



《论番剧区 UP 主的自我修养》

—— B 站番剧播放量影响因素分析

复旦第三区交通委

2019.6

目录

一、 背景介绍.....	2
二、 数据预处理.....	3
2.1 概述.....	3
2.2 衍生变量.....	3
三、 变量分析.....	4
3.1 因变量.....	4
3.2 番剧剧名.....	4
3.3 对数集均播放量与剧集来源关系.....	5
3.4 对数每集播放量与声优关系.....	6
3.5 对数每集播放量与 IP 关系.....	7
3.6 番剧种类.....	8
3.7 制作公司.....	9
四、 模型拟合.....	10
4.1 线性回归.....	10
4.2 决策树模型.....	11
五、 模型预测.....	11
六、 结论.....	12
七、 小组分工.....	12

一、背景介绍

“Bilibili(°- °)つロ,干杯!”

哔哩哔哩，即 B 站，创立于 2009 年，原名 MikuFans，意为初音未来粉丝站，于 2018 年正式上市，成长为中国最大的二次元文化视频、社群网站和年轻人文化社区，据其 2019 年第一季度的财报，B 站月活人数 1 亿以上，活跃用户平均一天使用 B 站的时间达 82min。¹

其中，番剧区以其巨大的观看流量和开创式的弹幕系统成为 B 站生态中的核心业务。“番剧”一词是一个外来语词汇，意为日本连载动画电视剧，也常泛化为所有的连载动画电视剧。B 站每季度都会从海外购进一批番剧版权，在当季度播放量和人气最高的番剧被称作“霸权番”，而番剧本身除了播放功能，其讨论、推荐等也形成了一条独特的产业链。番剧区的 UP 主的视频内容是对下一个季度的新上映的番剧进行预测和推荐。

类似于电影和电视剧，番剧的人气基于番剧本身的质量好坏，如果我是一名 B 站的番剧区 up 主，我要怎么给我的粉丝推荐和预测下一个季度的霸权番呢？而一个番剧的播放量究竟和哪些因素息息相关呢？我们从一个番剧区 UP 主的角度出发，从数据的角度去探究 B 站番剧播放量的影响因素，建立对番剧播放量的预测模型，尝试对番剧的播放量进行预测，进而选择出播放量最高的番剧作为我们预测的“霸权番”。

二、数据预处理

2.1 概述

爬取的数据由以下几栏组成：播放量、弹幕数、评分、评分人数、长评数、短评数、追番人数、集数、收费方式、番剧简介等，各特征如下表。

变量类型	变量名	详细说明	取值范围	备注
因变量	对数集均播放量	连续变量	[0, 19.989]	Log(总播放量/集数 + 1)
自变量	总弹幕数	连续变量	[0, 10104000]	
	评分	连续变量	[0, 9.9]	
	评分人数	连续变量	[0, 115555]	
	长评数	连续变量	[0, 540001]	番剧播放结束后的评价数量（类似豆瓣）
	短评数	连续变量	[0, 1575]	同长评数(最大值：紫罗兰永恒花园)
	追番人数	连续变量	[0, 6086000]	一部番剧的关注人数
	集数	连续变量	[1, 6674]	
	收费方式	类别变量 (3个水平)	大会员抢先看， 大会员观看，免费	番剧的收费方式
	制作人员	文本信息		对于番剧制作人员的描述
	番剧简介	文本信息		对于番剧剧情和设定的简单描述

表 2.1 B 站番剧相关特征概述

¹ 摘自方正证券研究报告《哔哩哔哩（BILI.O）118 页深度报告——具二次元属性的垂直在线视频领导者，PUG 生态构建 Z 世代泛娱乐社区》与 Bilibili 财报

2.2 衍生变量

随后我们对数据中“番剧简介”和“制作人员”两栏中的文本进行提取，得到各番剧的制作人，声优、制作商等变量，尝试寻找影响番剧播放量的外部、内部因素。

变量类型		变量名	详细说明	取值范围	备注
因变量		对数集均播放量	连续变量		
自变量	外部因素	开播年份	连续变量	[1928, 2018]	从开播日期中提取
		开播季度	分类变量 (4个水平)	春夏秋冬四个季度	从开播日期中提取
		制作地区	分类变量 (11个水平)	日本，中国，美国等	
	内部因素	声优	分类变量 (2个水平)	参演，未参演	提取了20位知名且在数据集中出现多的声优
		制作人	分类变量 (2个水平)	参与制作，未参与	提取了10位知名且在数据集中出现多次的制作人
		制作商	分类变量 (9个水平)	A-1 Picture，骨头社，SHAFT，其他等	
		IP	分类变量 (4个水平)	Fate系列，哆啦A梦，偶像大师系列，其他	番剧是否属于某知名的IP（知名系列）

