

基于XGBoost的电影票房影响因素分析及预测研究

韩淑淑

山东工商学院统计学院, 山东 烟台

收稿日期: 2024年3月28日; 录用日期: 2024年4月23日; 发布日期: 2024年4月30日

摘要

在我国电影产业中, 电影票房是整个产业收益的主要来源, 对票房进行准确预测对优化电影投融资, 助力电影经营主体决策优化, 促进整个电影产业的健康发展起着重要作用。本文主要构建了一种基于XGBoost算法筛选指标以及加入粒子群优化算法的BP神经网络的票房预测模型。首先, 构建一个更全面的电影票房影响因素体系, 加入微博因素和电影首日影评作为票房影响因素指标, 同时结合电影特征和市场因素构建电影票房影响因素体系; 其次, 对各指标因素预处理量化后, 为简化后期运算和提高模型的精度, 构建基于XGBoost的影响力测量模型, 并以此为依据进行筛选; 最后, 将筛选后的指标体系划分两部分即训练集和测试集, 并在此基础上分别构建了BP神经网络模型、RBF模型以及PSO-BP模型, 并引入评价指标和对案例电影预测精度进行分析, 结果表明本文构建的PSO-BP模型具有更高的预测精度。本研究所构建的模型在电影上映期间预测最终票房具有一定的参考意义, 可为有关部门提供决策参考。

关键词

电影票房预测, 粒子群算法, XGBoost算法, PSO-BP神经网络模型

Analysis and Prediction of Influencing Factors of Movie Box Office Based on XGBoost

Shushu Han

School of Statistics, Shandong Technology and Business University, Yantai Shandong

Received: Mar. 28th, 2024; accepted: Apr. 23rd, 2024; published: Apr. 30th, 2024

Abstract

In China's film industry, the movie box office is the main source of revenue for the entire industry, and the prediction of the movie box office can optimize movie investment and financing, help the movie business entities to optimize their decision-making, and promote the whole film industry. This paper mainly constructs a box office prediction model based on XGBoost screening index and BP neural network with particle swarm optimization algorithm. Firstly, a more comprehensive box office influencing factor system is constructed, in which Weibo factor and the first day film review are added as box office influencing factor indicators, and at the same time, a box office influencing factor system is constructed by combining film characteristics and market factors; Secondly, after preprocessing and quantifying each index factor, in order to simplify the later operation and improve the accuracy of the model, the influence measurement model of influencing factors based on XGBoost algorithm is constructed, and the box office influencing factors are screened on this basis; Finally, in the empirical part of the movie box office prediction model, the data includes two parts: the training set and the test set, and the BP neural network final model, RBF model and PSO-BP model are constructed. The evaluation index is introduced and the prediction accuracy of case movies is analyzed. The results show that the PSO-BP model constructed in this paper has higher prediction accuracy. The model constructed in this study has certain reference significance in predicting the final box office during the film release period, and can provide decision-making reference for relevant departments.

Keywords

Box Office Forecast, Particle Swarm Optimization, XGBoost Algorithm, PSO-BP Neural Network Model

Copyright © 2024 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

近年来,中国的电影产业正在进入快速发展时期。电影产业作为我国文化产业的重要组成部分,促进了我国国民经济消费水平的发展。电影作为一种文娱项目,不仅丰富了人们的娱乐生活和精神世界,同时还带来了巨大的社会效益和经济收益。对国产电影总票房进行有效预测,有利于帮助影视投资人提高决策水平,从而提升电影制作效率,以推动中国电影事业的健康发展。

结合当前国内外文献,关于电影票房的研究方向主要有两个,一是票房影响因素的研究;二是票房预测模型的研究。西方学术界早在上世纪80年代便已开始电影票房的研究,经过长达几十年的探究分析,其在电影票房的研究方面相对成熟。虽然我国起步时间晚,但我国拥有庞大的市场规模,经过近几十年的发展,我国已在2020年超越北美,成为世界第一大电影市场,吸引了越来越多的学者广泛关注并开展研究。

