

Multiple Linear Regression Analysis of the Animation Score

Xianrui Fu *

Department of Science, Beijing Forest University, Beijing, 100000, China

* Corresponding Author Email: fuxianrui@bjlydx.wecom.work

Abstract. With the development of China's economy, people's living standards continue improving. This makes the cultural industry flourished. As an important part of the cultural, the animation industry is entering a golden age of development. The rapid development of the animation industry has led to a large number of national comics of varying quality. Based on the cultural atmosphere of domestic animation is not strong, the originality of animation production is not strong, the animation industry chain needs to be improved and other problems. This paper analyzed the significant factors affecting animation ratings through multiple linear regression models to make some suggestions for the future development of domestic animation. The summary analysis shows that name, storyboard and protagonist have a significant positive influence on the score. From the results, China's animation industry should pay attention to plot development and character building. It is hoped that the findings of this paper can have some guiding significance for the development of national comics.

Keywords: Multiple linear regression models; animation ratings; China's animation industry development.

1. Introduction

With the development of technology, the entertainment industry for people's daily needs is gradually growing. We are in an era when the entertainment industry is taking off. Since the mid to late twentieth century, the animation industry has increasingly become a global industry that promotes cultural exchange between countries due to its cultural and entertainment attributes. In recent years, China's animation industry is developing rapidly, "Nezha: Birth of the Demon Child", "Monkey King: Hero is Back" and "The Legend of Luoxiaohai" and other excellent animation films continue to emerge, Bilibili, Tencent animation, Kuaikan cartoon and other new media animation platform is also gradually emerging. Therefore, relevant manga data is collected to analyze the relationship between labels such as adaptation source and animation rating to find the development direction of Chinese animation industry [1].

China's animation has been born for more than 100 years, but most of the works are made for children, the proportion of adult-oriented animation is small, but the returns are higher [2]. From previous studies, it is easy to see that the domestic animation industry chain is difficult to establish and maintain, and the brand image is difficult to establish and maintain [3]. The best way to solve this problem is to update the concept and strengthen the creativity, so that the animation meets the aesthetics of the Chinese people [4]. There is no denying the popularity of manga in Asia so if we want to develop domestic animation, we need to learn from the successful experience of highly rated manga [5]. On the one hand, the rating of an animation work is an objective criterion to evaluate whether an animation is attractive or not. On the other hand, there is a significant positive correlation between online ratings and the number of viewers [6]. The extensive use of the Internet has penetrated into every aspect of life. "Animation ranking" has been widely accepted by the public as an expression of animation quality content evaluation. There are six common animation rating sites. Bangumi is a professional animation rating website with more objective ratings. From this website, it is possible to obtain attributes such as soundtrack, director, screenwriter, and tags of the animation, where tags contain richer information like the source of the adaptation, style, etc. Therefore choose some high scoring anime to analyze which factors can make an anime score higher, which means more people like the anime.

According to the existing literature, no one has done the research on animation score analysis. However, there are many researches on movie score analysis similar to anime score analysis. Karl

Person collected the attribute and feature information of 3376 Hollywood movies from IMDB and constructed the Random forest model and support vector machine model to predict movie scores [7]. It can be seen that there is still room for improvement in the research of animation score prediction model.

Nowadays, with the rapid development of vertical websites like audio and video, books, etc. The rating of products or services by users who have already purchased or watched the user rating evaluation of products or services, has gradually developed into an important information carrier, According to the data on video websites, the number of users who express their opinions and ideas through social media platforms is growing exponentially [8].

More and more users will pay attention to the rating of the animation and other users' evaluation of the animation as a strong basis for whether to watch the animation or not. This is a quick and effective way to help users provide other users' opinions in the process of deciding whether to watch an animation or not, by providing visual access to the animation ratings. This is a fast and effective way to help users provide other users' opinions in the decision process of whether to watch a movie. The point of view of this paper is to take the top 200 animations on Bangumi as a data set. Nine factors such as director, soundtrack, genre, projection method, adaptation source, voice actor, studio and age of these 200 animations were collected by manual means and converted into corresponding indicators. Try to find the animation through multiple regression analysis the linear relationship between the score and the director, the score, the genre, the release mode, the adaptation source, the voice actor, the studio, the decade and so on [9].

2. Materials and Methods

2.1. Data Source and Preparing

The data comes from Bangumi (<https://bgm.tv/anime/browser/?sort=rank>), a professional animation rating website, as the score rankings are refreshed daily at 0:00, this article represents the rankings until January 17, 2023. Based on the scoring criteria, the top 200 Japanese manga were selected for the study, and tags with the same meaning were grouped together. Then the tags with less than 5 occurrences were ignored. Finally, multiple linear regression was used to analyze the data. The per-trained language model proposed by Devlin is referenced in this paper [10].

