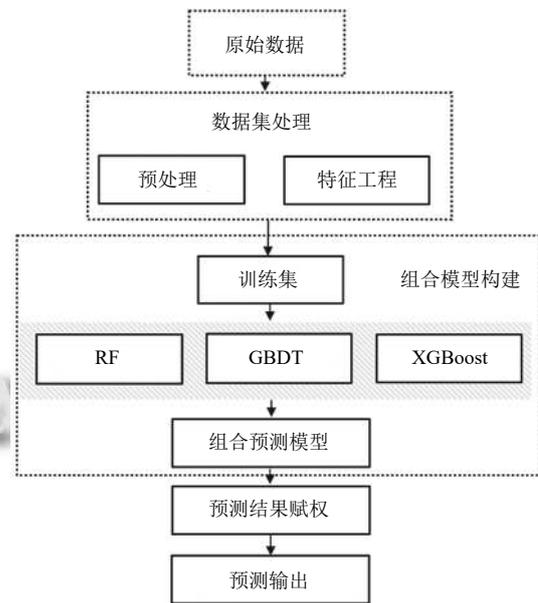


图1 预测流程图

2.2.1 数据来源及预测特点介绍

阿里巴巴旗下电商拥有海量的买家和卖家交易场景下的数据.本文研究的数据集来源于天池平台,由全国仓数据、区域仓数据和商品成本数据组成.数据集的基本信息如下:

(1) 全国仓数据:包含了从2014.10.10–2015.12.27共442天963种商品的210549条销售数据.其中,商品的分类特征有4个,如类目ID、品牌ID等,商品的用户行为特征有25个,如浏览人数、加购物车人数、成交人数等,以及日期、商品ID、仓库CODE特征,一共32维特征.

(2) 区域仓数据:和全国仓没有本质区别,涵盖了从2014.10.10–2015.12.27的相同963种商品的864772条区域仓的销售数据.

(3) 商品成本数据:记录每个商品在全国和区域仓的缺货和库存管理成本,即商品的补少成本 a 和补多成本 b ,用于计算总成本.

电商零售的所有工作都是让正确的产品在正确的时间正确的地点满足正确的消费者.而在多年的电商平台实践中得知,商品销量受众多主客观因素影响.总的来说,电商销量预测问题有以下的特点:

(1) 快速变化的用户需求.消费者需求会随着地

点、时间、特殊事件、个人偏好等各种各样因素而改变,有时一个热点就会导致商品销量的激增或暴跌.这就产生了高度非平稳的销量时间序列.

(2) 集中库存.为了针对消费者需求的快速变化,品牌电商平台会建设或租赁大型仓库对各类商品集中库存以提供对不确定性的缓冲,这使电商销量预测一般具有固定的时间周期.

(3) 成千上万的产品.不同消费者的消费偏好是不同的,为了尽可能满足所有用户需求,大型电商平台会在各种垂直品类间部署不同类型的大量商品,就产生了成千上万的时间序列.

2.2.2 数据预处理

此数据集经过了脱敏处理,和商品的实际销售量、成交金额等有一些差距,但是对整体的数据特性没有影响.但由于数据来源于真实场景,原始数据的完整性、稳定性等有所缺失,会对模型性能造成一定的影响,为此,需要对数据进行预处理以匹配模型的需求.

(1) 缺失值处理

有些商品信息在数据收集的过程中,由于商品的预售或团购等活动,存在一段时间内只有商品的用户行为特征而没有销量信息的情况.此外,后台日志记录的缺失也造成部分商品没有用户行为特征的现象.鉴于不同类型商品的销售周期、用户行为偏好都各不相同,本文统一对缺失值进行补“0”处理,既保证了数据的完整性,也符合电商商品的实际情况.

(2) 异常值处理

图2展示了所有商品销量按天进行统计的结果.