在电影票房影响因素方面,研究学者在早期研究中侧重于电影特征和市场因素,主要包括:电影的制作方式、类型、主演、导演、编剧、电影上映档期、电影发行公司等。胡晓红和王红(2018)在预测电影票房主要围绕分析导演影响力、主演影响力、是否是黄金档期、想看人数、发行国家、影片类型、影片时长等因素展开的[1]。何晓雪等人(2019)对比研究电影票房与首日票房、微博想看人数、豆瓣评分之间

的影响[2]。Sochay (1994)从三个角度即创意、发行模式和市场推广来分析电影票房影响因素[3]。陈邦丽等人(2018)从电影制作和营销宣传两方面做了系统分析[4]。Julian Hofmann (2016)主要探讨了明星影响力对电影票房影响,得出顶级流量明星和高票房之间呈现一种正向性关系的结论[5]。另外,随着网络的普及和蓬勃发展,在线评论等消费者因素借助于互联网应运而生并且发挥的作用日益重要,成为占据票房影响因素体系的一席之地。申林和王靖舒(2020)实证分析了网络评价和票房呈正相关关系的趋势[6]。Minhoe Hur (2016)通过影评情感和独立子空间方法研究票房预测[7]。谢治海(2020)构建了一种 MRS-VIS 模型研究影评情感对票房影响及预测[8]。吴珏(2018)将用户在社交平台上的阅读行为,作为电影上映前后衡量观影期待与消费意向的量化指标,发现用户关于特定电影的深度互动行为与票房结果显著相关[9]。史伟(2015)考虑到微博影响因素的重要性,从微博评论中挖掘情感信息从而建立微博情感计算方法来提高预测票房的精度[10]。

在电影票房预测模型研究中。最开始时主要通过线性回归来进行预测,例如 Litman (1989)采取多元线性回归的方法,是预测模型中早期最为典型的,其研究提供的方法为后来电影票房的预测研究奠定了基础[11]。但随着技术发展,目前线性回归模型主要用于验证新变量的引入。Ting Liu (2014)等人采用线性回归模型与非线性回归模型来进行电影票房预测,结果表明非线性模型的票房预测结果相较于线性模型的效果好[12]。张雪(2017)使用了线性回归和神经网络方法进行票房预测,证明神经网络的预测效果更好[13]。但随着技术的发展,越来越多的学者把机器学习应用于在票房预测上,例如申林(2020) [6]和 Minhoe Hur (2016) [7]。有部分学者侧重于研究机器学习和线性回归模型两种模型对票房预测的精度,通过对比实验发现,机器学习模型要优于线性回归。例如杨威(2015)建立了多元线性回归、SVM 和 BP 神经网络模型对票房进行预测,分析得出机器学习的模型优于线性回归,证实了上述观点[14]。目前,大多数学者在进行票房预测研究时多以神经网络为主,尤以 BP 神经网络最为流行。如米传民(2019)通过构建一种基于随机森林影响因素测力算法以及局部 BPNN 的票房预测模型[15]。

通过对其相关文献的梳理和总结可以发现,一方面电影票房预测模型的研究从最早的多元线性回归到机器学习方法再到现在的深度神经网络,模型的预测准确率与泛化能力不断提升,但是随着模型逐渐复杂,模型中的参数需要根据实际情况进行优化,进而有望实现该模型在针对不同数据时达到最优化的目标;另一方面,研究者不断完善电影票房影响因素指标体系,以期来提高预测的准确度,但是随着票房市场影响指标的增多,模型的预测精度并没有达到理想效果以及模型的指标因素的可解释性明显降低。针对上面问题,首先,本文在构建模型的同时引用粒子群优化算法进行优化模型;其次,在引入影响指标体系时,通过利用 XGBoost 模型的重要性特征图筛选主要影响因素,从而简化后期模型运算,提高模型特征变量的可解释性;最后,在电影票房预测模型的实证部分将数据划分训练集和测试集两部分,构建了 BP 神经网络模型、RBF 模型以及 PSO-BP 模型,并引入评价指标和对案例电影预测精度进行分析,结果表明本文构建的 PSO-BP 模型具有更高的预测精度。

