Evaluation of influencing factors of TV series on network based on multivariate statistical analysis

Yufei Hou1*, Jing Hu1, YangYang1, Lulu Dong1

1Department of Statistics, Communication University of China,100024, China  
*Corresponding author’s e-mail: 1156193628@qq.com

Abstract. This paper aims to make a systematic study on the influencing factors of network broadcast quantity of TV drama by means of multivariate statistical method. Based on the multiple primary data indexes of TV series, this paper uses factor analysis to redefine three influencing factors-publicity degree, participation degree and topic degree, and builds the algorithm of each factor model. Then, through the bivariate correlation analysis, it can be concluded that the participation degree is most related to the network broadcast quantity of TV series. After evaluating the limitations of the research process reasonably, the revelation and conclusion of promoting the development of TV dramas are drawn.

1. Introduction

The TV series industry is an important part of the media industry. With the opening of the multi-screen era, the Internet has become an important communication channel for all kinds of TV series. Network broadcast quantity has also become an important index to evaluate the influence of TV series. For the influencing factors of the network broadcast quantity of the TV drama, the domestic and foreign scholar also made various researches. For example, Liu Xuan researched the TV series itself, the playing platform, and the influence of the audience rating; Hu Bing et al. analyzed the positive correlation between the dissemination of content on weibo and audience rating; Zhou Yong made a research on the influence process of multi-stage communication on audience rating under the media environment. [1-3]

In addition, many scholars have also applied the multivariate statistical analysis method to network communication and heat research. For example, Sun Jianghua and Zhang Shu used principal component analysis (pca) and cluster analysis to establish an evaluation method for the influence of microblog. Zhang Jingtian used pca to evaluate the popularity of TV dramas. Du Shiyu and Qi Jiayin extracted the influence index of hot topics on microblog through dimensionality reduction. [4-6] Combined with the previous research results, this paper will use the multivariate statistical analysis method to make a quantitative analysis of the correlation between the network broadcast quantity and its influencing factors.

Firstly, based on the data of several TV series collected on various platforms of the network, the common factors in the original data indicators were extracted by factor analysis, which is then redefined to obtain the influencing factors studied in this paper. Secondly, the correlation analysis method is used to quantitatively discuss the correlation between each influencing factor and the network broadcast quantity of TV drama.

The innovation of this paper is reflected in two aspects: first, multiple primary data indicators are used for factor analysis to reduce dimensions, and the interaction of each primary indicator is removed, and several independent influencing factors are redefined. The second is to quantitatively discuss the
correlation between the influencing factors and the network broadcast quantity of TV series, which provides reference and guidance for expanding the spread of TV series on the Internet and improving the influence of TV series.

2. Extraction of influencing factors of TV play on network

2.1. Data collection and preprocessing
This paper selects 10 TV series broadcasted from June to August of 2018 as the research object, which include different types of subjects, play modes (only played at the network end or synchronized in the network end and TV end), and extracted the various kinds of data within 10 days from the beginning of the playing period. The data taken in the study were from all kinds of portal websites and the results given by the search engines.

2.2. Assumptions before analysis
- Ignoring the uncontrollable factors that cause data instability in the Internet, such as network water army.
- The data provided by various search engines are authentic and reliable.

2.3. Use factor analysis to deal with each index
After standardization of the original data, factor analysis is used to reduce the dimensions of the original indexes.

First, KMO and Bartlett sphericity tests are performed for each primary indicator, and the significance level of the data in this group is less than 0.05, and the KMO value is 0.772, which is suitable for factor analysis.

| Table 1. KMO and Bartlett's Test |
|----------------------------------|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy: | .772 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 1438.855 |
| df | 66 |
| Sig. | .000 |

In factor analysis, three components can be extracted according to feature value greater than 1, and the rotated composition matrix can be obtained as shown in the table below.

| Table 2. Rotated Component Matrixa |
|-----------------------------------|
| Component | 1 | 2 | 3 |
| Number of fans on weibo | .510 | .748 | .411 |
| Number of weibo posts | .200 | .833 | .452 |
| Douban "looking" number | .900 | .341 | .257 |
| Douban "wants to see" number | .663 | .720 | -.015 |
| Douban "looked" number | .944 | .307 | .114 |
| Total discussion number of douban (bar) | .736 | .498 | .424 |
| Short comments on douban | .940 | .308 | .129 |
| Reviews on douban | .934 | .331 | .119 |
| Number of related videos | -.338 | -.890 | -.243 |
| Baidu index | .568 | .684 | -.094 |
| Related news and information | -.039 | .162 | .968 |
| Discussions of topics on video website | .500 | .211 | .828 |
Based on the composition matrix, it can be concluded that: douban "looked" number, short comments on douban, reviews on douban and total discussion number of douban (bar) are all strongly related to component 1, with the correlation coefficient reaching above 0.7 and the first four data indicators reaching above 0.9. Therefore, the component 1 is defined as a new variable named as the participation index. Number of weibo posts, number of fans on weibo, douban "wants to see" number and number of related videos are strongly correlated with component 2, and the correlation coefficient is also above 0.7. Therefore, composition 2 is defined as a new variable named as the publicity index. Related news and information, discussions of topics on video website are associated with component 3 with the correlation coefficient of 0.968 and 0.828, respectively, and the correlation coefficient between other data index and composition 3 are less than 0.5, so related news and information, discussions of topics on video website are selected to describe component 3 defined as a new variable, named as the subject index.

