

大学生抖音直播体验及其影响因素分析

黄维嘉

北方工业大学理学院, 北京

收稿日期: 2023年9月12日; 录用日期: 2023年10月9日; 发布日期: 2023年10月16日

摘要

近年来, 网络直播赛道呈火爆之势, 人们逐渐形成了观看抖音直播的习惯, 其用户规模不断壮大。本文采用粗糙集的简约算法, 通过python的数据编程和0-1归一化分析, 实现可视化界面, 以此对大学生使用抖音的情况进行分析。本次数据分析主要针对用户对抖音app的忠诚度, 以及对于观看抖音直播的影响程度, 找出与之有强关联性的变量因素, 并得出结论: 1) “是否只有抖音一个直播app”与“是否驻足过抖音直播间”强相关; 2) 直播间氛围、直播内容、直播是否有内容有价值、弹幕交流是否友好4个价值取向因素对抖音直播有显著影响。最后根据结论提出切实可行的意见, 即增加抖音app使用率、提升抖音直播间内在价值。

关键词

抖音直播, 忠诚度, 影响程度, 粗糙集简约算法, 0-1归一化

Research on College Students' TikTok Live Streaming Experience and Its Influencing Factors

Weijia Huang

College of Science, North China University of Technology, Beijing

Received: Sep. 12th, 2023; accepted: Oct. 9th, 2023; published: Oct. 16th, 2023

Abstract

In recent years, the network live broadcast track is popular, and people gradually form the habit of watching TikTok live broadcast, and its user scale is growing continuously. The text adopts the parsimonious algorithm of rough set, and realizes the visual interface through python data programming and 0-1 normalization analysis, so as to analyze the use of TikTok by college students.

This data analysis focuses on users' loyalty to the TikTok app and the degree of influence on watching TikTok live broadcasts, identifies variable factors with strong correlations, and concludes that 1) "whether there is only one live broadcasting app for TikTok" is strongly correlated with "whether one has ever stopped at a TikTok live broadcasting room"; 2) four value orientation factors: live broadcasting room atmosphere, live broadcasting content, whether live broadcasts have valuable content, and whether live broadcasts are friendly, have a significant impact on TikTok live streaming. Finally, according to the conclusion, practical advice is proposed, for example, methods to increase the utilization rate of TikTok app and enhance the intrinsic value of TikTok Live Streaming.

Keywords

TikTok Live Broadcast, Loyalty, Influence Degree, Rough Set Simplicity Algorithm, 0-1 Normalization

Copyright © 2023 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

1.1. 研究背景

根据《第 49 次中国互联网络发展状况统计报告》显示[1], 截止至 2021 年 12 月, 我国短视频用户规模达 9.34 亿人, 相较于 2020 年 12 月增长了 6080 万人, 占全体网民的 90.5%。并且据艾媒咨询发布的《2022 年中国直播行业发展现状及市场调研分析报告》数据显示, 截止至 2022 年 12 月, 我国抖音直播用户规模超过 7 亿人, 将近八成用户每天的抖音直播平均在线时长在 1 小时左右。由此可见, 抖音 app 作为一个短视频社交软件已经被大众所熟知和喜爱, 用户规模不断扩大。越来越多的人开始下载和使用抖音来分享自己的生活, 并且由于直播功能的日渐完善和普遍, 直播卖货在抖音 app 上也是十分常见。因为互联网的快速发展和抖音平台的火爆之势, 分析用户使用抖音 app 的体验感和影响抖音平台的发展程度十分重要, 可以帮助我们看见未来互联网时代的发展趋势和抖音平台的优势与劣势。

1.2. 研究目的

目前, 各大网络短视频的使用者主要集中在 00 后与 90 后, 在移动互联网迅速发展的影响下, 他们可以较为迅速地适应媒体环境的不断变化。例如在使用抖音 app 时, 他们可以熟练地将短视频的制作、互动、直播等功能与学习和生活相互转化、相互融入。众所周知, 抖音 app 对大学生的影响有利有弊, 大学生可以通过抖音平台从中调节情绪、拓展眼界以及实现自我价值, 但也会产生一些负面影响。因此本文通过问卷星收集问卷, 在此基础上研究大学生对抖音 app 的忠诚度与用户特征的关系, 以及影响大学生观看抖音直播的主要因素, 对抖音平台和大学生而言都具有重要意义。