2.2. Data Assignment of “Type”

In this paper, we adopt the genre classification method within the bangumi rating system, and use cluster analysis to greatly simplify the original two dozen anime genres to arrive at a basic genre distribution (see Table 1). The assignment is then made according to the distribution.

Table 1. Bangumi top100 Animation type pivot table

Healing	Hot Blood	Daily	Spit	Youth	SF	Life	Psychology
28	19	16	8	22	30	13	5
Fantasy	Cute	Music	Love	Combat	tear	Connotation	Suspense
22	5	11	24	30	6	5	6

This avoids the disadvantage of neglecting the meaning behind the genre in the simple word frequency statistics, and also avoids the excessive gap between the equal weight assignment and the real situation. The genre of the animation as a kind of artificial classification is also taken into account. The word frequency of genre as a "word" is also taken into account.

2.3. “Director”, “Screenwriter”, “Voice Actor” Assignment

The data in this paper is based on Bangumi animation data mining, and the purpose of this study is to establish a multivariate linear regression relationship between animation ratings and various elements. Therefore, the rating system for "director," "screenwriter," and "voice actor" chose the

dimension of the average score of Bangumi's best work, without taking into account the commercial value potential, artistic influence, social responsibility, public image, etc.

2.4. Linear Regression Equation and Test

Firstly, the variable name and value range was determined as it in Table 2.

Table 2. Variable comparison table

variable name	Symbol	Value range (i = 1,2 ... 13)
script	X_1	$0.00 \leq X_i \leq 10.00$
character	X_2	
storyboard	X_3	
original works	X_4	
name	X_5	
era	X_6	
studio	X_7	
voice actor	X_8	
Adaptation source	X_9	
Screening method	X_{10}	
type	X_{11}	
music	X_{12}	
director	X_{13}	

Since different animation works have different numbers of participants, resulting in different magnitudes, the data of each type is normalized by equation 1.

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

After assigning and standardizing the indicators of, the above variables meet the basic conditions of multiple linear regression, and the linear relationship is logically valid.

Table 3. Linear regression analysis (n=198)

	B	SE	T Value	P Value	VIF
Constants	8.223	0.029	282.619	0.000**	-
X_1	-0.227	0.189	-1.203	0.230	2.304
X_2	0.182	0.150	1.219	0.224	1.281
X_3	0.230	0.144	1.593	0.113	1.847
X_4	-0.288	0.127	-2.269	0.024*	1.548
X_5	0.450	0.137	3.277	0.001**	1.699
X_6	-0.187	0.133	-1.407	0.161	3.111
X_7	-0.137	0.134	-1.022	0.308	2.178
X_8	-0.509	0.170	-2.999	0.003**	1.402
X_9	-0.039	0.148	-0.261	0.795	2.147
X_{10}	0.269	0.154	1.755	0.081	1.903
X_{11}	0.265	0.136	1.953	0.052	2.301
X_{12}	0.531	0.147	3.608	0.000**	1.496
X_{13}	0.211	0.146	1.444	0.150	2.443
R^2			0.255		
Adjustment R^2			0.203		

From the Table 3, script, character, storyboard, original works, name, era, studio, voice actor, Adaptation source, Screening method, type, music, director were taken as independent variables, while ratings were taken as dependent variables for linear regression analysis, and from the above table, it can be seen that the model equation is:

$$Y = 0.873 - 0.227X_1 + 0.182X_2 + 0.230X_3 - 0.288X_4 + \dots + 0.211X_{13} \quad (2)$$

The R-square value of model is 0.255, which means these factors can explain 25.5% of the variation of the score. The summary analysis shows that name, storyboard and protagonist have a significant positive influence on the score. And the voice actor will have a significant negative impact on the score. However, director, genre, screening method, adaptation source, studio, period, original work, script, design, and artist have no influence on the rating. The result shows that if you want to make a good animation, you should focus on improving the quality of the script and shaping plump characters.

From the data, we selected 42 animations with a rating of 8.5 or higher than 8.5 to make sure what are the main factors that determine a high rating for an animation. Then we got Figure 1 .