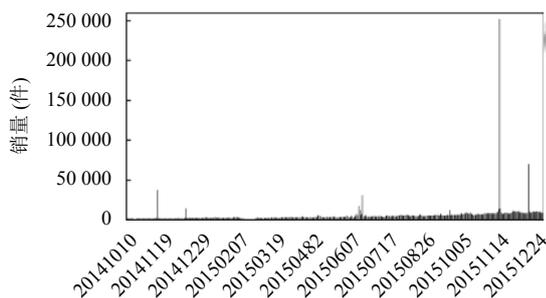


图2 销量直方图

(1) 本数据集的商品销量在该时间序列内较为平稳,没有表现出明显的季节性和周期性波动.可以理解为本数据集所选取的商品具有全面性,销量在时间维度上的变化互相抵消并呈现总体上升的趋势.

(2) 商品销量具有若干突出的局部峰值,这种表现

在2015.11.11尤为明显.结合店商品台的活动可知,峰值出现的日期均为平台促销活动“618”“双11”“双12”当天.因此本文将这几日的数据识别为异常值并予以删除.

2.2.3 样本组织与特征工程

目前,商品销量预测周期多以周销量为最小预测单元^[18,19],考虑到电商商品的需求量大、物流过程复杂等特点,本文以两周作为一个预测单元重新对原始数据进行样本组织,并在此基础上构建新特征集和标签.

时间滑窗法能消除数据噪音并扩展训练集^[20].如图3所示,样本重组织是利用了时间滑窗的方法,以两周为一个滑窗单元,在原始时间序列上按时间逆序进行滑动窗口操作.

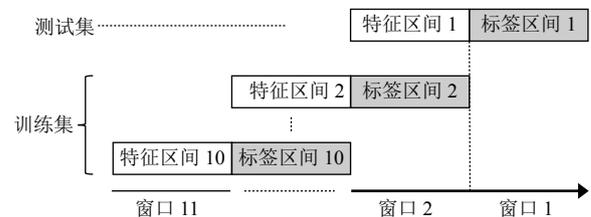


图3 时间滑窗法

由于数据的预测价值会随着离最终预测时间距离的增大而衰减^[19],本文将全国仓和区域仓数据合并,进行了11次前向滑窗,加和每个窗口后两周内各个商品的非聚划算支付件数,即总销量作为标签,使用窗口1,2内数据作为测试集,其余为训练集.

商家信誉、商品性价比与商品历史数据在较大程度影响了商品的销量^[21,22].结合以上因素,本文特征集的构建方法如下:统计窗口前 $N=1, 2, 3, 5, 7, 9, 11, 14$ 天内25维用户行为数据的总和与平均值为用户行为特征;以商品分类特征做聚合,统计各类特征的销量平均值、总和、标准差为商品分类特征;统计窗口内的部分转化率为商品比率特征.在利用时间滑窗法数据集处理并进行样本组织和特征构建后,共产生11个时间窗口,424维特征,46938条数据.新特征集结果如表1所示.

2.3 构建组合销售预测模型

考虑到实际销量的影响因素较多,本文在原始数据的基础上扩展特征集,从而使新数据集获得了更多可以被算法学习的信息.如果采用单一模型预测,可能面临精度下降或是过拟合的风险. Dietterich 指出,模型

结合可能会从统计、计算和表示 3 个方面带来好处^[23]. 因此, 本文利用不同算法的结构差异构建组合销售预

测模型, 在提升精度的同时减少单个模型过拟合的风险. 具体的步骤如下.

