What Sustains Individuals’ Participation in Crowdsourcing Contests?

Xuan Wang  
The University of Texas Rio Grande Valley  
xuan.wang@utrgv.edu

Hanieh Javadi Khasraghi  
University of Delaware  
hanieh@udel.edu

Helmut Schneider  
Louisiana State University  
hschnei@lsu.edu

Abstract

Crowdsourcing contests have become widely adopted for idea generation and problem-solving in various companies in different industries. The success of crowdsourcing depends on the sustained participation and quality-submissions of the individuals. Yet, little is known about the factors that influence individuals’ continued participation in these contests. We address this issue, by conducting an empirical study using data from an online crowdsourcing contest platform, Kaggle, which delivers data science and machine learning solutions and models to its clients. The findings show that the community activities and team activities do not contribute to motivating the continued participation, but tenure does significantly affect the continued participation. We also found statistically significant effects of amount of prize, number of competitions, previous team performance, and competition duration on individuals sustained participation in crowdsourcing contests. This research contributes to the literature by identifying the factors influencing individuals’ sustained participation in crowdsourcing contests.

1. Introduction

Crowdsourcing has hundreds of years of history which started with the British Government’s “The Longitude Problem” in 1714 [1]. There are other examples of Crowdsourcing in the history such as Oxford English Dictionary’s “cataloging words by the crowd” in 1884, Toyota’s “Logo Contest” in 1936, and The Sydney Opera House’s “Architectural Contest” in 1955. In all of these examples, crowds of people have been used to solve a problem. The idea of crowdsourcing has been existed for a long time, but its usage increased after the evolution of Web 2.0 and Web 3.0 technologies. Web 2.0 enables organizations to have access to a large-scale workforces in order to use the power of the crowd to get their tasks done [2]. In the early 2000s, collective intelligence started to gain recognition. During this period, a lot of processes, not yet termed crowdsourcing, launched that harnessed the efforts of a crowd of people for various tasks from innovation to implementation. Examples include iStockphoto and Threadless in 2000, InnoCentive and TopCoder in 2001, Amazon Mechanical Turk and Kodak’s “Go for the Gold” contest in 2005.

In 2006, Jeff Howe, the editor at Wired magazine, coined the term “crowdsourcing” in his article “The Rise of Crowdsourcing” that revolutionized the idea of crowdsourcing. Their definitions of crowdsourcing is as follows:

“Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call”[3].

Various crowdsourcing mechanisms are being used by companies [4], [5]. Some crowdsourcing platforms, such as Amazon Mechanical Turk, allow an individual to be the only provider of the solution [6], some crowdsourcing platforms, such as TopCoder, Kaggle, and TaskCn, are structured as contests to allow more people to provide solutions. In crowdsourcing contests, any user can submit solutions to the task, but the participant who has provided the solution of the highest quality is awarded [7], [8]. Some of the crowdsourcing platforms have both collaborative and competitive elements [9]. These platforms allow the individuals to simultaneously collaborate and compete with each other.

Companies are increasingly using crowdsourcing contests for solving problems, yet the success and sustainment of these crowdsourcing contests depend on individuals’ continued participation and high quality submissions. Previous research on crowdsourcing has paid considerable attention to crowdsourcing contests and more specifically to individuals’ behavior within these contests. This stream of research mainly focuses on identifying the factors that motivate individuals to join the competition and the factors that affect their performance in these competitions. Yet, very few studies have investigated individuals’ sustained participation in these platforms. None of these studies have examined how individuals’ prior participation
experience in competitions and their knowledge sharing activities in the community affect their continued participation in crowdsourcing contests. This study aims to address this research gap by using data collected from Kaggle.com, a Web-based platform which delivers data science solutions and models to its clients through problem solving contests. Kaggle platform has both collaborative and competitive environment in which individuals can team up and compete against the other teams in the competitions. Moreover, individuals within a team can communicate and share knowledge with the other members in the community. This simultaneous collaboration and competition environment makes Kaggle unique among the other platforms.

