

2018-2021 豆瓣综艺评分数据研究

传播 1901 吴宜檬

U201917095

日期: 2021.12

1 研究背景与意义

随着互联网的蓬勃发展和大众文娱需求的扩张,综艺节目正成为许多人生活中重要的一部分。而伴随主流网络视听平台纷纷加入综艺制作领域,国产综艺节目也从各大卫视独占、盲目引进国外形式转向到符合当代本土审美的积极自创的状态。《腾讯娱乐白皮书》[1]指出,2013年起,中国的综艺节目市场呈现百花齐放的态势,数量和类型都在不断增加。理念新、水平高的制作团队,与海量需求者背后的巨幅流量相呼应,最终使得国产综艺形式日益丰富,综艺质量不断提升,总体快速发展。

而在其中,2018年可以称得上是国产综艺发展中极为关键的一年。这一年,《偶像练习生》、《创造101》两档偶像选秀节目打开了中国的“偶像”新市场,留下了选秀元年的称号。优酷的王牌节目《这就是…》系列推出《这就是街舞第一季》,引领后续的“精细领域”综艺风潮。同时,老牌卫视综艺《奔跑吧》、《极限挑战》也都在这一年的网综压力之下走向了电视转网络的转型。

综艺节目作为依赖广告投资与所在平台用户会员充值盈利的一种文娱产品,其收益与节目曝光量、观众的关注度、以及口碑好坏等有密切的关系。但早在2018年,由于大量剧集被指出存在播放量灌水等伪造流量热度的行为,爱奇艺、优酷两大知名视频网站关闭了其视频前台的播放量数值显示[2]。这也使得当下的网络综艺热度难以用标准且统一的数值进行衡量与比对。在综艺口碑这一领域,虽然近年来也出现了诸如注水、刷分等相关争议,但基于平台对恶意评分行为的有效管制,其可信度相对播放量更高。而豆瓣平台作为国内知名的影视评分平台,拥有较大的体量与较高的用户素质,其中的综艺评分经常作为许多文章中“推荐”综艺的标准,也依然是许多综艺爱好者搜寻新综艺时会参考的重要因素,这些用户行为也在一定程度上证明了豆瓣评分在“口碑”这一评价维度上的代表性。

综上,为了能更好的了解当下的国产综艺生态,发掘影响综艺口碑的重要特性,进而为制作者进行综艺选题、观众选择综艺提供一定的参考与帮助,本研究将以2018-2021有评分记录的国产综艺为样本,观察综艺本身与评分导向,深度挖掘各个特性的关联与差异,并最终尝试归纳总结影响综艺的口碑的重要要素。

2 文献综述

在查询过往文献的过程中发现,对于综艺节目的点评数据分析研究数量较少。因此这里将这一主题拆分成对网络影音点评数据的分析和对综艺节目相关效果的研究,并分别简述部分具有代表性的研究方法与结果。

2.1 评分网站数据研究现状

在点评数据方面,国内外都有较为充分的挖掘、研究经验。其中 IMDB 和豆瓣分别作为国外与国内的重点分析平台,而电影则是这类分析的主要研究对象。在国外,Duai(2008)[3] 等人为 71 部电影收集了平均 1350 个用户评分,用于预测评分与点评人数的关系。K Persson(2015)[4] 使用 IMDB 上的 3000 余部好莱坞电影的特征信息,如导演、演员、拍摄成本及类别等,分别采用随机森林回归模型和支持向量回归模型进行建模,预测在线评分,并比较了模型精度。而在国内,苏炯铭,刘宝宏 (2014)[5] 构建了基于观点动力学的电影在线评分影响力模型,预测豆瓣网站上的影片评分。刘明昌 (2017)[6] 基于豆瓣电影数据,分别使用电影特征、评分结果与相似电影推荐建立模型,并进一步建立了混合预测模型。陆君之 (2018)[7] 将导演、编剧、主演、类型、发行地区的过往评分均值作为影响电影评分的特征,建立了基于随机森林回归模型的电影评分预测模型。谭家柱 (2021)[8] 使用 IMDB 电影评分数据,加入了如导演的社交媒体点赞量等特征,评价了随机森林回归模型与其他模型的区别。

2.2 国产综艺发展研究现状

而在综艺的相关效果分析方面,由于本次研究选取的对象是国产综艺,且由于国内外综艺存在发展平台、商业模式以及观众文化需求不同等多方面的差异,相通点较少,主要选择了国内学者关于国产综艺的相关研究进行整理。其中收视率与点播量作为主要的分析对象,而豆瓣评分等口碑相关项则作为中间项或影响收视的变量参与到研究中。万颖迪 (2016)[9] 对 2013-2016 年间的 27 部国产明星真人秀节目进行回归分析,得到了节目百度热度、豆瓣评分、电视平台是影响收视率的重要因素的结论。庄玉婕,刘娟 (2020)[10] 通过单因素方差分析和逐步回归法筛选影响因素,得到了影响收视率的线性模型。郭明欣 (2021)[11] 使用多元逐步加权回归和随机森林的组合模型,对腾讯视频综艺节目点播量进行研究,发现节目类型、衍生节目、节目时长、节目网络搜索指数和豆瓣评分人数都对节目播放量存在相对显著的影响。

2.3 文献评述

国内外学者在过往对点评数据的研究内容上,较多地青睐电影而忽略了综艺节目,这主要是因为此前的综艺生态发展不完善,很难作为一种“文化”或一个“产业”来看待,而电影则作为一种具有文化象征意义且利益生产与回收过程相对清晰的领域,对其评分的分析与研究能较好的与票房等明确的数据相绑定,也能一定程度上印证其艺术上的成就。但正如研究背景中指出的,目前的综艺也正在朝向产业、文化艺术的方向发展,其在许多层面上正在越来越接近电影在人们心中的文化与商业地位。而在研究方法上,则多采用数据采集与机器学习模型相结合的方式,实现自动化的数据分析与预测。

而在综艺效果的相关研究分析中，**收视率**作为主要的分析内容，研究方法则更多偏向传统的**回归分析**。另外在结果上，多位学者都曾指出豆瓣评分是影响节目最终收视与关注度的重要因素，这一结论也拓展了本次研究的最终意义。

综上所述，虽然国内外学者在点评数据分析以及综艺效果研究方面已有较多成果，但仍存在一定缺陷，因此本研究将尝试对其做出补充。本文将以近四年的国产综艺为例，对其在豆瓣网上的评分数据进行搜集与分析，并通过**可视化**、**情感分析**、**相关性分析**等方法发掘综艺节目的数据特性。同时，在研究中也尝试完成初步的**特征工程**，使用**机器学习模型**对所建立的特征进行简单的学习与预测，检验当前特征选择的效果。

3 数据收集与整理

3.1 数据来源

研究中使用的**所有数据**均来源于豆瓣网，由于技术原因，整体爬取分为了两次进行，其中有分数的综艺列表爬取，即综艺个体确定完成于11月28日，具体信息爬取则截止于12月8日。最终收集到**596条**综艺信息，每一条信息内部**28个**字段。

数据收集过程中，主要使用了Python中的Request库进行页面信息获取，同时为了应对豆瓣的反爬机制，使用Github上的Proxypool+Redis动态数据库搭建动态代理池辅助爬取。而对于原始网页数据，则使用了Beautifulsoup和正则表达式相结合的方式进行提取（详见代码文件夹前期爬取）。

