

基于视频网站会员制相关问题的调查研究

常芳锦, 高娟, 高晴, 林永祺, 王林, 王渊

安徽理工大学力学与光电物理学院, 安徽 淮南
Email: 1159215462@qq.com

收稿日期: 2021年5月13日; 录用日期: 2021年5月27日; 发布日期: 2021年6月8日

摘要

通过建立全面的调查框架, 对所收集到的调查数据进行数据清洗, 对样本的基本特征进行统计分析。采用描述性分析、列联分析等分析手段得到初步的结论。然后, 对网络视频网站会员制现状与可接受度, 进行定性和定量的分析。最后, 利用SPSS通过主成分分析确定网络视频网站会员制可接受度的影响因素构建Logistic模型, 探求对网络视频网站会员制可接受度最大的影响因素, 调控各服务的权重比, 并基于该模型完善网络视频网站会员制产品。

关键词

数据清洗, SPSS, 主成分分析, Logistic模型

Investigation and Research Based on the Related Issues of the Membership System of Video Websites

Fangjin Chang, Juan Gao, Qing Gao, Yongqi Lin, Lin Wang, Yuan Wang

School of Mechanics and Optoelectronic Physics, Anhui University of Science and Technology, Huainan Anhui
Email: 1159215462@qq.com

Received: May 13th, 2021; accepted: May 27th, 2021; published: Jun. 8th, 2021

Abstract

Through the establishment of a comprehensive survey framework, the collected adjustment data is cleaned, and the basic characteristics of the sample are statistically analyzed. Using descriptive analysis, contingency analysis and other analytical methods to get preliminary conclusions. Then,

qualitatively and quantitatively analyze the current status and acceptability of the membership system of online video websites. Finally, SPSS was used to determine the influencing factors of the acceptability of online video website membership through principal component analysis to construct a Logistic model, to explore the largest influencing factors on the acceptability of online video website membership, to control the weight ratio of each service, and based on this model Improve membership products for online video sites.

Keywords

Data Cleaning, SPSS, Principal Component Analysis, Logistic Model

Copyright © 2021 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. 调查背景与意义

1.1. 调查背景

随着网络化的发展, 视频网站播放终端的普及、视频资源的丰富, 各大平台争相提供优质资源, 不仅使得盈利增长, 也成为了学术界争相讨论的话题[1]。随着互联网技术的不断更新, 视频网站改变盈利模式引发了消费者的普遍关注, 同时也暴露出自身存在的问题。

爱奇艺在 2019 年率先宣布会员数过亿, 随后腾讯视频会员数也过亿, 国内视频网站的亿级会员时代随即到来[2]。然而, 在视频行业蓬勃发展的背后, 视频网站的巨额亏损还在继续。

1.2. 调查意义

国内外学者对视频网站会员制的相关研究较少, 通过对此方面的研究可以使相关理论体系更为完善。对于视频网站经营者, 研究视频网络会员制不仅可以增加盈利空间也可以为观影用户提供优质的体验。

本次调查主要围绕视频网站会员制用户普及度和用户体验情况展开。

2. 数据处理与分析

2.1. 调查方法

为了解视频网站会员制使用体验及用户普及度, 此次研究采取问卷调查与实地访问相结合的调查方式, 增强了市场调查的可信度。通过对视频网站会员制用户体验初步定性研究, 对可能有关的影响因素进行初步定量分析, 建立相应的模型, 对未来市场做出评估。最后, 基于调查结果与对数据的研究分析, 总结出视频网站会员制的市场现状及相应的建议。

2.2. 数据预处理

2.2.1. 数据的收集与清洗

首先, 采用线上问卷调查, 利用社交平台发送问卷链接来收集数据; 其次, 为保证数据的准确性与科学性, 运用 SPSS 对预调查收集到的 200 份线上问卷的数据进行清洗, 去除相关的缺失值、异常值、奇异点等不符合的数据。此次收集到的问卷数据没有缺失值, 有少量异常值, 直接将其删除。

2.2.2. 信度及效度分析

1) 信度分析

度量综合评价体系是否具有稳定性和可靠性的方法为信度分析方法[3]。信度越高，说明预调查使用的问卷获得的结果越可信。SPSS 对数据进行信度分析，得到克隆巴赫系数评价量表，根据此表可以进一步判断该问卷信度的高低，得到的克隆巴赫系数评价表 1 如下：