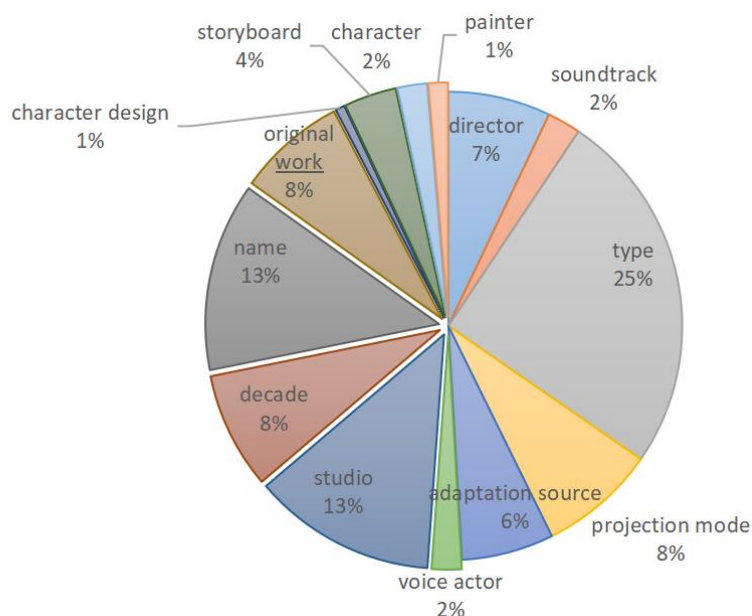


Fig 1. the Proportion of each Score

3. Results and Discussion

3.1. From A Style Point of View

There are ten popular styles that have a significant impact on animation ratings. Look at Table 4.

Table 4. Results of regression analysis of animation type and rating

	Coefficient	Statistical significance
Healing	0.26	***
Passionate	0.25	***
Daily	0.23	***
Roast	0.30	***
Youth	0.22	***
Science fiction	0.13	***
Sprout	0.15	***
Harem	-0.09	**
Music	0.16	**
Maiden	-0.13	*

As the results, the combination of healing, daily and sprout is very strong. This style group can improve the animation rating by 0.64 points, which is a proper code for top score. This is because this type of animation is smoother and less prone to plot thunderstorms, and viewers are willing to give this type of animation a higher score. On the contrary, harem and maiden genres received lower ratings due to their inconsistent quality and not being suitable for everyone. Other popular genres such as fantasy, battle, campus and romance were judged to have no significant impact on the ratings.

3.2. From a Screenwriter Point of View

There are five popular screenwriters that have a significant impact on animation ratings. Look at Table 5.

Table 5. Results of regression analysis of animation screenwriter and rating

	Coefficient	Statistical significance
Ma Z.Z	-0.68	***
Gu N.T.J.Z	-0.63	***
Tian Z.R	0.68	**
Chong F.D	0.74	**
Xiao T.D.Y.J	-0.68	**

The plot determines the upper limit of the series to a certain extent. It would be beneficial to have a strong scriptwriter on the production staff of the anime.

4. Conclusion

Through this analysis of the animation ratings, we can find to three laws. First of all, the determination of the type of animation overwhelmingly determines whether an animation can gain a larger audience. In today's fast-paced society, people tend to choose healing, daily, gossip, youth animation. When the type of animation is well chosen, pay more attention to the direction of the plot and do not appear minefield. Neither the source of the adaptation nor the scriptwriter has much influence on an animation rating. Therefore, animators should focus more on choosing a good genre and writing a good script. The accuracy of this model is not high because the data set is too small, and in future work more data will be added to improve the accuracy of the model.

References

- [1] Chengsen Chi, Yuhua Zhang, Anzhi Wang. China's animation industry development prospects Research. Technology and Economy Market2020, 8: 145-147.
- [2] Sifan Zhang. The current situation and suggestions for the development of China's animation industry. Shanghai Business, 2022, 17-19.
- [3] Zixuan Yang. Analysis of the development and current situation of Chinese animation. World of News, 2015, 76-80.
- [4] Ning Kuang. Experimental discussion on the development of China's animation industry status and countermeasures. Cradle of Journalists, 2021, 31-32.
- [5] Ling Li. Japanese manga in power-the development of Chinese animation should learn from the experience. Young Literary Artists, 2016, 155.
- [6] Gang Lei, Lu Liu, Changhui Zhou. An Empirical Study on the Effect of Online Movie Ratings on Box Office. Film Art, 2019, 146-153.
- [7] Juan Geng, Mingxing Guo. Douban Top 250 movie data mining and rating prediction. Hebei Enterprises, 2021, 2: 11-13.
- [8] Xiangjun Li, Xiaoling Xiao. A study on movie rating prediction based on machine learning. Computer Knowledge and Technology: Academic Edition, 2021.

- [9] Rubiao Zhou, Xiaoxia Lin, Yihua Wang. Multiple regression-based analysis of Douban movie ratings. *Art Technology*, 2019, 32(01): 67-68.
- [10] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2019.