3.2 数据处理

3.2.1 数据预览

数据收集整理完毕后使用info函数进行对数据进行初步预览，并在表1中整理变量并进行简单的解释。

3.2.2 数据清洗

本小节将基于前期对数据进行的初步观察，对存在缺失值或需要进一步拆分至可量化值的数据进行处理。

导演、编剧

导演与编剧作为两个性质类似的变量，具有以下两个特点：

1. 导演编剧与单个综艺节目或系列综艺节目具有强绑定关系，整体流动性弱，一个导演或编剧呈现出来的“评分高”的结果，很可能事实上只是一档节目评分高，不具有代表性。
2. 区别于电影，综艺导演与编剧很难为节目的质感与风格带来极强的个人色彩。现实生活中，因为导演而去看一部电影的人不在少数，但为了导演或编剧而去观看综艺的极少出现，因此也可以猜想导演与编剧并不会成为影响节目进而影响评分的关键因素。

另外，在收集的数据中，导演与编剧整体缺失情况较为严重，且难以用现有数据进行推理补充，而通过百度百科等网络数据进行查询时也发现仍存在大量缺失以及错误标注的情况，可能造成

字段名称	字段格式	含义	数据缺失情况
Score	数值	豆瓣评分，数值范围为 2-10。	
Name	字符串	综艺名称	
导演/编剧/主演:	字符串列表	演职人员姓名	存在缺失
类型	字符串列表	值集中在 [脱口秀、真人秀] 两大范围	存在缺失
首播	年-月-日	节目首播时间	
集数/单集片长	字符串，内部格式不确定	节目播出集数与时长	存在缺失
又名	字符串列表	综艺节目的其他名称	大量缺失
简介	字符串	综艺官方给出的简介语	存在缺失
标签	字符串列表	用户常用于标记节目使用的标签	
短评	字符串列表	豆瓣综艺页 (非短评详情页) 显示出的短评内容	
相关推荐	字符串列表	豆瓣官方基于节目性质给出的相关推荐综艺	
在看/看过/想看	数值	不同观影状态的用户数值	
季数	数值	当前综艺属于该综艺 IP 的第几部作品	大量缺失
评分人数	数值	参与评分的人数	
5/4/3/2/1 星	数值	豆瓣用户标记不同星级的人数	
讨论/短评/剧评数	数值	豆瓣用户对于综艺进行不同形式的评价条数	
平台	字符串列表	目前可以在网络上观看到该节目的平台名称	存在缺失

表 1: 数据预览

不够准确的风险，最终选择**放弃这两个变量**。

演员

在文献阅读中，演员、明星常成为对于综艺、电影品质分析的关键点。而演员作为一个不可量化的变量，常常被转化为演员相关的可量化特质进入模型，如：演员过去参演作品均分、演员社交媒体热度、演员搜索热度等。另外，由于综艺的特殊性质，综艺中的演员并不完全是固定，往往存在类似主持的长期参演演员和只在少部分片段中被邀请出席的不同嘉宾。因此其出现时长，对节目的代表性与影响力度都会出现差别。尤其目前还存在大量的选秀节目，其中的演员在节目中经历了从素人至明星的转变，其大众印象往往与节目紧密挂钩，很难研究个体单独的性质。

孙扶阳 (2020)[12] 中也曾对电视综艺节目中的明星贡献度做出分析，通过对于视频逐帧弹幕的情绪进行分析。最终得到飞行嘉宾，即仅在节目的一集或少数片段中出现的演员往往比固定嘉宾贡献度更高的结论。而本次研究中由于前期数据采集的失误，在豆瓣中并没有对飞行嘉宾与固定嘉宾的区分，最终选择**舍弃该特征**。

类型

通过对于类型数据的初步观察，可以发现类型数据的值较为集中，绝大部分节目都被“真人秀”、“脱口秀”两个标签覆盖，并不能很好地概括当前节目的特征，区分度极低，这里**舍弃该变量**。

首播

首播作为时间格式的数据，首先需要对其进行数值上的拆分，经观察可以发现，首播数据中的“日”单位粒度较细，且有大量缺失。因此这里将首播属性拆分为“节目首播年”和“节目首播月”两个属性。

集数/单集片长

两个数据都存在着缺失和格式不统一的问题。首先通过将单位统一转化为分钟，数值区间转化为上界和下界的平均值，得到单个数值变量。

绘图观察集数和单集片长数据之间的相关性与其本身的分布，可以发现两者都相对集中，其中对于不同片长的综艺，其集数也基本集中于一条水平线上下，因此对集数的缺失值可采用中位数直接填充。而对于单集片长，考虑到现实因素如人们生活习惯和科技的发展等，猜测可能与节目出现的年代存在一定关系，但验证后发现关系并不明确。最终单集片长同样采用**中位数**进行**缺失值的填充**。

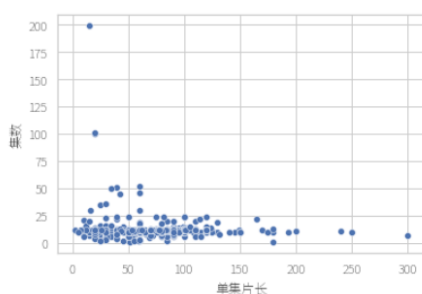


图 1: 片长与集数

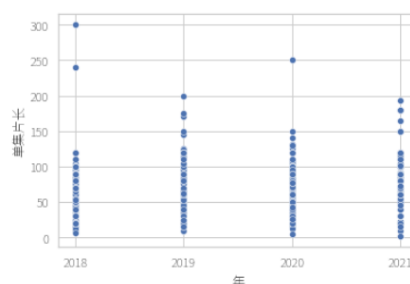


图 2: 片长与时间

季数

在收集到的原始数据中，季数存在大量缺失值，一方面可能由于绝大多数并不构成 IP 的综艺，可能仅出现一次而并不会作为第一季被标记。另一方面，豆瓣本身的数据标注并不规范。部分综艺名称中明显包含了“XX 季”字样，其“季数”属性却并未被标注。

对于是“第 N 季”却并未被标记的数据，考虑从“综艺名称”和“综艺又名”两个属性入手，提取其所属的 IP，即系列节目的统称。在处理过程中发现该字符串可能以年份结尾、或有直接的“第 N 季”、“XX 季”等标识，并不能进一步直接从名称中得到其季数特征。因此对这些数据进行提取并归类到其所属的 IP 后，再根据年份进行排序，回填季数。

而对于无续作的综艺节目或系列节目第一季，统一将“季数”标记为 1。

平台

人工随机抽样 30 条数据检查后发现，豆瓣上对于综艺节目及其主流播放平台的标注相对完整，“平台”属性中内容缺失的综艺的确大多数都是已被在线平台下架或只能通过小众平台观看。因此不对缺失值进行填充，仅将平台值拆分成 [优酷视频, 腾讯视频, 爱奇艺视频, 芒果 TV, 央视网, 哔哩哔哩, 咪咕视频, 西瓜视频] 的**多维二分值**，每个特征下的值**仅为 0/1** 标记当前综艺是否能在该平台播放。

4 数据分析

本小节将对前期爬取并清洗过后留下的特征进行进一步分析。预期通过统计图表等方式进行初步观察，使用相关性分析等方式进行重复特征筛选，判断各个属性与最终评分分数之间的关系，并从中提取出一些潜在的规律。

当前的剩余特征主要可以分为两个类别：综艺节目自带的属性与用户使用平台行为生成的属性。其中，综艺节目自带的属性可以用于最后的特征筛选，而用户使用平台产生的属性则主要用于观察其特性，进而辅助特征的筛选。

4.1 用户行为属性分析

4.1.1 分数

在进行后续的数据分析前，首先需要对分数的整体分布有整体的了解。通过描述性统计分析与分数密度图可以发现，评分值在评分范围内整体朝高分段倾斜，均分达到 6.9 将近 7 分。其中 6-8 分为评分集中区。

