The Influence of Internet Celebrities on the Purchase Intention of College Students

Aliyaguli Waili

Beijing Technology and Business University, Beijing, China

Aliya2001127@163.com

Abstract. With the rapid development of Internet technology, Internet celebrity marketing has become an emerging value creation model of the Internet.Nowadays, the Internet celebrities have become an important factor affecting consumers' purchase intention.In order to promote the continuous expansion and development of Internet celebrity marketing.It is necessary to research how the Internet celebrities influence the purchase intention of consumers.This study selected college students in first-tier cities as samples, collected relevant data through online questionnaire, and studied the influence of four characteristics of Internet celebrities on the purchase intention of college students based on principal component regression model and K-means clustering algorithm. Through the principal component regression model, the linear regression equation affecting the purchase intention of college students and the popularity of Internet celebrities have the greatest influence on the purchase intention. At the same time, the k-means clustering algorithm is used to cluster the original indicators, and the conclusion that the popularity of Internet celebrities has the greatest influence on the purchase intention of college students is verified again.

Keywords: Live streaming; Internet celebrity characteristics; Principal component regression model;K-means clustering algorithm;Purchase intention.

1. Introduction

According to the 48th Statistical Report on China's Internet Development released by China Internet Network Information Center (CNNIC) in Beijing[1] .According to the report,As of June 2021, the number of Chinese netizen has reached 1.011 billion, an increase of 21.75 million compared with December 2020, and the Internet penetration rate has reached 71.6%.With the rapid development of mobile Internet technology, a series of new groups of Internet celebrities and new marketing models of live streaming with goods have emerged.According to the survey[2], By June 2022, the number of live streaming users in China reached 716 million.Among them, the number of e-commerce live broadcast users was 461 million ,accounting for the whole Internet user’s 44.6%.

Internet celebrity marketing has become an emerging value creation model of the Internet, and consumers' purchasing behavior in the context of Internet celebrity economy has attracted more and more attention from all fields.Compared with traditional e-commerce, Internet celebrity marketing has its own "fan" traffic and no longer needs to relocate its audience group.Xu Jingyi [3] mentioned in his research that among the characteristics of online celebrity live streaming, the interest, intelligibility and commodity attraction of Internet celebrity.It can significantly improve the inner satisfaction of college students' consumers and further promote their purchase intention.Chen scholars[4] have make the research that the influence of Internet platform and Internet celebrity characteristics on consumers' purchase intention,and explore the role of self-construction. The research results suggest that consumer characteristics, credibility, professionalism and attractiveness of Internet celebrities will affect consumers' purchase intention.At present, most studies on the influence of Internet celebrity on college students' purchase intention lack of corresponding model and theoretical support, thus weakening the scientific results and inferences of the research.

From the perspective of the four characteristics of Internet celebrity, this paper analyzes how the Internet celebrities affect the purchase intention of college students.It is based on the principal component regression analysis model and K-means clustering algorithm.And then,it uses the model results to verify the specific impact and effect of Internet celebrities on college students' online

shopping behavior.In addition,this research makes certain theoretical contributions and inspiration for related research fields.

2. Variable definition and data source

2.1 Variable definition

2.1.1Popularity of Internet celebrities

The popularity of Internet celebrities[5] refers to the public's familiarity with Internet celebrities and the social influence they have.MengFei[6] scholars use grounded theory to study the influence of online opinion leaders on purchase intention,and measure the popularity of opinion leaders from three aspects: social status, public familiarity and celebrity effect.

2.1.2Professionalism of Internet celebrities

The professionalism of Internet celebrities[5] refers to the Internet celebrity has the professional knowledge and ability to bring goods products.And being able to skillfully introduce product information to consumers.The measurement of professionalism refers to the five dimensions of knowledge, ability, experience, experts and training proposed by Gilly[7] and other scholars when they studied traditional word-of-mouth communicators by using the information source theory.In combination with the questions proposed by non-dream scholars according to the characteristics of Internet celebrities, the professionalism of Internet celebrities is finally measured from the three dimensions of knowledge, ability and experience.

2.1.3Attractiveness of Internet celebrities

The attractiveness of Internet celebrities[5] refers to Internet celebrities have unique personal charm and physical features,this will promote affection for them.Attractiveness Based on the scales developed by Ohanian[8] scholars,and combined with the scales of information source credibility dimension proposed by Ding Xia qi[9] scholars.

2.1.4Credibility of Internet celebrities

The credibility of Internet celebrities[5] means that Internet celebrities are real and reliable, the products of Internet celebrities are guaranteed.Credibility was modified according to the scales developed by Ohanian's study, and combined with the items of this dimension proposed by Liu Feng jun[10], Wang Xiao min and other scholars according to the Internet celebrity scenario. Finally, the credibility of Internet celebrities is measured from three aspects:credibility, honesty and reliability.