表 2.2 B 站番剧相关衍生特征概述

三、 变量分析

3.1 因变量

由于不同番剧集均播放量差距极大，如最高的《Fate：Grand Order》达到了 17157000 次/集的播放量，而《蓝猫淘气三千问》的海洋世界和运动系列均播放量都没超过 300 次/集。故我们对集均播放量取对数进行研究并绘制直方图如下。

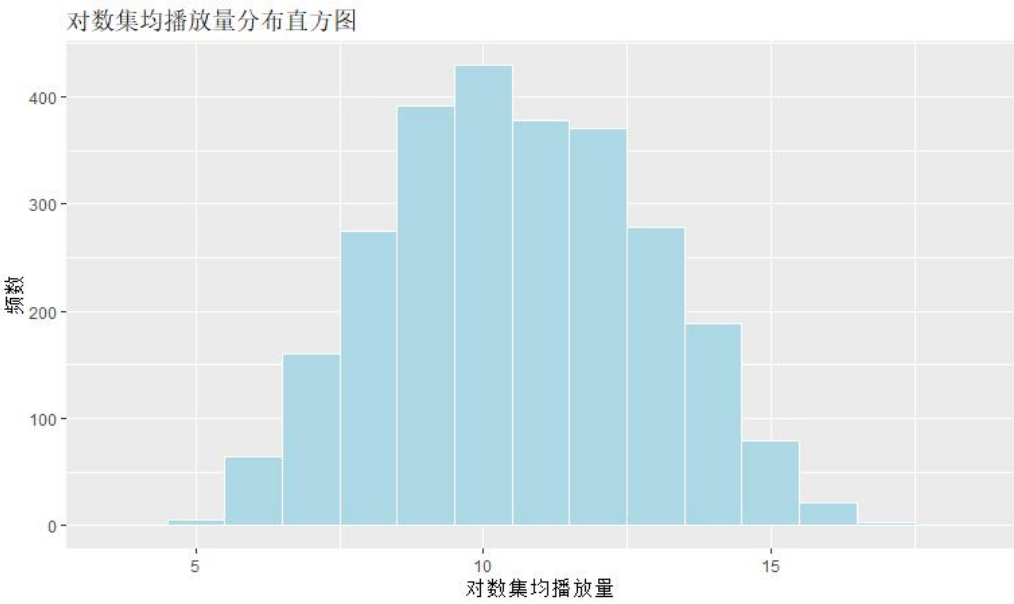


图 3.1.1 B 站番剧对数集均播放量直方图

观察对数价格的直方图，我们可以看出对数集均播放量的分布相对集中，样本之间的差距也被缩小，同时大部分样本的对数价格位于 8 至 13 之间，通过查询表中数据，可以找到各个对数水平中的代表性番剧：对数集均播放量水平高的如：《紫罗兰永恒花园》、《魔法少女小圆》、《哆啦 A 梦》；低的如《海螺小姐》、《蜘蛛骑士》和《怪医黑杰克》。

3.2 番剧剧名

为探究番剧的剧名是否具有某种一般性，我们对番剧的剧名进行词云的绘制，查看其中出现的高频词句。

从番剧的剧名中，我们可以一窥番剧的主要剧情结构：故事往往发生在日本，剧情则常是少年和少女开始在新世界的生活；或者为了拯救地球或者成为某人而进行战斗，往往还有魔法或者学校中组织的参与。从我们对词云的分析中，已经可以刻画出不乏耳熟能详番剧的轮廓，而这些番剧的特征从名字中也可见一斑。

