ISSN: 2688-9323

How Question Characteristics Affect Sales of Paid Viewership in Paid Q&A

Peng Liu[1], Yubei Wu[2]

[1]ChongQing University of Post and Telecommunications, Chongqing 400065, China

[2]chongqing Acoustics-Optics-Electronics Co., Ltd. of China Electronics Technology Group Corporation, Chongqing 401332, China

Abstract

Knowledge sharing economy subverts the way people acquire knowledge. With the gradual establishment and improvement of online payment system, the improvement of copyright protection system and the formation of users' habit of paying for knowledge, online paid Q&A shows great development potential. How to promote the innovation of paid question answering and maintain the sustainable development of paid question answering has become the common concern of paid question answering platforms and scholars. The existence of paid viewership in Q&A is beneficial to promote the participation of all kinds of users, enhance the competitiveness of the platform and maintain the sustainable development of the platform. Therefore, relevant researches on the paid viewership have attracted the attention of scholars, among which some achievements have been made on the influencing factors of user payment behavior. However, in the paid Q&A with text as the main form, there are few in-depth studies on the text characteristics of the question. In this context, it is helpful to solve the sales problem of paid Q&A products and provide enlightenment for the sustainable development of paid Q&A by exploring the factors influencing the sales of paid viewership in the paid Q&A.

Keywords

Paid Q&A; Sales of Paid Viewership; Readability; Sentiment.

1. Introduction

Traditional online Q&A platforms provide an open environment where people can ask and answer any question of their own free will. However, while the free online Q&A community is full of high-quality Q&A, it is also full of redundant, distorted and even wrong information [1]. This makes it expensive for users to obtain valid information. In the context of information overload, people with learning anxiety have stronger motivation to learn, and thus have a stronger willingness to pay for quality content. At the same time, with the improvement of users' personalized needs, online Q&A communities have introduced payment mechanisms [2].

The growth of online paid Q&A has spawned numerous business models such as subscription and pay-per-item[3]. This study focuses on a manifestation in paid Q&A which first launched in 2016 by 'Fenda'. The basic characteristics of the model are summarized as follows :1) it includes three types of users: askers, answerers and viewers; 2) askers can obtain personalized answers from the answerer at a certain question price; 3) viewers can view a question and answer session by paying a relatively low flat fee; 4) askers and answerers are allowed to shares the revenue paid by views.

The revenue sharing mechanism provides the possibility of providing additional financial incentives for answerers and askers, further covering as many paying users as possible, and also allowing potential questioners to view the quality of the question. Although paid Q&A as

ISSN: 2688-9323

an important aspect of knowledge payment has experienced rapid development, knowledge products as experience products, often due to information asymmetry and low sales, delayed the development of the platform. Therefore, for the sustainable growth of social question and answer websites, it is crucial to clarify the influencing factors and mechanism of sales from paid viewership.

2. Theoretical Background

2.1. Prior Research on Paid Q&A

Research on paid Q&A can be summarized into three streams. The first stream focus on the user's willingness to participate, why users are willing to participate in paid Q&A, and what are the drivers of users' motivation to participate. For example, Lee et al. 's study showed that respondents were mainly motivated by financial and intrinsic motivation rather than social factors[4]. The second stream focus on the assessment of question quality. The measurement of Q&A quality is of great significance to the in-depth research of user experience and user willingness to pay. Scholars have developed different methods to evaluate the quality of answers from different theoretical perspectives combined with research methods, for example, Wang et al. used the prediction method of deep learning to predict the quality of answers mainly through the characteristics of answers including the ranking of text answers and the length of answers[5], Fu et al. considered the question and answer topic as a potential influencing factor for the prediction of answer quality[6]. The online Q&A community has looked at question descriptions and responses as a whole, using question descriptions to predict the quality of responses. the third stream focus on the user's willingness to pay, for example, scholars have explored the factors influencing sales volume from the dimensions of answer characteristics and price[3, 7].

