5-28-2024

Effect of Knowledge sharing on Participant Performance in Bigdata Science Contests: A Quasi-experiment

Lingling Wang
School of Economics and Management, China University of Petroleum (East China), Qingdao 266580, China, linglingwang@upc.edu.cn

Sen Li
School of Economics and Management, Shandong Youth University of Political Science, Jinan 250014, China

Meijian Yang
School of Economics and Management, Anhui Polytechnic University, Wuhu 241000, China

Effect of Knowledge sharing on Participant Performance in Bigdata Science Contests: A Quasi-experiment

Lingling Wang[1] and Sen Li[2] and Meijian Yang[3]

[1] School of Economics and Management, China University of Petroleum (East China), Qingdao 266580, China

[2] School of Economics and Management, Shandong Youth University of Political Science, Jinan 250014, China

[3] School of Economics and Management, Anhui Polytechnic University, Wuhu 241000, China

linglingwang@upc.edu.cn (Lingling Wang,corresponding author)

1. INTRODUCTION AND RESEARCH QUESTIONS

Advancements in digital technology enable the storage, processing, and analysis of large-scale open data. Big data science contests have become the crucial paths of artificial intelligence and digital technology industrialization. How to efficiently utilize open data and shared knowledge to better serve enterprises' digital innovation has become a hot topic in the research field of cross-integration of ma-chine learning and management, especially in the context of digital development and the construction of a 'digital China.'

Within the big data science community, shared knowledge has both positive and negative effects on participants' innovative behavior. Knowledge exchanges during competitions can lead to more innovative solutions, while the presence of shared solutions can restrict contestants' innovative ideas. Dealing with shared knowledge during a competition can incur opportunity costs, dispersing contestants' focus and impacting their performance. However, the submission of diversified solutions can improve the quality of competition results due to the parallel path effect.

Existing research primarily focuses on knowledge-sharing mechanisms in collaborative communities, such as Wikipedia or open-source projects. However, in big data science competitions, knowledge sharing is influenced by the "cooperation-competition" mechanism, resulting in distinct degrees and behaviors of influence. Competition levels can affect contestants' efforts and performance, making the quality of their solutions dependent on the inherently competitive nature of the competition.

While many platforms have features for participants to share knowledge, research specifically exploring knowledge-sharing features in machine learning competition platforms is scarce. Previous studies have shown that the existence of certain knowledge-sharing features can impact contestants' performance (Liu and Kim 2023). To fill this gap, this study investigates the impact of the Code knowledge-sharing feature on competition performance using a quasi-experimental design and the PSM-DID model. The study aims to answer the following research questions:

Q1. Does knowledge-sharing affect participants' competition performance in the context of "competition-cooperation"?

Q2. Does the influence of the knowledge-sharing feature vary among different types of participants?

2. THEORY AND RESEARCH MODEL

2.1 Theoretical background and Hypotheses

Research suggests that knowledge sharing in online communities can enhance creativity through iterative cognitive processes such as posting, commenting, and polling. Competition platforms with dedicated discussion boards facilitate communication among participants and provide timely feedback. When these platforms incorporate knowledge-sharing features, it expedites information and knowledge acquisition, reducing barriers and costs of participation. Knowledge sharing encourages collaboration, reevaluation of ideas, and the

generation of creative solutions. Collaborative efforts in competitions further enhance participants' performance(Hutter et al. 2011; Jin et al. 2021a). Combining competition and cooperation promotes innovative output beyond individual cooperation or competition alone. Therefore, we propose:

H1. Knowledge sharing in Bigdata contests positively impacts participants' performance.

While implementing knowledge sharing in bigdata competitions generally benefits participants, disparities in backgrounds and knowledge levels can lead to varied impacts. Misinterpretations or misuses of shared knowledge can negatively affect efficiency and nullify its benefits. Therefore, further study is needed to understand the diverse effects on different participant types. Thus, we propose:

H2. Knowledge sharing has heterogeneity in participants' performance in competitions.

Specifically, the convenience of knowledge sharing enhances the creativity of individual participants, especially those disadvantaged by a lack of internal team communication. Therefore, we propose:

H2a. The knowledge-sharing function has a more significant influence on individual participants compared to team participants.

Newcomers actively engage in knowledge-sharing processes to acquire professional expertise and emulate existing solutions. Studies suggest that active knowledge-sharing leads to better task completion. Therefore, we propose:

H2b. Participants with less competition experience are more significantly influenced by the knowledge-sharing function.

H2c. Participants with shorter tenure are more significantly influenced by the knowledge-sharing function.

Team members utilize knowledge sharing to learn from others and improve their solutions through peer feedback. Valuable feedback can elevate contest performance for teams with superior solutions. Therefore, we propose:

H2d. Participants with a higher number of submissions are more significantly influenced by the knowledge-sharing function.

2.2 Quasi-experiment Design

New functions or services introduced on digital platforms or virtual communities are often treated as quasi-experimental scenarios. This study is designed as a quasi-experiment based on the introduction of the Code knowledge sharing function on the Kaggle platform on November 7, 2019 (As shown in Figure 1).

