Predictive Study on Question Response Rate in Social Q&A Platforms: A Case Study of Baidu Zhidao (Postprint)

Authors: Deng Shengli, Fu Shaoxiong, Liu Jin

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] Based on the low response rate phenomenon in social Q&A platforms, predicting question response rates can provide valuable references for enhancing user engagement, improving retention, and optimizing user experience. [Method/Process] Utilizing "Baidu Zhidao" as the research platform, this study collected 10,640 question records across 14 platform-defined topics. A research framework for factors influencing question response rates was constructed from the perspectives of question characteristics and asker characteristics. Binary Logistic regression was employed to empirically validate these factors, develop a predictive model for question response rates, and assess the model's accuracy. [Results/Conclusion] Research on question response rates in social Q&A platforms contributes to improving information service quality and fostering user knowledge contribution behaviors. Experimental results demonstrate the effectiveness of the proposed model in predicting question response rates on social Q&A platforms.

Full Text

Preamble

ChinaXiv Cooperative Journal, Volume 63, Issue 10, May 2019

Predicting Response Rates in Social Q&A Platforms: A Case Study of "Baidu Knows"

Deng Shengli, Fu Shaoxiong, Liu Jin

Center for Studies of Information Resources, Wuhan University, Wuhan 430072

Abstract

[Purpose/Significance] Given the low response rates currently observed in social Q&A platforms, predicting question response rates can provide valuable insights for these platforms to enhance user activity, improve retention rates, and optimize user experience. **[Method/Process]** Using "Baidu Knows" as the research platform, we collected 10,640 question records across 14 topic categories established by the platform. We constructed a research framework examining factors influencing question response rates from two perspectives: question characteristics and questioner characteristics. Binary logistic regression was employed to validate these factors statistically and build a predictive model for question response rates, whose accuracy was subsequently verified. **[Result/Conclusion]** Research on predicting response rates in social Q&A platforms can improve platform information service quality and promote user knowledge contribution behaviors. Experimental results confirm the effectiveness of our proposed model for predicting question response rates in social Q&A platforms.

Keywords: Social Q&A platform; Knowledge contribution behavior; Response rate; Logistic regression; Prediction

Classification Number: G252

DOI: 10.13266/j.issn.0252-3116.2019.10.011

Social Q&A platforms (social question & answer communities) are Internet platforms centered on users asking, answering, and discussing questions, incorporating community attributes of social networking services, social media, and knowledge management systems [?, ?]. Compared with traditional information seeking and sharing methods, although social Q&A platforms enable self-filtering and strict quality control of answers, they still suffer from low response rates. For instance, the cancer section of Yahoo! Answers has only a 6% response rate [?], and fewer than half of questions receive answers on Google Answers [?]. Building upon analyses of factors affecting response rates, we identify question media richness, wealth value, urgency expression, and politeness expression, along with questioner reliability, influence, and network centrality as key determinants of response rates. Using Python to crawl 10,640 question records across 14 topics from "Baidu Knows," we applied binary logistic regression to validate these factors, ultimately constructing and validating a predictive model for question response rates.

1 Related Research

1.1 Research on User Response Motivation in Social Q&A Platforms

Given the low response rates in social Q&A platforms [?, ?], effectively enhancing user willingness to respond and thereby improving platform activity has become an urgent issue. To ensure user activity and retention, social Q&A platforms must effectively promote user knowledge contribution intentions and focus on improving question response rates. Related research primarily concentrates on two aspects: first, exploring users' intrinsic and extrinsic motivations for responding [?]; second, predicting user information behaviors such as best answerers, forwarding behavior, and movie rating behavior [?]. Through literature review, we find that existing studies lack predictive research on response rates in social Q&A platforms, despite the value of such research for platform improvement.

To address this gap, we investigated intrinsic and extrinsic motivations for knowledge contribution in social Q&A platforms. Most studies employ questionnaires or interviews (see Table 1). Intrinsic motivation refers to positive psychological feelings users hope to obtain from completing behaviors, while extrinsic motivation refers to instrumental or material rewards users desire after participating in tasks [?]. In social Q&A platforms, users' motivations for answering questions primarily reflect self-value realization. Intrinsic motivation involves users participating in Q&A through autonomous choice or personal preference—when users seek adoption or likes to increase their sense of identification, they participate in Q&A. Extrinsic motivation mainly manifests in knowledge contribution behaviors when users hope to obtain experience or wealth values [?].