2. 文献综述

目前国内外存在关于大学生观看网络直播的研究, 大多数研究者是从网络直播这把双刃剑为出发点进行分析, 主要针对于直播平台的优势与弊端。从整体而言, 消极影响居多, 并且大部分是根据消极影响提出相应的建议与解决方法。于建波、马金宝和朱宏(2017)三人指出, 在网络直播如抖音等平台受到大学生热烈欢迎的同时, 各种不良信息也在快速传播, 这对大学生的身心健康存在不利影响; 有的网络直

播中还涉及猎杀动物、酒驾等各种违法犯罪的内容；大学生的心智不够成熟，法律知识不健全，这些违法乱纪的内容会对大学生的价值观和世界观造成严重的负面影响[2]。

黄京, 杨雪莹等人(2018)通过发放问卷调查发现, 经常和偶尔观看直播的人数比例接近八成, 观看生活娱乐类网络直播的人数占总人数的 58.30%, 男生和女生在“喜欢观看的类型”中存在明显差别。不过, 无论男生还是女生均认为网络直播最大的作用是增加生活乐趣[3]。刘珊珊(2017)根据国家统计局发布的 2016 年国民经济和调查公布数据显示, 大学生是网络直播的主要受众群体, 直播平台一方面给大学生提供了个性化的生活方式, 另一方面, 大学生可以赚取一定的生活费[4]。所以, 网络直播很受大学生的欢迎。

从上面的梳理可以看出, 当下网络直播在大学生群体中的接受度和认可程度非常高。不止观看网络直播的人数较多, 而且进行网络直播的大学生群体也十分庞大。

3. 问题分析

根据问卷星收集的数据显示, 在原始数据中, 调查问卷数据总计 205 份, 即 205 行数据。包括“提交答卷时间”、“所用时间”等在内, 一共 44 个属性, 即 44 列数据, 构成 205 行*44 列矩阵。

考虑到矩阵中的具体值大多为字符型数据, 不可采用连续型变量方法处理, 需转变为离散变量。如问题“6、您现阶段对于抖音直播的态度?”中, 可将“跳过”、“喜欢”、“中立”编码为 0、1、2, 按离散型变量处理。

44 个问题存在一定因果关系, 可用于分析抖音直播体验。如可分析“什么样的人群特征更可能只有抖音一个 app”。

为分析以上信息, 综合问卷数据结构特点, 拟采用粗糙集理论分析[5]。粗糙集理论是一种刻画不完整性和不确定性的数学工具, 能有效地分析不精确, 不一致、不完整等各种不完备的信息, 还可以对数据进行分析 and 推理, 从中发现隐含的知识, 揭示潜在的规律。粗糙集理论是建立在分类机制的基础上的, 它将分类理解为在特定空间上的等价关系, 而等价关系构成了对该空间的划分。粗糙集理论将知识理解为对数据的划分, 每一被划分的集合称为概念[6]。粗糙集理论的主要思想是利用已知的知识库, 将不精确或不确定的知识用已知的知识库中的知识来(近似)刻画。该理论与其他处理不确定和不精确问题理论的最显著的区别是: 它无需提供问题所需处理的数据集合之外的任何先验信息, 所以对问题的不确定性的描述或处理可以说是比较客观的。

考虑到粗糙集方法具有以下特点, 本分析使用了该方法: 1) 它能处理各种数据, 包括不完整的数据以及拥有众多变量的数据; 2) 它能处理数据的不精确性和模棱两可, 包括确定性和非确定性的情况; 3) 它能求得知识的最小表达和知识的各种不同颗粒层次; 4) 它能从数据中揭示出概念简单, 易于操作的模式。

而本调查问卷中 1) 数据包括了多条“跳过”的不完整数据, 且不同属性都是由众多离散性的变量组成的。2) 答卷内容包含了不精确性和模棱两可。如 12 题用户填报的为“有影响”、“一般”、“无影响”、“有一点影响”等。3) 可以通过问卷分析如“您是否只有抖音一个直播 app”与哪些用户特征关联密切等关于知识表达的内容。4) 可以从数据中通过约简揭示因果关系。

