

**内蒙古师范大学计算机科学技术学院**

**毕业设计（论文）开题报告**

**题 目： 基于电影票房季节效应的分析与预测**

**专 业 信息管理与信息系统（云计算大数据分析）**

**学 生 窦玉琢**

**学 号 20161104542**

**指导教师 苗忠义**

**日 期 2019.12.16**

**计算机科学技术学院制**

1. **课题来源及研究的目的和意义**

2018年，中国电影年票房收入额达到606.9652亿元，环比去年增长了8.6%，电影产业快速发展，《电影产业促进法》的实施也在法治上确保和规范了电影产业的健康稳定高速发展，中国已经成为了仅次于美国的全球第二大电影市场。随着电影市场的不断成长，电影票房也越来越受到关注和重视。电影票房，作为电影观众进入院线影院观看电影的购票消费，是衡量一部电影商业价值的最重要指标，也是电影艺术家们以及相关从业人员劳动的积极回报。

电影是一种明显具有季节性的商品。这里的季节，不单单是指一年四季，而是指一年中具有某些特点的一段时期。不过在现实生活中，人们常常更愿意用“档期”来代表电影市场的季节性。

和全世界的电影市场一样，中国电影市场的季节性也相当明显。例如，2018年春节档仅七天票房总收入达63.94亿元，占全年票房收入的10.53%，成为全年的第一大票仓。其中《红海行动》和《唐人街探案2》分别获得36.51亿元和33.98亿元，成为2018年中国电影票房榜上的前两名。节日、假期和季节对票房的影响如此显著，使得其对电影产业的各个链条都有影响。制片方需要跟根据季节因子估计票房总量、确定上映日期和制定营销战略；电影从业人员为某个具体节假日创作相关的类型电影也是十分普遍的现象，例如近年来春节档上映的合家欢电影，和每到情人节和七夕节上映的爱情类型片；观众由于节假日休闲娱乐的需要，在节假日集中进入院线观影；甚至对电影投资方来说，电影上映季节的选择会使他们对一个电影项目的投资方案和投资额加以权衡。

综上所述，可知根据电影季节性的研究，可大大提高电影行业的运行效率，使得电影票房尽可能的最优化。因此季节性分析和预测对于电影票房极具现实意义，非常值得研究。

1. **国内外在该方向的研究现状及分析**

中国电影现状，产业呈现出良好的发展势头。一方面，中式大片在产量和社会效果两方面都越来越好，重新赢得国内观众的信任；另一方面，它在海外市场的拓展也出现持续上升趋势，对于推广我国文化价值体系及其软实力起到积极作用。

中式大片的国际影响力不断增强，受到外国观众的欢迎，国际市场有了新的开拓，商业价值也趋于提高。为了加快电影行业发展的脚步，对于本课题的研究分析也显得至关重要。在我国，对于电影票房的季节性研究的课题，虽然存在，但并不多，这就越发体现了我选择此次课题的必要性，非常有实用参考价值。相反在国外，尤其是美国这样的电影行业巨头国家，这方面的研究已日趋成熟，整个电影行业链已经相当完备，所以作为世界电影产业第一阵营的我们，更要努力迎头赶上。

1. **主要研究内容**

本课题计划通过对2012年~2018年中国电影市场票房的数据分析，通过python语言编写爬虫脚本对http://www.films.cn/boxoffice进行爬取。在利用Python进行数据的处理、分析以及展示工作时，通常使用一些第三方的Python库。Numpy是Python科学计算的基础包，提供数组支持以及相应的快速高效的处理函数。因此我选用numpy。我将利用Python对采集到的电影数据进行分析。使用scikit-learn提供的朴素贝叶斯分类算法GaussianNB。使用Matplotlib将数据可视化。

1. **研究方案**

技术指标：python，Office，matplotlib，numpy，

主要内容：

（一）数据爬取

通过python语言编写爬虫脚本对http://www.films.cn/boxoffice进行爬取

（二）处理数据

在利用Python进行数据的处理、分析以及展示工作时，通常使用一些第三方的Python库。Numpy是Python科学计算的基础包，提供数组支持以及相应的快速高效的处理函数。所以我选用numpy

（三）数据探索

Pandas是Python下最强大的数据分析和探索工具，它包含高级的数据结构如Series和DataFrame以及精巧的工具，使得在Python中处理数据快速而且简单。我将利用Python对采集到的电影数据进行分析。

（四）数据建模

使用scikit-learn提供的朴素贝叶斯分类算法GaussianNB，该算法假设特征的先验概率为正态分布，主要用于连续变量中。

（五）数据展示

Matplotlib是Python最著名的绘图库，它主要用于二维绘图，让我们可以快捷地可视化数据。

1. **进度安排**

筹备阶段：在9月-11月，进行数据爬取整合的工作，并同时开展相关知识储备学习

操作阶段：11月-1月进行项目操作，用python进行代码编写，整合成果，并同时开始进行论文的攥写，

收尾阶段：2月开始进行论文成果的整理收尾工作。

1. **研究过程中可能遇到的困难和问题，解决的措施**

研究中的困难就是专业性方面相关知识储备不足，我会努力克服，积极学习相关知识，认真思考勤于动脑，积极与导师沟通，向相关专业人士积极问询，不能不懂装懂，为自己负责。

1. **主要参考文献**

[1]陈胜可，刘荣.SPSS统计分析入门到精通(第三版)[M].北京:清华大学出版社，2015.

[2]黄冬虹. 中国电影产业发展现状研究[J]. 传媒, 2014(17):14-16.

[3]郑中元. 电影产业的季节性、周期性变化特征分析[J]. 现代商贸工业, 2019(20):8-9.

[4]丁守銮, 康家琦, 王洁贞. ARIMA模型在发病率预测中的应用[J]. 中国医院统计, 2003, 10(1):23-26.

[5]吴家兵, 叶临湘, 尤尔科. ARIMA模型在传染病发病率预测中的应用[J]. 数理医药学杂志, 2007, 20(1):90-92.

[6]https://zgdypw.cn/#/root.html