2. 指标体系的构建

本研究从电影特征因素、市场因素、消费者因素和微博因素 4 个维度共选取 18 个指标构建电影票房影响因素体系,各指标数据均来自于艺恩网、中国票房网、豆瓣网、时光网、微博官网以及百度指数等网站。由于我国电影市场是从自 2010 年以后迅速发展,2010 年之前的电影数据质量层次不齐,收集数据难度大,研究分析也存在一定难度;而 2010 之后的电影,不管在数量方面还是票房方面都有了极大的提升。所以本文研究样本为 2012 年 1 月 1 日至 2022 年 12 月 31 日期间在中国上映的电影,通过对数据筛选、预处理以及分析后最终选取了 1305 个样本,数据量有保证,横跨年份较长,具有较强的代表性。

在电影特征方面,电影产品的感知易用性主要体现在观影体验与情感价值方面,综合了电影的审美

性、娱乐性等功能给消费者带来直观的体验与感受，电影本身的制作水平与观看方式直接决定了观众的感知易用性，在指标选取时，将电影类型、电影时长、上映档期、电影制式、影片类型以及阵容等因素加入评估指标中。在市场因素方面，电影作品的感知有用性主要体现在电影市场和消费者因素两方面，市场因素主要受电影的宣传力度、银幕数和潜在消费者的影响，在选取特征变量时，加入想看人数、银幕数、发行公司以及首日票房加入评估指标中。在消费者体系方面，在我国电影产业中，电影票房是整个产业收益的主要来源，消费者的态度直接影响着电影票房的收入，而且消费者对电影的网络检索数值越高，则说明该电影的热度越高，也就代表着其潜在消费者多，进而影响着电影的票房。目前网络搜索指数众多，国外学者使用的网络搜索数据大多来自 Google 搜索引擎，而对于中文内容的搜索而言，百度指数在一众指数搜索中脱颖而出，更具代表性。所以选择把消费者影评的情感分析和百度搜索检索量这两个因素加入这一影响体系中。在微博因素方面，近几年，微博在我国凭借其独特信息的简短性和发布的实时性等优点在一众社交平台中崭露头角。随着其地位的提升，越来越多的学者在研究票房因素指标中愈发重视这一因素，本文也考虑到这一现实因素，把微博中关于影片的话题讨论量也加入到票房影响因素体系中。

在对指标进行量化时，需要对文本指标进行转化。受篇幅影响本文只列举前两位指标的具体计算。电影类型量化：考虑到各种电影题材的质量层次不齐，进而导致票房的高低，如果只采取简单的取均值来代表这一类型的影响力，无法保证类型影响力的合理性，对此采取将各题材电影平均票房与各题材电影占比相乘的方法对电影题材因素进行量化，量化公式为：

$$G_i = \frac{\sum_{j=1}^{n_i} box_j}{n_i} \times p_i \quad (1)$$

电影档期处理过程：过对前人文献总结的基础上，把电影档期划分为五类，即春节档(1/20~2/20)、暑假档期(7/1~9/1)、国庆档期(9/30~10/10)、跨年档(2/25~1/5)和普通档。前人在研究档期时大多引入哑变量，该方法固然易懂便于计算，但没有充分考虑到档期之间的差异性，因此可以引用电影票房对档期变量进行量化，计算方法如下：

$$D_i = \frac{\sum_{j=1}^{n_i} box_j}{n_i} \quad (2)$$

其中， G_i 表示第 i 个电影类型的影响力， D_i 表示的是第 i 个档期的影响力， box_j 表示第 j 个电影的票房， p_i 为该题材电影的占比， n_i 表示第 i 个电影的档期数量。

3. 预测模型的构建

BP (back propagation)神经网络是由 Rumelhart 和 McClelland 提出的，是一种按照误差逆向传播算法训练的多层前馈神经网络。BP 神经网络的算法包含两个过程：前向传播和反向传播。BP 网络是一种具有三层或三层以上的神经网络，包括输入层、中间层(隐层)和输出层。上下层之间实现全连接，而每层神经元之间无连接。粒子群优化算法(Particle Swarm optimization, 简称 PSO)，最初是由 Eberhart 博士和 Kennedy 博士提出并成功地用于函数优化。PSO 算法凭借其所具有的十分简单的原理，较少的参数，快速的收敛和操作简单易实现等优点，目前已经被广泛应用于神经网络训练等多个领域。