1.2 Research on Predicting User Information Behavior Online

Since current research lacks predictive studies on question response rates in social Q&A platforms, we examined research on predicting online user information behavior, as shown in Table 2 . Prediction research objects primarily focus on social service websites and shopping websites, constructing predictive models based on characteristic data of user behaviors. Prediction methods mainly include logistic regression [?], random forests [?], complex network analysis [?], fuzzy neural networks [?], and decision trees [?]. Related studies report prediction accuracies between 50%-94% [?]. Based on comprehensive analysis of user behavior prediction methods, since question response rate is a typical binary classification problem, logistic regression is the most widely used multivariate quantitative analysis method for regression analysis of binary dependent variables, as it avoids structural defects of linear probability models through logit transformation of target probability [?]. Therefore, we adopted binary logistic regression to predict response rates in social Q&A platforms [?].

2 Model Construction and Hypotheses

Motivations for user knowledge contribution behavior in social Q&A platforms can be divided into extrinsic and intrinsic motivations [?, ?]. Intrinsic motivation typically stems from inner satisfaction or altruism and is primarily psychological. Unlike intrinsic motivation, which is mainly studied through questionnaires and interviews, we employed text analysis methods to explore question response rates from an extrinsic motivation perspective. Using Baidu Knows as our research object, we extracted factors influencing question response rates in social Q&A platforms. These indicators were categorized into question characteristics (entity features and non-entity features) and questioner characteristics (reliability, influence, and network centrality). Entity features include media richness and wealth value, while non-entity features include urgency expression and politeness expression. The research framework for factors influencing question response rates in social Q&A platforms is shown in Figure 1 [Figure 1: see original paper].

2.1 Question Characteristics

2.1.1 Entity Features Entity features affecting question response rates include media richness and wealth value. (1) **Media Richness**: The media through which questions are presented (text, audio, graphics, video, etc.) affects users' visual and auditory experiences [**?**]. We found that online information can be expressed through plain text, images, videos, audio, symbols, links, and their combinations. Media richness refers to the use of tables, images, and external links in question content. Content containing multiple media types enables more intuitive user understanding and stronger knowledge contribution willingness. Additionally, external links or images can influence answer quality prediction [**?**]. Therefore, we propose:

H1: The richer the media in question content, the higher the question response rate.

(2) **Wealth Value**: Wealth value is a unique indicator in social Q&A platforms. Users can obtain wealth values through check-ins and answering behaviors. Questioners can also set separate wealth values for questions, with answerers selected as best answers receiving additional corresponding wealth values [**?**]. Accumulated wealth values can be exchanged for gifts in the platform's marketplace, and increased wealth values can improve platform status and recognition. According to social exchange theory, users tend to obtain more rewards in social activities [**?**, **?**]. Paid questions receive more responses [**?**, **?**], and setting wealth values in questions attracts more user attention, potentially improving response rates. Therefore, we propose:

H2: The higher the wealth value set for a question, the higher the question response rate.

2.1.2 Non-Entity Features Non-entity features primarily refer to emotional support, a form of sharing sadness, happiness, or care that provides users with warmth, love, and help. Emotional support can positively influence the adoption of online community information services [**?**]. External emotional states can affect user decision-making [**?**], while emotional information in social Q&A platforms can enhance users' platform identification [**?**] and strengthen information adoption behavior [**?**]. Through platform investigation, non-entity features in questions mainly include urgency expression and politeness expression.

(1) **Urgency Expression**: Urgency expression refers to demands for immediate response or attention in questions. Measures include urgency statements, repeated punctuation, and repeated exclamations [?]. Users tend to use urgent words for emergency events and repetitive statements to express eagerness in message dissemination [?]. In social Q&A platforms, users often use repeated punctuation and exclamations to emphasize anxiety. When questions contain urgency expressions, the tension may infect other users, and emotions strongly influence user behavior [?], potentially prompting users to respond. Therefore, we propose:

H3: Urgency expression in questions positively influences question response rate.

(2) **Politeness Expression**: Politeness is a primary motivation for users' prosocial behavior [?]. According to social exchange theory, gratitude expression can enhance user self-efficacy and social value, thereby encouraging prosocial behavior [?]. Positive emotional information such as politeness expression can enhance users' platform identification [?] and promote participation. Simultaneously, politeness expression can influence answerers' emotions [?], and friendly attitudes can enhance their sense of achievement and willingness to communicate and respond on the platform. Therefore, we propose:

H4: Politeness expression in questions positively influences question response rate.