Table 1. Kronbach's coefficient evaluation table

表 1. 克隆巴赫系数评价表

克隆巴赫系数	项数
0.753	47

2) 效度分析

效度分析可以反映结果和内容的吻合程度。SPSS 对数据进行效度分析，得到 KMO 和巴特利特检验表，以此判定问卷效度的高低。表 2 为 KMO 和巴特利特检验表。

Table 2. KMO and Bachlet test table

表 2. KMO 和巴赫利特检验表

KMO 取样适切性量数		0.735
	卡方度	2855.442
Bartlett 的球形度检验	自由度	896
	显著性	0.000

KMO 值为 $0.735 > 0.7$ ，Bartlett 检验的显著性水平 $p\text{-value} = 0.000 < 0.01$ ，可得原始变量存在相关性，样本数据可以做因子分析。

本次以网络问卷调查法的方式收集数据。在腾讯问卷平台回收样本数据 593 份。首先利用 EXCEL 对所有问卷数据进行编号、加工、汇总，利用 SPSS 对该问卷所收集到的数据进行数据处理。最后得到有效问卷 593 份。

2.3. 样本的描述性统计

对清洗后的有效数据进行定量统计分析，得到男女性别比约为 0.946:1，性别比例分布较为合理。18~22 岁占总体比例最高。家庭常住地比为 0.234:1。家庭常住地分布不太均衡，在校学生占总体的比例最高。原因可能是调查范围与数量过小导致此次回收问卷的年龄分布存在偏差。

3. 模型的建立与求解

3.1. AISAS 模型

根据以上分析我们建立 AISAS 模型对视频网站会员制进行市场营销分析，从用户对视频网站的关注度、兴趣度、需求度、传播度进行分析。

1) 对视频网站会员服务的关注度

我们建立 AISAS 模型对视频网站会员制进行市场营销分析，首先对视频网站用户对这一项服务的了

解度进行了汇总分析。

表 3 和表 4 分别为用户对超前点播的了解程度和关注度，已知“超前点播”是在会员的基础上再付费。因此用户对超前点播这一服务的了解度、关注度在一定程度上可以反映出用户对视频网站会员服务的市场关注度。

Table 3. User's understanding of advanced on-demand

表 3. 用户对超前点播的了解程度

选项	百分比(%)
非常了解	34.50%
听说过但不了解	45.30%
没听说过	20.20%

Table 4. User attention to advanced on-demand

表 4. 用户对超前点播的关注度

选项	百分比(%)
非常了解	43.20%
听说过但不了解	56.80%

2) 对视频网站会员服务的兴趣度

图 1 为超前点播随着年份变化的热度趋势，可以看到热度逐年提升。由于视频网站推出会员这一服务已过数年，不容易调查用户对视频网站会员服务的兴趣度，因此研究人员决定分析用户对“超前点播”这一会员增值服务的兴趣度，从而侧面反映出用户对视频网站会员服务的兴趣度。

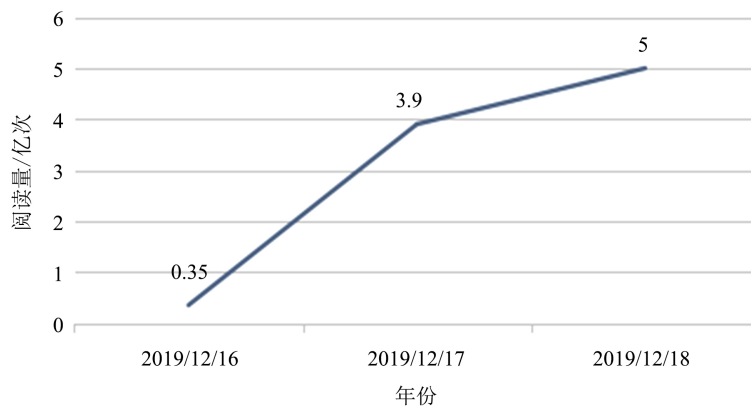


Figure 1. Trend chart of the popularity of advanced on-demand broadcasting

图 1. 超前点播的热度趋势图

3) 对视频网站会员服务的需求度

据《中国互联网络发展状况统计报告》显示 2018 年 6 月底，中国网络视频用户规模达到 6.12 亿，较 2017 年底增加 3309 万，占网民整体的 73.9%，用户规模非常庞大[4]。