The objective of this study is to determine the factors affecting individuals’ sustained participation in crowdsourcing contests. The results of this study point to the important factors affecting individuals’ continued participation in the crowdsourcing contests. The motivation for individuals’ sustained participation in the contests is different from that of the community. The findings show that community activities and team activities do not contribute to motivating individuals’ continued participation, but tenure is a significant factor that affects individuals’ sustained participation. The remaining of the paper is organized as follows: section 2 contains the literature review for individuals’ participation behavior in crowdsourcing; section 3 illustrates the theoretical background and develops the hypotheses; section 4 explains the data collection and variable measurement in this study; section 5 applies a negative binomial model and ordinal logistic regression for comparing and enhancing the findings that we obtained in this study; section 6 provides a conclusions and implications for helping crowdsourcing platform sponsors to design crowdsourcing contests in a way to facilitate crowdsourcing processes and motivate individuals to participate.

2. Literature

Previous research in crowdsourcing investigated three different components of crowdsourcing process: requesters, crowdsourcing platforms, and solvers. Some research in crowdsourcing focuses on the requesters’ and crowdsourcing sponsors’ attitudes and behavior toward crowdsourcing, including: the motivation for crowdsourcing, the crowdsourcing mechanism, organizing the crowdsourcing process, the types of the tasks to crowdsource, the strategy for choosing the best solution, and the quality assurance of the solutions [10], [11]. Another stream of research focus on the individuals’ (solvers) attitudes and behavior in crowdsourcing platforms including: the motivations for initial participation, the motivations for continued participation, and the factors affecting their performance [7], [10]–[12]. Understanding the solvers’ (or individuals) behavior is very important since it can help crowdsourcing sponsors to use appropriate mechanisms and strategies to build successful crowdsourcing platforms.

Many research studies have focused on individuals’ motivational factors for participation in crowdsourcing. Djelassi and Decoopman stated that the type of incentive depends on the type and mechanisms of crowdsourcing [13]. Previous research has applied theoretical lenses from various reference disciplines. The most notable theoretical lens is motivation theory (classic motivation theory and work motivation theory) [14]. Most of these studies drawing on motivation theory investigated the intrinsic and extrinsic motivations for participation in crowdsourcing contests [15]. Scholars of motivation and self-determination theory distinguish intrinsic and extrinsic motivation based on different reasons or goals that give rise to an action [16]. Intrinsic motivation “refers to doing something because it is inherently interesting or enjoyable” and extrinsic motivation “refers to doing something because it leads to separable outcome” [16]. Intrinsic motivation has two components: enjoyment based motivation and community based motivation; extrinsic motivation has three components: immediate payoffs, delayed payoffs, and social motivation [14].

Kaufman et al. in their study on 431 workers of Amazon Mechanical Turk indicated that the extrinsic motivation categories (immediate payoffs, delayed payoffs, social motivation) have a significant effect on the time individuals spent on the platform. They found that intrinsic motivation (fun, enjoyment, social interaction) is more important for some individuals to join and spent time in the platform [14]. Kazai et al. found that individuals with higher-order intrinsic factors such as fortune and fulfillment provide high-quality work while individuals with lower-order intrinsic/extrinsic factors such as fun and fame provide low-quality work [17]. One of the important motivational factors (and also design features) that have been examined in the literature is monetary awards. Archak in his study on a multiple simultaneous crowdsourcing contest (TopCoder) found that project payment is a significant determinant of the final project quality [7]. On the other hand Walter and Back in their empirical study on an idea contest (Atizo) found that monetary incentives only have an effect on the quantity of submitted ideas, not on the quality [18]. While Yang et al. in one of their study on Taskcn found that monetary awards are not a
significant incentive for individuals to participate in a task on the TaskCn site [19], in their later study they found that higher a reward induces both greater participation and higher submission quality [5]. The importance of monetary incentives have been identified in other studies as well [20], [21] [22]. Boudreau et al. found a significant relationship between cash incentives and continuous level of effort of individuals, but they could not find a significant relationship between cash incentives and the individuals choice to participate or the relationship of cash incentives and collaboration across team members [23].

