

加入情感分析的 Stacking 模型在网络剧播放量预测中的应用^①



李明珠, 米传民, 肖琳, 许乃元

(南京航空航天大学 经济与管理学院, 南京 211106)
通信作者: 米传民, E-mail: cmmi@nuaa.edu.cn

摘要: 随着网络剧近年来的飞速发展, 对播放量的研究也逐渐受到关注. 网络剧播放量反映了网络剧的口碑和受欢迎程度, 这与制作方和投资方的收益密切相关. 但目前的研究尚未考虑观众评论的情感态度对播放量的影响, 并且预测模型也较为简单, 预测精度有待进一步提高. 本文在对用户评论进行情感分析的基础上, 构建 Stacking 集成学习模型对我国网络剧的播放量进行预测. 首先基于 SO-PMI 算法构建网络剧领域情感词典, 并结合基础情感词典以及点赞数权重计算出评论情感得分, 加入预测指标体系中; 接着以随机森林 (random forest, RF), GBDT, XGBoost 以及 LightGBM 为基学习器, MLR 为元学习器, 构建 Stacking 网络剧播放量分阶段的预测模型, 使用当前数据对下一周的播放量进行预测; 最后进行模型比较分析, 并得出预测变量的重要性分值. 实验结果显示, 本文所构建的模型判定系数 R 方值达到了 0.89, 高于基学习器单独的模型预测 R 方值 (最高 0.84) 以及未加入情感得分变量的 Stacking 模型预测 R 方值 (0.81). 可以得出加入情感得分变量后, 本文构建的 Stacking 集成学习模型在一定程度上可以提高网络剧播放量的预测精度.

关键词: 网络剧; 播放量预测; 情感分析; 集成学习; Stacking 模型融合; 深度学习

引用格式: 李明珠, 米传民, 肖琳, 许乃元. 加入情感分析的 Stacking 模型在网络剧播放量预测中的应用. 计算机系统应用, 2022, 31(6): 315-323. <http://www.c-s-a.org.cn/1003-3254/8510.html>

Predicting Network Drama Broadcast Volume Based on Sentiment Analysis and Stacking Model

LI Ming-Zhu, MI Chuan-Min, XIAO Lin, XU Nai-Yuan

(College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: With the rapid development of network dramas in recent years, the research on broadcast volume has gradually attracted attention. Broadcast volume reflects the reputation and popularity of a network drama, which are closely related to the profits of producers and investors. However, current research rarely considers the impact of the sentiments in viewers' comments on broadcast volume, and the forecasting models are simple. Consequently, the accuracy of prediction needs to be further improved. After a sentiment analysis of users' comments, we construct a stacking ensemble learning model to predict the broadcast volume of network dramas in China. Using the SO-PMI (semantic orientation-pointwise mutual information) algorithm, we build a sentiment dictionary in the network drama domain. A basic sentiment dictionary and the number of likes are also taken into account to calculate the comment sentiment scores, which are then added into the prediction index system. With random forest (RF), GBDT (gradient boosting decision tree), XGBoost (extreme gradient boosting), and LightGBM (light gradient boosting machine) as base learners and MLR as a meta learner, a stacking prediction model is constructed to predict the broadcast volume of a network drama in stages. The broadcast volume of the next week can be forecasted with data of the current week. Finally, the results of different models are compared and analyzed, and the importance scores of predictive variables are obtained. The experimental results show

^① 基金项目: 国家自然科学基金 (72001106)

收稿时间: 2021-08-28; 修改时间: 2021-09-26; 采用时间: 2021-10-09; csa 在线出版时间: 2022-02-21

that the determination coefficient R-square of the proposed model reaches 0.89, which is higher than that of a single base learner (maximum 0.84) as well as that of the stacking model without sentiment score variables (0.81). It can be concluded that with sentiment score variables, the proposed stacking ensemble learning model delivers better prediction accuracy on the broadcast volume of network dramas than that of traditional models.

Key words: network drama; broadcast volume prediction; sentiment analysis; ensemble learning; stacking model fusion; deep learning

1 引言

网络剧是通过互联网播放的一类连续剧,以手机、平板、计算机等网络设备为播放媒介.我国网络剧诞生于21世纪初,在近20年的时间内经历了由多元化向精品化方向发展的蜕变,越来越受到观众青睐.骨朵传媒数据显示,从2015年起,网络剧平均播放量从2015年的0.72亿次上升至2018年的7.18亿次,平均年增长率为224%^[1].在学术领域,很多学者从制作、播放、营销等方面对网络剧的发展现状和发展路径进行分析^[2,3].从以往的研究可以看出,相比于电视剧,网络剧呈现了灵活性、碎片化、创新性、互动性、多元化等特征,宣传力度、作品质量、明星效应、口碑效果等是影响网络剧播放的重要因素^[2].虽然很多学者对网络剧的研究提出了许多建议,但只是从定性或思辨的角度进行分析.目前已有一些学者从定量分析的角度对网络剧的播放量展开研究:王大员等用多元线性回归MLR模型对低成本网络剧播放量进行预测,最终拟合优度R方为0.455^[4];肖兴辉等从营销节奏的角度,使用微博转发数、评论数和点赞数对营销效果进行量化,运用BP神经网络、RBF神经网络和决策树CHAID算法进行预测,但都呈现出过拟合的趋势^[5];李聪等在对弹幕进行情感分析的基础上,构建网络剧播放量的贝叶斯概率预测模型,但只是提出了构想,没有具体实现^[6].目前对于网络剧播放量的研究还处于初级阶段,尚未考虑观众评论的情感态度对播放量的影响,并且预测模型的选择也有待优化.与此同时,以往的研究都是对最终的播放量进行预测,但由于互联网数据的滞后性以及播放量的动态变化性,无法根据播出时期的实时动态数据对播放量进行调整,因此有必要对网络剧播放量进行分阶段的预测.

