Winter 12-30-2021

How perceived personalities of earlier contributors influence the content generation on online knowledge-sharing platforms?

Yu Feng

Cong Wang

Qiaowei Shen

How perceived personalities of earlier contributors influence the content generation on online knowledge-sharing platforms?

Complete Research

Yu Feng
Peking University
1801110964@gsm.pku.edu.cn

Cong Wang
Peking University
wangcong@gsm.pku.edu.cn

Qiaowei Shen
Peking University
qshen@gsm.pku.edu.cn

Extended Abstract

It is common to see that similar words spoken by different people will generate different impacts on others. Although they contribute the same knowledge, other people may perceive it and respond to it differently. One important implication is in the knowledge-sharing Peer-to-Peer (P2P) platforms, where solving a problem needs the wisdom of the crowd. Such platforms are increasingly important with a growing passion of people on knowledge acquisition (Forbes, 2020, 2021). One typical example is the online Peer-to-Peer (P2P) Q&A platforms, like the Quora in U.S. and Zhihu in China, where individuals can exchange their knowledge through asking and answering questions. They have successfully built large growing knowledge repositories, valued at USD 281 million in 2019 for the whole global market and projected to reach USD 520 million by 2027 (Verified Market Research, 2019).

One critical issue such platform facing is to keep contributors engaged (Kraut and Resnick, 2012; Ray et al., 2014), which is important to the platform's market value and social impact (Berger and Milkman, 2012; Susarla and Tan, 2012; Zhou and Alstyne, 2019). In online P2P Q&A platforms, one important engagement behavior is content generation (Kraut and Resnick, 2012; Ray et al., 2014), i.e., the decision of whether to contribute and the new content generation.

People engaging in knowledge-sharing activities to contribute their knowledge for free are driven by some intrinsic factors, e.g., reciprocity (Zhou et al., 2011; Chang and Chuang, 2011), as well as other social influence factors like existing contributors and contents. In general, contributors are affected more by earlier contributors in terms of their characteristics and the content they generated before (Mudambi and Schuff, 2010; Chang and Chuang, 2011; Milkman and Berger, 2014; Yin et al., 2014; Zhao et al., 2018; Lee et al., 2019; Peng et al., 2020; Al-Ramahi and Alsmadi, 2021). The linguistic choice is informative in addition to the objective value of the content. The choice of words, the structure of sentences one adopts to formulate a question, writing answers, give clues to her emotional, and cognitive states (Funder, 2012; Maity et al., 2018) as well as her ability (Gorbaniuk et al., 2015). The perception generated from one's language choice is among the influencing factors of other contributors' choices (Maity et al., 2018). In this paper, we focus on how earlier contributors influence later ones' content generation through their linguistic choice.

Contributors start a conversation as questioners, share their knowledge voluntarily as responders, critics and collectors in different contexts (Li and Bernoff, 2008). Although they may use different language under different contexts, e.g., different topics, audience groups and emotions (Pillutla and Chen, 1999; Dolan, 2002; Charness et al., 2007; Chen and Bond, 2010; Qiu and Kumar, 2017), there exists inherent language habits can be extracted to infer one's traits, e.g., the personality (Hirsh and Peterson, 2009; Pennebaker and King, 1999; Gupta, 2008; Yarkoni, 2010; Liu et al., 2020). According to psychology literature, personality plays a central role in describing a person and is considered to be the most fundamental dimension of variation among humans (McCrae and John, 1992). We can also use the language to predict one's self-reported personality with the help of the machine learning method (Agarwal and Karahanna, 2000; Gupta, 2008; Fast and Funder, 2008; Poropat, 2009; Hirsh and Peterson, 2009).

Despite that some factors may impede others to identifying one's personality, e.g., economic status and relationship (Youyou et al., 2015; Leckelt et al., 2018), people still respond based on their judgments on one's personality, i.e., personality perception, in social livings (Letzering, 2008; Funder, 2012). The personality perception is proven to be the antecedent of various outcomes like political achievements (Gorbaniuk et al., 2015), fraud behavior (Netzer et al., 2019), livestreaming popularity (Zhao et al., 2019). Thus, we propose our research question: How perceived personalities of earlier contributors influence the later content generation through different linguistic choices? Following previous literature, we use Big Five personality traits as the framework of personality (Goldberg, 1992; Pennebaker and King 1999; Pennebaker and Graybeal 2001; Kosinski et al., 2013).

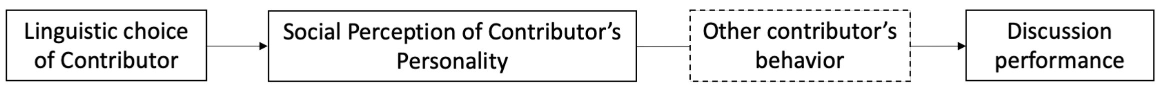

Figure1. Theoretical Framework

To answer our research question, we apply text-mining and machine learning methods to a data set of 4,056 questions and 51,573 answers from a professional and open-to-public forum DingXiangYuan (DXY) in medical science. It boasts about 5.5 million professional users and provides academic discussions, continuing medical education and job recruitment services for medical professionals (China Daily, 2021).