Brabham in his qualitative study, based on the interviews with 23 Next Stop Design project’s participants, found that learning new skills and knowledge, career advancement and peer recognition (delayed payoff), contributes to collaborative effort (community-based motivation), and having fun (enjoyment based motivation) are motivators for individuals’ participation in that project [24]. Brabham in another qualitative study based on 17 interviews with members of the crowd at Threadless found five primary motivators for participation at Threadless: the opportunity to make money (immediate payoff), the opportunity to develop one’s creative skills (delayed payoff), the potential to take up freelance work (delayed payoff), the love of community at Threadless and addiction (community based motivation) [25]. Brabham in his qualitative study found that the opportunity to make money is a motivator for individuals’ participation [25].

Some researchers went beyond the motivation theory to examine other factors that affect individuals’ participation in crowdsourcing contests. These factors are classified into four main categories: (1) task-specific factors (reward, task type, task complexity, and contest duration for task, etc.) [5], [7], [26]–[30]; (2) individual-specific factors (extrinsic motivations, intrinsic motivations, individuals’ strategy, and individuals’ experience) [4], [7], [27], [29], [31], [32]; (3) environment-specific factors (competitors’ rating, number of competitors, number of super-star competitors, number of non-super-star competitors, collaboration) [33]–[35]; and (4) organization-specific factors (brand-strength and marketplace maturity) [36].

Most of the previous studies concentrated on the factors affecting individuals’ initial participation in crowdsourcing and did not distinguish between initial and sustained participation. However, Sustaining the individuals’ participation is essential to the success of crowdsourcing [37]. Sun et al. in a field survey with 205 subjects in TaskCn found that extrinsic and intrinsic motivations significantly influence individuals’ sustained participation [12]. They also found that task complexity negatively moderates the relationship between extrinsic motivation on sustained participation and self-efficacy positively moderates the relationship between intrinsic motivation and sustained participation [12]. Studies have argued that individuals felt rewarded for their participation when they receive feedback from the requesters regardless of whether their solution was selected or not [37]. This kind of reward (non-financial– knowledge acquisition, enhancing skills, having fun, and sense of accomplishment) gives the impression that future success is possible and strongly affect the chance of future participation by individuals [37]. Boons et al. in their field study found that feeling of pride drive ongoing member activity in crowdsourcing platforms [38]. Platform management by engaging in communication practices can increase members’ feelings of pride and respect [38], [39]. Feller also found that the periodic success or a belief that future success is possible strongly influence the chance of future participation by individual innovators [37].

In this paper we focus on crowdsourcing contests that have both collaborative and competitive components. We investigate the effect of individuals’ community activities, team activities, and their tenure on their sustained participation in crowdsourcing contests.

3. The Theoretical Background and Hypothesis Development

3.1 Community activities and continued participation in the contests

Intrinsic motivation has been shown to have an important role in facilitating crowd’s participation in crowdsourcing platforms [25]. One of the important components of intrinsic motivation is community based motivation. Community-based crowdsourcing contests offer the possibilities of knowledge sharing and collaboration among the individuals [4]. In these community-based contests, individuals collaborate while simultaneously compete with each other to submit the best idea or design and win the contests [35]. The effectiveness of a collaboration and competition coexistence approach has been supported by previous studies [35], [41].

In crowdsourcing communities, ideas and solutions are shared among members via posting
topics, commenting on the topics, and sometimes sharing part of or the whole solutions. This knowledge sharing behavior allows individuals to communicate, interact, discuss, and share their ideas [4]. Individuals who actively interact with other members in crowdsourcing communities, tend to feel a greater sense of community and take their contributions more seriously [4]. We hypothesize that individuals who share their knowledge with the other members of the community and have interaction with them are more willing to contribute to the future contests as well. This discussion is summarized in the following hypothesis:

Hypothesis 1: individuals’ continued participation in the contests is positively related to their community interactions.

3.2 Team activities and continued participation in the contests

Previous literature on teamwork indicates that by teaming up individuals can evolve their knowledge and expertise rapidly [4]. Highly specialized professionals need teaming up with other people to carry out integrative development projects [42]. Teamwork has a critical role in creating psychological safety in teams that face significant learning challenges [43]. Literature on teamwork on virtual communities also shows that teamwork expands individuals’ perspectives of problems [44], [45]. Thus, individuals’ teamwork activities will improve their expertise level which results in self-efficacy [12] and sustained participation. Thus we hypothesize:

Hypothesis 2: individuals’ continued participation in the contests is positively related to their team activities