表 1 商品特征表

特征因素	特征名称	说明
用户行为特征	sum()_Nmean()_N	窗口前N天内25维用户行为数据加和窗口前N天内25维用户行为数据平均
商品类目特征	supplier_id_qty_alipay_njhs (sum/avg/std)cate_id_qty_alipay_njhs (sum/avg/std)cate_level_id_qty_alipay_njhs_sum(sum/avg/std)brand_id_qty_alipay_njhs (sum/avg/std)	以商品对应供应商ID聚合后销量的统计特征以商品对应类目ID聚合后销量的统计特征以商品对应叶子类目ID聚合后销量的统计特征以商品对应品牌ID聚合后销量的统计特征
商品比率特征	deal_rate success_dealdiscount_ratealipay_njhs_rateqty_njhs_rateamt_njhs_rate	成交人次/流量UV 成交笔数/拍下笔数成交金额/拍下金额非聚划算支付人次/流量UV非聚划算支付件数/成交件数非聚划算支付金额/成交金额

步骤 1. 基础模型训练. 将重新构建样本与特征后的训练集, 分别用随机森林、GBDT、XGBoost 模型训练, 输入测试集特征集进行预测, 得到每个基础模型的预测值, 如式 (10) 所示.

$$\widehat{y}^{(m)} = [\widehat{y}_1^{(m)}, \widehat{y}_2^{(m)}, \dots, \widehat{y}_n^{(m)}]^T \quad (m = 1, 2, 3) \quad (10)$$

步骤 2. 组合预测. 对于电商平台来说, 商品销量预测结果与实际销量差异直接决定了使用补少还是补多成本计算成本, 而不同商品的补少补多成本不同. 因此, 本文将补少补多成本加入销量预测模型建模.

基于每个商品的补多补少成本, 本文提出一种成本厌恶偏向性组合预测方法, 具体如下: 对于每一个商品, 若该商品的补少成本 a_i 小于补多成本 b_i , 说明预测结果比真实值大时会承担更多的成本风险, 则将 3 个模型中的最小预测值做作为组合预测结果, 使模型有预测少的偏向性; 反之, 则将 3 个模型中最大值作为预测结果, 使模型有预测多的偏向性, 如式 (11) 所示.

$$\widehat{y}_i = \begin{cases} \min(\widehat{y}_i^{(1)}, \widehat{y}_i^{(2)}, \widehat{y}_i^{(3)}), & a_i < b_i \\ \max(\widehat{y}_i^{(1)}, \widehat{y}_i^{(2)}, \widehat{y}_i^{(3)}), & a_i \geq b_i \end{cases} \quad (11)$$

步骤 3. 预测赋权. 在组合预测的基础上, 为了进一步提升预测结果对成本厌恶偏向性, 本文对每个样本进行赋权. 如图 4 所示, 由于 Sigmoid 函数有着良好的单调递增等性质, 常被用作神经网络的激活函数.

本文以式 (12) 的 Sigmoid 函数为基础构建赋权函数. 在定义域大于 0 时, 该函数取值范围为 (0.5, 1). 由第 2.2.2 节的分析可知, 商品整体销量较为平稳, 故赋予样本的权重值不应过大, 在式 (12) 基础上加 0.5 构建赋权函数, 如式 (13) 所示. 显然, 补少补多成本间较大与较小的比值始终大于 1, 由此计算出的权重值域为

($S(1) + 0.5, 1.5$), 即约(1.23, 1.5), 满足权重值的稳定性要求.

$$S(x) = \frac{1}{1 + e^{-x}} \quad (12)$$

$$\omega_i = 0.5 + \frac{1}{1 + e^{-\left[\max\left(\frac{a_i}{b_i}, 1\right), \max\left(\frac{b_i}{a_i}, 1\right)\right]}} \quad (13)$$

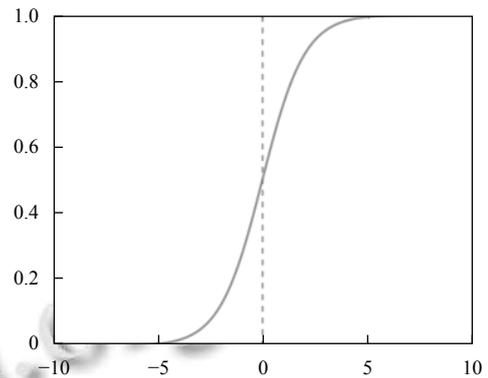


图 4 Sigmoid 函数

步骤 4. 在计算出每个样本的权重值后, 对式 (11) 中成本厌恶偏向性组合预测结果进行赋值, 最终的预测结果如式 (14) 所示.