PSO 优化 BP 神经网络的基本步骤是：首先通过利用 PSO 算法的优点——全局搜索能力对 BP 神经网络初始的权值和阈值进行优化，以寻取最优值；其次再通过适应度函数来判断粒子是否到达最佳的位置，如果未达到最佳位置则利用粒子寻优不断优化位置，以此达到最佳位置从而得到最佳的权值和阈值，最

后将求取的最优值通过 BP 神经网络算法求解。粒子群算法改变了传统 BP 神经网络的梯度下降法原理，通过粒子群算法增强了个体之间的学习能力和信息分享能力，达到更快更好得到最优解的目的。

本研究构建的 XGBoost 和 PSO-BP 票房预测模型，模型的步骤介绍如下：

步骤 1：贡献度。对样本 (x_i, y_j) ，其中 $i=1, 2 \dots m$ ，计算负梯度 r_i ，从而求出 (x_i, r_i) ，然后拟合分类与回归树得到第 t 颗回归树，其对应叶子节点区域为 R_j ，其中 $j=1, 2 \dots J$ ，其中为 J 回归树 t 的叶子节点个数。为使目标函数最小，通过贪心算法遍历所有特征的划分点，在叶子节点的分裂过程中，选取最大的特征及其切分点作为最优特征和最优切分点进行分裂。对叶子区域 $j=1, 2 \dots J$ ，计算最佳拟合值 c_{ij} 。进而更新强学习器，求出各个自变量对因变量的贡献度。其作为衡量变量重要性的指标，值越高说明该变量对因变量越重要，从而进行指标筛选。

步骤 2：初始化。由原始数据为起点设置神经网络的参数值先初始化各数值以及粒子个数。再根据公式 $H=(n+1) \times (m+1) \times s$ 计算 BP 神经网络所需优化的权值阈值，其中 n 、 s 、 m 分别代表输入层、输出层、隐含层的神经元个数。确定粒子群算法适应度值，确定粒子群内粒子最优个体位置以及全局最优位置。

步骤 3：主循环。依据从小到大顺序排序粒子群内粒子，保留适应度值较高的数量为 $N/2$ 的粒子，用 N 表示所设定粒子总数量，当 N 与所保留粒子总数量相同时，利用公式 $V_{ij}(t+1)=\omega V_{ij}(t)+C_1 R_1(P_j(t)-X_{ij}(t))+C_2 R_2(G_j(t)-X_{ij}(t))$ 和 $X_{ij}(t+1)=X_{ij}(t)+V_{ij}(t+1)$ 更新粒子位置以及粒子运动速度，其中 V_{ij} 、 X_{ij} 和 ω 分别表示粒子速度、粒子位置获取粒子和惯性权重，从而获取相应适应度值；判断误差是否低于 ε 以及是否达到最大迭代次数，当满足以上要求时，将所输出粒子信息设置为 BP 神经网络初始阈值以及初始权值。

步骤 4：预测。输出计算出的 BP 神经网络权值与阈值，并将数据带入训练网络中，可得到最终的票房预测结果。

4. 实证分析

4.1. 影响因素筛选

利用 XGBoost 模型筛选主要影响因素，主要是通过计算出各个自变量在 XGboost 模型中对电影票房预测的贡献度，其值越高说明该变量对因变量越重要，把贡献度值作为衡量变量重要性的指标，根据该指标数值的高低进行筛选变量。在进行票房影响因素的选择时只选取影响力较大的因素，剔除贡献数值相对较小的变量，从而在输入层对预测模型进行简化[15]。但由于 XGBoost 随机性的特征，当对因变量的重要性分数进行求解时，各个自变量的重要性分数在多次试验中得出的结果是有差异的，但通过观察多次实验结果发现各个指标的贡献度数值在一定范围内上下波动。因此，本文为使数据更具有稳定性采取多次试验求均值的方法对指标因素重要性特征进行选取，具体票房影响因素筛选思路如下所示：

Step1：通过对 XGBoost 进行多次构建以求解指标体系中各自变量的多组贡献度值；

Step2：按照上述步骤所求出的多组指标因素的重要性分数再求取各指标的平均值，并对并按照从大到小的顺序对得到的最终重要性分值排序；

Step3：根据筛选原则和现实情况选取重要性数值较大的若干个影响因素，从而完成基于 XGBoost 的影响力指标的筛选，得出最终的票房影响因素集合，并在此基础上进行后续票房的预测。