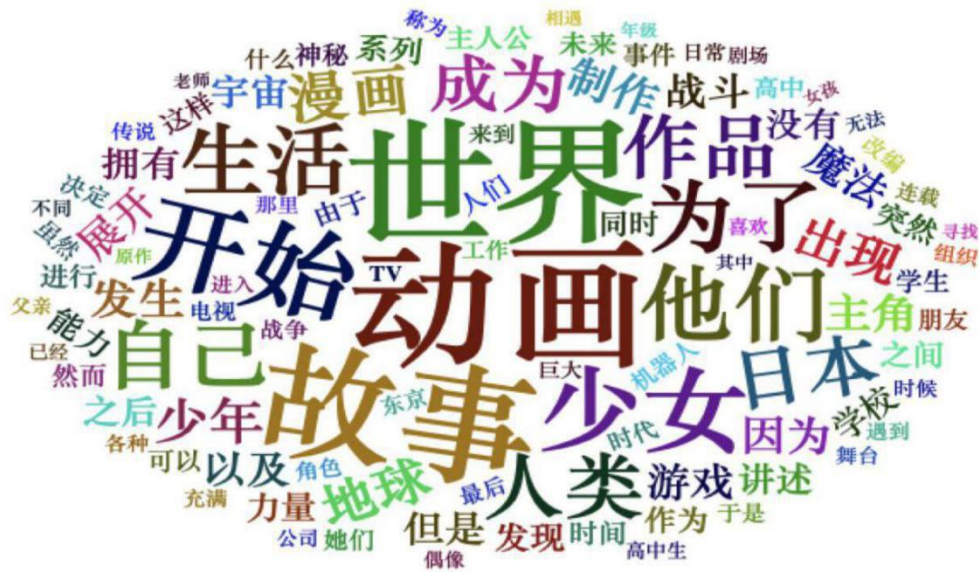


图 3.2 番剧剧名词云

3.3 对数集均播放量与剧集来源关系

从柱状图中可以看出，漫画改编的平均播放数最高，其次是小说改编，原创作品与游戏改编的动画均较低。小说和漫画的播放数高，源于高质量原作吸引的一批原作党，其次故事有了前期长时间的构思、积累、读者反馈，一般来说世界观比较充实，也经历了市场的选择，因而改编为动画，具有一定的吸引力。

而对于游戏和漫画的改编，两者呈现的东西不同，一个是作为主角的体验，另一个是故

事。游戏改的作品往往吃力不讨好，花费大量的版权费，且剧情到番剧剧本的转换困难而容易脱离本来的主旨，难以传达出游戏给人的体验，导致本来游戏的粉丝不买账，新看番的粉丝也无法被剧情吸引，故游戏改编的作品不仅在市场上很少而且普遍人气偏低。