$$\widehat{y}_i' = \begin{cases} \widehat{y}_i / \omega_i, & a_i < b_i \\ \widehat{y}_i \cdot \omega_i, & a_i \geq b_i \end{cases} \quad (14)$$

3 结果分析

3.1 评价指标

误差度量通过将预测值与实际值进行比较来度量模型的质量. 一般地, 对于预测问题, 有均方误差、平均绝对误差等指标.

由于原始数据中给出了每个商品的补多成本 a_i , 补

少成本 b_i , 因此本文选用预测的总成本作为评价指标, 与传统指标无本质差别, 且使结果更直观具有实际价值。

$$C = \sum_i [a_i \cdots \max(y_i - \hat{y}_i, 0) + b_i \cdot \max(\hat{y}_i - y_i, 0)]$$

3.2 结果分析

为了评价组合模型的预测能力, 本文使用了 RF、GBDT、XGBoost 三个基础模型, 加入特征工程的模型, 以及组合预测模型对实验数据进行预测, 使用总成本作为评价指标, 结果如表 2 所示。

表 2 各模型下的预测总成本

模型	总成本(万元)
RF	177.35
GBDT	191.53
XGBoost	200.25
特征工程+RF	170.78
特征工程+GBDT	174.56
特征工程+XGBoost	180.34
特征工程+组合模型	149.04
特征工程+组合模型+预测赋权	127.43

由表 2 可知:

(1) 特征工程能提升预测准确性。加入特征工程后的模型与基础模型相比, 总成本都有所下降。说明 2.2 节中的特征构建流程从原始数据中增加了有效特征, 并帮助提升了模型准确性。

(2) 组合预测模型有较好的预测准确性。加入特征工程后的组合模型预测为 127.43 万元, 相比于加入特征工程后的基础模型约下降了 42–52 万元。其中, 式 (13) 的预测赋权贡献了约 22 万元的成本下降, 式 (11) 的组合方法贡献了 20–30 万元的成本下降。这展现了组合模型相较于基础模型的优越性和预测赋权方法的有效性。

4 结语

为了更准确预测电商商品的销量, 本文提出了基于机器学习的组合预测模型。通过数据的预处理重新组织样本, 从用户行为、商品分类和商品比率 3 个方面构建了 424 维新特征集, 并有效地使用了商品成本数据, 在本文提出的成本厌恶偏向性组合模型的基础上对预测结果赋权, 得到最终的预测结果。该方法在实现了对平台数据的精细化运用的基础上, 降低了总体仓储成本, 对电商平台提前了解商品销量情况, 合理制