2.2 Questioner Characteristics

Questioner characteristics include reliability, influence, and network centrality.

(1) **Questioner Reliability**: Questioner reliability is an important factor for establishing trust and increasing influence in social networks. It can be confirmed through user background and expertise. Baidu Knows' voting mechanism and reputation system provide references for identifying questioner reliability. Information source reliability affects information acceptance—less reliable information receives lower acceptance [?], and users hope to obtain reliable knowledge [?]. Therefore, questioner reliability may affect question acceptance. We propose:

H5: Higher questioner reliability leads to higher question response rate.

(2) **Questioner Influence**: Influence is a unique indicator in Baidu Knows that comprehensively evaluates users' contribution levels and recognition. Every Baidu Knows user's personal homepage displays their influence value [?]. According to the official system, influence factors include answer count, activity level, likes received, professional certification, and cheating status. Higher user influence indicates higher activity, more interaction with other users, and greater platform status and impact [?].

Questions from more influential users are more likely to receive responses. Therefore, we propose:

H6: Higher questioner influence leads to higher question response rate.

(3) **Network Centrality**: Questioner network centrality includes out-degree centrality (number of users followed) and in-degree centrality (number of followers). Higher out-degree centrality may indicate more responding behavior; higher in-degree centrality increases question visibility [**?**]. Additionally, interaction between friends on the platform is significantly higher than between strangers, and friends are more likely to answer each other's questions [**?**]. Higher network centrality may increase the probability of receiving responses. Therefore, we propose:

H7: Higher questioner network centrality leads to higher question response rate.

3 Research Methods

3.1 Data Source

Baidu Knows topics include 14 categories such as "Economics & Finance," "Healthy Living," "Entertainment & Leisure," and "Sports." Using Python, we crawled questions under each topic, ensuring non-empty question content, removing duplicates, and supplementing with new non-duplicate questions to maintain consistent sample sizes across topics. By October 20, 2017, we had collected 10,640 questions (760 per topic). Raw data were stored in a MySQL database for preprocessing. For question data, we recorded question time, title, content, view count, wealth value, questioner nickname, questioner influence, historical answer count, and media type. For answer data, we recorded answer count, best answer content, best answer time, other answer content, other answer times, and answer media.

3.2 Prediction Method

We first used SPSS 22.0 for descriptive and correlation analysis to examine knowledge contribution behavior characteristics and factors influencing response rates from perspectives of overall resource distribution, user questioning and answering behavior, and high-quality answers. Since users can choose to respond or not to questions in social Q&A platforms, whether a question receives a response is a typical binary classification problem. Logistic regression is a typical nonlinear regression analysis method that can classify statistical results and obtain probabilistic predictions [**?**]. Binary logistic regression, where the dependent variable takes only values 0 and 1, is commonly used for binary classification problems. Therefore, we used binary logistic regression to validate influencing factors, construct a predictive model for question response rates, and verify model accuracy. Specifically, we divided the dataset into test and training sets, randomly selecting some data as the test set to verify prediction accuracy and using the remaining data as the training set to validate factors and build the prediction model.

3.3 Variable Measurement

Based on our research model and hypotheses, the study includes dependent variables (response count) and independent variables (media richness, urgency expression, etc.). For the dependent variable, questions receiving responses were coded as 1, unanswered questions as 0. Combining Baidu Knows platform characteristics, independent variables were measured as follows:

(1) **Entity Features**: For media richness, questions containing tables, images, or links were coded as 1, otherwise 0. For wealth value, we directly obtained values from the question interface, coding as 0 when no wealth value was set and using the specific value otherwise.

(2) **Non-Entity Features**: For urgency expression, we first wrote Python programs to segment question text into sentences, then used ICTCLAS for word segmentation, and automatically matched against urgency words from the CNKI emotion word list, predefined repetitive punctuation, and meaningless emotional words [**?**]. Sentences containing urgency expressions were coded as 1, otherwise 0. For politeness expression, we predefined a list of 14 related words such as "thank you," "grateful," and "repay," then matched against segmentation results. Presence of politeness expressions was coded as 1, otherwise 0.