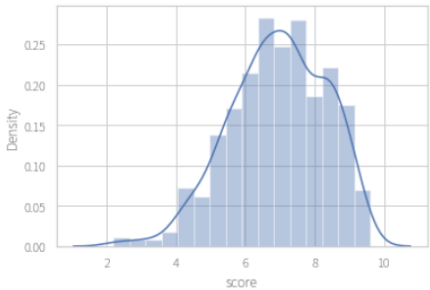


图 3: 分数分布

count	595.000000
mean	6.939832
std	1.380556
min	2.200000
25%	6.000000
50%	7.000000
75%	8.000000
max	9.600000

图 4: 分数特征

4.1.2 短评

短评事实上并不作为综艺节目本身具有的属性，但该特征可以一定程度上反映观众对不同评分节目的态度与关注点，能有效帮助后期选取与重构特征，因此这里预先对短评情感与词频分别进行分析。

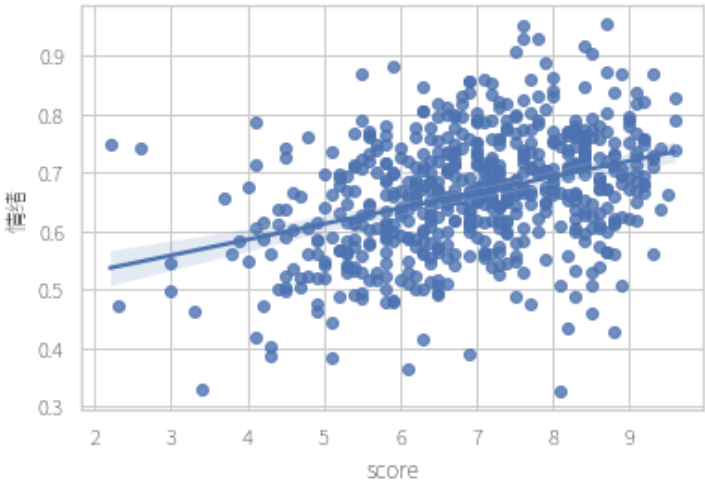


图 5: 节目分数与短评情感

人数则相对较多。进一步地，我们将 y 轴变化为“想看”与“在看”以及“看过”之和的比例发现差异更加明显。这一结果初步证实了之前的猜想，即**综艺的高分会至少在一定程度上为综艺增加关注**。

4.1.4 不同星级人数

该系列变量能反映分数呈现背后的不同情绪。如同样分值的剧，星级分布方差大即极高分与极低分分布较为密集，则说明可能该节目存在争议，有助于发掘潜在的影响特征。

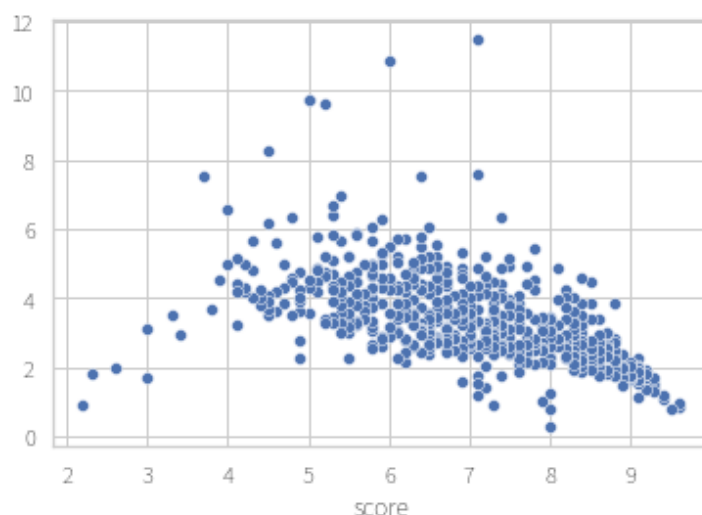


图 10: 不同评分的节目的星级人数方差

不同分段与星级人数方差间的关系如上图所示，整体呈现出二次函数形状，与极端分数往往是大量极端分数分的叠加结果，而中间分数则往往存在不同层次的分数的一般认知相符。但相对来说偏低分段的方差比对称位置的高分段更高，表示可能存在更多争议。

考虑进一步对离群点进行分析并观察其原始数据。如最上方的离群点，表示的是综艺节目《这就是中国》，其评分 7.1，方差值高达 11，5 星与 1 星各占 0.5 与 0.3。在网络上对该综艺节目检索后发现由于该综艺涉及一定的**政治因素**，出现了大量基于不同政治观点的打分行为，另外的离群点《信·中国》也是类似。另有综艺节目《我们的乐队》，评分 6.4，5 星与 1 星各占 0.2，观察其评论相关内容后可知主要矛盾点在于“**流量明星**”的**粉丝文化**和综艺节目爱好者之间的冲突，最终使得极端评分数量增加。

近年来，豆瓣上屡次出现了“追捧”和“抵制”型的恶意刷分行为。但从上述结果中也能看出，尽管难以完全遏制这一大环境下的不理智行为，平台方对类似行为做出的限制还是基本达到了效果。因此在后续分析中，如果少数类似的争议题材节目数据出现异常，也可以考虑对其舍弃。

4.2 综艺节目自身属性分析

4.2.1 节目类型

节目的类型是节目主题的概括，往往决定了节目的调性与主要内容，也常常作为观众选择节目时的重要考量因素。在过往研究中，节目的题材类型也常被选作研究的对象，用于分析其对收视率、票房等的影响效应。而在本研究中，由于原始数据中类型值存在过于简略、区分度很低的缺陷，需要重新构造填充该特征。因此考虑两种方法：

1. 通过节目的其他特征进行归纳与聚类

经观察发现在爬取的节目数据中，存在“标签”和“简介”两个与节目的类型强相关的属性。标签是用户对于该节目的一种归类方式，能体现用户对当前节目所属类型的判断，而简介则作为节目制作方对节目的介绍，其中也往往包含制作方对节目类型的判定。

对于“标签”属性，直接将爬取的字符串列表中的单个词语进行词频统计，而对于“简介”属性，则使用 *jieba* 库进行分词后提取关键词进行同样的分析。在最终排序前，对两组提取出的词语中可能出现的年份数值、平台名称、动词等干扰值做了初步的清洗。得到如下结果：

	word	time
0	脱口秀	90
1	音乐	78
2	搞笑	76
3	文化	63
4	青春	39
..
227	歌舞	2
228	甜	2
229	蒙面	2
230	电子竞技	2
231	嘻哈	2

图 11: 标签词及其词频

	word	time
0	音乐	45
1	文化	28
2	体验	28
3	竞技	26
4	展现	25
5	视角	24
6	观察	24
7	播出	24
8	聚焦	23
9	观众	22
10	战队	21
11	美食	20
12	担任	20
13	赛制	20
14	情感	19

图 12: 简介分词结果及其词频

观察高频词发现与常识认知中的节目分类词语比较接近，但考虑到整体词量有 4000 多个，词量较大，而频率较高的词个数则显著小于整体词量，猜想在某一出现次数阈值之上的词语能概括绝大多数的节目分类信息。经过验证发现，多于 90% 的节目数据，都包含至少一个出现频次大于等于 2 的词标签，因此证明高频词具有代表性，使用这些出现频次大于等于 2 的词语，能覆盖 90% 以上的信息。

得到这些高频词后，考虑使用聚类方法，使其呈现出自然的“分类”。具体实现使用 *Word2vec* 词向量分析方法：由于并未检索到针对影视作品的词向量模型，使用相对权威且体量较大的 *TencentAILabChineseEmbedding* 即腾讯 AI 实验室的 15G 中文语料向量库进行初步建模，再将提取出的高频词输出至图像，结果呈现如图 13。