定库存水平有着重要的意义。

参考文献

- Verstraete G, Aghezzaf EH, Desmet B. A leading macroeconomic indicators' based framework to automatically generate tactical sales forecasts. *Computers & Industrial Engineering*, 2020, 139: 106169.
- Boone T, Ganeshan R, Jain A, *et al.* Forecasting sales in the supply chain: Consumer analytics in the big data era. *International Journal of Forecasting*, 2019, 35(1): 170–180. [doi: 10.1016/j.ijforecast.2018.09.003]
- 刘萌, 谷文林. 大数据时代下电子商务的发展刍议. *价格月刊*, 2016, (4): 51–54.
- 陈日进. 销售预测中指数平滑法与时间序列分解法的比较. *统计与信息论坛*, 2004, 19(4): 40–43. [doi: 10.3969/j.issn.1007-3116.2004.04.009]
- Papacharalampous G, Tyralis H, Koutsoyiannis D. Predictability of monthly temperature and precipitation using automatic time series forecasting methods. *Acta Geophysica*, 2018, 66(4): 807–831. [doi: 10.1007/s11600-018-0120-7]
- Ke JT, Zheng HY, Yang H, *et al.* Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation Research Part C: Emerging Technologies*, 2017, 85: 591–608. [doi: 10.1016/j.trc.2017.10.016]
- Karmy JP, Maldonado S. Hierarchical time series forecasting via support vector regression in the European travel retail industry. *Expert Systems with Applications*, 2019, 137: 59–73. [doi: 10.1016/j.eswa.2019.06.060]
- Duan DM. Research on hotel online sales forecast model based on improved WaveNet. *Journal of Physics: Conference Series*, 2020, 1544: 012067. [doi: 10.1088/1742-6596/1544/1/012067]
- Bandara K, Shi PB, Bergmeir C, *et al.* Sales demand forecast in E-commerce using a long short-term memory neural network methodology. *Proceedings of the 26th International Conference on Neural Information Processing*. Sydney: Springer, 2019. 462–474.
- Turgeman L, May JH, Sciulli R. Insights from a machine learning model for predicting the hospital Length of Stay (LOS) at the time of admission. *Expert Systems with Applications*, 2017, 78: 376–385. [doi: 10.1016/j.eswa.2017.02.023]
- Castillo PA, Mora AM, Faris H, *et al.* Applying computational intelligence methods for predicting the sales of newly published books in a real editorial business

- management environment. Knowledge-Based Systems, 2017, 115: 133–151. [doi: 10.1016/j.knosys.2016.10.019]
- 12 焦志伦, 金红, 刘秉镰, 等. 大数据驱动下的共享单车短期需求预测——基于机器学习模型比较分析. 商业经济与管理, 2018, (8): 16–25, 35.
- 13 叶倩怡, 饶泓, 姬名书. 基于 Xgboost 的商业销售预测. 南昌大学学报(理科版), 2017, 41(3): 275–281.
- 14 Yu LA, Wang SY, Lai KK, *et al.* Time series forecasting with multiple candidate models: Selecting or combining? Journal of Systems Science and Complexity, 2005, 18(1): 1–18.
- 15 Breiman L. Random forests. Machine Learning, 2001, 45(1): 5–32. [doi: 10.1023/A:1010933404324]
- 16 Friedman JH. Greedy function approximation: A gradient boosting machine. Annals of Statistics, 2001, 29(5): 1189–1232. [doi: 10.1214/aos/1013203450]
- 17 Chen TQ, Guestrin C. XgBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: ACM, 2016. 785–794.
- 18 吴建材. 零售业商品需求预测体系设计与实证检验. 商业经济研究, 2016, (1): 16–17. [doi: 10.3969/j.issn.1002-5863.2016.01.005]
- 19 张钠, 殷哲, 吕飞, 等. 面向连锁零售业的时间序列预测方法研究. 工业工程与管理, 2014, 19(3): 60–65, 71. [doi: 10.3969/j.issn.1007-5429.2014.03.010]
- 20 Weng TY, Liu WY, Xiao J. Supply chain sales forecasting based on lightGBM and LSTM combination model. Industrial Management & Data Systems, 2019, 120(2): 265–279.
- 21 谢琦, 邵晓峰. 电商大型促销活动下商品销量的影响因素研究——基于天猫“双十一”的实证探究. 上海管理科学, 2018, 40(3): 44–51. [doi: 10.3969/j.issn.1005-9679.2018.03.008]
- 22 孙永波, 刘晓敏. 电商新趋势下影响网络消费者购买行为因素研究. 北京工商大学学报(社会科学版), 2014, 29(4): 93–101.
- 23 Dietterich TG. Ensemble methods in machine learning. Proceedings of 1st International Workshop on Multiple Classifier Systems. Cagliari: Springer, 2000. 1–15.