4. 数据预处理

44 个属性列中, 大多数列属于离散型变量, 需用 0、1、2……替换编码, 便于后续编程分析。部分特殊列需特别处理:

4.1. 提交答卷时间

该列数据类型是时间变量。该属性不在本次数据分析所关心的问题中, 不进行深入处理。

4.2. 所用时间

该列是时间变量，可用于判断问卷数据的质量。考虑到答卷所用时间短的数据多是因为“未驻足过抖音直播间”，选择了“跳过”。而未选“跳过”的答卷中，其答卷时间频次图如图 1 所示。未选“跳过”的答卷中最短用时为 25 秒，而大多问卷答题用时集中在[25, 235]秒区间，无明显区分性。

故本次分析选用所有 205 条数据，不考虑答卷时间长短对问卷质量的影响。

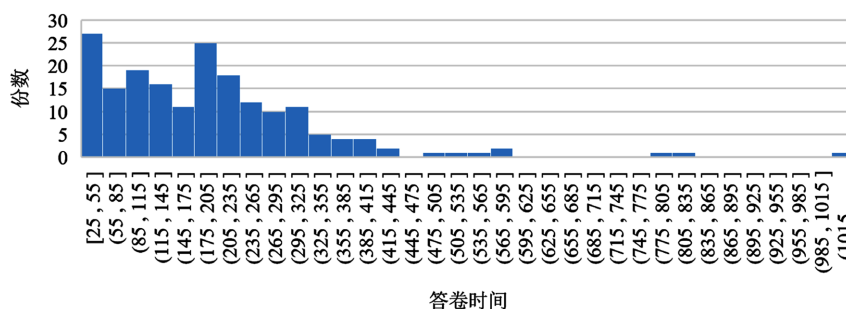


Figure 1. Time and frequency chart of the answer paper
图 1. 答卷时间频次图

4.3. 来自 IP

格式为“223.104.103.7 (河北 - 保定)”。该列包含省、市两层信息。在预处理中分解为“省”、“市”两列。

4.4. 多项问题

如格式“游戏直播(游戏直播, 比赛等) | 卖货直播(如东方甄选) | 搞笑类直播”。通过“|”符号，拆分为数组。

4.5. “跳过”的处理

“跳过”本身也反应了被调查者对抖音的看法，不做过滤处理。同样进行编码。

4.6. 其他编码

列数值多为字符型，根据字符，去重编号。

如“2、您是否只有抖音一个直播 app”的编码，可将“跳过”、“是”、“否”分别编码为 0, 1, 2。

4.7. 一题多问

如 12 题。“12、以下选项对于您观看抖音直播时的影响程度——价值取向”中，又涉及了直播间氛围、直播内容、主播颜值、才艺等多列。该情况自带了因果关系，不用再考虑谁是自变量、谁是因变量。单独成块，可单独分析，无需额外预处理。

5. 数据分析

5.1. 分析：“您是否只有抖音一个直播 app”与哪些用户特征关联密切

“您是否只有抖音一个直播 app”为问卷中的问题 2，该问题可作为决策变量，反应用户对抖音 app

的用户忠诚度。为分析“您是否只有抖音一个直播 app”与哪些用户特征关联密切，通过问题列表，先人为选取了问题 1, 3, 4, 5, 6, 7, 8, 20, 21 作为属性。通过粗糙集理论分析这些属于与“您是否只有抖音一个直播 app”的关联性。所使用的原始数据如图 2 所示。