而原创作品中质量更为参差不齐，国漫、实验作品等使得这个市场鱼龙混杂，作品中既有可能出现下一个爆款，也有可能就此沉寂，故样本量较大但是平均播放数低。

对数集均播放量与剧集来源柱状图

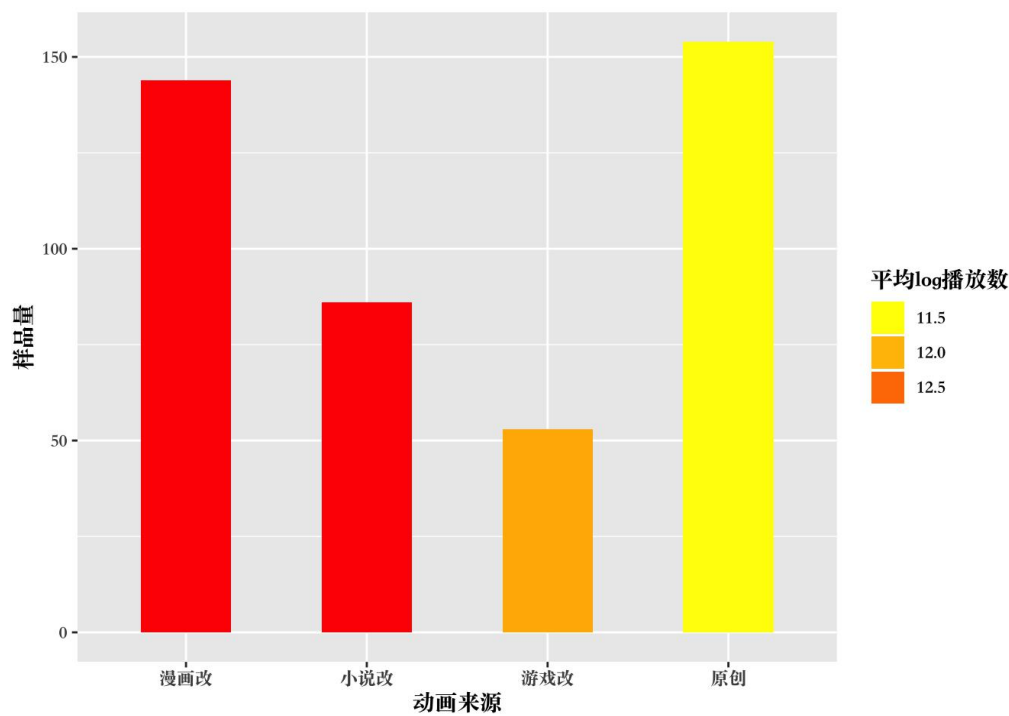


图 3.3 对数集均播放量与剧集来源柱状图

3.4 对数每集播放量与声优关系

番剧、动画片出名的很大一部分得益于声优们精湛的演绎，不同的声优有着不同特色鲜明的声线，而且往往一个声优就有着多种声线甚至一人役完成多个角色的演绎。

声优们极强的表现力和对角色的塑造让人印象深刻，参与的广播剧和综艺节目也带来偶像效应，同时在日本很多声优也作为歌手活跃着，比如《恋爱循环》的原唱花泽香菜。所以出名的声优能为一部番剧带来更多的播放量，这里我们的数据集中选择了 20 位知名度较高而且在本数据集中参与配音的声优作为变量，来考察声优对于一部番剧播放量的影响。

提取到的声优都是相对来说知名度比较高的，所以相对而言大家出演的番剧数都不少，但是各个人之间的平均对数播放量还是存在一定的差距。神谷浩史和川澄绫子，前者的平均

对数播放量较高，后者则是最低。神谷浩史，一个多声线的声优，真正能代表当季霸权的男人，代表是物语系列中的主角阿良良木历和《进击的巨人》里虽然只有 1 米 6 但是超帅的兵长利威尔。神谷浩史的多样声线为他带来了大量的粉丝，参演的几个系列都有极高的认可度和知名度，尤其是物语系列可以称得上是日本番剧里最热门的 IP 之一。

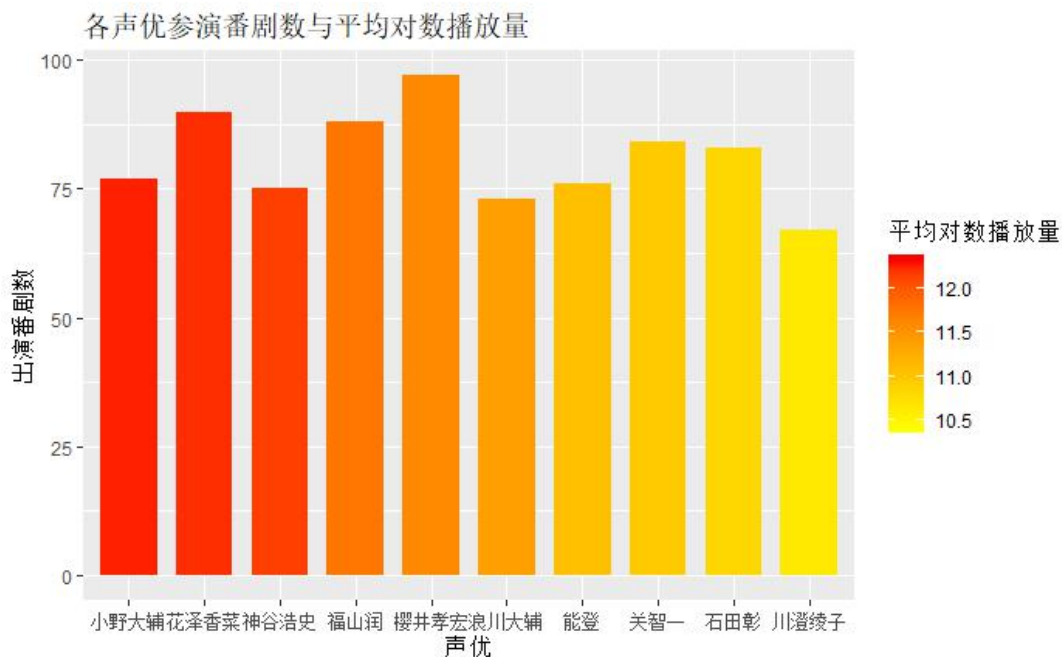


图 3.4 对数每集播放量与声优关系柱状图²

而川澄绫子小姐从人气上来说和神谷浩史差距就很大了，在她配音的角色中最著名的角色是 Fate 系列里的女主角阿尔托莉雅。同样人气第二高的小野大辅则是神谷浩史最好的搭档，他们俩也是很著名的一对 CP，两人在很多部番剧里都有过合作。