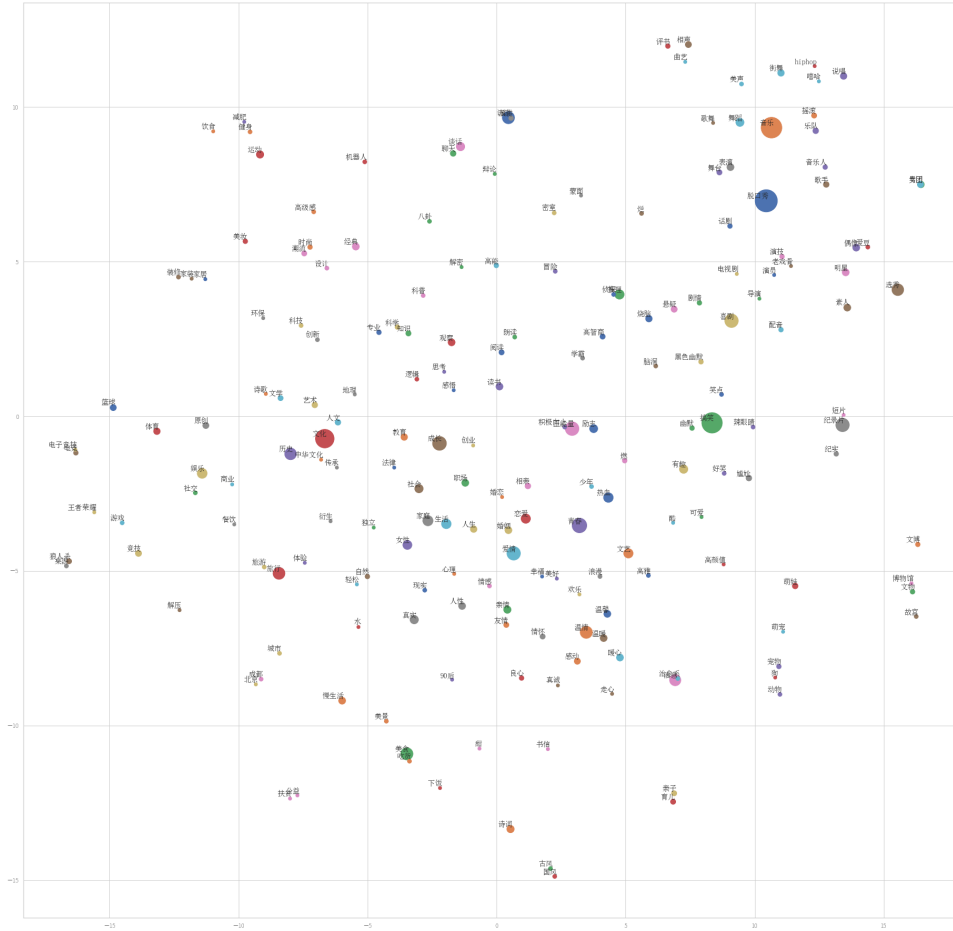


图 13: 词语聚类结果

可以发现，由于语料库基于日常语言，在计算距离时可能考虑词语词性以及所属的大类等，该聚类结果综艺分类并不适配，另外图中也并未表现出明显的词语聚集，整体分布较为均匀，因此最后判定：**无法使用词语本身聚类形成分类。**

2. 通过文献阅读与资料搜集，先行确定一个类型值的范围，再划分单个节目归属

孙宏姣, 葛进平 (2017)[13] 归纳了国内外对于电视综艺节目的分类，最终提出了针对国内综艺的四级分类方法，分别涉及播出时间、制作场地、节目形式，而最后一级则为针对真人秀节目的再分类：才智、生活、体验、艺术、游戏。这一分类对于电视综艺节目应用性较强，但对于目前快速发展的网络综艺则存在未能覆盖和划分不够清晰的问题。

与学界相对的业界对网络综艺节目的划分则相对更为贴合。这一方面与其能直接接触到生产与消费方有关，另一方面也出于业界分类往往是观众在使用视频平台时进行检索与节目定位时使用的分类。对市面上的多个业界分类标准进行筛选后，最终选取了腾讯综艺年度白皮书中的综艺分类方法，将其划分为[**竞技, 文化, 表演, 美食, 喜剧, 情感, 访谈, 生活, 亲子, 观察, 时尚, 萌宠, 偶像, 舞蹈, 旅游, 职场, 音乐**]共 17 个二分值变量。

对节目进行划分的标准依然是前文所提取的高频词。在这种思路之下，首先也同样尝试了计算上述 17 个词语与高频词语义距离来划分高频词分类的方式，效果依然不佳。考虑到当前词数较少，最终选择使用**人工标注**的方法对所提取出的高频词进行归类。标注过程中为了保证各个词的权重不出现差别，采用单分类的标注方式，并在标注过程中删去可能具有多重语义的词

以及与上述分类标准词无关的词，最终标注了 150 个词语，并使用这些词语对综艺信息进行回标。

尽管在单词标注的过程中，为了避免造成各词语不必要的权重偏移而采用了单分类，但在现实情况下一个综艺完全可能具有多重类别。因此，在使用标注好的词语对综艺信息进行回标时采用了多分类方法，即一条综艺可能拥有多个类型属性。对于完全没能被标注词语覆盖的信息，则标注“其他”属性。

为了验证各个属性的区分度，避免出现部分属性“一手遮天”和属性语义重复的情况，对属性标注的数量和各个分类属性之间的相关性进行分析，可观察到除了个别的属性如”萌宠“、“时尚”所标注的节目数量较少，整体属性标注节目数量比较均匀；另外，观察热力图可发现各个属性之间并未出现强相关性，属性之间具有一定的区分度。这也证明了前期标注的有效性。

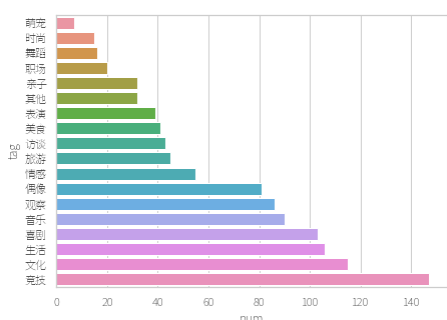


图 14: 各标签标注节目数量

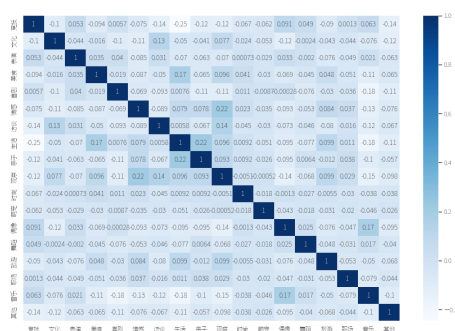


图 15: 各标签相关性热力图

进一步地，我们绘制不同类别标签节目的分数分布图像。观察可发现，不同属性综艺节目评分结果区间存在较大差异。其中能显著观察到拥有**文化**、**访谈**、**观察**、**萌宠**的分类属性的节目呈现出较高的点评分数分布，而**情感**、**偶像**、**音乐**类节目整体分数则相对较低。