| U/A | x1 | x2 | x3 | x4 | x5 | x6 | x7 | x8 | x9 | y |
|-----|----|------|------|------|------|------|----------|----|----|------|
| u1 | 是 | 是 | 否 | 否 | 中立,会 | 偶尔 | 10-30 分钟 | 男 | 大四 | 否 |
| u2 | 是 | 是 | 是 | 是 | 中立,会 | 经常 | 30-60 分钟 | 女 | 大二 | 是 |
| u3 | 否 | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | 男 | 大二 | (跳过) |
| u4 | 是 | 否 | 否 | 否 | 中立,会 | 偶尔 | 30-60 分钟 | 女 | 大一 | 是 |
| u5 | 是 | 是 | 是 | 否 | 中立,会 | 偶尔 | 10 分钟内 | 男 | 大二 | 否 |
| u6 | 否 | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | 女 | 大一 | (跳过) |
| u7 | 是 | 否 | 否 | 否 | 中立,会 | 很少 | 10 分钟内 | 女 | 大二 | 是 |
| u8 | 否 | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | 女 | 大三 | (跳过) |
| u9 | 是 | 是 | 是 | 是 | 中立,会 | 偶尔 | 10-30 分钟 | 女 | 大二 | 否 |
| u10 | 是 | 否 | 是 | 是 | 我很喜欢 | 经常 | 60 分钟以外 | 男 | 大一 | 是 |
| u11 | 否 | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | 女 | 大三 | (跳过) |
| u12 | 是 | 是 | 否 | 是 | 中立,会 | 很少 | 10 分钟内 | 女 | 大二 | 否 |
| u13 | 是 | 是 | 否 | 是 | 中立,会 | 很少 | 10 分钟内 | 女 | 大一 | 否 |
| u14 | 否 | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | 女 | 大二 | (跳过) |
| u15 | 是 | 否 | 是 | 是 | 中立,会 | 偶尔 | 10-30 分钟 | 女 | 大四 | 否 |
| u16 | 是 | 否 | 否 | 是 | 中立,会 | 偶尔 | 10 分钟内 | 女 | 大四 | 是 |
| u17 | 是 | 是 | 否 | 是 | 中立,会 | 偶尔 | 10 分钟内 | 男 | 大一 | 否 |
| u18 | 是 | 是 | 是 | 否 | 中立,会 | 偶尔 | 10-30 分钟 | 女 | 大一 | 是 |
| u19 | 是 | 是 | 是 | 是 | 我很喜欢 | 经常 | 10-30 分钟 | 男 | 大二 | 是 |
| u20 | 是 | 是 | 是 | 是 | 中立,会 | 很少 | 10-30 分钟 | 女 | 大一 | 是 |
| u21 | 是 | 否 | 是 | 是 | 中立,会 | 偶尔 | 30-60 分钟 | 男 | 大一 | 是 |
| u22 | 是 | 是 | 是 | 否 | 中立,会 | 很少 | 10 分钟内 | 男 | 大一 | 否 |
| u23 | 否 | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | 男 | 大一 | (跳过) |
| u24 | 否 | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | 男 | 大一 | (跳过) |
| u25 | 是 | 是 | 否 | 是 | 中立,会 | 很少 | 10-30 分钟 | 女 | 大一 | 否 |
| u26 | 是 | 是 | 是 | 是 | 中立,会 | 偶尔 | 10-30 分钟 | 女 | 大三 | 是 |
| u27 | 是 | 否 | 否 | 是 | 中立,会 | 偶尔 | 30-60 分钟 | 女 | 大一 | 是 |

Figure 2. Raw data diagram (1)

图 2. 原始数据图(1)

首先，将所有属性列以及决策列“您是否只有抖音一个直播 app”均进行 0、1、2、……编码。代码如下：

```
for col_name in [...] # 选取属性列于决策列
    key = df_original[col_name].unique().tolist()
    m = dict(zip(key, range(len(key))))
    df_original[col_name] = df_original[col_name].map(lambda x: m[x])
```

然后采用粗糙集方法计算属性列的约简情况[7] [8]。

该约简过程依赖于不可分辨关系。分类过程中，相差不大的个体被归于同一类，它们的关系就是不可分辨关系。如假定只用两种黑白颜色把空间中的物体分割两类，{黑色物体}，{白色物体}，那么同为黑色的两个物体就是不可分辨的，因为描述它们特征属性的信息相同，都是黑色。

采用的粗糙集算法过程如下，初始化属性集合为空集，此时约简集合下的正域为空集[9]，每次选取使正域中对象增加最多的属性加入到约简集合中，直至正域中的对象不再增加，输出集合。再算法的计算过程中，每次仅对还未判定为正域的样本进行正域计算。