中间的三位参演的番剧数最多，其人气也颇高，个人粉丝众多，都参与配音了多部知名的番剧，比如《物语系列》，《命运石之门》，去年暑假的《工作细胞》（B 站目前最高播放量番剧）等等。香菜人气很高还有一部分原因是因为她歌手的身份，而另外两人则更多是因为角色，福山润是因为鲁路修，樱井孝宏则就像之前提到过的，自己都在广播剧里向大家呼吁“不要总是把声优和角色关联到一起”。

3.5 对数每集播放量与 IP 关系

IP 是一个最近才流行起来的词，这个 IP 不是指网络的 IP，而是指知识产权，一般指一个系列的作品里面的世界观设定，人物设定等等。为理解方便，我们就把番剧的 IP 理解为

² “能登”指“能登麻美子”

一系列作品。一个好的 IP 能带来巨大的流量，最典型的例子是 Fate 系列——中国最热门的 IP 之一。因为这个 IP 效应，其衍生产品手游 FGO 也非常热门，在 B 站的财报里可以看出 FGO 几乎占了 B 站接近一半的收入。这里我们寻找了三个热门的系列来作为变量进行分析。

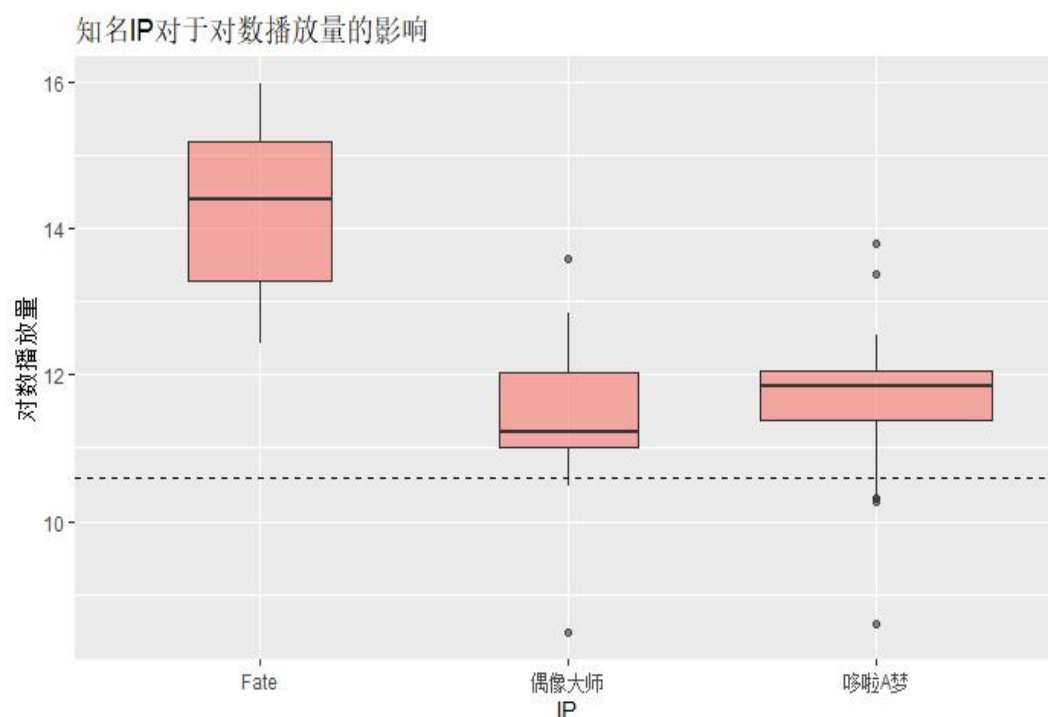


图 3.5 知名 IP 对对数集均播放量影响箱线图

图中黑色线代表市场的平均水平，可以看出知名 IP 的播放量都高于 B 站平均水平。从箱线图来看，哆啦 A 梦系列里的番剧数最多，其总体的播放量（人气）也不是很高，可能因为喜欢看哆啦 A 梦的人都已经长大。偶像大师相对而言在国内知名度较低。而对于 Fate 这个 IP，我们可以看出其在 B 站上的受欢迎程度，虽然作品数量较少但总体总体播放量都很高。去年 Fate 系列推出了一部新的作品——《Fate Apocrypha》，尽管剧情、画风以及战斗画面都饱受诟病，但是仍旧有 1 亿的总播放量，可见热门的 IP 为番剧带来巨大流量。