To obtain the social perception of contributors' personalities, we develop a new method combining natural language processing techniques (NLP) and the unsupervised learning method to predict one's perceived personality based on linguistic activities. To verify our method above, we run an online experiment to test if the algorithm result is consistent with people's personality perception by others, instead of the self-reported personality. We hire participants to write essays and apply our algorithm to make an inference. We ask participants to self-report personality using two scales (Saucier, 1994; Lu, 2020) and hire 5 raters to evaluate participants' personalities after reading the essays. In the end, we run a correlation analysis between the algorithm results and two kinds of personality traits. The results confirm the efficiency of our algorithm on inferring social perception of one's personality.

With the help of the proposed algorithm, we explore how earlier contributors' personality perception influences the content generation of later contributors on whether to join the discussion and the degree of new content generation. The individual choice determines the final participation size, discussion duration, popularity and new content generation (our focal dependent variables) of each question. Considering other confounding factors which may influence the discussion performance during the time window, we only estimate the effect of the first contributor (starter)'s perceived personalities of each question.

We arrive at several interesting findings. Firstly, starters with high conscientiousness will increase the total discussion size, popularity, and new content generation of the question. Secondly, a starter with high openness is found to play a contrary role, i.e., she will decrease the discussion size, duration, and new content generation. Thirdly, the starter's real identity in the offline world exhibits a significant effect on discussion performance. It shows a negative effect as more advanced starters reduce others' motivation to express their opinion and end the question quickly. Perceived impact (proxied by fans size) of starter also shows a negative avoidance effect on motivation to join the discussion. The mechanism behind this is the trade-off between persuasiveness of existing answers and exploration space remained or created by earlier contributors.

Our research contributes to related literature and real-world practice. Theoretically, we investigate contributor's engagement choice from the perspective of social influence, unpacking the path of how contributors influence each other through personality perception backed by linguistic choice instead of the objective value of the content, and its downstream consequences on content creation, complementing the literature of social engagement in content generation platforms. We highlight the important role of contributors' perceived personality on the online knowledge-sharing platform rather than self-reported personality, which can be widely applied to other content generation platforms. Among the five personalities, we figure out the importance of perceived openness and conscientiousness on discussion performance, enriching the literature of personality perception. Practically, our findings help platforms assign questions to the correct contributor. If the question is in urgent need, a starter with high openness will be more appropriate. If the question is an academic discussion, a starter with high consciousness will be more helpful. We also develop a new method to detect users' personality perception using their existing content, which is much more convenient and efficient for platform governors. Our algorithm requires less historical information, which helps the platform better acknowledge and manage new users or inactive users. The unsupervised learning method is good at exploratory analysis and it can be applied to other related scenarios to help platforms enrich their user's labels.

Reference

Agarwal, R., & Karahanna, E. (2000). Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. MIS Quarterly, 665-694.

Al-Ramahi, M., & Alsmadi, I. (2021). Classifying insincere questions on Question Answering (QA) websites: meta-textual features and word embedding. Journal of Business Analytics, 4(1), 55-66.

Berger, J., & Milkman, K. L. (2012). What makes online content viral?. Journal of marketing research, 49(2), 192-205.

Chang, H. H., & Chuang, S. S. (2011). Social capital and individual motivations on knowledge sharing: Participant involvement as a moderator. Information & management, 48(1), 9-18.

Charness, G., Rigotti, L., & Rustichini, A. (2007). Individual behavior and group membership. American Economic Review, 97(4), 1340-1352.

Chen SX, Bond MH. Two Languages, Two Personalities? Examining Language Effects on the Expression of Personality in a Bilingual Context. Personality and Social Psychology Bulletin. 2010;36(11):1514-1528.

China Daily (2021). http://global.chinadaily.com.cn/a/202101/06/WS5ff518b1a31024ad0baa0c45.html

Dolan, R. J. (2002). Emotion, cognition, and behavior. science, 298(5596), 1191-1194.

Fast, L. A., & Funder, D. C. (2008). Personality as manifest in word use: correlations with self-report, acquaintance report, and behavior. Journal of personality and social psychology, 94(2), 334.

Forbes (2020). https://www.forbes.com/sites/johnkoetsier/2020/09/26/global-online-content-consumption-doubled-in-2020/?sh=203d0d362fde

Forbes (2021). https://www.forbes.com/sites/quora/2021/07/22/how-the-pandemic-has-accelerated-demand-for-online-education/?sh=b70c19b47747

Funder, D. C. (2012). Accurate personality judgment. Current Directions in Psychological Science, 21(3), 177-182.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. Psychological assessment, 4(1), 26.