编码初始化后，再进行 0-1 归一化，此时数据如图 3 所示。

| | U/A | x1 | x2 | x3 | x4 | x5 | x6 | x7 | x8 | x9 | y |
|-----|------|-----|-----|-----|-----|-----|----------|------|-----|----------|-----|
| 0 | u1 | 1.0 | 1.0 | 0.5 | 0.5 | 0.4 | 0.368421 | 0.25 | 1.0 | 0.333333 | 0.5 |
| 1 | u2 | 1.0 | 1.0 | 1.0 | 1.0 | 0.4 | 1.000000 | 0.75 | 0.0 | 0.222222 | 1.0 |
| 2 | u3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.00 | 1.0 | 0.222222 | 0.0 |
| 3 | u4 | 1.0 | 0.5 | 0.5 | 0.5 | 0.4 | 0.368421 | 0.75 | 0.0 | 0.000000 | 1.0 |
| 4 | u5 | 1.0 | 1.0 | 1.0 | 0.5 | 0.4 | 0.368421 | 0.50 | 1.0 | 0.222222 | 0.5 |
| .. | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 200 | u201 | 1.0 | 1.0 | 1.0 | 0.5 | 0.4 | 0.368421 | 0.50 | 0.0 | 0.222222 | 0.5 |
| 201 | u202 | 1.0 | 1.0 | 0.5 | 0.5 | 0.4 | 0.368421 | 0.25 | 0.0 | 0.222222 | 0.5 |
| 202 | u203 | 1.0 | 1.0 | 1.0 | 0.5 | 0.4 | 0.736842 | 0.25 | 0.0 | 0.111111 | 1.0 |
| 203 | u204 | 1.0 | 1.0 | 0.5 | 0.5 | 0.4 | 0.368421 | 0.50 | 0.0 | 0.222222 | 0.5 |
| 204 | u205 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 0.368421 | 0.75 | 1.0 | 0.222222 | 0.5 |

Figure 3. 0-1 Normalization plot (1)

图 3. 0-1 归一化图(1)

按文献的计算结果如图 4 所示。约简后的剩余属性为['x1']，不需要的属性为['x2' 'x3' 'x4' 'x5' 'x6' 'x7' 'x8' 'x9']。

| 决策属性下的分类: | |
|---|--|
| {0.0: | [2, 5, 7, 10, 13, 22, 23, 32, 65, 105, 106, 123, 136, 141, 148, 152, 196, 197], 0.5: |
| [0, 4, 8, 11, 12, 14, 16, 21, 24, 28, 29, 30, 36, 39, 41, 44, 46, 49, 50, 51, 55, 57, 59, | 60, 63, 66, 68, 69, 78, 79, 80, 82, 84, 87, 90, 92, 93, 94, 96, 108, 109, 113, 114, 116, |
| 117, 120, 121, 124, 127, 128, 131, 132, 133, 134, 135, 137, 138, 139, 140, 143, 146, 147, | |
| 约减后的属性 | |
| ['x1'] | |
| 不需要的属性: | |
| ['x2' 'x3' 'x4' 'x5' 'x6' 'x7' 'x8' 'x9'] | |

Figure 4. Partracted residual attribute diagram (1)

图 4. 简约剩余属性图(1)

将编码结果返回至原文，该结果表面，y 列“您是否只有抖音一个直播 app”在所选的属性中与 x1 “1、您是否驻足过抖音直播间？”强相关，其他属性是弱相关或不相关，因而被粗糙集算法约简。这一结论与直觉相符。比如“您是否只有抖音一个直播 app”确实不应与“4、您在抖音是否有关注并常去的直播间？”有强相关性。

此外由于选择了 x8 “性别”，x9 “年级”作为属性，这两列也被约简掉，说明在调查问卷范围中，“您是否只有抖音一个直播 app”没有表现出与性别或大一还是大四有强关联性。