3.6 番剧种类

在对几个影响明显的特征进行描述性分析后，我们对番剧的种类进行聚类，将所有的番剧按照叙事和题材分为数类，尝试探究什么样的番剧更受观众青睐。

从箱线图中可以看出，如果将其他水平视为 B 站番剧播放平均水平，推理智斗番、音乐偶像番和日常番较受观众的欢迎。一些观众观看番剧可能是为了在忙碌日常生活之外得到放松，故叙事节奏较为和缓的日常番和音乐番较为符合这种观众的口味。而相比起来，场面火爆，叙事紧凑的机战番和战斗番受众人群众较小，不那么“下饭”。

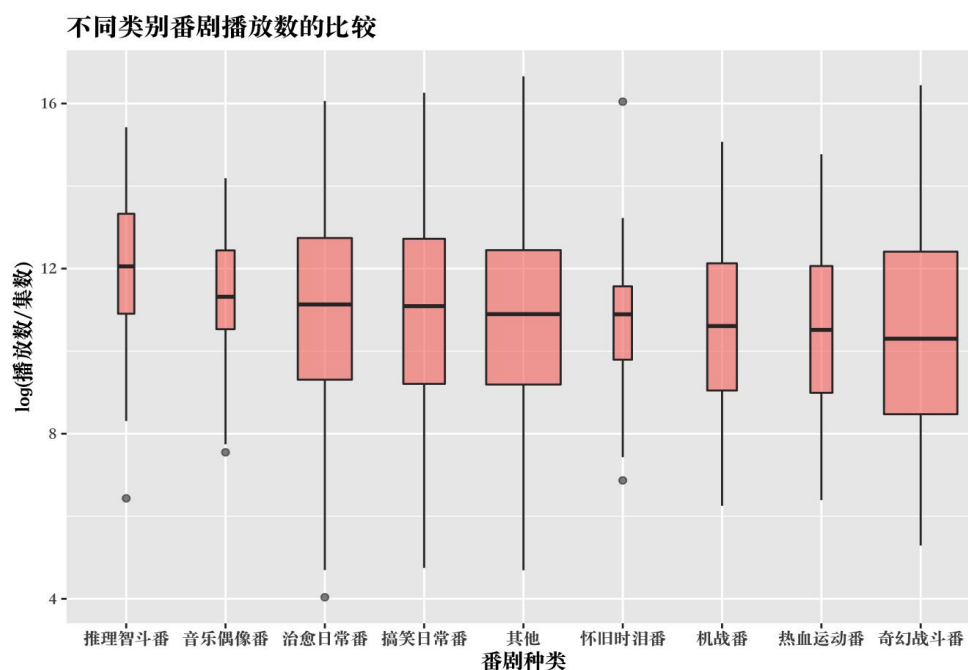


图 3.6 番剧种类对对数集均播放量影响箱线图

3.7 制作公司

以下考察不同制作公司是否会有制作动画类别的倾向性。取每个公司总播放量前三位的番剧类别画图。

前五位制作公司的播放量第一的类别都是奇幻战斗番,可以看出日本制作公司的确对奇幻战斗类作品情有独钟,虽然奇幻战斗类作品平均的播放量不高,但由于这类作品多,公司最多的收视率还是由奇幻战斗类的作品提供的。

比如 J.C.STAFF 龙之子工作室和 Production.I.G 都主打奇幻类作品,设计的类型比较广泛。J.C.STAFF 代表作有灼眼的夏娜、魔法禁书目录。Production.I.G 也擅长做热血运动番,主要的作品有进击的巨人、黑子的篮球等都是非常热血向的高人气作品,兼顾校园向热血体育片和奇幻向热血战斗剧情。

AIC 和 studio DEEN 的主力作品都是治愈日常番,如 AIC 的我的妹妹哪有这么可爱,蜡笔小新等,与 studio DEEN 的哆啦 A 梦、水果篮子、搞笑漫画日和等。