2.2. Signaling Theory and Antecedents of the Sales of Paid Viewership

The signal theory was proposed by Spence in his pioneering research on the labor market. It mainly describes how two transaction subjects reduce information asymmetry through their behaviors in the case of information asymmetry in the market, so as to improve their choice ability[8]. Generally, there are sender and receiver of signals, and signals are often important information to distinguish product quality. For example, in the labor market, high-quality applicants try to distinguish themselves from low-quality applicants with the signal of strict higher education, so as to improve the selection ability of employers. This theory has been applied in the selection scenarios of many disciplines and obtained abundant research literature. For example, management scholars in the field of corporate governance use signal theory to explain how executives release the invisible quality of corporate development to the market and investors through financial statements.

The trust sensitive feature of knowledge consumable means that consumers have higher risk perception in consumption[9], In knowledge payment, there is a widespread situation of product information asymmetry. Consumers need to identify the value signal of products to eliminate the risk of such information asymmetry as far as possible. A competent, sincere and reputable content creator can gain the trust of users, trust in content creators further boosts users' willingness to pay[10].

As for the question characteristics, past research focuses on problems of the external characteristics of text, such as time, speed of response, number of responses, number of likes, etc., and few quantitative research question text itself features and impact on the user's system. However, previous Q&A community studies prove that text is the important basis of judging Q&A quality[6]. Jan et al. 's research shows that online product description helps to reduce product uncertainty[11].Text can extract a lot of information, sentiment analysis of text is one

ISSN: 2688-9323

of the most popular methods. Zhao et al. believe that problem emotion can stimulate readers' empathy, thus stimulating readers to participate in the follow-up discussion and feedback high-quality answers[12]. Text readability is a long-standing research problem, it is widely accepted that readability is an important factor in knowledge product trading. Text readability is defined as the sum of all elements that affect readers' comprehension, reading speed and interest level in a text[13].

According to the research of Lee et al., financial incentive are one of the main influencing factors of respondents' participation in Q&A[4]. The quality of answers in paid Q&A is significantly higher than social Q&A[14]. Therefore, price can be a powerful indicator of the quality of answer.

3. Hypothesis Development and Research Model

3.1. Sentiment

Sentiment refers to the positive or negative emotions shown by the question text and its intensity. Sentiment analysis can extract emotional information from the text and classify it according to its polarity. This method is widely used in empirical studies involving text to help understand how consumers understand product information[15]. Aristotle's narrative theory points out that emotion is one of the elements of verbal persuasion. How emotional information in product text affects consumer attitudes is a complex study. Research from online word of mouth shows that emotional information in text affects actual sales through word-of-mouth effect[17]. Research has shown that emotionally appealing text language styles are detrimental to project success. Hence, we hypothesize:

Hypothesis 1 (H1): Sentiment negatively affects the sales of paid viewership.

3.2. Readability

In the process of readers' reading, the extraction of basic information of different texts, its influencing factors and measures are collectively referred to as readability research. Forming value judgments about text content is an exhausting process, and low-quality content may further increase the cost of searching, browsing and processing information for users[17]. In order to capture the value signal released by the text, the information receiver needs to fully understand the measured information conveyed by the text. For the information receiver with limited energy, the bloated text is easy to distract attention, while the clear and concise text is more readable and more likely to cause the response of the information receiver. Hence, we hypothesize:

Hypothesis 2 (H2): Readability has an inverted U relationship with the sales of paid viewership.

3.3. Social Endorsement

Social endorsement refers to accurate and resource claims of user support in social networking services, such as user's likes for the question[18]. Social adoption indicators of Q&A, such as likes, provide credibility for the quality of Q&A, thus attracting the attention of viewers and increasing the likelihood that they will adopt Q&A. Hence, we hypothesize:

Hypothesis 3 (H3): Social endorsement positively affects the sales of paid viewership.

3.4. Picture

Picture refers to the picture that the asker attaches to explain the question. With the rapid development of digital communication technology, the content of Internet questions and comments is no longer limited to text, and the type and quantity of visual information and text information have different influence paths on consumers' response to experiential products[19]. Paid Q&A is still mostly text-based, However, the function of picture overlaps with text. When pictures incompletely replace text for question description, it takes more

ISSN: 2688-9323

energy to obtain question information from pictures than from text description. Therefore, the presence of pictures hinders viewers from obtaining question information and further inhibits sales of paid viewership. Hence, we hypothesize:

Hypothesis 4 (H4): Picture negatively affects the sales of paid viewership.