(3) **Questioner Characteristics**: We obtained questioner influence values, answer counts, number of people helped, following count, and follower count from user profile pages. Measurements included: Questioner reliability, defined as the ratio of answer count to number of people helped; Questioner influence, categorized as 0 (value=0), 1 (0-50), 2 (50-100), 3 (100-150), 4 (150-200), and 5 (>200); Network centrality, including out-degree centrality (following count) and in-degree centrality (follower count).

4 Research Results

4.1 Descriptive Statistics

(1) **Overall Resource Distribution**: Among all samples, 10,640 questions received 34,048 answers, averaging 3.2 answers per question. The distribution of questions by answer count is shown in Figure 2 [Figure 2: see original paper]. Questions with zero answers accounted for 20.3%, while those with one answer accounted for 25.9%. Only 1.4% of questions received eight or more answers, indicating that questions obtaining large numbers of responses constitute a low proportion on Baidu Knows.

(2) **Response Time Distribution**: The overall distribution of answer times is shown in Figure 3 [Figure 3: see original paper]. Response volume

peaked between 11:00-14:00 and 21:00-23:00 daily, reaching its maximum at 21:00. These time slots likely represent non-working hours when users have time to contribute knowledge. Response counts declined significantly between 14:00-19:00 and after midnight, possibly corresponding to work and rest periods.

(3) **Question Topics**: Answer counts by topic type are shown in Table 3 , with 760 questions per topic. Healthcare, social livelihood, and psychological analysis topics had the highest response percentages at 10.37%, 12.74%, and 11.88% respectively, indicating higher user attention. Conversely, enterprise management, entertainment/leisure, and administrative regions had lower response percentages at 1.73%, 4.32%, and 4.32%. Overall, response rates vary significantly by topic type, though topic data were not used as sample data for model construction.

(4) **Client and Anonymous Behavior**: From the questioner perspective, 64.8% of questions came from mobile clients and 35.2% from computer clients. Conversely, only 38.6% of responses came from mobile clients, with 60.4% from computers. Mobile clients may be more convenient for asking questions, while answerers prefer computers for researching and composing responses. Regarding anonymity, 29.1% of questions were anonymous versus 70.9% non-anonymous, while 20.8% of answers were anonymous versus 79.2% non-anonymous. Slightly more users chose anonymous questioning, possibly to protect privacy when asking about personal or health-related sensitive information, while answerers tend to use non-anonymous status to gain recognition. Client and anonymous statistics are detailed in Table 4 .

4.2 Model Construction and Hypothesis Testing

4.2.1 Predictive Model Construction Based on the above analysis, 29.1% of users chose anonymous questioning. Since these users' information such as like counts and influence values cannot be obtained, and their question response status is unknown, anonymous data were excluded. This yielded 7,544 valid question records. We randomly selected 200 records as the test set to verify prediction accuracy, with the remaining 7,344 records used as the training set for model construction.

Using the binary logistic regression module in SPSS Statistics 22.0, we constructed the prediction model and evaluated model parameters and significance. After calculating variable values according to measurement indicators, we performed binary logistic regression to complete model training. Model variables and statistical indicators are shown in Table 5 .

4.2.2 Model Parameters and Hypothesis Testing After establishing the logistic regression model, we first tested model fit. The -2 log likelihood value (20.832) exceeds the chi-square critical value (5.991), indicating good model fit.

Additionally, Cox & Snell R-square and Nagelkerke R-square values approaching 1 indicate better fit, and results show satisfactory model fit. We then conducted omnibus tests of model coefficients, finding chi-square values for step, block, and model all exceeded the critical value (5.991) with significance far below 0.05, passing coefficient tests. Finally, we performed the Hosmer-Lemeshow goodness-of-fit test, which accepts the null hypothesis when chi-square statistics are below critical values and sig. > 0.05. Results show the model fits well overall, with no significant differences between predicted and observed values (see Table 6).

Based on these results, we tested hypotheses. Logistic regression findings show:
H1 is supported: Questions containing external links, images, or tables have higher response rates. External links, images, and tables enhance visual perception, improving user understanding efficiency and response rates [**?**].
H2 is supported: Setting wealth values increases answerer attraction and response rates. Users in social activities, including social Q&A, tend to seek more rewards [**?**, **?**], and wealth values represent such rewards, improving user recognition and response rates [**?**, **?**, **?**]. **H3 is supported**: Questions containing urgency words or symbols have higher response rates. Urgency expressions convey question immediacy, influencing other users' response behavior [**?**, **?**].
H4 is supported: Questions containing politeness words have higher response rates. Politeness expression enhances interpersonal closeness [**?**] and positively influences answerer emotions, increasing response participation [**?**, **?**]. **H5 and H6 are not supported**: Questioner reliability and influence did not significantly affect response rates. On social Q&A platforms, question pages only display user avatars and nicknames; users must visit questioner profiles to analyze influence values. Therefore, users focus more on question content than questioner reliability or influence, resulting in minimal impact on response rates.
H7 is supported: Higher questioner following and follower counts increase response rates. Higher following and follower counts indicate stronger network centrality and greater interaction with other users [**?**, **?**], increasing question visibility and corresponding response rates.