2.1.5Purchase intention

Purchase intention[5] refers to how likely consumers are to buy products recommended by Internet celebrities while watching live broadcasts.In view of the study on the purchase intention of consumers with live-streaming delivery, many scholars have put forward relevant measurement scales.This research based on the scale items proposed by Gilly, Bansa, and Chen Hai quan scholars.

Table1 variable item and sources

Variable	Item	Reference sources
Popularity	The Internet celebrity has a dominant and influential position in the field of live streaming	Mengfei Chenhaiquan
	The Internet celebrity is a well-known figure in the field of live streaming goods	
	The Internet celebrity has a large number of fans and a high degree of activity	
Professionalism	The network celebrity has the relevant knowledge in the field of live delivery and product promotion	Gilly Mengfei

	he network celebrity has the corresponding professional ability in the field of live delivery and recommended products	
	The network celebrity has rich practical experience in the field of live streaming with goods	
Attraction	The Internet celebrity is very beautiful.	Ohanian Dingxiaqi
	The Internet celebrity is humorous and charming	
	The Internet celebrity is popular with live viewers	
Credibility	The products recommended by the Internet celebrity are reliable	Ohanian LiuFengJun WangXiaoMin
	The Internet celebrity is trustworthy	
	The Internet celebrity is honest	
Purchase intention	The Internet celebrity enriched my understanding of this kind of product	Gilly Bansal Mengfei Chenhaiquan
	The Internet celebrity helped me when making purchasing decisions	
	The Internet celebrity's product recommendation stimulated my desire to buy	

2.2 Data Source

This study is carried out for college students in first-tier cities, and the reasons for selecting such groups are as follows:①First-tier cities have a high degree of acceptance of live delivery, and there are more consumer groups in live broadcast rooms.②Contemporary college students are the main consumers of live broadcast studios, and this group can give a more objective and comprehensive evaluation of the characteristics of live broadcast internet celebrities.

According to the survey results, a 585 questionnaires were released,and 449 copies were recovered,by screening a 360 valid questionnaires have been selected.What is more,80 percent of women and 20 percent of men participated in the survey. 41% of respondents have monthly living expenses between 1001-1500.It has been 1-3 years since they bought commodities through the live broadcast room, which also indicates that the survey samples have a complete understanding the live broadcast goods and Internet celebrities.This can ensure the validity and availability of the questionnaire results.

3. Principal component regression model

3.1 Principle of principal component analysis regression model

Principal component analysis[11] is an important method of multivariate statistical analysis. As a multivariate statistical analysis method, multivariate analysis is widely used in the field of mathematical statistics.It mainly uses all the independent variables to run the principal component analysis, reduces the dimension of the independent variables, and obtains a set of unrelated principal components, which bear the variance of the original variables.Therefore, it is unnecessary to consider the interaction between various parameters in the establishment of linear regression equation, which relies on the same parameters of the matrix corresponding to the parameters of the equation, so there will be no case of parameter multiplication or flat placement.

3.2 Steps of principal component analysis regression model

The matrix of original data in this study is established and the original data is standardized. In this study, there are 5 index variables of principal component analysis, and each index variable has 3 main impact factors. The original data is converted into the following matrix, as shown in the formula:

$$X= \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} = \quad (1)$$

$$[X_1, X_2, X_3, \ldots, X_p]$$

Since the unit of each variable in this study is uniform, no standardization is needed. Next, directly calculate the correlation coefficient matrix, as shown in the formula:

$$R= \begin{bmatrix} r_{11} & \cdots & r_{1p} \\ \vdots & \ddots & \vdots \\ r_{p1} & \cdots & r_{pp} \end{bmatrix} \quad (2)$$

What is more,identify principal component indicators.The eigenvalues are calculated in detail the information contribution rate and cumulative contribution rate of $\lambda_i (i = 1,2,\ldots\ldots,p)$ .Therefore, cumulative contribution rates of principal component indicators of data are obtained according to the above steps as shown in the table below:

Table 2 Cumulative contribution rate

compositions		Initial eigenvalue			Extract the sum of loads squared		
		Total	Percentage of variance	Accumulation %	Total	Percentage of variance	Accumulation %
1		5.203	43.362	43.362	5.203	43.362	43.362
2		1.427	11.893	55.255	1.427	11.893	55.255
3		1.209	10.077	65.332	1.209	10.077	65.332
4		.887	7.389	72.721	.887	7.389	72.721
…		…	…	…			
12		-2.350E-16	-1.958E-15	100.000			

