Poster on *Trustworthiness in Enterprise Crowdsourcing: a Taxonomy & evidence from data*

ICSE SEIP 2016, FSS Fall 2018

Shrikanth N C
NC State University
snaraya7@ncsu.edu

ABSTRACT

In this poster, we discuss this paper [1] on crowdsourcing that presents a taxonomy around the trustworthiness context, effects of various factors and guidelines to mitigate certain risks. However, in the context of this class we will focus mainly on the effects of various factors around trustworthiness using years of empirical data collected from two leading crowdsourcing vendors Upwork[4] and Topcoder[2].

1 INTRODUCTION

"Crowdsourcing is an emerging trend where a group of geographically distributed individuals contribute willingly, sometimes for free, towards a common goal." [1].

This paper highlights 3 key benefits to crowdsourcing a) Faster time-to-market due to parallel execution of tasks, b)Lower cost due to access to the right skills and c) Higher quality through the creativity competitive nature of the crowd. This paper predominantly deals with macro tasks. Macro tasks are time consuming and complex in nature as opposed to micro tasks like language translation, image tagging etc.

2 PROBLEM

Despite crowdsourcing's success in software development, complex problem solving, this paper analyzes the key issue with crowdsourcing, which is the apparent lack of 'trust' on the crowd. Due to this enterprises are skeptical to adopt this model completely for full-fledged software development. At numerous instances like the prominent DARPA shredder challenge [3], where the crowd workers turned rogue and sabotaged the submissions. In another instance, the social media campaign of Henkel (refer Figure 1) where humorous designs got top voted. Below is a comment from a task poster in Upwork about a crowd worker.

"This guy is a scumbag. Didnt do any work and wasted my time. Then begged for money when I ended the contract. When I refused a virus ended up on my site.".

This paper from various literature identifies 6 untrustrworthy taxonomy elements. They are Quality of the Submission,Timeliness of the Submission,Ownership Liability of the Submission,Network Security Access Control,Loss of Intellectual Property *IP* and Loss of Data.

3 ABOUT DATA

We collected data (macro tasks & user comments) from actual crowdsourcing tasks in two prominent vendors - Upwork & TopCoder using their public API's refer Table 1.

Figure 1: The most voted designs from the crowd.

	Upwork	Topcoder
Time period	1-Mar-2006 to 31-Aug-2015	1-Jun-2012 to 31-Dec-2014
No. of tasks completed	86,160	7488
No. of task submissions	86,160	18,195
No. of crowd workers	34,445	15,64
Comments	1,037,690	-

Table 1: Data collected

4 REASONS FOR UNTRUSTWORTHINESS

This paper used Latent Dirichlet Allocation (LDA) as a topic model and identified the top set of issues reported by crowdsourcers among a total of 1,037,690 open ended comments. We mapped few comments to the 6 taxonomy elements. One such comment in the 'Quality of Submission' taxonomy element is :

"Overall, not a good working knowledge of MVC design patterns"

5 EFFECT OF MONETARY BENEFITS ON TRUSTWORTHINESS

We performed a linear regression between the task score and the monetary benefits of a task. The correlation coefficient was close to 0 (refer Figures 2 and 3). As the range of the prize was large, we also performed the regression till the median prize. The median for Upwork was $97.23 & for TopCoder it was $900.
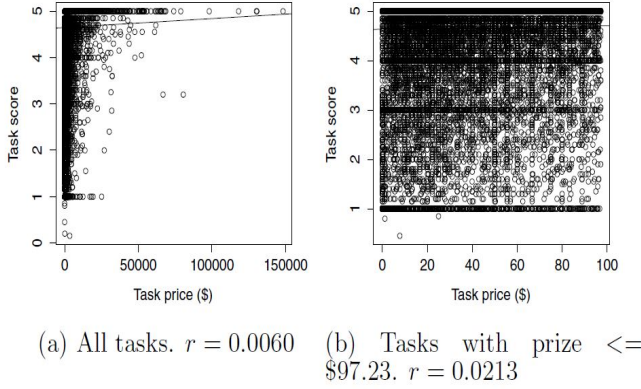
(a) All tasks. $r = 0.0060$

(b) Tasks with prize $<=$ $97.23. $r = 0.0213$

Figure 2: Task score versus monetary benefits in Upwork.