对于网络剧播放量预测模型的选择,可以参考票房预测的相关模型.网络剧播放量与票房都属于影视范畴,且有一些共同的影响因子,如主演影响力、豆瓣

评分、相关微博数等^[4].目前票房预测模型可以分为3大类:统计模型,例如线性模型^[7]和概率模型^[8];时间序列模型,例如向量自回归VAR模型^[9];机器学习模型,例如支持向量机SVM^[10]和神经网络^[11].随着集成学习算法的不断发展,集成学习模型也被用于票房预测,如属于Bagging集成思想的Random Forest^[12],属于Boosting集成思想的梯度提升树(gradient boosting decision tree, GBDT)^[13],轻量级梯度提升机(light gradient boosting machine, LightGBM)^[14],极端梯度提升树(extreme gradient boosting, XGBoost)^[15,16]以及Stacking模型融合算法^[14].可以得出,集成学习算法可以通过集成各个机器学习模型的优势,获得比传统的机器学习模型更佳的预测结果.

观众评论的情感倾向是影响其他观众观看的重要因素,一定程度上会对网络剧的播放量带来影响,而以往研究的影响因素并未纳入评论的情感特征.在票房的预测中,很多学者对评论进行情感分析,对评论情感倾向量化从而对票房进行更准确的预测.对于量化方法,有对社交媒体上的评论进行情感分类,直接将正面、负面、中性情感的评论数目作为预测指标^[17-19];也有在得出正面、负面、中性情感数目的基础上通过计算相应的比率(如占总评论的数目,正负情感比)作为预测指标^[20,21];也有通过对情感词和程度词的权重进行赋值计算评论的情感得分对评论情感进行量化^[22,23].评论情感得分的量化方法相比于单纯计算数量和比率更加体现了评论的情感强度大小,扩大了情感数据的差异性.然而,很少有研究将评论的点赞数纳入情感得分的计算中.点赞数代表了一条评论的认可度,已有研究表明用户的点赞行为对票房会产生显著的正向影响^[24].将点赞数融入情感得分的计算中,更能体现所有评论用户的整体情感倾向,提高数据的代表性.

综上所述,对于网络剧播放量的研究仍然处于初级阶段,预测的准确性有待提高.本文通过借鉴票房预

测的相关算法和模型,以国产网络剧播放量预测为研究对象,在对评论进行情感分析得出情感得分的基础上,使用 Stacking 集成学习模型进行预测.在情感分析上,本文基于 SO-PMI 算法构建网络剧领域情感词典,并与基础情感词典合并得到综合情感词典,对评论进行情感分类,再结合修饰词权重和点赞数权重构建了情感得分的计算规则.情感得分与其他的预测变量一起作为网络剧播放量预测的输入变量,从而构建了多维度的网络剧播放量预测指标体系.在预测模型上,本文将集成学习算法 XGBoost, LightGBM, Random Forest 和 GBDT 作为基学习器,多元线性回归 MLR 作为元学习器,进行 Stacking 模型融合,通过整合各个集成学习模型的优势,提高模型的预测性能.

2 相关工作

2.1 SO-PMI 算法

情感倾向点互信息算法 (semantic orientation-pointwise mutual information, SO-PMI) 通常被用于构建领域情感词典^[25],它结合语料库中词的出现概率和句子中词与词同时出现的概率来判断词的情感倾向 (semantic orientation, SO).

在 SO-PMI 算法中,需要先构建两个情感种子词库:积极情感种子词库 $Pwords$ 和消极情感种子词库 $Nwords$,通过计算某个候选词与 $Pwords$ 和 $Nwords$ 的 PMI 的差值 $SOPMI(word)$,将差值的正负性作为情感倾向的判断依据.若 $SOPMI(word) > 0$,则为积极情感词;若 $SOPMI(word) < 0$,则为消极情感词;否则为中性情感词. SO-PMI 计算公式如下:

$$SOPMI(word) = \log_2 \frac{\prod_{Pword \in Pwords} \frac{N \times df(word, Pword)}{df(word) \times df(Pword)}}{\prod_{Nword \in Nwords} \frac{N \times df(word, Nword)}{df(word) \times df(Nword)}} \quad (1)$$

其中, N 表示语料库的总词数, $df(word)$ 表示 $word$ 在语料库中出现的次数, $df(word, Pword)$ 表示 $word$ 和 $Pword$ 在语料库中共同出现的次数,其余类似.

2.2 集成学习模型

集成学习主要分为 Boosting、Bagging、Stacking 三种类型,其中 Boosting 可用于降低偏差, Bagging 可用于降低方差, Stacking 可用于提升预测结果.

随机森林 (random forest, RF) 属于 Bagging 集成算法,它具有原理简单、预测精度高、泛化能力强以

及支持并行计算等优势.但是随机森林的抗噪能力不强,当数据集中出现噪声点时,容易陷入过拟合. GBDT 属于 Boosting 集成算法,适合处理非线性数据;与此同时,通过使用诸如 huber、quantile 等损失函数可以增加模型的鲁棒性.但是由于决策树之间是相互依赖的,难以对数据并行训练,因此训练速度难以提升. XGBoost 和 LightGBM 是对 GBDT 算法的改进, XGBoost 在目标函数中加入了正则项提高泛化能力, LightGBM 采用了基于梯度的单边采样 (GOSS)、互斥特征捆绑 (EFB) 以及基于 leaf-wise 的决策树生长策略等方法,降低时间复杂度,提高模型的预测性能.