3.3 Individuals’ tenure and continued participation in the contests

Research on organizational employees show that employee’s tenure (length of time on the job) has negative relationship with their turnover (Mobley et al.). Research on online customer community indicates that bidders became more selective in their behavior as a result of their experience in online communities. Therefore, their participation in online communities has null or negative effect on individual-level bidding volume [46]. Crowdsourcing contests are different from traditional organizations’ environment and as the individuals’ tenure increases, they gain more experience and become selective in choosing contests to participate in. Specially, in crowdsourcing contests that include collaboration element, individuals have more chance to interact with other members of the community, receive feedback on their content, and understand the weakness and strength of their content [40]. The individuals spend some time for learning from the community and increasing their knowledge. Therefore, we hypothesize that it takes longer time for individuals to find an appropriate contest to participate in. This discussion is summarized in the following hypothesis:

Hypothesis 3: individuals’ continued participation in the contests is negatively related to their tenure.

4. Research Methodology

4.1 Empirical Context: Contests at Kaggle.com

Data for this study comes from Kaggle.com, a web-based platform for data science competitions in which crowd of people compete to produce the best models for predicting and describing the datasets. Kaggle competitions are open to all data scientists registered on the site and for competitions the rewards vary from $0 to $500,000 depending on the contest. Since its launch in 2010, Kaggle has served many companies, including General Electric, Allstate, Merck, Ford, and Facebook [47]. In April 2015, Kaggle implemented the first version of their Kernels product in their platform. Kernels allow users to write, run, and publically share their code on Kaggle. This product helps members to communicate with each other, share their solutions, and receive feedback from the other members.

4.2 Data collection and analysis

For this study, we collected data on the Kaggle’s public contests since the launch of the platform in April 2010 through August 2016. Since Kaggle implemented Kernels in the platform in April 2015, we only considered the contests from April 2015 to August 2016 that include Kernels. Our final sample include 2155 observations which consists of 875 users that participated in 23 contests, and each user has attended at least 2 contests of the 23 contests in the dataset.

4.3 Variable measurement

To test our hypothesis, we measured the dependent variable (continued participation) and independent variables (community activities, team activities, and tenure). The variables are summarized and described in Table1.
Table 1. Variable descriptions

<table>
<thead>
<tr>
<th>Variables</th>
<th>Variable descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continued participation</td>
<td>The number of days that have passed from the individual’s last participation in the competition until the current participation.</td>
</tr>
<tr>
<td>Community activities</td>
<td>The total number of the votes that the individual has received for solution sharing activities in the last competition.</td>
</tr>
<tr>
<td>Team activities</td>
<td>The number of team members that the individual has had in the last competition.</td>
</tr>
<tr>
<td>Tenure</td>
<td>The number of days that have passed from the individual’s registration date in the platform until the current competition.</td>
</tr>
</tbody>
</table>

5. Research Model and Results

5.1 Dependent variables and Independent variables

The dependent variable in this analysis is the number of days that have passed from the individual’s last participation in the competition until the current participation. “Recent total votes” (community activities), “recent team members” (team activities), and “Tenure” are the three covariates for testing the hypothesis, which are described in Table 1.

Four other variables are included as control variables for model adjustment. We controlled for current competition’s “prize”, namely the amount of the prize for the competition measured in $10,000. This variable is included because the literature identified monetary reward as one of the important motivators for individuals’ participation in crowdsourcing contests [25]. We controlled for the “number of competitors” (the number of teams competing for the contest) since the literature shows individuals react negatively to an increase in the total number of competitors [33]. We controlled for “previous performance” (the rank the individual’s team received in the previous competition). The literature indicates that the individual’s performance can affect her/his self-efficacy. Sun et al. in their study on crowdsourcing showed that self-efficacy moderates the relationship between motivation and sustained participation [12]. We also controlled for the competition “duration” because duration is one of the factors that have received considerable attention in crowdsourcing research that affects individuals’ participation in crowdsourcing contests [36].