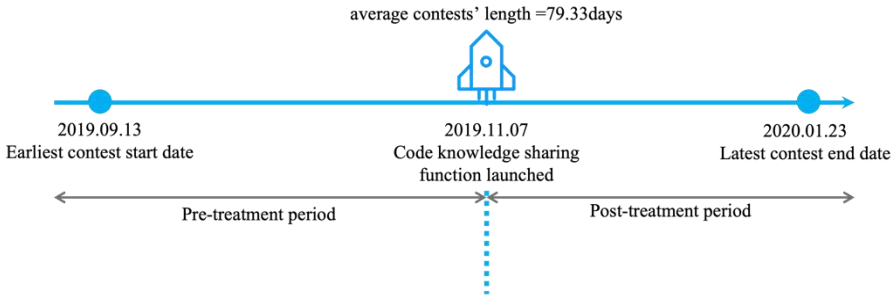

Figure 1. Timeline diagram of the quasi-experiment

To investigate how knowledge sharing pertains to participants' performance in bigdata competitions, we employ a PSM-DID approach. We construct a two-way fixed effects model to analyze panel data, set as:

$$Stdscore_{ijt} = \beta_0 + \beta_1 post_{jt} \times treat_{ij} + \beta_2 treat_{ij} + \beta_3 post_{jt} + \lambda X' + u_{ij} + v_{jt} + \varepsilon_{ijt} \quad (1)$$

3. MATERIALS, RESULTS AND MAJOR FINDINGS

To carry out the study, we crawl a unique dataset with 2477 participants of machine learning contests from

the Kaggle platform. The findings reveal that the launch of the knowledge-sharing feature significantly improves overall competition performance, offering preliminary support to hypothesis H1. However, there is a time lag effect, with the performance initially affected positively, followed by a non-significant effect, and then a significant positive effect from the fourth week onwards.

Notably, the prerequisite for utilizing the DID method is that it satisfies the parallel trends assumption, that is, there is no significant difference between the treatment group and the control group before policy effects. We use the event study to test the parallel trends assumption. The results of the parallel trend test show that the coefficients of the virtual w-3 and w-2 weeks are not significant at the 5% statistical level, which indicates that the changing trend of the treatment group and the control group before the launch of the knowledge sharing function is consistent, the parallel trend assumption is satisfied, and the benchmarking analysis results are robust. Further, the placebo test is used to verify the robustness of DID results.

Then, we conduct the heterogeneity analysis of team size, tenure, numbers of submissions, and competition experience. The results show that hypothesis H2 is supported, that is, the launch of the knowledge-sharing feature has heterogeneity in participant performance; competition experience positively moderates the impact of knowledge sharing on the competitors' performance, not supporting Hypothesis H2b; the launch of the knowledge-sharing function has a more significant impact on the com-petition performance of competitors with longer tenure, not supporting hypothesis H2c; the knowledge-sharing function negatively affects the competition performance of competitors who make more submissions, not supporting Hypothesis H2d.

4. MAIN CONTRIBUTIONS

These results provide insights for big data science competition platforms on leveraging the knowledge-sharing feature effectively. This includes establishing a dynamic reputation mechanism to encourage cooperation and sharing among participants with extensive experience and tenure. For newcomers and less-experienced participants, the platform may consider implementing other features such as beginner competitions, dataset and course training, or unlocking difficult tasks after completing beginner ones.

ACKNOWLEDGEMENT

This study was supported by The Ministry of Education of China Youth Project of Humanities and Social Science project under Grant 23YJCZH216, Shandong Postdoctoral Science Foundation under Grant SDBX2023089, and Qingdao Postdoctoral Foundation Project under Grant QDBSH20230102070.

REFERENCES

1. Du Shuili, Babalola Mayowa, D'Cruz Premilla, Doci Edina, Garcia-Lorenzo Lucia, Hassan Louise, Islam Gazi, Newman Alex, Noronha Ernesto, van Gils Suzanne. The Ethical, Societal, and Global Implications of Crowdsourcing Research. Journal of Business Ethics 2024: 1-16.
2. Hutter Katja, Hautz Julia, Füller Johann, Mueller Julia, Matzler Kurt. Communitition: The tension between competition and collaboration in community-based design contests. Creativity and innovation management 2011; 20 (1): 3-21.
3. Jin Yuan, Lee Ho Cheung Brian, Ba Sulin, Stallaert Jan. Winning by Learning? Effect of Knowledge Sharing in Crowdsourcing Contests. 2021b; 32 (3): 836-859.
4. Liu Jialu, Kim Keehyung. Designing contests for data science competitions: Number of stages and the prize structures. Production and Operations Management 2023; 32 (11): 3752-3772.
5. Majchrzak Ann, Malhotra Arvind. Effect of knowledge-sharing trajectories on innovative outcomes in temporary online crowds. Information Systems Research 2016; 27 (4): 685-703.
6. Wooten Joel O. Leaps in innovation and the Bannister effect in contests. Production and Operations Management 2022; 31 (6): 2646-2663.