4.3 Model Validation

4.3.1 Experimental Process Based on model parameters and hypothesis testing results, we derived the following logistic regression equation, where X1 = media richness, X2 = wealth value, X3 = urgency expression, X4 = politeness expression, and X5 = network centrality:

$$P(Y=1|X)=f(X)=\frac{1}{1+e^{-g(X)}}=\frac{1}{1+e^{-(-22.797+49.593X_1+14.589X_2+16.057X_3+12.230X_4+1.021X_5)}}$$

We calculated observation probabilities for prediction: probabilities < 0.5 predicted unanswered questions, while probabilities > 0.5 predicted at least one answer. Using 200 test set records, we calculated variable values and compared actual versus predicted response status.

4.3.2 Model Evaluation After multiple iterations, regression parameters converged and stabilized, yielding final model parameters. Based on the final logistic regression model, we predicted question response rates. The prediction classification table is shown in Table 7 . From this, we constructed formulas for prediction accuracy, sensitivity, specificity, miss rate, and false alarm rate (see Table 8).

Test set calculations show a prediction accuracy of 91.1%, indicating high predictive performance for social Q&A platform response rates. The miss rate (actual answered but predicted unanswered) was 9.5%, and the false alarm rate (actual unanswered but predicted answered) was 8.3%. Overall, our model effectively predicts question response rates.

The ROC curve provides effective evaluation for binary classification problems. The ROC curve for our prediction model is shown in Figure 4 [Figure 4: see original paper], where larger area under the curve indicates better prediction performance. The area under the curve (AUC) is 0.693, significant at the 0.05 level, confirming the model's predictive capability (see Table 9).

Conclusion

Low response rates in social Q&A platforms neither effectively meet users' information needs nor promote sustainable platform development. Our study addresses this issue and provides references for improving platform information services and knowledge contribution behaviors. For platforms like "Baidu Knows," "Zhihu," and "Yahoo! Answers," predicting response rates can inform strategy development and question rate improvement. By referencing characteristics of high-response-rate questions, platforms can provide targeted suggestions to improve response rates. By identifying low-response-rate questions, platforms can enhance personalized recommendations to domain experts. Analyzing external factors affecting response rates can inform tool and service design. For users, focusing on improving reliability, influence, and network centrality while refining questioning techniques (e.g., enhancing media richness) can increase response rates.

Limitations: First, we only explored extrinsic motivations; future research could combine questionnaires and interviews to measure intrinsic motivations [**?**, **?**] and analyze interaction effects between motivations. Second, our research object, "Baidu Knows," is a first-generation social Q&A platform with weak social connections; future studies could examine second-generation platforms like Zhihu. Third, subsequent research could conduct empirical studies across different platform sections to analyze prediction mechanisms for different topics. Additionally, our dependent variable used answer count (answered/unanswered) as a binary measure; future research could consider nuanced measures distinguishing between single versus multiple answers.

References

[1] Fu Shaoxiong, Chen Xiaoyu, Deng Shengli. Research on the transformation of user information behavior in social Q&A communities: constructing a theoretical model from information adoption to continuous information seeking [J]. Library and Information Science, 2017(4): 80-88.

[2] Chen Xiaoyu, Fu Shaoxiong, Deng Shengli. Research on factors influencing user information seeking in social Q&A: a mixed-methods perspective [J]. Library and Information Service, 2018, 62(20): 102-111.

[3] ADAMIC L A, ZHANG J, BAKSHY E, et al. Knowledge sharing and yahoo answers: everyone knows something [C]//Proceedings of the 17th international conference on world wide web. New York: ACM, 2008: 665-674.

[4] RAFAELI S, RABAN D, RAVID G. How social motivation enhances economic activity and incentives in the google answers knowledge sharing market [J]. Social science electronic publishing, 2007, 3(1): 1-11.