商品广告的存在给视频网站带来了很大一部分的会员数量。以国内主流的视屏网站为例，其VIP会员服务中一定包括了会员可去除广告和跳过广告这一特权。据艾瑞咨询发布的报告显示，73.6%的VIP会员看中的会员权益是去除广告和跳过广告[5]。

图2和图3为视频网站VIP会员看中的VIP权益，可以得出视频网站逐渐成为网络广告的投放渠道，广告已成为不可缺少的营销和推广手段。网络视频广告的收入目前来说是逐年增长，视频网站投放广告是其收入的一大来源[6]。正因为广告的大量投入，也大大刺激了视频网站用户开通会员的需求。

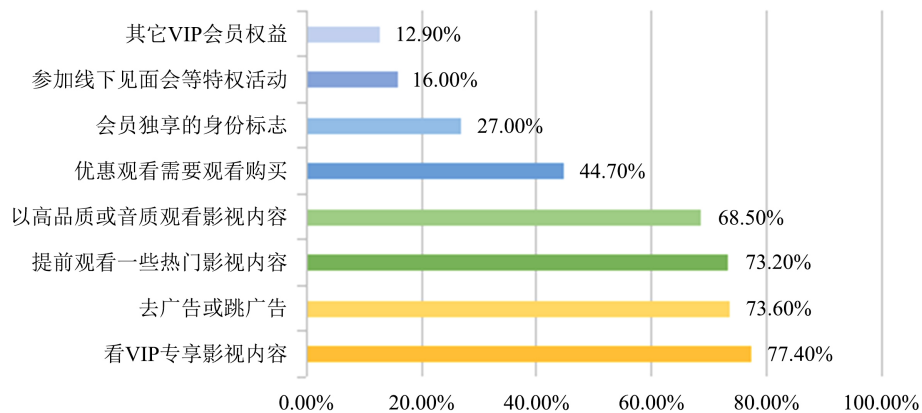


Figure 2. Rights and interests valued by video site members in this survey

图2. 本次调查中视频网站会员看重的权益

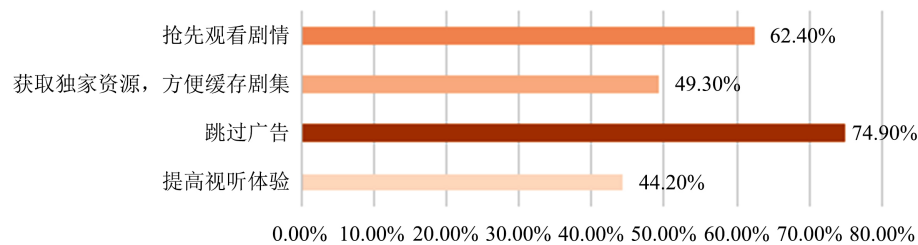


Figure 3. VIP rights in VIP members of video websites

图3. 视频网站VIP会员中的VIP权益

4) 对视频网站会员服务的传播度

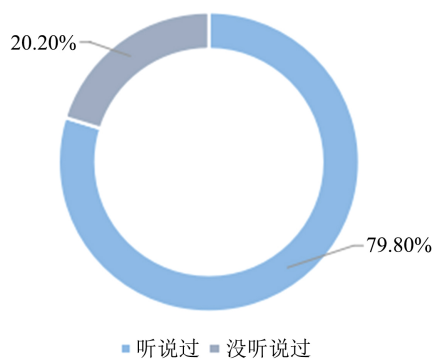


Figure 4. Awareness of advanced on-demand

图4. 超前点播的知晓度

图 4 为超前点播的知晓度，图中可见近八成的人听说过超前点播，而在这一群体中，对关注超前点播的约占五分之二。与此同时，观众也在微博、知乎、抖音等社交平台讨论和分享剧情，引发了广泛网友热议，更大程度上加深了人们对超前点播的认识，成功引起了一部分处于犹豫阶段的人的兴趣。