Stacking 模型融合是通过将各个算法进行集成,实现优势互补.一般有两层,第 1 层是基学习器预测,第 2 层是元学习器预测.其中元学习器的输入是基学习器的输出.通过元学习器综合基学习器的学习结果,并修正模型预测误差.在 Stacking 集成学习中,通常采用交叉验证的办法防止过拟合.

3 网络剧评论情感得分计算

本文使用基于情感词典的情感分析方法,并构建了情感得分的计算规则.

3.1 网络剧选择

通过查阅猫眼和骨朵影视相关数据,有关网络剧播放过程中的详细数据主要集中在 2017–2021 年.由于网络剧制作成本、营销节奏以及类型等差异,播放量存在很大的差距,为了提高预测的精确性和代表性,本文选择 2017–2021 年列入“骨朵热度指数排行榜 (年榜)”的网络剧,并剔除存在部分缺失值的网络剧,最终得到 128 部网络剧.

3.2 基于 SO-PMI 算法的领域情感词典构建

在构建领域情感词典之前要对评论数据进行爬取.本文通过使用 Python 代码爬取每部网络剧所对应播放平台的评论数据.通过对 128 部网络剧的评论进行爬取,总共获得 1 450 200 条短评及对应的点赞数.领域情感词典构建的流程图如图 1 所示.

由第 2.1 节可知,使用 SO-PMI 算法构建领域情感词典之前需要构建情感种子词库.一般情感种子词库都是人工筛选,为了提高筛选效率,本文使用 TF-IDF 算法先筛选出具有代表性的前 700 个词语,再从中选择情感种子词,降低人工筛选的成本^[26].通过计算候选词与情感种子词的 SO-PMI 值,得到词语的情感倾向.

3.3 综合情感词典构建及权重赋值

通过结合 HowNet 知网词典以及 NTUSD 简体中文情感词典, 以及王勇等创建的否定词词典^[27], 得到综合情感词典, 如表 1 所示. 其中, 下文提到的修饰词即下表中的程度词和否定词.

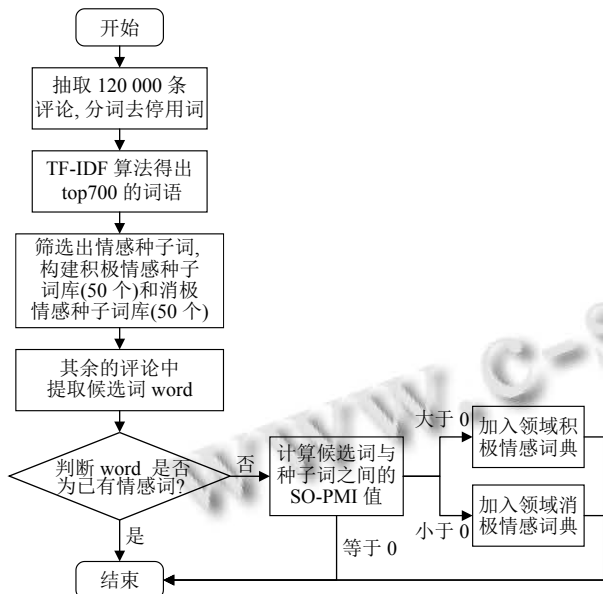


图 1 基于 SO-PMI 算法的领域情感词典生成

表 1 综合情感词典构造

情感词典	分类	示例	词语个数	初始权重
基础情感词典	积极	喜欢、好看	2 872	1
	消极	讨厌、无聊	2 624	-1
程度词典	极度	完全、万分	64	4
	高程度	好、很、太	40	3
	中程度	更加、较	37	2
	低程度	一点、略	30	0.8
	微程度	有点、稍微	11	0.6
否定词词典	否定词	从不、不用	41	-1
领域情感词典	积极	良心剧、重刷	351	1
	消极	弃剧、拖沓	234	-1

通过将基础情感词典和领域情感词典进行汇总, 得到积极情感词 3 223 个, 消极情感词 2 858 个.

3.4 基于情感词典的评论情感分析

在对视频评论数据进行情感分类时, 既要考虑到不同情感词的情感倾向, 也要结合程度词以及否定词的情感权重, 以及评论点赞数对评论代表性带来的影响, 图 2 展示了评论情感分析的总体流程.

本文在构建评论情感词典并进行权重赋值的基础上, 结合点赞数权重构建出评论情感得分的计算方法. 点赞数权重的设置 (如表 2 所示) 借鉴了刘彦虬的用户

点赞数值表^[28].

网络剧评论情感得分的算法流程图如图 3 所示.

假设第 i 条评论经过预处理后得到的词序列为 $W_i = (w_1, w_2, \dots, w_n)$, 在评论情感词典中查找是否存在情感词, 若存在情感词 w_j , 则确定与上一个情感词之间是否存在修饰词. 若有, 则将情感词的初始权重乘以修饰词对应的权重 (若有多个修饰词则一一相乘) 得到情感得分; 否则情感得分即为情感词的初始权重. 情感得分存入临时变量 $Wordscore$ 中.