为检查分析 x1 与 y 的强关联性，通过源数据统计了调查问卷中该两项问题的频次数据。在问题“2、您是否只有抖音一个直播 app”中，选项有“是”、“否”。在问题“1、您是否驻足过抖音直播间？”中，有选项“是”、“否”、“跳过”。统计数据表明，选择了“驻足过抖音”的用户中 97 人只有抖音一个 app，90 人不单只有抖音，且无人选择跳过问题 2。而在问题 1 中选了“未驻足过抖音”的用户共 18 人，且均跳过了问题“2、您是否只有抖音一个直播 app”。说明源数据上确实两者有强相关性，算法结果与源数据结果以及直观分析结果相一致。如表 1 所示。

Table 1. The strong association of x1 and y
表 1. x1 与 y 的强关联性

| | 驻足过抖音 | 未驻足过抖音 |
|--------------|-------|--------|
| 只有抖音一个 app | 97 | 0 |
| 跳过 | 0 | 18 |
| 不单只有抖音一个 app | 90 | 0 |

5.2. 分析：对于观看抖音直播时的影响程度

基于问卷 12 题。以“12、以下选项对于您观看抖音直播时的影响程度 - 价值取向”主列信息作为决策列 y。“直播间氛围”、“直播内容”、“主播颜值、才艺等”、“主播素质(话术、态度等)”、“直播是否有内容、有价值”、“弹幕交流是否友好”、“直播间热度(在线人数)”、“直播间福利”、“直播预告(或直播时间是否稳定)”为属性列 x。共 9 列属性，1 列决策。源数据整理表格如图 5 所示：

| U/A | x1 | x2 | x3 | x4 | x5 | x6 | x7 | x8 | x9 | y |
|-----|-------|------|------|-------|-------|-------|-------|-------|-------|------|
| u1 | 有影响 | 影响很大 | 影响很大 | 一般 | 一般 | 一般 | 有一点影响 | 有一点影响 | 无影响 | 有影响 |
| u2 | 有影响 | 有影响 | 有影响 | 有影响 | 有影响 | 一般 | 有影响 | 有影响 | 有影响 | 一般 |
| u3 | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) |
| u4 | 有影响 | 有影响 | 有影响 | 有影响 | 有影响 | 有影响 | 有影响 | 有影响 | 影响很大 | 有影响 |
| u5 | 一般 | 一般 | 有影响 | 有影响 | 有影响 | 一般 | 有影响 | 有一点影响 | 有一点影响 | 一般 |
| u6 | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) |
| u7 | 一般 | 有影响 | 一般 | 有影响 | 一般 | 有影响 | 有影响 | 有影响 | 有影响 | 一般 |
| u8 | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) |
| u9 | 有影响 | 有影响 | 有影响 | 有影响 | 有影响 | 一般 | 有影响 | 有影响 | 有影响 | 有影响 |
| u10 | 影响很大 | 影响很大 | 影响很大 | 影响很大 | 影响很大 | 影响很大 | 影响很大 | 影响很大 | 影响很大 | 影响很大 |
| u11 | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) |
| u12 | 有影响 | 有影响 | 有影响 | 有影响 | 有影响 | 有影响 | 有影响 | 有影响 | 有影响 | 有影响 |
| u13 | 有影响 | 影响很大 | 无影响 | 影响很大 | 有影响 | 有影响 | 无影响 | 有一点影响 | 无影响 | 无影响 |
| u14 | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) |
| u15 | 无影响 | 无影响 | 无影响 | 有一点影响 | 有一点影响 | 有一点影响 | 无影响 | 无影响 | 无影响 | 无影响 |
| u16 | 一般 | 一般 | 一般 | 一般 | 一般 | 一般 | 一般 | 一般 | 一般 | 无影响 |
| u17 | 有一点影响 | 有影响 | 无影响 | 影响很大 | 影响很大 | 一般 | 无影响 | 无影响 | 无影响 | 影响很大 |
| u18 | 一般 | 一般 | 一般 | 一般 | 一般 | 一般 | 一般 | 一般 | 一般 | 一般 |
| u19 | 有影响 | 有影响 | 有影响 | 有影响 | 有影响 | 有影响 | 有影响 | 有影响 | 有影响 | 有影响 |
| u20 | 有影响 | 有影响 | 有影响 | 影响很大 | 有影响 | 影响很大 | 无影响 | 一般 | 有影响 | 无影响 |
| u21 | 有一点影响 | 有影响 | 有影响 | 有影响 | 有影响 | 有影响 | 无影响 | 一般 | 一般 | 无影响 |
| u22 | 无影响 | 无影响 | 无影响 | 无影响 | 无影响 | 无影响 | 无影响 | 无影响 | 无影响 | 无影响 |
| u23 | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) | (跳过) |