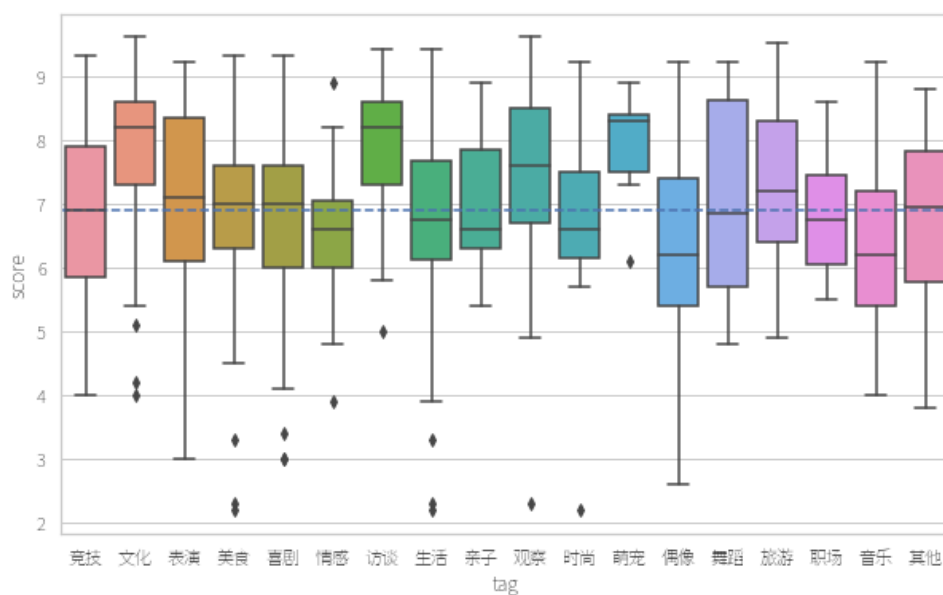


图 16: 节目类型与分数

4.2.2 首播时间

在前面的清洗重构过程中已对首播时间进行拆分，因此这里分别对年、月两种时间单位进行分析。之所以考虑首播时间，与电影分析中常出现的“春节档”、“国庆档”有关，因此这里的分析并不只是探讨时间与分数的关系，也期望得到关于综艺播出时间规律的总结。

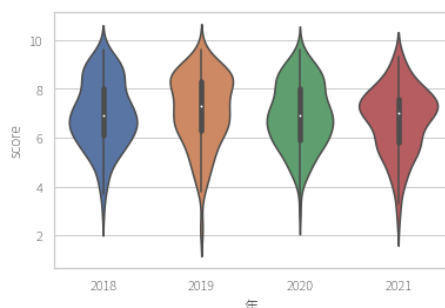


图 17: 首播年份与分数

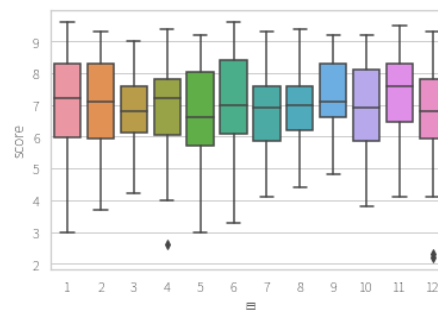


图 18: 首播出现月与分数

以年、月为单位观察年份与分数关系，发现整体的分数分布并未出现较大的差异，也并不存在明显特异值，因此在后续影响因素分析中不再选用这两个变量。

4.2.3 集数与单集片长

考虑到在前期数据清洗过程中已发现“集数”具有较强的聚集性，区分度不高，这里直接弃用该特征，只考虑单集片长的可能影响。

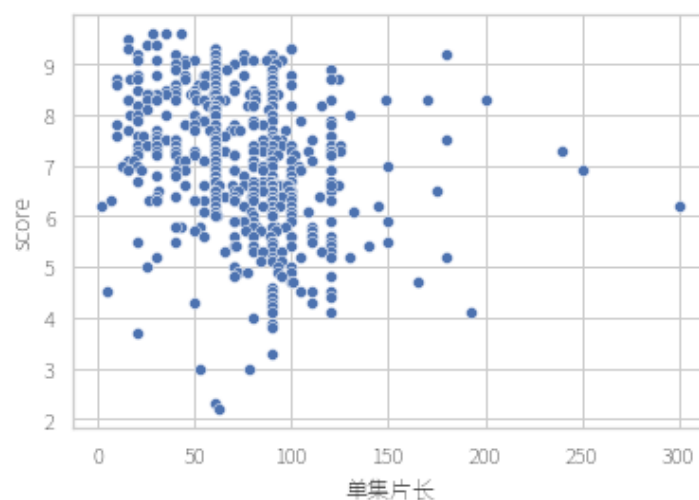


图 19: 单集片长与分数

绘制单集片长与分数的二维散点图，可发现整体差异不明显，但相对而言片长偏短的，尤其是 50 分钟以内的节目得分整体更高。结合现实情况，可能存在片长与节目类型之间的相关性，如有故事线的综艺节目往往时长较长等。在一般的特征工程建造过程中，也需要有对相关性较强的特征的筛选，因此这里对节目类型这一定类变量和单集片长这一定距变量进行相关性测量。操作上使用 SPSS27.0 软件，并通过 η^2 平方进行判断。

最终的结果有部分节目属性与单集片长呈现较强 ($\eta^2 > 0.06$) 的相关性：竞技、访谈、偶

像，也有部分存在强相关性 ($\eta^2 > 0.16$)：文化、情感。结合之前对于节目类型的分析可以发现，这几种类型也表现出了与最终评分结果较强的相关性。

综上，单集片长这一属性自身的区分度较弱，且与节目类别这一属性有较强的重合，可以舍去。

4.2.4 季数

在之前的数据清洗中，已经将“季数”属性的空缺值进行了填充，但在直接观察季数与点评分数的关系时发现并未呈现出明显的相关关系。另外，季数值较大的节目事实上数量很少，代表性不强。因此考虑将季数这一属性进行进一步的转换。

在电视业中，常以“综 N 代”来描述一档综艺节目的持续性播出。^[14] 据艾瑞咨询研究报告显示，综 N 代所代表的 IP 能量对当下的综艺市场产生了极大的影响，尤其在商业价值上表现极为突出。

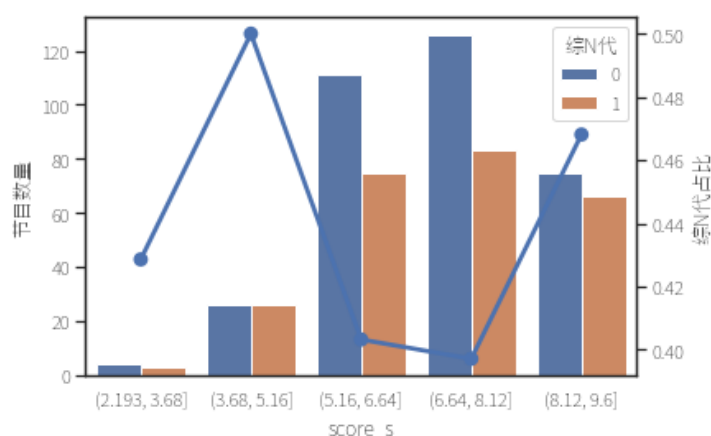


图 20: 不同分数段“综 N 代”与非“综 N 代”节目数量

观察不同分数段下“综 N 代”与非“综 N 代”的节目数量及其比例关系可发现，“综 N 代”节目在评分上并不存在绝对的优势。但相对“非综 N 代”节目来说其表现得更为极化，在极低分和极高分处的占比相对更高。

这说明，尽管“综 N 代”已经体现出了其巨大的商业价值，但在口碑方面依然可能是作为双刃剑的存在。一方面，综 N 代属性本身可能就是口碑的“结果”，高人气、好口碑的节目更可能做成 IP；另一方面，综 N 代属性也加码了观众的严苛程度，吴炜华, 张守信 (2018)^[15] 曾指出：在长久的品牌消耗、模式与嘉宾固化、审美疲劳的背后，“综 N 代”节目的品牌在效果与影响力上并未得到有效的延续。根据上述数据图像我们也可以猜测：当综 N 代节目配不上之前的节目品质时，甚至可能出现口碑的反噬。