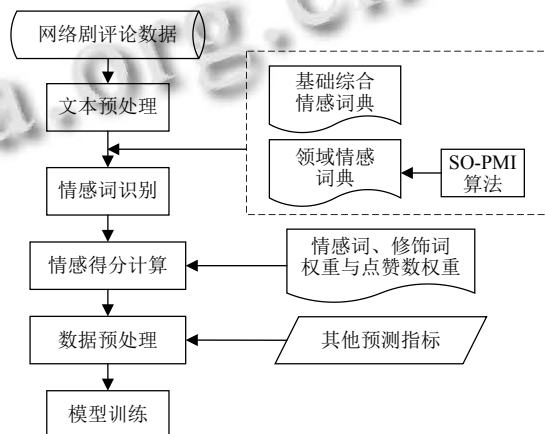


图 2 基于情感词典的评论情感分析流程图

表 2 用户点赞数值表

用户点赞数 n	权重
$10000 \leq n$	2.5
$1000 \leq n < 10000$	2
$100 \leq n < 1000$	1.5
$0 \leq n < 100$	1

如果 $Wordscore > 0$, 则加入积极情感得分变量 $PosScore_i$ 中, 否则加入消极情感得分 $NegScore_i$ 中. 若该评论的词序列全部遍历后既无积极情感词也无消极情感词, 则判断为中性评论, 中性评论情感得分 $NeuScore_i$ 为 1. 将该评论各个情感词情感得分求和即可得到网络剧第 i 条评论的 3 种情感倾向得分.

将每条评论的情感得分乘以该评论点赞数的权重即可得到该评论的综合情感得分: $ComPosScore_i$, $ComNegScore_i$, $ComNeuScore_i$. 整个网络剧的情感得分为网络剧每条评论的综合情感得分之和.

4 网络剧播放量预测模型构建

4.1 预测指标体系构建

由于网络剧播放量是动态数据, 本文对播放量进

行分阶段的预测,即利用前 $n-1$ 周的数据以及前 n 周的集数对第 n 周的播放量进行预测.对于每一阶段的预测模型,都有11个预测变量,其中3个静态变量(与网络剧自身因素有关的变量),8个动态变量(不同预测阶段需要更新的变量).由于网络剧播出前不存在相关播放数据,预测第1周的播放量时缺乏相应信息,因此播放量的预测从第2周开始.表3展示了网络剧播出 $n(n \geq 2)$ 周后播放量的预测指标,表4是对每个预测指标的详细说明.

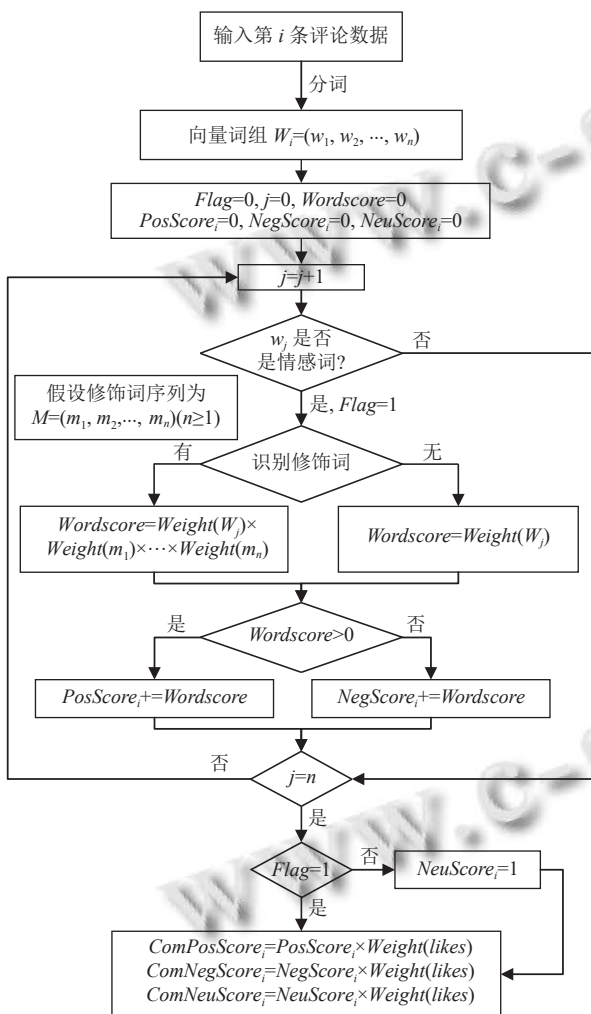


图3 网络剧评论情感得分计算流程图

表3 网络剧播出 $n(n \geq 2)$ 周后的播放量预测指标

指标类别	预测指标
静态变量	Platform, SequelOrIP, WbFans Episodes _n , DramaView _{n-1} , WbNum _{n-1} ,
动态变量	BaiduIndex _{n-1} , DoubanScore _{n-1} , ComPosScore _{n-1} , ComNegScore _{n-1} , ComNeuScore _{n-1}

4.2 Stacking 集成学习预测模型构建

本文基于 Stacking 集成学习模型对网络剧播放量进行预测,期望获得比基学习器单独预测更好的预测结果.本文构建的 Stacking 预测模型的结构如图4所示.将 Random Forest, GBDT, XGBoost, LightGBM 作为 Stacking 集成学习模型的基学习器,线性回归 MLR 作为元学习器.其中 Random Forest, XGBoost, GBDT, LightGBM 模型通过采用集成学习的方法可以提高预测性能,而 MLR 的简单性可以防止预测结果出现过拟合.本文在得出各个基学习器的最优超参数组合的前提下,使用 Python 3.7 进行 Stacking 集成学习模型的建模,具体操作过程如下所示:

(1) 首先将所有数据按照接近 3:1 的比例划分训练集 Training Set 和测试集 Testing Set, 设训练集条, 测试集条, 训练集的标签值记为 y_{train} , 测试集的标签记为 y_{test} .