3.5. Price

Question price refers to the fee paid by the asker to the answerer for consultation. In the trade of knowledge products, it is generally believed that the price of the product significantly restrains the sales of the product[20]. However, no matter what the price of Q&A is, viewers only need to pay a small fixed fee to view Q&A. For viewers, the price paid by questioners is positively correlated with the length of answers provided by respondents. The more information the answers contain, the higher the quality of the answers among similar questions. Hence, we hypothesize:

Hypothesis 5 (H5): Social endorsement positively affects the sales of paid viewership.

3.6. Control Variables

We also included control variables that may impact the dependent variable, including the number of inquiries that an answerer receives, the number of favorites that an answerer receives on social Q&A platforms, and the number of favorites that an answerer receives. Research model shown in fig.1.

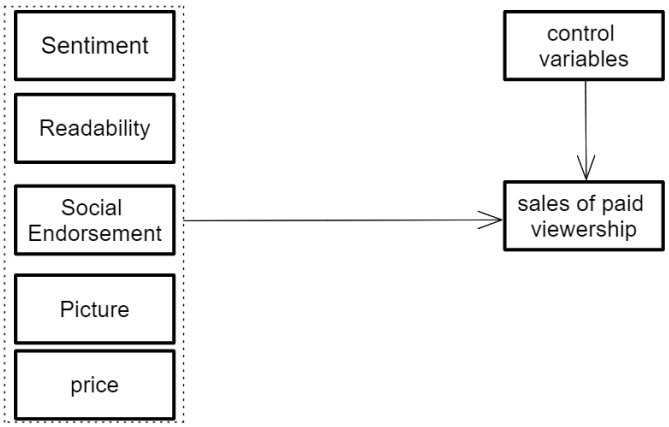

Fig 1. Research model of paid Q&A

4. Research Methodology

4.1. Data Collection

We collected data from Zhihu Q&A. We crawled the web data by By deploying crawler from 2022 to January 21, 2022, obtained a total of 65,300 questions from zhihu paid Q&A on all subject areas.

After removing abnormal crawling records and empty records, a total of 54007 valid records were obtained, including the ID of the host, the number of endorses, the number of thanks, the number of favorites, the stars of the answerers and the times of inquiries, as well as the sales of paid viewership to the question, the number of likes, pictures and the text of the question.

4.2. Variable Measurement

Text readability is a long-discussed problem to measure the complexity of text. It was first proposed by educationologists. However, the development of Chinese text analysis technology is relatively slow. At present, there is no universally accepted method to measure the

ISSN: 2688-9323

readability of text information. We use readability formula to extract the readability of the text in question[21].

Sentiment analysis, also known as opinion mining, is a computational study of opinions, emotions and attitudes of entities such as product and service issues, events, topics and their attributes. In this study, the Python natural language processing library PaddlenL P was used to calculate the emotional polarity score of the text in question. This method covers mainstream Chinese pre-training models such as BERT, GPT, XLNet, BART and baidu's self-developed pre-training models such as ERNIE series, PLATO, SKEP, etc. After the calculation, the score range is [-1,1], and the closer the score is to the left end of the range, the more negative the emotional tendency of the question text is. On the contrary, the closer the score is to the right end of the range, the more positive the emotion tendency of the question text is.

As for the processing of picture variables, we adopt the method of recognizing whether there are pictures and processing dummy variables. The code of question without pictures is 0, and the code of question with pictures is 1. And for variables that do not require additional processing, we use their statistical characteristics directly. Variable description and statistics shown on table1.