4.3 播出平台

网综时代，播出平台方往往就是节目的制作、合作方。播出平台的差别不仅一定程度上反映了受众的差别，也与通过平台调性影响了节目选材、品质等。

观察上述节目播放平台和评分的关系图可以发现，播出平台差异的确可能伴随着整体评分

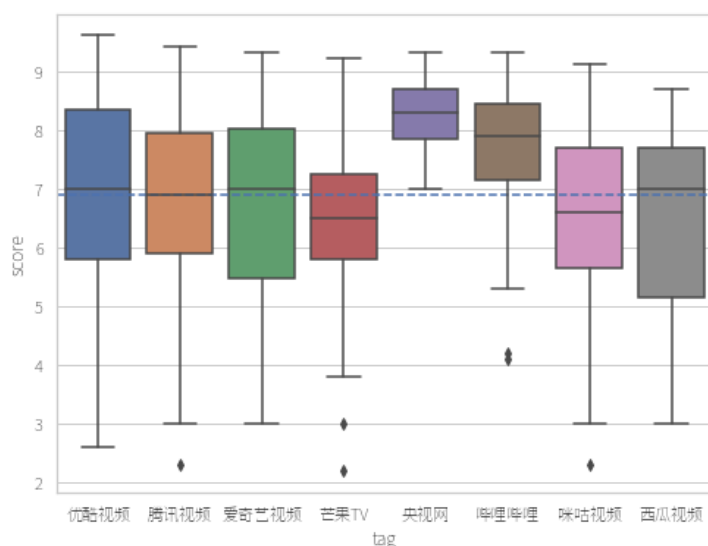


图 21: 不同平台播出节目分数分布

情况的差异，如央视网、哔哩哔哩平台播出的综艺节目整体评分较高，而芒果 TV、咪咕视频播出的节目则表现出较低的整体评分。

5 部分特征交叉分析与验证

在前期对数据进行初步的观察与分析后，我们可以发现：节目的类型标签、节目的播放平台、节目是否为“综 N 代”这三个属性的差值可能会伴随着节目评分的差异。因此，在这一部分，我们将对这三个属性进行进一步的交叉分析，并使用由该三个属性建立的特征工程，通过随机森林这一在过去的点评数据分析中被大量使用的机器学习模型进行简单的验证，通过其回归准确率结果来判断这三个属性是否具有一定的代表性与区分度。

5.1 类型与综 N 代

在分析前，根据日常生活经验，我们可以大致猜测这两个属性可能会存在一定的相关性，即部分题材可能相对其他类型节目更适合做成“系列”。但鉴于近年国产综艺对创新的追求与对“IP”建设的狂热，仍需使用数据来做进一步的判断。

观察占比的分布情况，可以注意到整体来看，不同类型中，综 N 代节目数量总是明显少于非综 N 代节目。结合图 20 中出现的综 N 代节目在极高分和极低分处出现的比例明显上升的情况，更加可以确认这一极化现象主要是由“分数”这一划分方式导致的。

而在分数分布上，可以发现，综 N 代相对非综 N 代整体分数更高和整体分数更低的情况都有出现，这证明了不同属性节目的“可持续发展性”可能存在一定的差异。

有趣的是，在占比分布中，综 N 代占比接近甚至反超非综 N 代节目的是“竞技”、“美食”、“喜剧”这三个属性，而在分数分布中也可以观察到，这三个属性对应的节目的分数分布在是否为综 N 代上并没有出现明显的差异，也许这种口碑上的稳定正是这几类节目能以 IP 形式存活下去的重要原因。

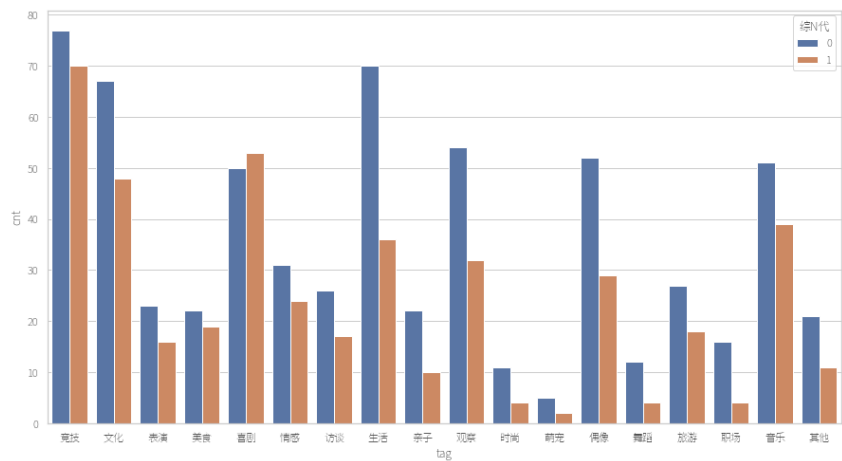


图 22: 不同类型节目中综 N 代节目的占比

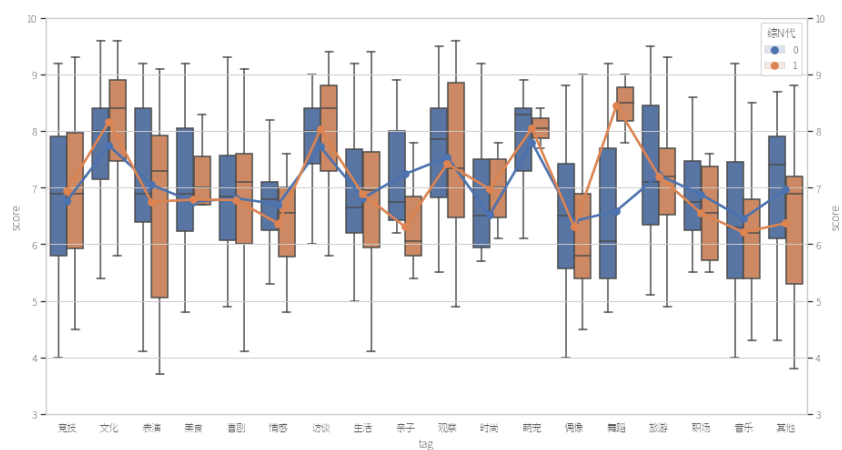


图 23: 不同类型节目中综 N 代节目的分数分布

5.2 平台与综 N 代

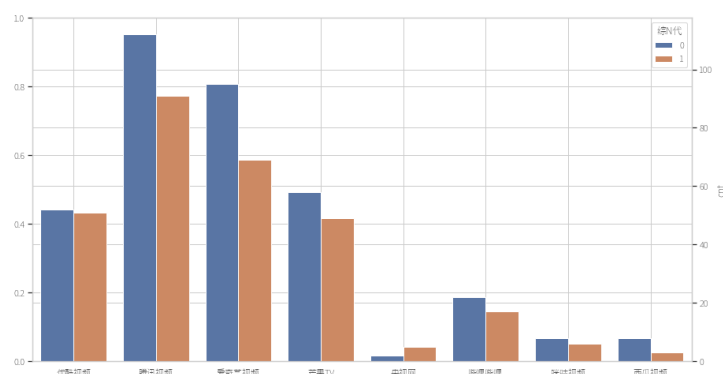


图 24: 不同平台播出节目中综 N 代节目的占比

在综 N 代节目数量分布方面，各平台并未表现出明显的差异。其中只有央视网出现了综 N 代数量明显超出非综 N 代数量的情况，这可能与央视网作为国家级媒体平台，对节目把关严，精品节目多，且不需要承受过多的市场压力，可以深耕于部分 IP 的特性有关。