(2) 对于训练集 Training Set, 使用 5 折交叉验证的方法对每个基学习器进行训练, 将训练集 Training Set 平均分成 5 份, 每次取其中 1 份作为验证集 Testing Data (图4中的 Predict), 其余 4 份作为子训练集 Training Data (图4中的 Learn), 使用各个基学习器训练 5 次, 将 5 次子训练集的预测结果 Predict1 汇总作为元学习器的训练集, 取 5 次对测试集 Testing Set 预测值的平均值 Predict2 作为元学习器的测试集.

(3) 将 4 个模型的预测结果进行汇总, 得到元学习器的测试集和训练集, 此时训练集是一个 $m \times 4$ 的矩阵, 测试集是 $n \times 4$ 的矩阵, 一共 4 个特征值. 再结合标签值 y_{train} 进行模型训练, 对测试集进行预测得到条预测结果 Predictions, 并与实际值 y_{test} 进行比较, 用对应的评估指标 (MSE、MAE、MAPE、R 方值等) 进行预测结果分析.

5 实验验证与结果分析

5.1 数据收集与预处理

由于网络剧从上映到完结的平均时间为一个月, 因此本文收集网络剧播出第1周至播出第4周一共4个阶段的相关数据.对128部网络剧4个阶段的数据进行收集, 总共得到512条记录, 如表5所示.

由表5可知, 由于量纲的差异, 原始数据差别较大, 因此需要在模型训练之前对数据进行归一化, 得到 $[0, 1]$ 范围内的数据, 转化函数为:

$$X^* = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2)$$

表4 网络剧播放量预测指标体系

分类	变量名	含义	数据来源	m 取值范围
自身因素	<i>Platform</i>	网络剧播放平台,在腾讯、优酷、爱奇艺播出的网络剧取值为1,否则为0	骨朵影视数据	静态变量无须 m
	<i>SequelOrIP</i>	是否续集或IP	骨朵影视数据	
	<i>WbFans</i>	领衔主演微博粉丝数	新浪微博	
阶段性特征因素	<i>DramaView_m</i>	截至 m 周的累计播放量	猫眼专业版	$1 \leq m \leq 4$
	<i>Episodes_m</i>	截至 m 周的播放集数	相应播放平台	$2 \leq m \leq 5$
营销因素	<i>WbNum_m</i>	截至 m 周累计提及完整剧名的微博数量	猫眼专业版	$1 \leq m \leq 4$
	<i>BaiduIndex_m</i>	截至 m 周的提及完整剧名的百度指数平均值	百度指数网	$1 \leq m \leq 4$
口碑因素	<i>DoubanScore_m</i>	截至 m 周的豆瓣评分平均值	豆瓣网	$1 \leq m \leq 4$
	<i>ComPosScore_m</i>	截至 m 周的评论积极情感得分	情感得分由第3节得出	$1 \leq m \leq 4$
	<i>ComNegScore_m</i>	截至 m 周的评论消极情感得分		$1 \leq m \leq 4$
	<i>ComNeuScore_m</i>	截至 m 周的评论中性情感得分		$1 \leq m \leq 4$

注:骨朵影视数据是国内领先的网络影视数据分析平台,可以为网络剧提供第三方数据监测和数据评估;猫眼专业版是一款影视行业大数据专业应用工具,在影视界具有相当高的权威性和专业性。