Gorbaniuk, O., Kusak, K., Kogut, A., & Kustos, M. (2015). Dimensions of political party "personality" perception. Journal of Political Marketing, 14(1-2), 35-63.

Gupta, B. (2008). Role of personality in knowledge sharing and knowledge acquisition behavior. Journal of the Indian Academy of Applied Psychology, 34(1), 143-149.

Hirsh, J. B., & Peterson, J. B. (2009). Personality and language use in self-narratives. Journal of research in personality, 43(3), 524-527.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. Proceedings of the national academy of sciences, 110(15), 5802-5805.

Kraut, R. E., & Resnick, P. (2012). Building successful online communities: Evidence-based social design. MIT Press.

Leckelt, M., Richter, D., Schröder, C., Küfner, A. C., Grabka, M. M., & Back, M. D. (2019). The rich are different: Unravelling the perceived and self-reported personality profiles of high-net-worth individuals. British Journal of Psychology, 110(4), 769-789.

Lee, S. Y., Rui, H., & Whinston, A. B. (2019). Is best answer really the best answer? The politeness bias. MIS Quarterly, 43(2), 579-600.

Letzring, T. D. (2008). The good judge of personality: Characteristics, behaviors, and observer accuracy. Journal of research in personality, 42(4), 914-932.

Li, C., & Bernoff, J. (2008). Groundswell: Winning in a world transformed by social technologies. Harvard Business School Pr.

Lu, J. G., Liu, X. L., Liao, H., & Wang, L. (2020). Disentangling stereotypes from social reality: Astrological stereotypes and discrimination in China. Journal of Personality and Social Psychology, 119(6), 1359.

Maity, S. K., Kharb, A., & Mukherjee, A. (2018). Analyzing the linguistic structure of question texts to characterize answerability in quora. IEEE Transactions on Computational Social Systems, 5(3), 816-828.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. Journal of personality, 60(2), 175-215.

Milkman, K. L., & Berger, J. (2014). The science of sharing and the sharing of science. Proceedings of the National Academy of Sciences, 111(Supplement 4), 13642-13649.

Mudambi, S. M., & Schuff, D. (2010). Research note: What makes a helpful online review? A study of customer reviews on Amazon. com. MIS Quarterly, 185-200.

Netzer, O., Lemaire, A., & Herzenstein, M. (2019). When words sweat: Identifying signals for loan default in the text of loan applications. Journal of Marketing Research, 56(6), 960-980.

Peng, C. H., Yin, D., & Zhang, H. (2020). More than Words in Medical Question-and-Answer Sites: A Content-Context Congruence Perspective. Information systems research, 31(3), 913-928.

Pennebaker, J. W., & Graybeal, A. (2001). Patterns of natural language use: Disclosure, personality, and social integration. Current Directions in Psychological Science, 10(3), 90-93.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: language use as an individual difference. Journal of personality and social psychology, 77(6), 1296.

Pillutla, M. M., & Chen, X. P. (1999). Social norms and cooperation in social dilemmas: The effects of context and feedback. Organizational behavior and human decision processes, 78(2), 81-103.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. Psychological Bulletin, 135(2), 322.

Qiu, L., & Kumar, S. (2017). Understanding voluntary knowledge provision and content contribution through a social-media-based prediction market: A field experiment. Information Systems Research, 28(3), 529-546.

Ray, S., Kim, S. S., & Morris, J. G. (2014). The central role of engagement in online communities. Information Systems Research, 25(3), 528-546.

Saucier, G. (1994). Separating description and evaluation in the structure of personality attributes. Journal of Personality and Social Psychology, 66(1), 141.

Susarla, A., Oh, J. H., & Tan, Y. (2012). Social networks and the diffusion of user-generated content: Evidence from YouTube. Information systems research, 23(1), 23-41.

Xia Liu, A., Li, Y., & Xu, S. X. (2021). ASSESSING THE UNACQUAINTED: INFERRED REVIEWER PERSONALITY AND REVIEW HELPFULNESS. MIS Quarterly, 45(3).

Verified Market Research (2019). https://www.verifiedmarketresearch.com/product/qa-platforms-market

Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. Journal of research in personality, 44(3), 363-373.

Yin, D., Bond, S. D., & Zhang, H. (2014). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. MIS Quarterly, 38(2), 539-560.

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. Proceedings of the National Academy of Sciences, 112(4), 1036-1040.

Zhao, K., Hu, Y., Hong, Y., & Westland, J. C. (2019). Understanding Characteristics of Popular Streamers on Live Streaming Platforms: Evidence from Twitch. tv. Journal of the Association for Information Systems, Forthcoming.

Zhou, T. (2011). Understanding online community user participation: a social influence perspective. Internet research.

Zhou, Z., & Van Alstyne, M. (2019). Platform Valuation: Profit or User Base?. working paper.