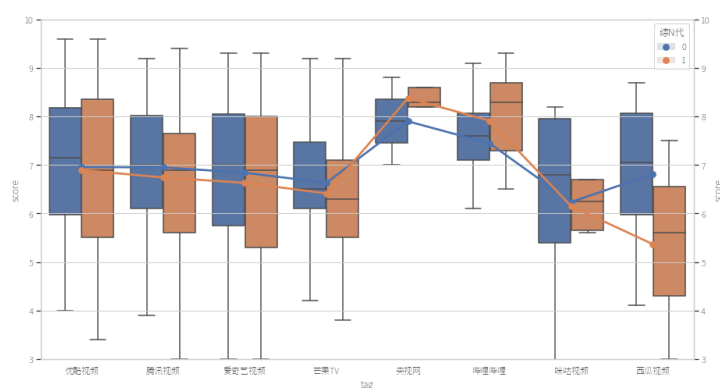


图 25: 不同平台播出节目中综 N 代节目的分数分布

而在分数分布上，则仅有央视网和哔哩哔哩出现了综 N 代节目分数更高的情况，其他平台中，芒果 TV 和西瓜视频的综 N 代节目分数明显低于非综 N 代节目。

5.3 节目类型与节目平台

在电视年代，各个电视台往往会在名称中体现自己在节目类型上的偏好，而随着互联网视频产业的发展，各大视频网站作为整合平台，往往会呈现出更加兼容并收的对节目类型的包容。但在这其中，也依然存在一定的偏好与倾向，这可能与该平台过去的定位（芒果 TV 由传统媒体演化而来、哔哩哔哩发家于青年群体），也与该平台签约的制作团队和该平台在 IP 大战中抢夺优秀综艺节目播放权的能力有关。

通过雷达图呈现的比例和热力图呈现的数值分布，我们可以发现两个明显的特征：

1. 优爱腾芒四大平台整体分布相对均匀，而体量较小的视频平台则呈现出明显的偏好偏移。如央视网在文化和生活类节目比重高，哔哩哔哩在喜剧和萌宠类目上产出多，咪咕视频和西瓜视频则分别在音乐和观察类上占比较多。

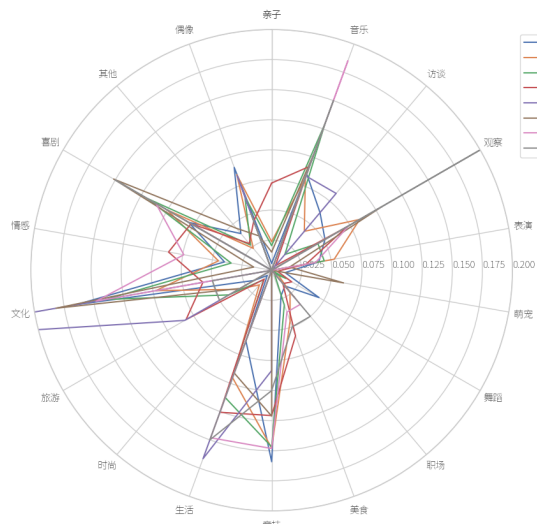


图 26: 不同平台播出不同属性节目占比

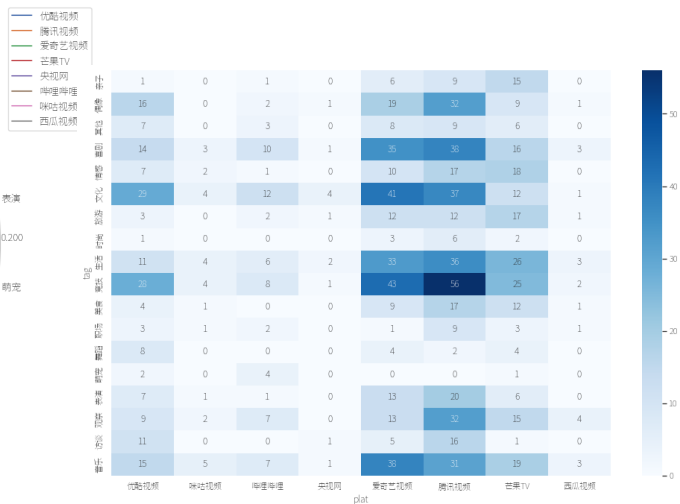


图 27: 不同平台播出不同属性节目数量

2. 而在优爱腾芒四大平台中，爱奇艺和腾讯的综艺品类相对接近，优酷视频在文化和舞蹈类尤为突出，芒果TV则在亲子、情感类有较多布局。结合现实的具体情况，的确在这两个提及这两个平台时，可以想到如《这就是街舞》、《爸爸去哪儿》等与平台绑定极其密切的爆款综艺。

而在分数分布上，整体观察可以发现，央视网、哔哩哔哩、优酷视频分数较高颜色较深，而芒果TV、咪咕视频分数较低颜色较浅，这与我们前期观察箱型图得出的结论是一致的。

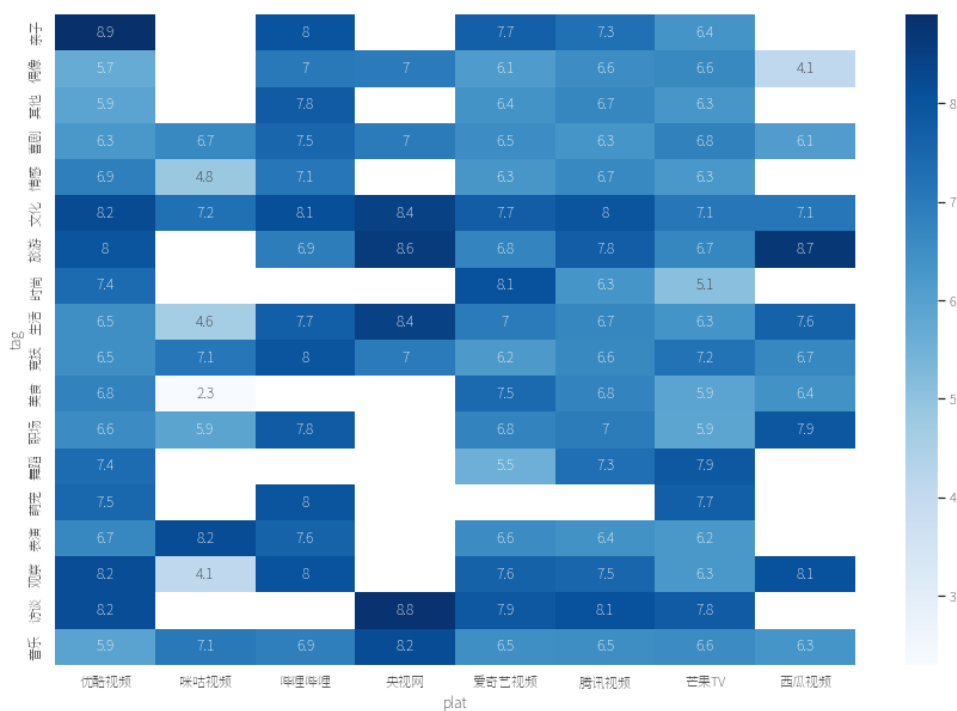


图 28: 不同平台播出不同属性节目的分数

5.4 随机森林简单回归

在对这三个属性进行进一步的观察与交叉分析后，我们考虑使用其作为机器学习输入特征，并测试模型的回归效果，进而检验特征选择的效用。在处理上，由于这三个属性全为二分值，不涉及进一步拆分，本身维度也较少，并不需要降维处理。因此可直接使用。操作上直接调用 *sklearn* 库中的 *RandomForestRegressor*，将 30% 的样本划为测试集，经过反复测试，在决策树数量为 50，最大深度为 20 时，**测试准确率即稳定在 0.75 左右**，其中一次的误差分布如图 29 所示。