[5] CHOI E, KITZIE V, SHAH C. "10 Points for the best answer!"? baiting for explicating knowledge contributions in online Q&A [C]//Proceedings of the 76th ASIS&T annual meeting: beyond the cloud: rethinking information boundaries. Montreal: Wiley, 2013: 103-106.

[6] Xu Peng, Zhang Dan. Exploring knowledge sharing motivation in online Q&A communities: from the perspective of social exchange theory [J]. Library and Information Knowledge, 2018(2): 105-112.

[7] Chen Xing, Zhang Xing, Zeng Shuyun. Research on factors influencing knowledge sharing willingness in health Q&A platforms [J]. Modern Information, 2017, 37(4): 62-71.

[8] OH S. The characteristics and motivations of health answerers for sharing information, knowledge, and experiences in online environments [J]. Journal of the Association for Information Science & Technology, 2012, 63(3): 543-557.

[9] DEARMAN D, TRUONG K N. Why users of Yahoo! Answers do not answer questions [C]//Proceedings of the 2010 SIGCHI conference on human factors in computing systems. New York: ACM, 2010: 329-332.

[10] NAM K K, ACKERMAN M S, ADAMIC L A. Questions in, knowledge in: a study of naver's question answering community [C]//Proceedings of the 2009 international conference on human factors in computing systems. New York: ACM, 2009: 779-788.

[11] DU Q. A relationship-based social question and answer system in social network [J]. Journal of information & computational science, 2015, 12(10): 3783-3798.

[12] Xu Anying, Ji Zongcheng, Wang Bin. Research on predicting answer quality based on user answer order [J]. Journal of Chinese Information Processing, 2017,

31(2): 132-138.

[13] JEON J, CROFT W B, LEE J H, et al. A framework to predict the quality of answers with non-textual features [C]//Proceedings of the 29th annual international ACM SIGIR conference on research and development in information retrieval. New York: ACM, 2006: 228-235.

[14] Tian Lei, Ren Guoheng, Wang Wei. Predicting microblog user forwarding behavior for reading promotion [J]. Modern Library and Information Technology, 2017, 1(8): 48-58.

[15] Zhang Pengyi, Wang Danxue, Jiao Yifan, et al. Predicting mobile purchase behavior based on user browsing logs [J]. Data Analysis and Knowledge Discovery, 2018, 2(1): 51-63.

[16] Zhang Hongli, Liu Jiying, Yang Sinan, et al. Rating prediction based on online user reviews [J]. Data Analysis and Knowledge Discovery, 2018, 2(1): 51-63.

[17] Deng Shengli, Fu Shaoxiong. Analysis of online rumor characteristics and prediction model design: based on user trust perspective [J]. Information Science, 2017, 35(12): 28-33.

[18] Chen Jun, He Mengting. Research on instant/continuous online word-of-mouth in virtual communities from a motivational perspective [J]. Information Science, 2017, 35(12): 28-33.

[19] Chen Shu, Dou Yongxiang, Zhang Qingjie. Research on factors influencing microblog user forwarding behavior based on theory of reasoned action [J]. Information Magazine, 2017, 36(11): 147-152.

[20] An Lu, Yi Xingyue, Yu Chuanming, et al. Prediction research on microblog influence during public health emergencies [J]. Information Theory and Practice, 2017, 40(8): 76-81.

[21] Wang Linsen, Wang Xueyi. Research on microblog introverted transmission hotspot discovery and prediction algorithms [J]. Information Science, 2017(11): 126-131.

[22] Hu Yue, Wang Yamin. Microblog public opinion trend prediction method based on fuzzy neural network [J]. Library and Information Service, 2018, 62(3): 71-77.

[23] AGICHTEIN E, CASTILLO C, DONATO D. Finding high-quality content in social media [C]//Proceedings of the 2008 international conference on web search and data mining. New York: ACM, 2008: 183-194.

[24] FLEISCHMANN M, BLOEMHOF-RUWAARD J M, DEKKER B R, et al. Quantitative models for reverse logistics: a review [J]. European journal of operational research, 1997, 103(1): 1-17.

[25] MCMILLAN S J. Effects of structural and perceptual factors on attitudes toward the website [J]. Journal of advertising research, 2004, 43(4): 400-421.

[26] TIAN Q, ZHANG P, LI B. Towards predicting the best answers in community-based question-answering services [C]//Proceedings of the seventh international AAAI conference on weblogs and social media. Palo Alto: AAAI Press, 2013: 725-728.