3.2. 主成分分析模型

3.2.1. 理论分析

问卷的效度及信度较好，可以进行因子分析。利用 SPSS 对这些因素进行统计分析，选出部分相关性较大的因素，但若将这些因素都加入其中，可能会出现多重共线性，也可能出现对该模型不显著的部分因素，为了避免多重共线性与其他误差的出现，提高模型的可信度与科学性，故对这些因素进行降维处理。

假设有 n 个样本， p 个指标，则可构成大小为 $n \times p$ 的样本矩阵 x ：

$$x = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} = (x_1, x_2, \cdots, x_p)$$

1) 首先对其进行标准化处理：

按列计算均值 $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$ 和标准差 $S_j = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}}$ ，计算得标准化数据 $x_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j}$ ，原始样本矩阵经过标准化变为：

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} = (X_1, X_2, \cdots, X_p),$$

原始样本矩阵经过标准化变化为：

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}。$$

2) 计算标准化样本的协方差矩阵

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix},$$

其中，

$$r_{ij} = \frac{1}{n-1} \sum_{k=1}^n (X_{ki} - \bar{X}_i)(X_{kj} - \bar{X}_j) = \frac{1}{n-1} \sum_{k=1}^n X_{ki} X_{kj},$$

$$R = \frac{\sum_{k=1}^n (x_{ki} - x_i)(x_{kj} - x_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - x_i)^2 \sum_{k=1}^n (x_{kj} - x_j)^2}}$$

3) 计算 R 的特征值和特征向量

特征值: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ (R 是半正定矩, 且 $tr(R) = \sum_{k=1}^p \lambda_k = p$)。

$$\text{特征向量: } a_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{p1} \end{bmatrix}, a_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{p2} \end{bmatrix}, \dots, a_p = \begin{bmatrix} a_{1p} \\ a_{2p} \\ \vdots \\ a_{pp} \end{bmatrix}.$$

4) 计算主成分贡献率以及累计贡献率

$$\text{贡献率} = \frac{\lambda_i}{\sum_{k=1}^p \lambda_k} (i=1, 2, \dots, p),$$

$$\text{累计贡献率} = \frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^p \lambda_k} (i=1, 2, \dots, p)。$$

3.2.2. 模型求解

影响大众对视频网站会员制的可接受度的因素有很多, 如年龄 X_1 、职业 X_2 、最高学历 X_3 、平均月收入 X_4 、平均每天使用视频类 APP 的时长 X_5 、平均每天使用视频类 APP 的次数等因素 X_6 。变量较多不容易分析且变量之间可能存在着较强的相关性, 故首先对这些变量进行主成分分析。

将数据导入 SPSS 之后, 进行主成分分析。得到的相关系数矩阵可视化结果如图 5 所示:

1	0.79	0.36	0.76	0.25	0.51
0.79	1	0.31	0.55	0.17	0.35
0.36	0.31	1	0.35	0.64	0.58
0.76	0.55	0.35	1	0.16	0.38
0.25	0.17	0.64	0.16	1	0.63
0.51	0.35	0.58	0.38	0.63	1

Figure 5. Visual representation of the correlation coefficient matrix
图 5. 相关系数矩阵的可视化表示

得到变量的相关系数矩阵的特征值、相应的特征向量以及贡献率如表 5 所示:

Table 5. Results of principal component analysis
表 5. 主成分分析的结果

特征向量	a_1	a_2	a_3	a_4	a_5	a_6
X_1	-0.112	-0.365	0.092	-0.122	-0.080	-0.321
X_2	0.142	0.123	0.524	0.342	0.027	0.443
X_3	0.023	0.352	-0.278	0.543	0.405	-0.125
X_4	0.012	0.547	-0.345	-0.108	-0.235	0.321
X_5	0.427	-0.223	0.182	-0.013	-0.731	0.012
X_6	0.532	-0.123	0.153	-0.231	0.485	0.128
特征值	3.826	1.457	0.531	0.432	0.289	0.141
贡献率	0.578	0.265	0.063	0.042	0.030	0.022
累计贡献率	0.578	0.843	0.906	0.948	0.978	1.000