As can be seen from the above table, the first four principal components explain 72.721% of the total variance.According to the principle of principal component accumulation, the first four principal component factors can basically reflect the main characteristic types of Internet celebrities.Therefore, the first 4 principal component factors are selected for the next step.The coefficient of each index was solved according to the principal component score and component matrix coefficient, and the principal component regression linear formula of the data in this study was obtained:

GM= $0.681 \times F_1 + 0.544 \times F_2 + 0.204 \times F_3 + 0.248 \times F_4$ According to the above model regression equation, the factors that have the greatest influence on college students' purchase intention are the popularity and professionalism of Internet celebrities.Among them, the popularity of Internet celebrities has the greatest impact on college students' purchase intention. This is because in today's information network era,famous Internet celebrities not only have many fans, but also have a high popularity compared with other Internet celebrities.Moreover, the objects of this survey are contemporary college students, who are young and leading groups under the "fan flow". Therefore, the higher the popularity of Internet celebrities in this study, the stronger the purchase intention of college students.

Table 3 Principal component matrix regression coefficient

	Regression model	Standardization coefficient	t	Significance
Argument	constant		203.30	.000
	F1	.681	24.270	.000
	F2	.544	19.360	.000
	F3	.204	7.251	.000
	F4	.248	8.843	.000

According to the above principal component matrix regression coefficient,there is a positive correlation between the four principal components and the purchase intention of the dependent variable. At the same time, the greater the regression coefficient of internet celebrity of popularity and professionalism,while the smaller contribution rate of attractiveness and credibility to the purchase intention.It can be seen that the popularity and professional characteristics of Internet celebrities are strong variables affecting college students' purchase intention.

4. K-means clustering algorithm

4.1 Basic principles of K-means clustering algorithm

K-means clustering algorithm[12], namely, k-means clustering algorithm, was first proposed by MacQueen[8]. Its basic principle is to divide a given data set into K different types, find out the clustering center of each class, and minimize the non-similar features (such as distance) between different types.

4.2 The basic steps of K-means clustering algorithm

First, the data set is redistributed in order to ensure that all the data in each dimension are arranged from small to large, forming a new set Mia.Mia of the data set was divided to obtain K intervals.The central value of each interval was calculated within the interval or serial number.The data corresponding to the middle serial number was taken out and identified as the initial clustering center:

$$q=\left(\frac{n*j}{k}+\frac{n*(j-1)}{K}\Big/2\right) \qquad (3)$$

In addition, the above initial clustering center results were iteration processed, and 15 indicators were clustered in the process of 7 iterations, and the following results were obtained:

Table4 Initial cluster center

Initial cluster center				
	Clustering			
	1	2	3	4
ZM1	4.00	3.00	5.00	2.00
ZM2	4.00	4.00	5.00	2.00
ZM3	3.00	3.00	5.00	4.00
ZY1	2.00	5.00	4.00	4.00
ZY2	4.00	4.00	3.00	5.00
ZY3	4.00	5.00	4.00	5.00
XY1	1.00	4.00	4.00	4.00
XY2	1.00	5.00	3.00	1.00
XY3	1.00	5.00	2.00	3.00
GM1	3.00	5.00	3.00	1.00
GM2	4.00	5.00	5.00	4.00
GM3	1.00	4.00	4.00	4.00
KX1	1.00	5.00	3.00	1.00
KX2	1.00	5.00	2.00	3.00
KX3	1.00	4.00	4.00	4.00

Table 5 Finally clustering center

Finally clustering center				
	Clustering			
	1	2	3	4
ZM1	3.85	4.13	4.25	3.50
ZM2	3.31	4.29	4.30	3.48

ZM3	3.69	4.45	4.39	3.93
ZY1	3.00	4.10	4.05	3.65
ZY2	3.15	4.10	4.14	3.58
ZY3	3.77	4.34	4.23	3.93
XY1	2.08	4.21	4.25	3.10
XY2	1.54	4.23	3.09	2.80
XY3	2.00	4.37	3.45	3.10
GM1	2.38	4.30	3.20	3.00
GM2	2.92	4.32	3.66	3.30
GM3	2.08	4.22	4.25	3.13
KX1	1.54	4.23	3.09	2.80
KX2	2.00	4.37	3.45	3.10
KX3	2.08	4.21	4.25	3.10

Table6 Results of the number of cases in each cluster

Number of cases in each cluster		
Clustering	1	13.000
	2	82.000
	3	44.000
	4	40.000
Effective		179.000
Deletion		1.000

This iterative algorithm is adopted to make a cluster analysis on the influence of Internet celebrity characteristics on college students' purchase intention. The initial 15 indicators are classified into four categories, and according to the scores of the four characteristics of Internet celebrity in each category, the following categories are classified:the first category belongs to the credibility of Internet celebrity, the second category belongs to the popularity of Internet celebrity, the third category belongs to the professionalism of Internet celebrity,and the fourth category belongs to the attractiveness of Internet celebrity. The results show that there are more cases in the category of Internet celebrity popularity, which further verifies the result that Internet celebrity popularity has the largest influence (coefficient) on purchase intention in the above principal component regression model.