[27] Baidu Knows. Wealth value [EB/OL]. [2018-07-30]. http://help.baidu.com/question?prod_{id}=9&class=

[28] HOMANS G C. Social behavior as exchange [J]. American journal of sociology, 1958, 63(6): 597-606.

[29] HSIEH G, KRAUT R E, HUDSON S E. Why pay?: exploring how financial incentives are used for question & answer [C]//Proceedings of the 2010 SIGCHI conference on human factors in computing systems. Atlanta: ACM, 2010: 305-314.

[30] Wu Jiang, Li Shanshan. Research on online health community user information service usage intention [J]. Information Science, 2017(4): 119-125.

[31] Xu Jian. Research on prediction methods based on online user sentiment analysis [J]. Journal of Library Science in China, 2013, 39(3): 96-107.

[32] JOYCE E, KRAUT R E. Predicting continued participation in newsgroups [J]. Journal of computer-mediated communication, 2006, 11(3): 723-747.

[33] Jin Jiahua. Research on factors influencing user knowledge behavior in social Q&A platforms [J]. Journal of the China Society for Scientific and Technical Information, 2017(11): 1175-1182.

[34] BALDWIN C L, MAY F. Verbal collision avoidance messages of varying perceived urgency reduce crashes in high-risk scenarios [C]//Proceedings of the third international driving symposium on human factors in driver assessment, training and vehicle design. Rockport: Public Policy Center of University of Iowa, 2017: 128-134.

[35] Shen Wang, Guo Jia, Li He. Research on online community information quality and reliability evaluation from user perspective [J]. Data Analysis and Knowledge Discovery, 2013, 29(1): 69-76.

[36] HEIM J. Why people use social networking sites [C]//Proceedings of the 3rd international conference on online communities and social computing. Berlin: Springer, 2009: 143-152.

[37] Baidu Knows. User influence [EB/OL]. [2018-07-30]. https://jingyan.baidu.com/article/84b4f5659102b4f6d

[38] Peng Lihui, Li He, Zhang Yanfeng. Research on online public opinion leader identification and influence ranking based on grey relational analysis: taking the "8 · 12 Binhai explosion" on Sina Weibo as an example [J]. Information Theory and Practice, 2017, 40(9): 90-94.

[39] WENG J, LIM E P, JIANG J, et al. TwitterRank: finding topic-sensitive influential twitterers [C]//Proceedings of the third ACM international conference on web search and data mining. New York: ACM, 2010: 261-270.

[40] CNKI. Emotion analysis word set (beta version) [EB/OL]. [2018-07-30]. http://www.keenage.com/html/c_{index}.html.

[41] Deng Shengli, Fu Shaoxiong. Analysis of the impact of social media additional information on user trust and sharing of health rumors [J]. Information Science, 2018, 36(3): 51-57.

[42] Fu Shaoxiong, Hu Yuan. Research on the impact of college students' health information behavior on actual health levels: from health literacy and health information seeking perspectives [J]. Modern Information, 2018, 38(2): 84-90, 105.

Author Contributions: Deng Shengli: conceptualized the research, revised the final manuscript; Fu Shaoxiong: designed the research framework, drafted and revised the manuscript; Liu Jin: collected and analyzed data.

The Prediction Research of Response Rate in Social Q&A Communities: A Case Study of Baidu Knows

Deng Shengli, Fu Shaoxiong, Liu Jin

Center for Studies of Information Resources, Wuhan University, Wuhan 430072

Abstract: [Purpose/significance] Based on the current situation of low response rate of social Q&A communities, the research can provide references for social Q&A communities to improve user activation, retention rate and user experience by predicting the response rate of questions. [Method/process] The paper took "Baidu Know" as the research platform, and grabbed 10640 question records under 14 topics set by the platform. From the perspective of question and questioner characteristics, the paper constructed the research framework of the factors affecting the question response rate. The binary logistic regression was used to verify the influencing factors, and then the prediction model of the question response rate was constructed. [Result/conclusion] The prediction research of response rate in social Q&A communities can improve the quality of platform information services and promote user knowledge contribution behavior. The experimental results have verified the validity of the model in the prediction of question response rate of the social Q&A communities.

Keywords: social Q&A community; knowledge contribution behavior; response rate; logistic regression; prediction

Note: Figure translations are in progress. See original paper for figures.