将六个变量的贡献率结果可视化，如下图 6 所示：

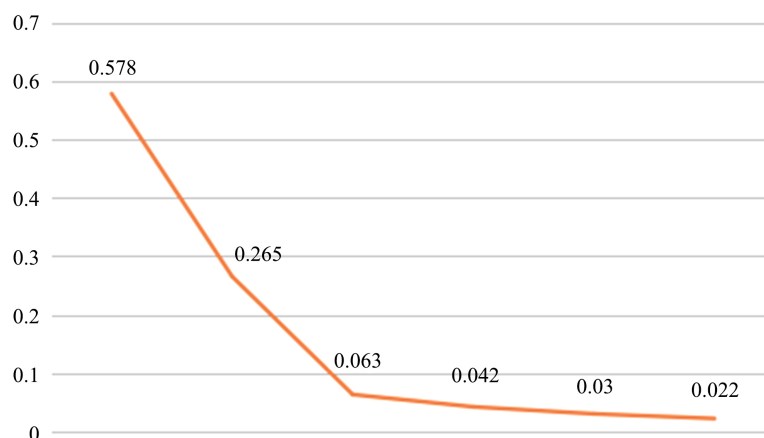


Figure 6. Contribution rate of variables
图 6. 变量的贡献率

从图 6 变量的贡献率可以看到前两个和前三个的主成分的累计贡献率分别达到 84.3% 和 90.6%。第一主成分 F_1 在平均每天使用视频类 APP 的时长 X_5 和平均每天使用视频类 APP 的频率 X_6 上有较大的正载荷，其余变量上有较小的正载荷或是负载荷，因此可认为这个主成分度量了对视频类 APP 的使用程度。第二主成分 F_2 在最高学历 X_3 和平均月收入 X_4 上有较大的正载荷，而在其余变量上有负载荷或较小的正载荷，因此第二主成分可反应了收入水平。第三主成分的贡献率仅有 6.3%，且第三主成分很难给出明显的解释，因此我们取前面两个主成分。

由此得出的主成分为：

$$F_1 = -0.112X_1 + 0.142X_2 + 0.023X_3 + 0.012X_4 + 0.427X_5 + 0.532X_6$$

$$F_2 = -0.365X_1 + 0.123X_2 + 0.352X_3 + 0.547X_4 - 0.223X_5 - 0.123X_6$$

因此, 通过主成分分析对变量进行降维之后选取两个指标作为影响视频网站会员制接受度的因素, 分别是收入水平 Y_1 和对视频类 APP 的使用程度 Y_2 。

3.3. 网站会员制前景预测—Logistic 回归模型

3.3.1. 理论分析

Logistic 分析方法主要应用于计算某些事件发生的概率 P , 所以 Logistic 适用于该选题的客户续保概率问题。当因变量只取 0 与 1 两个数值时, 即是一个二元变量, 因变量取 1 的概率 $p = (y=1)$ 便是所要分析的对象。根据上述思想, 以上述主成分作为新的自变量, 建立客户续保概率影响因素的 Logistic 回归模型的基本形式为:

$$P(Y=1|y_1, y_2, \dots, y_k) = \frac{\exp(\beta_0 + \beta_1 y_1 + \dots + \beta_k y_k)}{1 + \exp(\beta_0 + \beta_1 y_1 + \dots + \beta_k y_k)}$$

上式中, $\beta_1, \beta_2, \dots, \beta_k$ 等多元线性回归模型中的回归系数 ε_i 大致相同, 这个式子表达了当指标为 y_1, y_2, \dots, y_k 时, 自变量 P 为 1 的概率。对这个式子进行对数变换, 可得:

$$\ln \frac{P}{1-P} = \beta_0 + \beta_1 y_1 + \dots + \beta_k y_k$$

考虑到定性实践, 则 P 的取值只有 0、1, 为了使得该数学模型符合实际, 先定义一种单调连续的概率函数 π , 令 $\pi = P(Y=1|y_1, y_2, \dots, y_k)$, 其中 $0 < \pi < 1$, 则得到 Logistic 模型的变形为:

$$\ln \frac{\pi}{1-\pi} = \beta_0 + \beta_1 y_1 + \dots + \beta_k y_k,$$

最后得到的模型为:

$$\pi = \frac{\exp(\beta_0 + \beta_1 y_1 + \dots + \beta_k y_k)}{1 + \exp(\beta_0 + \beta_1 y_1 + \dots + \beta_k y_k)}。$$