Table 1. Variable description and statistics

Variable	Mean	SD	Min	Max
Sales of paid viewership(SPV)	25.30235	97.45766	0	4837
Sentiment(Se)	-.4310699	.7641002	-.99858	.99813
Readability(Re)	25.30405	4.323783	6.533741	46.59282
Social Endorsement(SE)	2.395967	16.74334	0	1652
Picture(Pi)	.3060714	.460864	0	1
Price(pr)	35.05375	47.96161	1	1999
Number of Inquires(NI)	1222.69	1321.266	5	8357
Number of Favorites(NF)	128673	255852	15	2948014
Number of Reviews(NR)	1212.308	1315.51	5	8300

4.3. Model Specification

The dependent variable is a counting variable, and its value can only be a non-negative integer. When using the linear regression model, it faces the problems of discontinuity and non-normality. Although the linear regression model can be used for least square estimation, serious heteroscedasticity problems will occur. In this case, the data mean and variance are not equal and have heteroscedasticity, which is no longer consistent with the basic hypothesis of Poisson regression. Negative binomial regression should be used to solve the heteroscedasticity problem. The regression model is shown in Eq. (1).

$$SPV=\beta_0 + \beta_1 Se + \beta_2 Re + \beta_3 Re * Re + \beta_4 SE + \beta_5 Pi + \beta_6 Pr + \beta_7 NI + \beta_8 NF + \beta_9 NR + \varepsilon \quad (1)$$

5. Analysis and Results

Stata16 was used to perform negative binomial regression on the data, and the modeling and testing results were obtained. The test results are shown in table 2.

According to the results in table 2, as the indicator of question sentiment, Se (β1= -0.137; $p < 0.001$) has significantly negative effects on sales of paid viewership, which supported H1. Re (β2=0.169; $p < 0.001$), which represents the readability of question, has significantly positive effects on sales of paid viewership, while its squared term Re*Re (β3=-0.003; $p < 0.001$), which means that readability has an inverted U relationship with the sales of paid viewership, H2 is

ISSN: 2688-9323

supported. SE (β1= 0.130; p < 0.001), which represents number of likes of a question, has significantly positive effects on sales of paid viewership, H3 is supported. Pi (β1= -0.349; p < 0.001), as the indicator of the exist of picture in question, has significantly negative effects on sales of paid viewership, H4 is supported. Pr (β1= -0.003; p < 0.001), which represent the price paid by the asker, has significantly positive effects on sales of paid viewership, H5 is supported. Thus, all hypotheses are supported.

Table 2. Hypothesis testing

SPV	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
Se	-.136795	.0084255	-16.24	0.000	-.1533087	-.1202813
Re	.1685935	.0130045	12.96	0.000	.1431052	.1940818
Re*Re	-.003113	.0002498	-12.46	0.000	-.0036027	-.0026234
SE	.1296461	.0016406	79.02	0.000	.1264306	.1328615
Pi	-.3494546	.0144893	-24.12	0.000	-.3778531	-.3210561
Pr	.0031116	.0002022	15.39	0.000	.0027153	.0035079
NI	.0017266	.0004522	-3.82	0.000	-.0026128	-.0008404
NF	5.98e-07	2.74e-08	21.80	0.000	5.44e-07	6.52e-07
NR	.0018461	.0004538	4.07	0.000	.0009567	.0027355
_cons	-.1674796	.1674618	-1.00	0.317	-.4956988	.1607396
/lnalpha	.7579684	.0062034			.74581	.7701268
Alpha	2.133936	.0132376			2.108148	2.16004

6. Discussion and Conclusion

In this paper, we conducted an empirical study based on the signal theory to examine the influence of question characteristics on sales of paid viewership in Zhihu.com. The results shows that question characteristics, include text feature (readability, sentiment), social endorsement, picture, and price, with the exception of sentiment, the rest have significantly positive effects on sales of paid viewership. Social endorsement significantly boosted sales of paid viewships, and the higher the likes of the question, the higher sales of paid viewships.

Based on the results, management suggestions are put forward. The paid Q&A platform should provide guidance on user question writing. First, the results show that positive emotion hinder sales of paid viewership, and negative emotion instead promote sales of paid viewership, as a result, the questioner should show more negative emotions, and decrease the positive emotion, in order to highlight the severity and importance of problems to attract viewers. In terms of readability, it can be seen from the research results that the question text with low readability hinder the viewer's adoption of the question. Therefore, the manuscript of question should be appropriately concise and should be explained clearly. In terms of pictures, picture hinders viewers to view the question, it takes more effort to understand the information through pictures, and thus lowered viewer's interest, therefore, askers should use text to explain questions whenever possible and minimize the use of pictures.