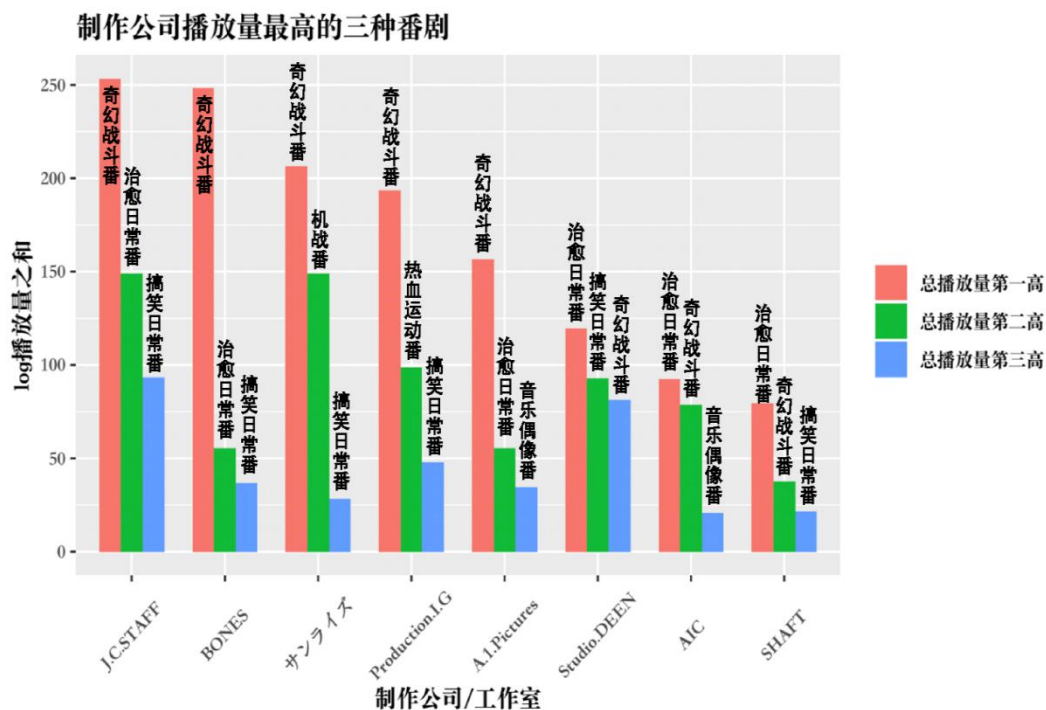


图 3.7 制作公司对不同种类番剧的播放量总和箱线图

四、 模型拟合

在对较为显著的特征进行描述分析，并根据数据集对其进行分类后，为在新番上架前对其播放量进行预测，我们使用不同的模型对进行拟合，其中效果最好的模型是线性回归模型和决策树模型，将其建模结果呈现于下：