其中, $\beta_1, \beta_2, \dots, \beta_k$ 为模型自变量的系数, y_1, y_2, \dots, y_k 作为自变量, 通过 Logistic 模型可计算此相关矩阵。

3.3.2. 模型求解

我们对 Logistic 模型进行求解, 求得视频网站会员制可接受度以及相关重要的影响因素。利用 SPSS 软件进行 Logistic 回归分析。

$$P(Y=1|y_1, y_2) = \frac{\exp(-1.567 + 1.209 y_1 + 0.564 y_2)}{1 + \exp(-0.567 + 1.209 y_1 + 0.564 y_2)}$$

最终, 求得的 Logistic 模型为:

$$\pi = \frac{\exp(-1.567 + 1.209 y_1 + 0.564 y_2)}{1 + \exp(-0.567 + 1.209 y_1 + 0.564 y_2)}。$$

根据所求得的模型可知, y_1 的相关系数最大, 由此可得影响网络视频会员制可接受度, 观看视频 APP 的频率是否产生变化的主要因素为广告的时长和对会员制的了解程度。

3.3.3. 模型的检验

对 Logistic 回归模型作系数的综合检验, 判断模型是否合理。Logistic 回归模型的拟合优度的卡方值为 121.542, 且 p -value < 0.05 , 即表明该模型的自变量与因变量 $\ln \frac{P}{1-P}$ 的线性关系显著在所有回归系数

不同时为 0 时, 模型合理。

计算出相关系数 $R^2 = 0.105$, 该值在一定的程度上受自变量和样本容量的影响; 对较小的样本容量容易获得较大的 R^2 , 而较大的样本容量求得的 R^2 往往偏小, 说明拟合优度不高, 即使样本容量和自变量数相等, 也有可能具有虚假性, 但是本次问卷收集的数据偏多, 故所得值对模型总体的影响不大。

4. 结论

从 2015 年起, 视频网站从不同的领域吸引着消费者的兴趣。在网络电影发展日趋成熟的背景下, 网络自制剧和自制综艺不断进入到了观众的视野[7]。会员服务逐步走进了人们的生活。国内视频网站通过大量的免费内容吸引流量, 并将流量引入其他相关的付费视频服务中。视频网站对此模式进行了创新, 提出了“超前点播”的播出方式, 同时也加强了用户的付费意识, 实现了多元化的收入。视频网站的收入来源于用户, 为了视频网站的长期稳定发展, 提升用户的体验感是不可缺少的。我国视频网站发展没有精准的定位, 这就需要各个视频网站根据自身特点做适当强化, 将特点发展成为优势。品牌的多元延伸是借助视频网站的版权优势将各个视频资源利用最大化, 多开设视频网站独有的综艺节目, 建立起一整套独具特色的自制节目系列。寻求更为合理的“超前点播”制度, 及时调查消费者的用户期待。对于不同热度的影视作品在“超前点播”的集数和费用上采取弹性制度, 合理进行, 形成长期并稳定的情感依赖。这需要通过长期的市场调研来制定出切实有效的会员制方案, 促进视频网站长效发展。

参考文献

- [1] 韩雪. 国内视频网站的盈利模式优化分析——以爱奇艺、腾讯视频、搜狐视频为例[D]: [硕士学位论文]. 吉林: 吉林大学, 2017.
- [2] 刘燕秋. 视频流媒体激战 2020: 优爱腾下沉求生, B 站以一敌三? [EB/OL]. <https://finance.sina.com.cn/chanjing/cywx/2020-01-01/doc-iihnzakh1267095.shtml>, 2021-03-05.
- [3] 薛薇. SPSS 统计分析方法及应用[M]. 北京: 北京电子工业出版社, 2009.
- [4] CNNIC. 第 43 次中国互联网络发展状况统计报告[R]. 北京, 2019.
- [5] 艾瑞咨询. 2018 年中国网络视频行业经营状况研究报告[R]. 2018-5-23.
- [6] 艾瑞咨询. 2019 年 Q3 中国网络广告市场数据发布报告[R]. 2019-12-27.
- [7] 央广网. 2018 年网络视频付费用户已达 3.47 亿人[EB/OL]. <https://baijiahao.baidu.com/s?id=1636105501077943050&wfr=spider&for=pc>, 2021-03-05.