按照上述步骤将特征重要性按照得分大小排序，结果如图 1 所示。在电影票房影响因素中，从总体来看，首日票房重要性分值最高，则说明其重要性程度远高于其他变量；电影评论因素、演员 1、导演和演员 2 的影响力分值名列前茅，由此可见，可见消费者参与的评论以及影片主创团队的创作核心能力在电影票房影响因素体系发挥着举足轻重的作用。经观察发现在图 1 中出现三个明显的断层，由上到下

依此定为断层 1、断层 2 和断层 3，其中在电影制式和演员 3 影响力之间出现的第 3 个断层，其累计重要性数值占据总值的 95%左右，能代表着因变量影响因素体系中绝大部分信息，所以以此为界限进行变量的筛选。因此筛选后的票房影响因素共包含首日票房、百度指数、导演、主演 1、想看人数、主演 2、短评情感倾向、微博话题讨论量、屏幕数、上映档期、编辑和电影制式等因素。

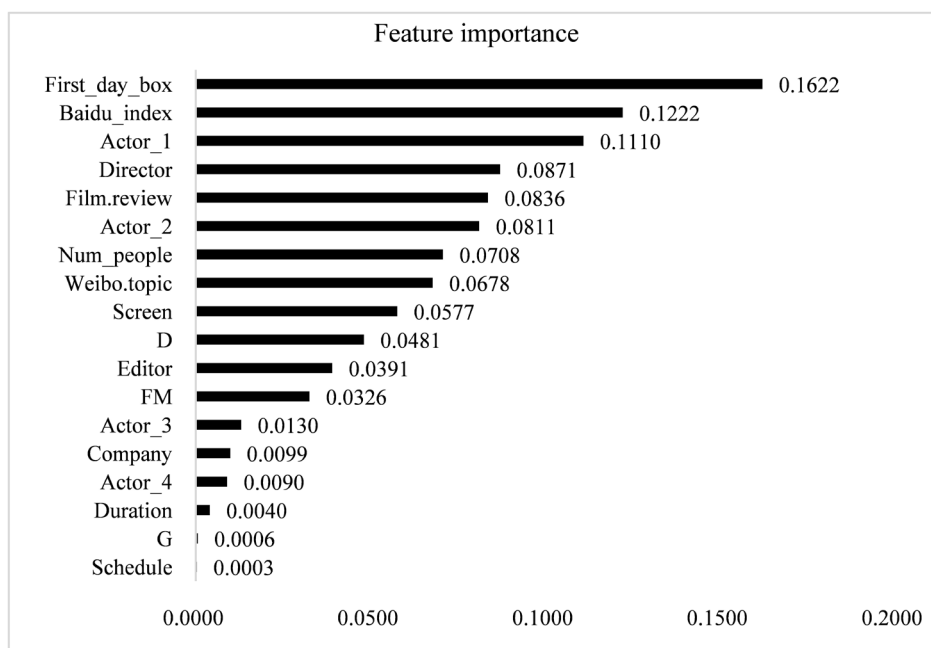


Figure 1. Importance score chart of movie box office influencing factor variables
图 1. 电影票房影响因素变量重要性分数图

4.2. 模型评估

(1) 模型的评价指标

评估函数进行模型评估工作也是进行机器学习预测研究中的关键一环。对于分类任务而言，通常选择计算准确率以及 F 值等作为评级指标；对于回归任务来说，通常使用的评价指标有均方误差(Mean Squared Error, MSE)、均方根误差(Root Mean Square Error, RMSE)、绝对百分比误差(Mean Absolute Percentage Error, MAPE)、平均绝对误差(Mean Absolute Error, MAE)以及可决系数(R-Squared)等。其中 MAPE、MSE、RMSE、MAE 等这些评价指标可以准确的计算出预测结果和真实结果的误差大小；R-Squared 的结果衡量模型的好坏程度。为了科学评估对比各模型的预测效果，本文参考了米传民[15]和杨威[14]的指标选取方法，本文采用三种评价指标来对比模型预测效果，即平均绝对误差(Mean Absolute Error)、平均绝对百分比误差(Mean Absolute Percentage Error)和可决系数(R-Squared)。