3. Establishment of influence factor model

The algorithm using the component score coefficient matrix to give the new variables is shown in table 3 below:

| Table 3. Component Score Coefficient Matrix |
|---------------------------------------------|
| Component | 1 | 2 | 3 |
| Number of fans on weibo | -.072 | .236 | .066 |
| Number of weibo posts | -.234 | .396 | .085 |
| Douban "looking" number | .244 | -.146 | .043 |
| Douban "wants to see" number | .021 | .250 | -.187 |
| Douban "looked" number | .281 | -.155 | -.037 |
| Total discussion number of douban (bar) | .113 | -.017 | .116 |
| Short comments on douban | .279 | -.157 | -.028 |
| Reviews on douban | .268 | -.137 | -.038 |
| Number of related videos | .188 | -.428 | .056 |
| Baidu index | -.003 | .277 | -.223 |
| Related news and information | -.115 | -.080 | .542 |
| Discussions of topics on video website | .099 | -.210 | .425 |

Let "number of fans on weibo” “number of weibo posts” “douban ‘looking’ number” “douban "wants to see" number” “douban ‘looked’ number" “total discussion number of douban (bar)” “short comments on douban” “reviews on douban” “number of related videos” “baidu index” “related news and information” “discussions of topics on video website” be $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}$, $x_{11}, x_{12}$, and “participation index” “publicity index” “subject index” be $y_1, y_2$, and $y_3$, respectively, then

$y_1=-0.072x_1-0.234x_2+0.244x_3+0.021x_4+0.281x_5+0.113x_6+0.279x_7+0.268x_8+0.188x_9-0.003x_{10}-0.115x_{11}+0.099x_{12}$

$y_2=0.236x_1+0.396x_2-0.146x_3+0.250x_4-0.155x_5-0.017x_6-0.157x_7-0.137x_8-0.428x_9+0.277x_{10}-0.080x_{11}-0.210x_{12}$

$y_3=0.066x_1+0.085x_2+0.043x_3-0.187x_4-0.037x_5+0.116x_6-0.028x_7-0.038x_8+0.056x_9-0.223x_{10}+0.542x_{11}+0.425x_{12}$
4. Correlation analysis of influencing factors and TV drama broadcast quantity on network

The three influencing factors, including participation index, publicity index and subject index, are respectively analyzed with the network broadcast quantity of TV series after standardization. As the three influencing factors after dimensionality reduction are not correlated with each other, the partial correlation analysis method is not appropriate here. Therefore, a bivariate simple correlation analysis is selected to determine the correlation degree between them and the network broadcast quantity of TV series, as shown in the table below.

| Table 4. participation index |
|-----------------------------|
| Network broadcast of TV series | Participation index |
| Network broadcast of TV series | Pearson Correlation | .827** |
| Sig. (2-tailed) | .000 |
| N | 40 | 40 |
| Participation index | Pearson Correlation | .827** |
| Sig. (2-tailed) | .000 |
| N | 40 | 40 |

| Table 5. publicity index |
|--------------------------|
| Network broadcast of TV series | Publicity index |
| Network broadcast of TV series | Pearson Correlation | .486** |
| Sig. (2-tailed) | .001 |
| N | 40 | 40 |
| Publicity index | Pearson Correlation | .486** |
| Sig. (2-tailed) | .001 |
| N | 40 | 40 |

| Table 6. subject index |
|------------------------|
| Network broadcast of TV series | Subject index |
| Network broadcast of TV series | Pearson Correlation | .149 |
| Sig. (2-tailed) | .359 |
| N | 40 | 40 |
| Subject index | Pearson Correlation | .149 |
| Sig. (2-tailed) | .359 |
| N | 40 | 40 |

According to the above table, the participation index is most related to the network broadcast quantity of TV series, followed by the publicity index with a low correlation degree, while the subject index fails the significance test.
Meanwhile, the correlation between the two variables can be quantitatively analyzed from the Pearson coefficient in the table:

The Pearson coefficient between the participation index and the broadcast amount of a single episode reached 0.827, which was a strong correlation. The Pearson coefficient between the publicity index and the broadcast amount of a single episode reached 0.486, which was a moderate correlation.

5. Conclusion
This paper quantitatively analyzes the three factors, participation degree, publicity degree and subject degree, affecting the network broadcast quantity of TV series by using the multivariate statistical method. The results show that there is a strong correlation between participation degree and network broadcast quantity, followed by publicity degree and the last is subject degree.

This conclusion can not only provide references for TV drama producers to improve network rating and audience influence, but also provide theoretical methods for the construction of TV drama influence evaluation system. In the allocation of index weight, multivariate statistical methods can be used, such as using factor analysis to evaluate the credibility and effectiveness of the index system construction, using sample data to measure the correlation coefficient between each index and to empower each path in the complex network of TV drama transmission.

References
[1] Liu Xuan. A study on the factors influencing the audience rating of the talent show mamma mia [D]. Shanghai normal university,2014.
[2] Hu Bing, Deng Ji. Research on the influence of microblog on TV rating [J]. Present media,2015,23(06):32-34.
[3] Zhou Yong, Chen Huiru. The construction of an evaluation system for the influence of online audio-visual information under the multi-level communication path [J]. Modern communication (journal of communication university of China),2013,35(03):123-128.
[4] Sun Jianghua, Zhang Shu. Research on the influence of traditional newspaper microblog based on principal component analysis and cluster analysis [J]. Modern communication (journal of communication university of China),2015,37(04):141-143.
[5] Zhang Jingtian. Research on TV drama popularity based on principal component analysis [J]. Science and technology economics guide,2018(02):158-159.
[6] Du Shiyu, Qi Jiayin. Research on evaluation of weibo topic influence index based on principal component analysis [J]. Journal of intelligence,2014,33(05):129-135.