Figure 5. Raw data diagram (2)
图 5. 原始数据图(2)

编码初始化后，再进行 0-1 归一化，此时数据如图 6 所示。

| | U/A | x1 | x2 | x3 | x4 | x5 | x6 | x7 | x8 | x9 | y |
|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 0 | u1 | 1.0 | 0.4 | 0.4 | 0.2 | 0.2 | 0.2 | 0.8 | 0.8 | 0.6 | 1.0 |
| 1 | u2 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 0.2 | 1.0 | 1.0 | 1.0 | 0.2 |
| 2 | u3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 3 | u4 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 0.4 | 1.0 |
| 4 | u5 | 0.2 | 0.2 | 1.0 | 1.0 | 1.0 | 0.2 | 1.0 | 0.8 | 0.8 | 0.2 |
| .. | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 200 | u201 | 0.2 | 1.0 | 0.2 | 0.4 | 0.2 | 0.2 | 0.4 | 1.0 | 0.2 | 1.0 |
| 201 | u202 | 1.0 | 0.4 | 0.2 | 1.0 | 1.0 | 0.2 | 1.0 | 0.4 | 1.0 | 1.0 |
| 202 | u203 | 0.2 | 1.0 | 0.4 | 1.0 | 1.0 | 0.2 | 1.0 | 0.2 | 0.2 | 1.0 |
| 203 | u204 | 0.2 | 1.0 | 1.0 | 0.4 | 1.0 | 0.2 | 1.0 | 1.0 | 0.2 | 0.2 |
| 204 | u205 | 0.4 | 1.0 | 0.2 | 0.4 | 1.0 | 0.2 | 1.0 | 1.0 | 0.2 | 0.4 |

Figure 6. 0-1 normalization plot (2)

图 6. 0-1 归一化图(2)

约简后剩余属性列为 x1 “直播间氛围”、x2 “直播内容”、x5 “直播是否有内容、有价值”、x6 “弹幕交流是否友好”。如图 7 所示。

| 数据对象个数: 205 | | 数据属性个数: 10 | | | | | | | | | |
|---|----|------------|-----|-----|-----|-----|----------|------|-----|----------|-----|
| 数值化与归一化处理后的决策表: | | | | | | | | | | | |
| U/A | x1 | x2 | x3 | x4 | x5 | x6 | x7 | x8 | x9 | y | |
| 0 | u1 | 1.0 | 1.0 | 0.5 | 0.5 | 0.4 | 0.368421 | 0.25 | 1.0 | 0.333333 | 0.5 |
| 1 | u2 | 1.0 | 1.0 | 1.0 | 1.0 | 0.4 | 1.000000 | 0.75 | 0.0 | 0.222222 | 1.0 |
| 2 | u3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.000000 | 0.00 | 1.0 | 0.222222 | 0.0 |
| 决策属性下的分类: | | | | | | | | | | | |
| [0, 0]: [2, 5, 7, 10, 13, 22, 23, 32, 65, 105, 106, 123, 136, 141, 148, 152, 196, 197], 0.5: [0, 4, 8, 11, 12, 14, 16, 21, 24, 28, 29, 30, 36, 39, 41, 44, 46, 49, 50, 51, 55, 57, 59, 60, 63, 66, 68, 69, 78, 79, 80, 82, 84, 87, 90, 92, 93, 94, 96, 108, 109, 113, 114, 116, 117, 120, 121, 124, 127, 128, 131, 132, 133, 134, 135, 137, 138, 139, 140, 143, 146, 147, | | | | | | | | | | | |
| 约减后的属性 | | | | | | | | | | | |
| ['x1' 'x2' 'x5' 'x6'] | | | | | | | | | | | |
| 不需要的属性: | | | | | | | | | | | |
| ['x3' 'x4' 'x7' 'x8' 'x9'] | | | | | | | | | | | |