Table 1. Comparison of model evaluation criteria

表 1. 模型评价标准的对比

模型	MAE	MAPE	R ²
BP	0.8656	0.4357	0.8086
RBF	0.8934	0.4471	0.8498
PSO-BP	0.7185	0.3536	0.8871

由表 1 是各模型评价指标的对比,从总体上看,PSO-BP 模型的预测结果在 MAE、MAPE 和 R^2 三项评估指标中均优于其他两种模型,即本研究所构建的 PSO-BP 电影票房预测模型具有更高的预测精度。

(2) 案例电影的对比

为了更进一步对本文模型以及另外两个对比模型的预测的准确度以及误差进行对比,随机选取了 10 部受消费者呼吁较高的电影(记为案例电影)作为测试数据,其余电影作为训练数据,再次基础上重新建模回归进而对比分析不同模型的预测效果。

Table 2. Comparison of box office forecast of case movies

表 2. 案例电影预测票房的对比

序号	电影名	真实值	BP 预测值	RBF 预测值	PSO-BP 预测值
1	唐人街探案 3	452234.30	495466.21	232317.15	429155.30
2	复仇者联盟 4: 终局之战	423881.70	334850.24	424674.21	412149.32
3	红海行动	364726.20	390304.54	355366.50	371584.06
4	唐人街探案 2	339768.80	301385.77	365652.19	368793.31
5	我和我的祖国	317118.90	328662.29	353790.53	327838.79
6	独行月球	310291.10	387186.88	166979.46	302907.91
7	唐人街探案 3	452234.30	495466.21	232317.15	429155.30
8	人生大事	171231.40	287827.95	131376.96	184749.50
9	送你一朵小红花	143246.00	163776.37	124738.58	130041.31
10	湄公河行动	118417.40	156384.29	143471.15	131475.29

由表 2 可以看出,PSO-BP 模型的预测值相较于其他两种模型更为接近票房的真实值。为更好地看出各个模型对案例电影票房预测值与真实值之间的差距,在对案例电影的预测效果进行对比时主要采用平均绝对百分比误差进行分别计算预测模型预测结果的误差,计算结果如表 3 所示:

Table 3. Error comparison of case movie prediction results

表 3. 案例电影预测结果的误差比较

序号	电影名	电影票房	BP 误差(%)	RBF 误差(%)	PSO-BP 误差(%)
1	唐人街探案 3	452234.30	9.56	48.63	5.10
2	复仇者联盟 4: 终局之战	423881.70	21.00	0.19	2.77
3	红海行动	364726.20	7.01	2.57	1.88
4	唐人街探案 2	339768.80	11.30	7.62	8.54
5	我和我的祖国	317118.90	3.64	11.56	3.38
6	独行月球	310291.10	24.78	46.19	2.38
7	中国机长	290004.10	11.09	10.63	12.60
8	人生大事	171231.40	48.09	23.28	7.89
9	送你一朵小红花	143246.00	14.33	12.92	9.22
10	湄公河行动	118417.40	32.06	21.16	9.03

表 3 是案例电影预测结果的 MAPE 比较,总体上看,本文构建模型的预测模型误差小于其他两种模型,表明本文所构建的 PSO-BP 在票房预测表现上优于其他模型。PSO-BP 模型在预测高票房影片时误差控制在 15% 以内,波动范围小且只有一部电影预测的误差值超过 10%,十部案例电影中有四部电影预测的误差值都在 5% 以内,尤其是在对《红海行动》进行预测时,票房绝对百分比误差仅有 1.88%。综上所述,PSO-BP 模型的拟合效果较好,误差更小,票房预测值更精准。

5. 结论

电影的票房收入作为国内电影市场最重要的指标，是每一部电影最终要追求的目标，关系到了制片方，出品方的决策和电影市场的投资方向，同时也关系到电影生产和营销的各个环节，包括电影题材的选择，剧本的设定，导演、演员、编剧团队的组建，电影宣发策略和营销方式等各个方面的因素。本文在对电影票房体系总结梳理的基础上从以下三方面展开分析本研究：