This study is no exception for limitations. First, due to data sources, the conclusions can be extended remains to be verified. Future research can start from expanding the external applicability of research conclusions. And the readability formula used in this study is only one of many readability calculation methods. Future research can start from different readability calculation methods, such as deep learning, or explore other ways to reflect text readability.

ISSN: 2688-9323

References

[1] Yoon, S., The antecedents and consequences of trust in online-purchase decisions. Journal of interactive marketing, 2002. 16(2): p. 47--63.

[2] Zhao, Y., et al., How knowledge contributor characteristics and reputation affect user payment decision in paid Q&A? An empirical analysis from the perspective of trust theory. ELECTRONIC COMMERCE RESEARCH AND APPLICATIONS, 2018. 31: p. 1-11.

[3] Ye, H.J., et al., Monetization of Digital Content: Drivers Of Revenue On Q&A Platforms. Journal of Management Information Systems, 2021. 38(2): p. 457-483.

[4] Lee, U., et al., Analyzing crowd workers in mobile pay-for-answer q\&a. 2013. p. 533--542.

[5] Wang, H.C., Y.H. Chiang and S.T. Lin, Spam detection and high-quality features to analyse question - answer pairs. ELECTRONIC LIBRARY, 2020. 38(5-6): p. 1013-1033.

[6] H, F., W. S and O. S, Evaluating answer quality across knowledge domains： using textual and non-textual features in social Q&A. 2015.

[7] Yang, X.P. and H. Ye, Commercialized Content on Social Media Platforms: Exploring the Drivers of the Viewership of Paid Q&A, in PROCEEDINGS OF THE 52ND ANNUAL HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES, T.X. Bui, T.X. Bui^ Editors. 2019: 52ndHawaii International Conference on System Sciences (HICSS). p. 2709-2718.

[8] Spence, M., Signaling in retrospect and the informational structure of markets. American Economic Review, 2002. 92(3): p. 434--459.

[9] Debo, L.G., L.B. Toktay and L.N. Van Wassenhove, Queuing for expert services. MANAGEMENT SCIENCE, 2008. 54(8): p. 1497-1512.

[10] Yoon, S., The antecedents and consequences of trust in online-purchase decisions. Journal of interactive marketing, 2002. 16(2): p. 47--63.

[11] Pavlou, P.A. and A. Dimoka, Understanding and mitigating product uncertainty in online auction marketplaces. 2008.

[12] Yiming, Z., et al., How Question Characteristics Impact Answer Outcomes on Social Question-and-Answer Websites. Journal of Global Information Management (JGIM), 2021. 29(6): p. 1-21.

[13] Dale, E. and J.S. Chall, The concept of readability. Consciousness and Cognition, 1949.

[14] Harper, F.M., et al., Predictors of answer quality in online Q\&A sites. 2008. p. 865--874.

[15] Hajiali, M., Big data and sentiment analysis: A comprehensive and systematic literature review. CONCURRENCY AND COMPUTATION-PRACTICE & EXPERIENCE, 2020. 32(14).

[16] Moore, S.G. and K.C. Lafreniere, How online word-of-mouth impacts receivers. Consumer Psychology Review, 2020. 3(1): p. 34--59.

[17] Gu, B., et al., Competition among virtual communities and user valuation: The case of investing-related communities. Information systems research, 2007. 18(1): p. 68--85.

[18] Cai, S., et al., What drives the sales of paid knowledge products? A two-phase approach. Information & Management, 2020. 57(5): p. 103264.

[19] Kim, M., et al., Impact of visual information on online consumer review behavior: Evidence from a hotel booking website. JOURNAL OF RETAILING AND CONSUMER SERVICES, 2021. 60.

[20] Cai, S., et al., What drives the sales of paid knowledge products? A two-phase approach. Information & Management, 2020. 57(5): p. 103264.

[21] Yang, S., A readability formula for Chinese language. 1971: The University of Wisconsin-Madison.