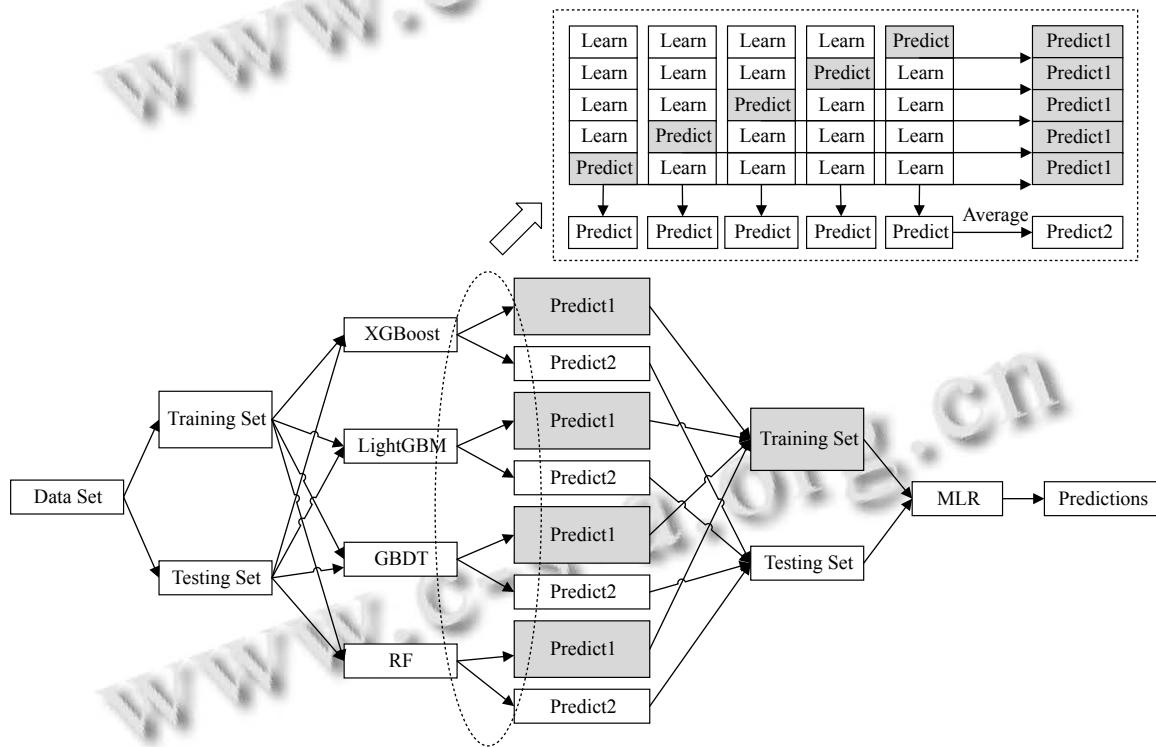


图4 Stacking集成学习预测模型结构

5.2 各单模型预测性能分析

对于 Random Forest、XGBoost、GBDT 以及 LightGBM, 在正式预测前需要对超参数进行调整, 以提高模型的预测性能. 由于本文的样本量不是很大, 因此采用网格搜索的办法调整, 在所有候选的超参数列表中, 通过遍历各种超参数组合形式, 选择预测效果最好的组合作为最终的参数组合^[29]. 将所有数据按照接

近 3:1 的比例划分训练集 Training Set 和测试集 Testing Set, 其中训练集 385 条, 测试集 127 条, 超参数调优使用训练集中的数据. 确定好各个模型的最佳超参数后, 对 127 条测试集中的数据进行预测. 由于本文对于网络剧播放量的预测是回归问题, 因此采用均方根误差 (MSE)、平均绝对误差 (MAE) 和平均相对误差 (MAPE) 以及 R^2 作为模型性能的评估. 各个集成学习模型的预

测结果如表6所示。

对比表中的数据可以看到, LightGBM 在 MSE、MAE、MAPE 以及 R^2 分别为 3.28、1.53、0.23、0.84, 相比于其他 3 个集成学习模型, 预测性能更优, 说明 LightGBM 在 GBDT 和 XGBoost 的基础上使用了 GOSS、EFB 等算法进行优化后, 有助于模型复杂度的降低和模型泛化能力的提高。与此同时, 各单模型预测结果的 R^2 都高于 0.65, 预测结果较好。

表5 变量描述统计结果

变量	个案数	最小值	最大值	平均值
Platform	512	0	1	0.8
SequelOrIP	512	0	1	0.6
WbFans	512	110	12 698	2 618
DramaView	512	3.2	263 000	47 539.5
Episodes	512	8	54	22.4
WbNum	512	1	19 549	1 614.9
BaiduIndex	512	426	577 083	60 842.5
DoubanScore	512	3.1	8.3	6.3
PosScore	512	2 312	46 848	13 695.8
NegScore	512	-16 230	-235.8	-2 990.4
NeuScore	512	347.5	3 967	1 166.8

表6 集成学习单模型预测结果比较表

评价指标	LightGBM	XGBoost	Random Forest	GBDT
MSE	3.28	3.93	4.55	6.28
MAE	1.53	1.65	1.88	2.17
MAPE	0.23	0.24	0.27	0.32
R^2	0.84	0.81	0.78	0.69

5.3 各单模型预测误差相关性分析

要获得好的集成效果, 基学习器需要“好而不同”, 即个体学习器要保证一定的准确性, 与此同时也要有多多样性, 在保证预测精度的同时也具备差异性。一般对各个模型预测结果和真实结果的差值(误差)进行 Pearson 相关系数的检验, 判断是否都是弱相关^[30]。本文使用 SPSS 21.0 进行相关系数的检验。

由表7可知, 上述4种集成学习模型的预测相关性都在 0.65 以下, 总体的相关性较低, 因此可以使用这4种集成学习模型作为 Stacking 集成学习模型的基模型进行训练。

表7 各个模型的 Pearson 相关系数检验

模型	XGBoost	LightGBM	Random Forest	GBDT
XGBoost	1	—	—	—
LightGBM	0.523	1	—	—
Random Forest	0.648	0.328	1	—
GBDT	0.359	0.562	0.319	1

5.4 输入特征重要性分析

为了提高模型的可解释性, 观察重要的特征是否符合业务直觉, 本文对输入变量进行特征重要性分析。特征重要性一般用于观察特征对模型的贡献度, 特征重要性越高说明这个特征是对模型的作用程度越大。对于回归问题, 根据分裂前后节点的纯度 (impurity) 减少量来评估特征重要性。由于 LightGBM 的 R^2 最高、预测效果最好, 因此用其内置函数 feature_importance_ 输出各个特征的重要性数值, 特征重要性的输出结果如图5所示。

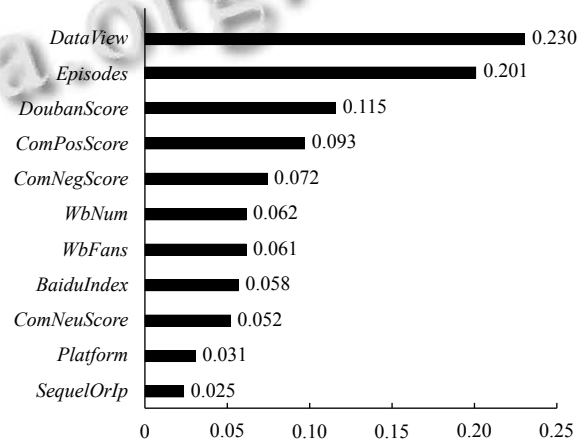


图5 基于 LightGBM 模型的特征重要性排序

对各个变量的特征重要性进行排序, 可以得出 DramaView、Episodes、DoubanScore、ComPosScore 和 ComNegScore 是影响网络剧播放量的关键因素。目前累计播放量和播放集数是预测下一周播放量的基础, 与此同时, 豆瓣评分以及用户评论情感得分代表了该网络剧播出之后的反馈和口碑, 也会对网络剧接下来的播放走势产生一定的影响。