1) 考虑到目前票房预测研究领域很少有学者将微博话题讨论量以及电影首日影评情感分析因素作为电影票房影响因素，所以本研究在加入电影首日影评情感因素的基础上，同时连结电影特征与市场因素构建了一种更为全面电影票房影响因素体系，并在总结文献的基础上对票房各影响指标采用了较为合理的量化方法。

2) 为达到简化后续预测模型的输入和提高模型的精度目的，通过构建一种基于 XGboost 算法的影响力测量模型来进行变量的筛选。

3) 通过对以往文献的梳理发现，BP 神经网络与常见的票房预测模型相比具有很多优势，但精度还不够准确，故本文对 BP 神经网络加入粒子群优化算法，通过对比发现，该模型对电影票房的预测精度得到了提升。

因此，研究我国电影票房收入的影响因素并对其进行预测研究，可以为国产电影投资商与制作发行商的投资决策提供建议，使其在投资过程中做出正确的价值判断，实现收益的稳步提升，有利于电影产业与其它产业实现良性循环，共同促进经济发展。

参考文献

- [1] 胡晓红, 王红. 基于多元线性回归的电影票房预测研究[J]. 信息技术与信息化, 2018(Z1): 183-185.
- [2] 何晓雪, 毕圆梦, 姜绳. 基于网络数据预测电影票房的多元线性回归方程构建[J]. 新媒体研究, 2018, 4(5): 41-48.
- [3] Sochay, S. (1994) Predicting the Performance of Motion Pictures. *Journal of Media Economics*, 7, 1-20. https://doi.org/10.1207/s15327736me0704_1
- [4] 陈邦丽, 徐美萍. 基于 LARS-SVR 的电影总票房预测模型研究[J]. 陕西师范大学学报(自然科学版), 2018, 46(1): 10-15.
- [5] Hofmann, J., Clement, M. and Völckner, F. (2017) Empirical Generalizations on the Impact of Stars on the Economic Success of Movies. *International Journal of Research in Marketing*, 34, 442-461. <https://doi.org/10.1016/j.ijresmar.2016.08.006>
- [6] 申林, 王靖舒. 从豆瓣电影看网络评价对电影票房的影响——以 2019 年院线电影为例[J]. 中国电影市场, 2020(8): 13-17.
- [7] Hur, M., Kang, P. and Cho, S. (2016) Box-Office Forecasting Based on Sentiments of Movie Reviews and Independent Subspace Method. *Information Sciences*, 372, 608-624. <https://doi.org/10.1016/j.ins.2016.08.027>
- [8] 谢治海, 朱敏, 牛红宇, 梁晶, 夏婷. 面向票房预测的影评情感可视分析[J]. 计算机应用研究, 2020, 37(10): 2945-2950.
- [9] 吴珏, 潘徐. 基于用户内容消费数据的电影票房预测模型探索[J]. 全球传媒学刊, 2018, 5(3): 96-107.
- [10] 史伟, 王洪伟, 何绍义. 基于微博情感分析的电影票房预测研究[J]. 华中师范大学学报(自然科学版), 2015, 49(1): 66-72.
- [11] Litman, B.R. and Kohl, L.S. (1989) Predicting Financial Success of Motion Pictures: The 80s Experience. *Journal of Media Economics*, 2, 35-50. <https://doi.org/10.1080/08997768909358184>
- [12] Liu, T., Ding, X., Chen, Y.H., Chen, H.C. and Guo, M.S. (2014) Predicting Movie Box-Office Revenues by Exploiting Large-Scale Social Media Content. *Multimedia Tools and Applications*, 75, 1509-1528. <https://doi.org/10.1007/s11042-014-2270-1>
- [13] 张雪. 基于深度学习卷积神经网络的电影票房预测[D]: [硕士学位论文]. 北京: 首都经济贸易大学, 2018.
- [14] 杨威. 基于微博数据的电影票房预测模型研究[D]: [硕士学位论文]. 合肥: 安徽大学, 2015.
- [15] 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[J]. 计算机系统应用, 2019, 28(2): 15-23. <https://doi.org/10.15888/j.cnki.csa.006709>