Figure 7. Partracted residual attribute diagram (2)

图 7. 简约剩余属性图(2)

该结果体现了 x1 x2 x5 x6 列与 y 的强关联性，即表明对于用户观看抖音直播时的影响程度，在价值取向方面，重要影响因素包含直播间氛围、直播内容、直播是否有内容有价值、弹幕交流是否友好。说明抖音从提升对于用户观看抖音直播时价值取向的影响程度方面应重视这些因素，不断提升平台价值。

6. 结论与建议

通过粗糙集的简约算法，可以帮助我们从庞大、复杂的数据中提取出有用的信息，更加清楚、简单地分析出影响因变量的主要因素，做出准确且有效的判断。从而建立简约概念，提高算法搜索效率。

6.1. 增加抖音 app 使用率

由 5.1 的数据分析可知，填写问卷者是否驻足过抖音 app 与是否只有抖音一个直播 app 有很强的关联性，而与“是否会因为某人/事关注某个直播”、“在抖音是否有关注并常去的直播间”、“是否参与过直播间的弹幕互动”、“现阶段对于抖音直播的态度”、“看直播的频次”、“每次使用抖音观看直播大概时间”、“性别”和“年级”均无显著相关性。这说明如果大学生曾经停留过或使用过抖音的直播平台，那么只拥有抖音一个直播 app 的概率会显著增大，可以进一步反映出大学生对抖音 app 的忠诚度。

因此抖音平台为巩固自身的地位，赢得更多的用户忠诚，可以关注于打造抖音直播平台的知名度，

让更多的人知晓抖音 app 并下载进行体验。尽量做到让用户在第一次使用、驻足抖音直播平台时, 就对其产生良好印象, 获得较好的感官体验, 增加用户对抖音的信任, 并逐渐成为用户手机里唯一的直播 app。

6.2. 提升抖音直播间内在价值

由 5.2 可以得出, 价值取向因素是影响大学生观看抖音直播的主要因素, 包括: “直播间氛围”、“直播内容”、“直播内容或价值”和“弹幕交流友好”, 具有强相关性。而“主播颜值、才艺”、“主播素质”、“直播间热度”、“直播间福利”和“直播预告”与影响用户观看抖音直播无显著影响, 不存在关联性。

进一步说明了抖音博主应该将重心放在提升直播间的内在价值方面, 关注可以对社会产生正向影响的事件和新闻, 从而引导观众和用户树立正确的价值观和世界观, 发挥个体的价值和力量。

参考文献

- [1] 中国互联网络信息中心. CNIC 发布第 49 次《中国互联网络发展状况统计报告》[J]. 新闻潮, 2022(2): 3.
- [2] 于建波, 马金宝, 朱宏. 网络直播对大学生的影响及应对策略[J]. 大学教育, 2017(11): 148-150.
- [3] 黄京, 杨雪莹, 李晓明, 等. 网络直播对大学生影响的研究——以桂林普通高校为例[J]. 高教学刊, 2018, 3(75): 197-199.
- [4] 刘姗姗. 大学生进行网络直播现状及问题[J]. 现代商贸工业, 2017(2): 144-146.
- [5] 吴婉琳, 张贤勇, 莫智文. 基于粗糙集不确定度的特定类属性约简[J]. 四川师范大学, 2021, 44(6): 88-91.
- [6] 危前进, 古天龙, 常亮, 等. 粗糙集多目标并行属性约简算法[J]. 软件学报, 2022, 33(7): 101-102.
- [7] 阿里云开发者社区. python 粗糙集约简算法 + 可视化界面[EB/OL]. <https://developer.aliyun.com/article/1297886>, 2023-08-07.
- [8] We2sWing.CSDN. 粗糙集相对属性约简 python 代码实现[EB/OL]. <https://www.jianshu.com/p/ef1d9321ba42>, 2018-12-20.
- [9] 周涛, 陆惠玲, 任海玲, 等. 基于粗糙集的属性约简算法综述[J]. 电子学报, 2021, 49(7): 52-54.