4.1 线性回归

我们在划分训练集、测试集、去除异常点并逐步优化拟合方程后，得到的拟合结果如下：

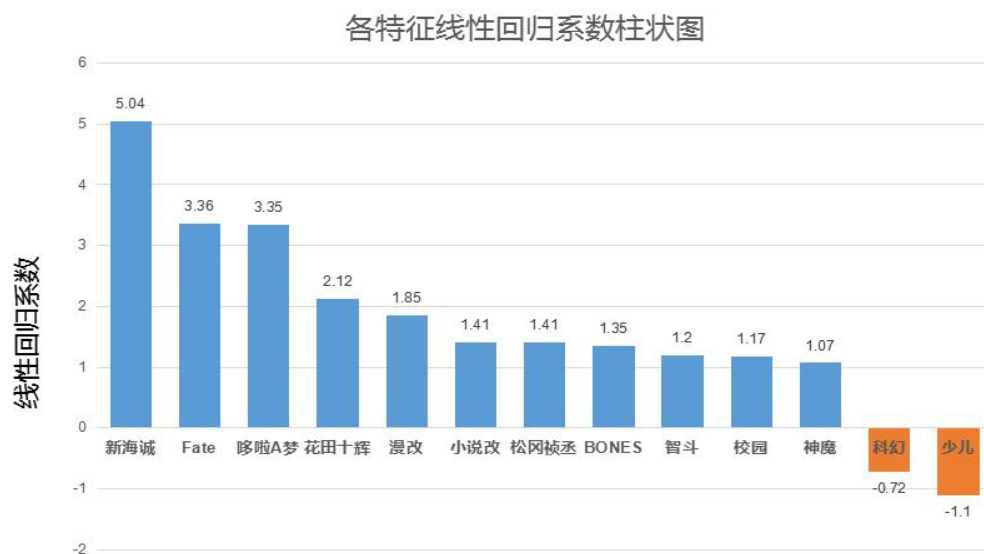


图 4.1—番剧集均播放量各特征线性回归系数（所示特征 p 值均小于 0.001）

从优化后的线性回归模型，可以看到热门 IP 能为番剧带来大量的播放量和热度，如可以被称为日本动漫界行业标杆的新海诚，与热门 IP 哆啦 A 梦和 Fate 系列，都能为番剧提供可观的品牌效应收益。同时，漫改和小说改的动画不仅有粉丝基础，且有完整的原作，质量有一定保障。制作人与声优同样能带来粉丝和番剧质量保证，为其参与制作的番剧带来流量。相比起来，科幻和少儿类型的番剧在 B 站市场反应一般，因为 B 站的主要用户都是青少年和成年人，科幻和少儿题材显得有些幼稚。

4.2 决策树模型

使用决策树模型对数据集进行建模，得到结果如下



表 4.2 番剧集均播放量决策树模型

从决策树模型，我们可以发现许多在线性模型中难以发现的有趣规律，比如在线性模型中被忽视的时间规律，2011 年中魔法少女小圆和银魂这两部高人气番剧开始播放，同时 B 站拜年祭这一番剧窗口也开始它的征程。在 2011 年前，哆啦 A 梦是毫无疑问的“霸权”动画，圆头圆脑的机器猫也是不少人的童年回忆，在 2011 年后，日本动画以其优良的制作，逐渐超越了其他地区的番剧，而漫改题材的引流作用也能拉抬番剧的收视情况。

五、 模型预测

当然，作为番剧区 UP 主，怎么能不对即将上映的番剧进行一次“有理有据”的预测呢？我们提取了七月新番的相关数据输入模型中进行预测，并与番剧区著名 UP 主 Lexburner 的预测结果进行对比，结果如下：

可以看出，我们模型的预测结果大部分与 Lexburner 的预测结果相近，不同的是我们预测《擅长捉弄的高木同学 第二季》的播放量高，而 Lexburner 预测《平凡职业造就世界最强》将成为新番霸权。这是因为我们的模型将 IP 作为了一个非常重要的因素加以考虑，故

会认为高人气番剧的续集往往也能获得较高的认可度。

	新人UP主小辉	知名UP主Lex
七月新番 霸权预测 前四名	擅长捉弄的高木同学 第二季	普通攻击是全体攻击...
	君主埃尔梅罗二世的记事簿	君主埃尔梅罗二世记事簿
	普通攻击是全体攻击...	在地下城寻求邂逅是否搞错了什么 第二季
	在地下城寻求邂逅是否搞错了什么 第二季	平凡职业造就世界最强

表 5 建模预测结果和对比

六、 结论

从建模结果上看，我们从 B 站上爬取番剧区的 2654 条数据，对影响番剧的内部、外部因素进行了分析，并对接下来的新番霸权进行了预测。

基于上述信息，我们对有志于成为下一个知名番剧区 UP 主的朋友提出以下建议：

- 1、漫改能为番剧带来大量的播放量，多看漫画原作，既能吸引到喜欢原作的粉丝，漫改作品的强势也会为你带来人气。
- 2、热门的 IP 也能增加番剧的播放量，所以一定要了解这些热门 IP，即使你觉得他们并不好看。
- 3、了解人气制作人(花田十辉)与高水准制作商(骨头社)，因为他们会是番剧质量的保障。
- 4、多涉猎各种类型的番剧，因为下一季的霸权，可能是高燃的异世界奇幻冒险，也可能是搞笑的校园日常。

“Bilibili(°- °)つロ，干杯！”希望复旦继老番茄之后，能够出现更多代表新时代青年人风貌的“水果”UP 主！

七、 小组分工

李泽君、谢炳辉：爬虫编写、数据预处理、描述分析

赵雅滢：数据预处理、描述分析

王维实：聚类建模

钟诚：撰写报告、PPT 制作