At the same time, the above results further provide corresponding inspiration and improvement points for e-commerce live streaming platforms.For college students, the popularity of Internet celebrities is the biggest factor influencing their purchase intention of commodities in the live broadcast room.E-commerce live broadcast platforms need to train and cultivate online celebrity anchors accordingly. Through corresponding measures such as fan discount and live delivery training, the characteristics of Internet celebrities should be fully cultivated at the same time. Professional live streaming network celebrities should also be cultivated according to the factors influencing the purchase intention of target groups.

5. Conclusion

With the booming development of e-commerce and mobile Internet, live streaming has become a new marketing mode. Especially for young college students, they have a high acceptance of the new live streaming marketing mode and a relatively strong perception of the characteristics of anchors.Therefore, this study selected college students as the research object, based on the principal component analysis method to determine the degree of influence of the four characteristics of Internet celebrities on college students' purchase intention, through the model results found that the popularity and professionalism of Internet celebrities on the purchase intention of college students to the greatest degree.In addition, the K-means clustering algorithm was used to cluster the research data, and four types of clustering were determined according to the final clustering center table and

the number of each cluster case, and it was verified again that the popularity of Internet celebrities had the greatest influence on the purchase intention of college students.

In addition to the above research results， there appear the insufficient effective sample data results in a weak positive correlation between the independent variable of Internet celebrity characteristics and the dependent variable of purchase intention.The psychological mechanism of consumer groups between the Internet celebrities and the purchase intention of college students is not taken into account.For example,the SOR theoretical model is not used to find out the organism between Internet celebrity and purchase intention, in order to make the study valid and scientific.Thus,based on the analysis of the limitations of the study, future research of this direction,not only should ensure sufficient and effective sample data, but also we should use structural equation model to find and verify the body mediating effect between Internet celebrities and purchase intention.

Through the above research model and algorithm results, specific analysis and verification are carried out on how the characteristics of Internet celebrities affect the purchase intention of college students. The above

conclusions and inferences can inspire and reflect on the related research direction, and also provide constructive suggestions for the targeted cultivation of the characteristics of live Internet celebrities.

References

[1] The data of China Internet Network Information Center (CNNIC) in Beijing retrieved from:https://finance.sina.com.cn/tech/2021-08-27/doc-ikqciyzm3870486.shtml

[2] The data of the survey retrieved from:https://3g.163.com/dy/article/HMF0EINH051481OF.html

[3] XuJingYi.The influence of online celebrities' live streaming of goods on college students' purchase intention.2096-0298.2021.13.043

[4] Chen; Jaeyeon, Sim， Influence of Internet Celebrity’s Characteristic on Consumer Purchase Intention: The Mediating Effect of Self-Construal ,Korean Management Consulting Review

[5] Xiaojiaqi.A study on the influence of characteristics of live streamers on purchasing intention[D].Central South University of Forestry and Technology,,2022.DOI:10.27662/d.cnki.gznlc.2022.000024.

[6] MengFei.Research on the influence of opinion leaders on purchase intention in socialized business environment.F274;F224

[7] Gilly M C,Graham J L,Wolfabarger M F.Yale L J.A dyadic study of interpersonal information search[J].Journal of the Academy of Marketing within a services purchase decision context[J].Journal of Service Research,2000(2)

[8] Ohanian R.Construction and validation of a scale to measure celebrity endorsers’ perceived expertise, trust worthiness and attractiveness [J].Journal of Advertising,1990(3):39-52

[9] DingXiaQi,Wang Huai Ming,Ma mou chao.The Influence of moral reputation of celebrity promoters on the effectiveness of celebrity advertisement.Acta psychologica Sinica,2005(03):382-389.

[10] Liu Feng Jun,Meng lu.Study on the influence and mechanism of online celebrity live streaming on consumers' purchase intention.Journal of management,2020,17(01)

[11] Zhang Mengsu, Liu Chunchun, LiXijin,scholars.Design of Fuzzy Comprehensive Evaluation System for Performance Appraisal Based on K-means Clustering Algorithm[J],Journal of Jilin University (Engineering and Technology Edition)2021,51(05):1851-1856.DOI:10.13229/j.cnki.jdxbgxb20201004.

[12] Ding Zhi-cheng, Ge Hong-wei, Zhou Jing.Density peaksclustering based on Journal of Chongqing University of Posts a Telecommunications(Natural Science Edition),2019, 31(3)