5.2 Analysis

Poisson regression models have been widely used in information systems to account for the discrete and non-negative nature of the response variable with count data [48]–[51]. However, the Poisson distribution requires that the variance to be equal to the mean of the response variable. In this study, the variance of the response variable is much larger than the mean of the response variable. In order to overcome the restriction of equi-dispersion that is imposed by the Poisson model, the Negative Binomial Model was used. Therefore, the continued participation, which is measured by the number of days between two competitions can be presented as:

\[
	ext{Continued Participation} = \exp(C + \beta_1 \cdot \text{recent total votes} + \beta_2 \cdot \text{recent team members} + \beta_3 \cdot \text{tenure} + \beta_4 \cdot \text{previous performance} + \beta_5 \cdot \text{duration} + \beta_6 \cdot \text{number of competitors} + \beta_7 \cdot \left(\frac{\text{amount of prize}}{10000}\right) + \epsilon)
\]

To investigate the robustness of the findings, we use an alternative to the Negative Binomial regression model by categorizing the response variable into a number of classes, depicted in Table 2, and use ordinal logistic regression. Ordinal logistic regression requires that the potential values of the independent variables have a natural ordering [48]. More specifically, we converted the dependent variable “continuous participation” to an ordinal variable with 4 categories as shown in Table 2.

Table 2. Ordinal variable for “Continuous participation”

<table>
<thead>
<tr>
<th>continued participation days</th>
<th>level</th>
</tr>
</thead>
<tbody>
<tr>
<td>days&lt;=30</td>
<td>0</td>
</tr>
<tr>
<td>30&lt; days&lt;=60</td>
<td>1</td>
</tr>
<tr>
<td>60&lt;days&lt;=90</td>
<td>2</td>
</tr>
<tr>
<td>90&lt; days</td>
<td>3</td>
</tr>
</tbody>
</table>
5.3 Results

The results for the Negative Binomial model in Table 3 show that the recent total votes and recent number of team members are not statistically significant contributors to individuals’ continued participation. In another word, there is no evidence that if an individual is highly active in a contest’s kernels and has received high number of votes for his/her kernels is motivated to participate in another contest after a short period of time. Moreover, if an individual attended a competition as a team with many team members, it does not meaningfully inspire the individual to come back quickly to participate in another competition. Therefore, receiving votes and having more team members does not increase individuals’ sustained participation. Therefore, the hypotheses 1 and 2 are not supported by these results. However, we found a significant positive relationship between the continued participation and tenure, which means if an individual has longer tenure and thus has been on the platform for a longer time, he/she will be less inclined to participate in another competition in a short period of time. Therefore, increased tenure is associated with returning to competitions less frequently. Thus, Hypotheses 3 is supported by these results. Moreover, all control variables, i.e., the amount of the competition prize, the number of competitors, previous performance, and competition duration have significant effect on the individuals’ continued participation. These findings are consistent with the previous studies on crowdsourcing contests and highlight the importance of contest-specific and individual-specific factors on individuals’ participation.

The results from the ordinal logistic regression, depicted in Table 4, are consistence with the findings from the Negative Binominal model. We found a significant positive relationship between the continued participation and tenure. We also found that a higher amount of competition prize, a lower number of competitors, a better ranking in previous competitions, and shorter competition duration are all associated with individual coming back more frequently to participate in the platform’s competitions.

Table 4. Results from Ordered Logistic Regression

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recent total votes</td>
<td>0.0011</td>
<td>0.569</td>
</tr>
<tr>
<td>Recent team members</td>
<td>-0.00003</td>
<td>0.533</td>
</tr>
<tr>
<td>Tenure</td>
<td>0.0003</td>
<td>&lt;0.001 ***</td>
</tr>
<tr>
<td>Prize</td>
<td>-0.011</td>
<td>&lt;0.001 ***</td>
</tr>
<tr>
<td>Number of competitors</td>
<td>0.0001</td>
<td>&lt;0.001 ***</td>
</tr>
<tr>
<td>Previous performance</td>
<td>1.6493</td>
<td>&lt;0.001 ***</td>
</tr>
<tr>
<td>Duration</td>
<td>0.0147</td>
<td>&lt;0.001 ***</td>
</tr>
</tbody>
</table>

Note: Dependent variable: Continued Participation

* Level of significance: p < 0.1.
**Level of significance: p < 0.05.
***Level of significance: p < 0.001