5.5 Stacking 集成学习模型预测性能分析与比较

确定好基学习器后, 用 Python 3.7 按照第 4.2 节的步骤编写 Stacking 集成学习建模程序。其中训练集 385 条, 测试集 127 条, 使用 5 折交叉验证的方法对每个基学习器进行训练。将 Stacking 模型的预测结果与基学习器单独的预测结果进行比较, 结果如表8所示。

表8 Stacking 集成学习模型与各单模型预测结果比较

评价指标	Stacking	LightGBM	XGBoost	Random Forest	GBDT
MSE	2.32	3.28	3.93	4.55	6.28
MAE	1.33	1.53	1.65	1.88	2.17
MAPE	0.19	0.23	0.24	0.27	0.32
R^2	0.89	0.84	0.81	0.78	0.69

Stacking 集成学习模型的 MSE、MAE、MAPE 均小于其他 4 个集成学习单模型, 且决定系数 R^2 达到了 0.89, 可见 Stacking 集成学习模型综合了其他 4 个基学习器的优势, 获得更高的鲁棒性和泛化能力. 和前人的播放量预测研究相比, 本文的决定系数 R^2 高于王大员等使用多元线性回归预测的 R^2 值 (0.455)^[4], 而肖兴辉等使用 BP 神经网络预测时在测试集上的线性回归系数为 0.976, 但在测试集上只有 0.288, 产生了过拟合^[5], 因此可以得出本文的模型与前人的模型相比具有更优的预测性能. 图 6 展示了 Stacking 集成学习模型的预测值和真实值的拟合曲线.

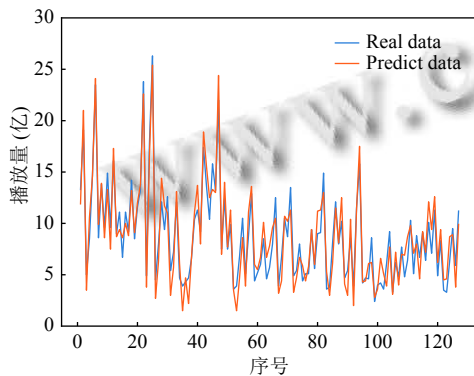


图 6 Stacking 模型预测实际值与预测值拟合曲线

Stacking 集成学习模型在特殊值上都取得了较好的预测效果, 改善了各个基学习器在部分预测结果上的不足, 取得较好的预测性能.

为了研究本文的情感得分输入变量对整个模型预测性能的影响, 本文使用剔除情感得分变量的 Stacking 集成学习模型进行训练和预测, 最终预测结果如表 9 所示.

表 9 加入情感得分变量前后 Stacking 预测结果比较

指标	未加入情感得分变量	加入情感得分变量
MSE	3.92	2.32
MAE	1.69	1.33
MAPE	0.24	0.19
R^2	0.81	0.89

从表中可以看到, 未加入情感得分变量的模型预测 MSE、MAE、MAPE 值均大于加入情感得分变量的值, 且 $R^2=0.81 < 0.89$, 这表明本文情感得分变量的加入有助于提高网络剧播放量的预测精度, 也说明了观众评论的情感倾向对播放量的大小起到一定的作用.

6 总结与展望

本文立足于当下网络剧兴起的时代大背景以及学术界对网络剧进行定量分析的空缺, 并借鉴对电影票房预测的方法, 对国产网络剧播放量进行预测. 通过创建情感词典对网络剧评论进行情感分析, 创新性地输入特征中加入情感得分变量 (积极情感得分、消极情感得分和中性情感得分), 运用 Stacking 集成学习模型进行网络剧播放量的预测. 通过比较集成学习模型与基学习器 (XGBoost、LightGBM、Random Forest、GBDT) 单独预测结果的差异, 总结出 Stacking 集成学习模型的预测精度更高, R^2 达到了 0.89, 并且高于未加入情感得分特征的预测模型 R^2 值 (0.81), 实验证明加入情感得分变量后使用 Stacking 集成学习模型进行预测取得了不错的预测效果. 本文从定量研究的角度分析网络剧的播放趋势以及影响网络剧播放量的关键因素, 提高了分析的客观性、准确性和针对性, 为网络剧的定量研究提供了新的思路.

由于有些数据并不公开, 增加了获得其他影响因素的数据的难度, 例如投入成本. 因此, 未来的研究可以努力与网络剧的相关机构合作, 获得尚未在相关平台公开的数据. 此外, 有些指标目前较难量化, 如演技、服装、音乐、特效等, 在后续研究中, 可以尝试通过调研将其转化为相应的打分或点赞数等量化指标, 对预测指标体系做进一步的完善.

参考文献

- 1 王康. 中国网络剧发展的动力机制. 青年记者, 2019, (16): 55-57.
- 2 吕曠如. 浅谈网络剧相对于电视剧的创新和发展. 科技传播, 2019, 11(11): 105-107. [doi: 10.3969/j.issn.1674-6708.2019.11.069]
- 3 朱恺劭. 媒介融合视角下网络剧如何发展. 戏剧之家, 2019, (8): 118. [doi: 10.3969/j.issn.1007-0125.2019.08.086]
- 4 王大员, 徐敏. 低成本网络剧的营销策略分析——基于网络剧播放量的实证研究. 新闻世界, 2019, (2): 60-64.
- 5 肖兴辉, 李波. 营销节奏对网络剧播放量的影响因素探析. 现代传播 (中国传媒大学学报), 2019, 41(7): 133-136.
- 6 李聪, 余梦, 汪之舟, 等. 基于弹幕大数据的网络视频精准营销模式研究. 市场周刊, 2020, (6): 90-91.
- 7 Litman BR. Predicting success of theatrical movies: An empirical study. Journal of Popular Culture, 1983, 16(4): 159-175. [doi: 10.1111/j.0022-3840.1983.1604_159.x]
- 8 Chance DM, Hillebrand E, Hilliard JE. Pricing an option on