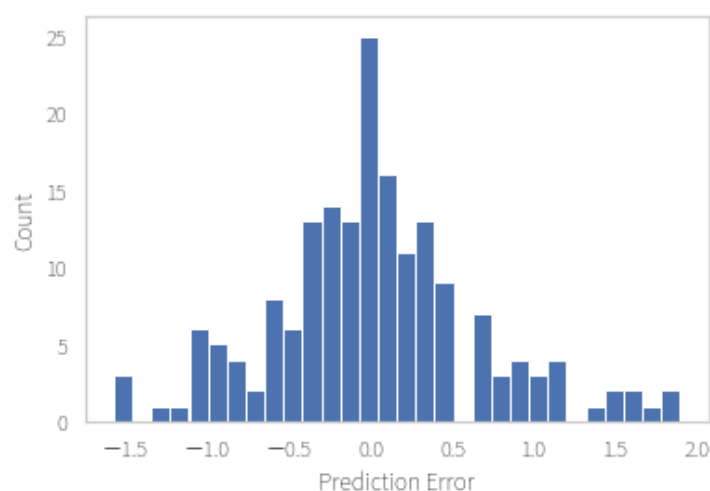


图 29: 随机森林回归误差分布

整体来看，与所阅读的论文在评分预测时的准确度基本接近，可认为特征选取有一定的价值。

6 研究结论

根据前期分析得到的部分统计与图像结果，结合现实的综艺制作生态，最终得到以下三方面的结论。

1. 综艺评分的高分的确能带动更多“想看”。而综艺的评分与综艺的内容类型和综艺的制作水平相关，其中内容好是高分的主要原因，而制作水平差则主要是低分的罪魁祸首。
2. 在内容类别上，文化、访谈、观察这类有深度的类别往往能收获高分，而情感、偶像、音乐这几类曾经“红过”的类别却在近几年口碑不佳。这类“看脸”、“看故事”、“看才艺”的，以娱乐为主要目的的综艺在电视时代曾作为综艺节目的重要组成部分，却并没有得到好的口碑。
3. 综 N 代发展火热，但能做成“N 代”的不一定就是好的。事实上，并不应该是综 N 代所代表的 IP 给予了节目良好的口碑，而是可能在综一代、综二代时制作出的内容，让节目有了持续制作的底气。央视网能深耕“文化”类内容，这才是能有大量综 N 代节目积累且评分高的原因。而芒果 TV 这样也有深厚综艺节目制作背景的平台，却更多只是在一味地延续 IP，这样的做法看似打造了长寿的综艺品牌，事实上却会损耗 IP 本身的美誉，甚至可能进

一步影响平台的综艺生态。

7 总结与展望

该课题研究是我第一次以个人身份独立完成从课题选取,到数据分析,到模型建立与结论解释的数据挖掘分析。由于时间较为紧张,整体呈现出的可以说仍是一个“未完成”的形态。出现这种未完成的原因主要在于我对数据挖掘各个工具使用不够熟练,且没有明确的问题意识,在刚开始的时候太想把所有自己辛辛苦苦爬到的数据都用上,反而浪费了大量的时间,没能在发觉到有价值的数据处进一步深挖,比如观察不同年份的节目类型分布、不同平台的综艺观看人数趋势等。如果一开始就能尽快地定位到平台、类型和综 N 代这三个属性,相信一定能挖掘出更多深层次的内容。

作为一名新传学子,同时也是一个很需要一些乐趣来“下饭”的年轻人,综艺节目一直是我个人很感兴趣的一个领域。也许有的人看综艺是为了学习知识,也许有的人看综艺只是为了找乐子打发时间,但每个看综艺的人都无法否认的是,综艺是一种不会给观众带来“时空压力”的存在,它脱离了电影院的场景限制,脱离了电视剧的冗长与连续,早已成为了我们等一会儿车、吃一会儿饭时的不二选择。经由这次分析,我初步描摹了当前国产综艺制作的形态,也对一些自己发现的问题提出了小小的看法。我也衷心地希望这次的分析能作为一个小小的切口,从数据的角度为综艺的生态发展带来一些新的思考。

参考文献

- [1] 腾讯综艺年度白皮书. 腾讯综艺年度白皮书[EB/OL]. <https://ent.qq.com/zt2020/whitePaper19/variety.htm>.
- [2] 36KE. 优酷继爱奇艺之后关闭前台播放量, 腾讯视频何时会跟进[EB/OL]. <https://36kr.com/p/1723156774913>.
- [3] DUAN W, GU B, WHINSTON A B. Do online reviews matter? —an empirical investigation of panel data[J]. Decision Support Systems, 2008, 45(4): 1007-1016.
- [4] PERSSON K. Predicting movie ratings : A comparative study on random forests and support vector machines[J]. 2015.
- [5] 苏炯铭, 刘宝宏, 李琦, 等. 基于观点动力学的在线评分人数预测[J]. 计算机工程, 2014, 000(010): 155-160,167.
- [6] 刘明昌. 豆瓣网站电影在线评分的混合预测模型研究[D]. 河北大学, 2017.
- [7] 陆君之. 基于随机森林回归算法的电影评分预测模型[J]. 江苏通信, 2018, 34(1): 4.
- [8] 谭家柱. 基于随机森林算法的 IMDB 电影评分预测研究[J]. 现代计算机, 2021, 27(30): 24-31.
- [9] 万颖迪. 中国明星真人秀收视率的影响因素及其实证分析[D]. 广西大学, 2016.
- [10] 庄玉婕, 刘娟. 我国综艺节目收视率的影响因素研究[J]. 现代商贸工业, 2020, 41(20): 2.
- [11] 郭明欣. 综艺节目点播量组合预测模型研究[D]. 河北经贸大学.
- [12] 孙扶阳. 电视综艺节目评价体系中的明星贡献度研究[D]. 浙江: 浙江传媒学院, 2020.
- [13] 孙宏姣, 葛进平, 章洁. 电视综艺节目多级分类及应用[J]. 浙江传媒学院学报, 2017, 24(04): 16-22.
- [14] 何天平. 光明日报: 综艺节目常变才能常青[EB/OL]. <http://opinion.people.com.cn/n1/2019/0227/c1003-30904296.html>.
- [15] 吴炜华张守信. 综艺文化与文化综艺的再发现——2018 年中国综艺节目的文化创新与理念重塑[J]. 电视研究, 2021, 12(17): 3.
- [16] 艾瑞咨询. 中国综 N 代 IP 价值研究 2020 年[C]//艾瑞咨询系列研究报告. 2020: 475-511.

附录

Data columns (total 29 columns):

#	Column	Non-Null Count	Dtype
0	id	595 non-null	int64
1	score	595 non-null	float64
2	name	595 non-null	object
3	导演	440 non-null	object
4	编剧	215 non-null	object
5	主演	593 non-null	object
6	类型	591 non-null	object
7	首播	595 non-null	object
8	集数	579 non-null	float64
9	单集片长	522 non-null	object
10	又名	357 non-null	object
11	简介	590 non-null	object
12	标签	595 non-null	object
13	短评	593 non-null	object
14	相关推荐	595 non-null	object
15	在看	595 non-null	int64
16	看过	595 non-null	int64
17	想看	595 non-null	int64
18	季数	213 non-null	float64
19	评分人数	595 non-null	int64
20	5星	595 non-null	object
21	4星	595 non-null	object
22	3星	595 non-null	object
23	2星	595 non-null	object
24	1星	595 non-null	object
25	讨论数	595 non-null	int64
26	短评数	595 non-null	int64
27	剧评数	595 non-null	int64
28	平台	498 non-null	object

图 30: 数据预览