6. Conclusions

6.1 Discussion and Implications

The objective of this study was to determine the factors affecting individuals’ sustained participation in crowdsourcing contests. In contrary with hypothesis 1, our findings indicate that there is no statistically significant relationship between individuals’ community activities and continued participation. One explanation could be that the individuals who are active in the community and share their solution with the community are in the learning stage and they do not want to get involved with the competitions until they acquire the skills and knowledge that enable them to compete with the other teams. Also, in contrary with hypothesis 2, the relationship between “team activities” and continued participation is not statistically significant. The number of teammates that an individual had in the previous competition does not affect his/her continued participation in the competitions. We hypothesized that the individual tends to come back earlier because he/she liked the experience of working in a larger team and learning from teammates. One possible explanation for the
contrary results could be that not every team member actually contributes to the teamwork and thus these individuals probably did not have a good experience from teaming up and did not benefit from the teamwork.

We found a statistically significant relationship between “tenure” and continued participation. While in the traditional organizations, employees with longer tenure are more motivated to continue their job, our findings show that in crowdsourcing contests individuals with longer tenure take longer time to come back to the platform and continue their participation in the other contests. One explanation for this finding is that tenured individuals are more selective in choosing the contests to participate in, and based on their experience on the platform they participate in contests that align with their skills and where they can perform better.

For control variables, we found a statistically significant effect of the amount of the monetary prize on individuals’ continued participation. Even though many teams compete with each other in Kaggle competitions and the chance of winning is very low, the amount of the monetary prize is an important incentive for individuals to continue their participation. The amount of a prize motivates individuals although they see low chance for winning the contests. This is similar to more people buying lottery tickets when the payout gets larger even though the chance of winning in the lottery is very small.

We also found a statistically significant relationship between competition duration and continued participation, namely, the longer a competition’s duration, the less inclined individuals are to come back to the platform and participate in the competition. This could be due to a burnout effect when participating in a long competition.

We also found a statistically significant relationship between the “number of competitors” and continued participation. This indicates that as the competitiveness of the contests increase individuals are less willing to continue their participation.

We also found significant relationship between previous performance and continued participation. The results indicate that having low performance in a previous competition is associated with longer time between participations in the contests. The explanation for this finding is that when individuals’ performance is not good, their perceived ability and competence to accomplish tasks diminishes, therefore it will take them a longer time to come back to the platform and participate in the contests.

This research has interesting and valuable theoretical and managerial implications. Prior research have demonstrated the importance of the individuals’ sustained participation and quality of submissions to the success of crowdsourcing processes. Although there is considerable research that examined individuals’ participation in crowdsourcing contests and the factors that affect their participation on these platforms, most of them have been mostly silent on identifying the factors affecting individuals’ sustained participation in crowdsourcing contests, specifically the contests that have both collaborative and competitive components.

Our research results highlight the importance of individual-specific and contest-specific factors on individuals’ sustained participation. Our research suggest that crowdsourcing platform sponsors should focus on reducing the competition duration and increase the amount of prize to attract more participants, especially participants who are new on the platform. Moreover, individuals should be motivated to submit quality solutions because their performance influences their continued participation. The crowdsourcing platforms should be designed in a way that facilitates individuals’ connection with more experienced participants in order to motivate their sustained participation in the contests.

6.2 Limitation and Future Research

This study contains some limitations that influence the potential generalizability of our findings. First, while there are many crowdsourcing platforms, our study only considers one platform, namely Kaggle. In order to generalize our findings, research on other platforms would be needed. For example, InnoCentive only offers single competition and Topcoder only allows at most 20 people in a competition. The restrictions on other platforms are different from Kaggle’s platform, and thus results might be different. Second, researchers can explore more factors that may affect individuals’ sustained participation in contests. For instance, the competition types and the contents of competitions may be associated with continued participation of individuals. Third, many factors may affect individuals’ continued participation. Crowdsourcing platform sponsors can focus on the effect size of each factor in order to select and design competitions to increase participation. Despite these limitations, this paper makes an important contribution to theory and practice. It provides empirical evidence that receiving more votes from kernel submission for a competition and attending competition as a team with more members do not necessarily contribute to individuals’ continued participation. However, reducing the duration of competitions and increasing the prize amount offer opportunities to increase participation.
7. References


2015, pp. 4609–4618, Hawaii, USA.