- revenue from an innovation: An application to movie box office revenue. *Management Science*, 2008, 54(5): 1015–1028. [doi: [10.1287/mnsc.1070.0826](https://doi.org/10.1287/mnsc.1070.0826)]
- 9 Feng N, Feng HH, Li DH, *et al.* Online media coverage, consumer engagement and movie sales: A PVAR approach. *Decision Support Systems*, 2020, 131: 113267. [doi: [10.1016/j.dss.2020.113267](https://doi.org/10.1016/j.dss.2020.113267)]
- 10 Quader N, Gani MO, Chaki D, *et al.* A machine learning approach to predict movie Box-Office success. 2017 20th International Conference of Computer and Information Technology (ICCIT). Dhaka: IEEE, 2017. 1–7.
- 11 Zhang L, Luo JH, Yang SY. Forecasting box office revenue of movies with BP neural network. *Expert Systems with Applications*, 2009, 36(3): 6580–6587. [doi: [10.1016/j.eswa.2008.07.064](https://doi.org/10.1016/j.eswa.2008.07.064)]
- 12 米传民, 鲁月, 林清同. 基于加权 K-Means 和局部 BPNN 的票房预测模型. *计算机系统应用*, 2019, 28(2): 15–23. [doi: [10.15888/j.cnki.csa.006709](https://doi.org/10.15888/j.cnki.csa.006709)]
- 13 Wu SY, Zheng YF, Lai ZK, *et al.* Movie box office prediction based on ensemble learning. 2019 IEEE Symposium on Product Compliance Engineering—Asia (ISPCE-CN). Hong Kong: IEEE, 2019. 1–4.
- 14 Liao Y, Peng YX, Shi SL, *et al.* Early box office prediction in China's film market based on a stacking fusion model. *Annals of Operations Research*, 2020: 1–18.
- 15 李旺泽. 基于监督学习的国产电影票房影响因素研究 [硕士学位论文]. 武汉: 湖北工业大学, 2020.
- 16 杨曼. 基于 XGBoost 算法的电影票房预测 [硕士学位论文]. 兰州: 兰州大学, 2020.
- 17 Apala KR, Jose M, Motnam S, *et al.* Prediction of movies box office performance using social media. 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013). Niagara Falls: IEEE, 2013. 1209–1214.
- 18 王晓耘, 袁媛, 史玲玲. 基于微博的电影首映周票房预测建模. *现代图书情报技术*, 2016, (4): 31–39.
- 19 朱梦娴. 口碑与票房: 社会化媒体电影评论的商业价值研究. *信息资源管理学报*, 2015, 5(4): 57–64, 80.
- 20 Ahmad IS, Bakar AA, Yaakub MR, *et al.* Sequel movie revenue prediction model based on sentiment analysis. *Data Technologies and Applications*, 2020, 54(5): 665–683. [doi: [10.1108/DTA-10-2019-0180](https://doi.org/10.1108/DTA-10-2019-0180)]
- 21 Hur M, Kang P, Cho S. Box-office forecasting based on sentiments of movie reviews and Independent subspace method. *Information Sciences*, 2016, 372: 608–624. [doi: [10.1016/j.ins.2016.08.027](https://doi.org/10.1016/j.ins.2016.08.027)]
- 22 Gaikar DD, Marakarkandy B, Dasgupta C. Using Twitter data to predict the performance of Bollywood movies. *Industrial Management & Data Systems*, 2015, 115(9): 1604–1621.
- 23 Qiu XY, Tang TY. Microblog mood predicts the box office performance. *Proceedings of the 2018 Artificial Intelligence and Cloud Computing Conference*. Tokyo: ACM, 2018. 129–133.
- 24 Ding C, Cheng HK, Duan Y, *et al.* The power of the “like” button: The impact of social media on box office. *Decision Support Systems*, 2017, 94: 77–84. [doi: [10.1016/j.dss.2016.11.002](https://doi.org/10.1016/j.dss.2016.11.002)]
- 25 Zhao M, Zhang TZ, Chai JP. Based on SO-PMI algorithm to discriminate sentimental words' polarity in TV programs' subjective evaluation. 2015 8th International Symposium on Computational Intelligence and Design (ISCID). Hangzhou: IEEE, 2015. 38–40.
- 26 Kim J, Moon N. Rating and comments mining using TF-IDF and SO-PMI for improved priority ratings. *KSII Transactions on Internet and Information Systems*, 2019, 13(11): 5321–5334.
- 27 王勇, 吕学强, 姬连春, 等. 基于极性词典的中文微博情感分类. *计算机应用与软件*, 2014, 31(1): 34–37, 126. [doi: [10.3969/j.issn.1000-386x.2014.01.010](https://doi.org/10.3969/j.issn.1000-386x.2014.01.010)]
- 28 刘彦虬. 基于 HowNet 和用户点赞的中文电影评论情感分析研究 [硕士学位论文]. 呼和浩特: 内蒙古大学, 2019.
- 29 张公让, 鲍超, 王晓玉, 等. 基于评论数据的文本语义挖掘与情感分析. *情报科学*, 2021, 39(5): 53–61.
- 30 朱国森, 郑晓亮. 基于 stacking 集成模型的网络流量预测研究. *重庆工商大学学报 (自然科学版)*, 2021, 38(2): 16–22.