(a) All tasks. $r = 0.0884$

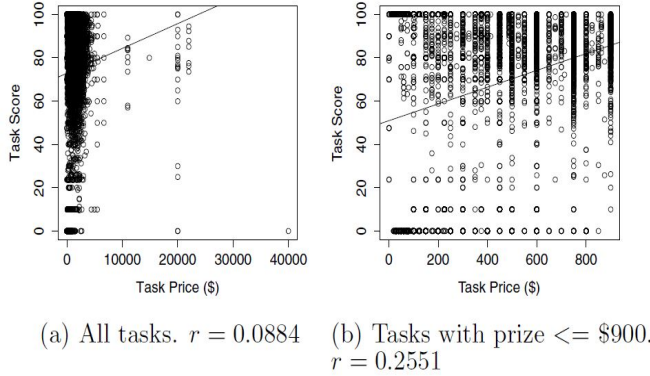(b) Tasks with prize $<= $900. $r = 0.2551$

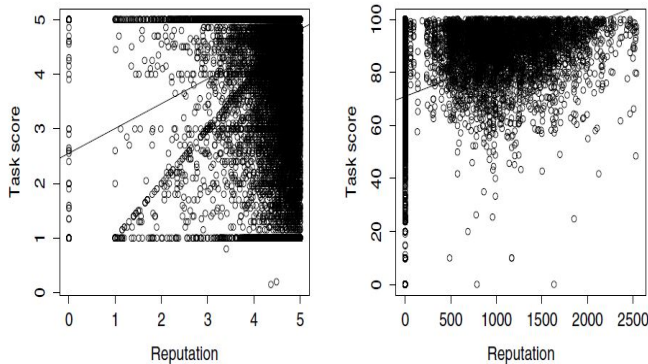**Figure 3: Task score versus monetary benefits in Topcoder.**

Figure 4: Task score vs. reputation. a) Upwork b) Topcoder

6 EFFECT OF REPUTATION ON TRUSTWORTHINESS

The linear regression between task score and reputation exhibited poor correlation refer figure 4, the corresponding correlation coefficients are 0.2457 for Upwork and 0.2072 for TopCoder. Additionally, because of skewness in Upwork we ran Welch two sample single sided t-test with null hypothesis that the population of crowd workers with a high reputation has the same mean of task score as the entire population (reputation >= 4.5). The p-value came to 2.2e-16 and the Cohen's d effect size was 0.076. This indicated although reputation increased but the effect size is too small to be beneficial.

7 EFFECT OF GEOGRAPHY ON TRUSTWORTHINESS

Most the crowd workers were from India (37%) in Upwork and China (32%) in TopCoder. India topped in the amount of untrustworthy workers. We ran the Welch two sample single sided t-test with null hypothesis that the population of crowd workers from the 'U.S.' geography has a higher average task score than the entire population. Here the p-value was 1.908e-12 for Upwork with an Cohen's d effect size of 0.10, and for TopCoder, the p-value was 1.

8 EFFECT OF CONTESTS ON TRUSTWORTHINESS

Crowdsourcing tasks are often posted as contents and we wanted to test whether contests produced trustworthy submissions, due to its competitive nature. It turned out to be true. There was a reduction amount of untrustworthy behavior from 25.8% to 3% in Upwork when tasks are posted as contests.

9 CONCLUSION

This paper identified and discussed a comprehensive taxonomy of trustworthiness in crowdsourcing for software development, specifically in the enterprise context. Later, we discussed the effects from the data collected from popular crowdsourcing vendors. This paper suggests a 'private verified crowd' since reputation has no significant impact on trustworthiness and importance of contests in crowdsourcing to yield trustworthy submissions.

REFERENCES

[1] A. Dwarakanath, S. N.C., K. Abhinav, and A. Kass. 2016. Trustworthiness in enterprise crowdsourcing: a taxonomy amp;amp; evidence from data. In *2016 IEEE/ACM 38th International Conference on Software Engineering Companion (ICSE-C)*. (May 2016), 41–50.
[2] Jack Hughes. [n. d.] website. (). Retrieved Sept. 4, 2018 from www.topcoder.com.
[3] Nicolas Stefanovitch, Aamena Alshamsi, Manuel Cebrian, and Iyad Rahwan. 2014. Error and attack tolerance of collective problem solving: the darpa shredder challenge. *EPJ Data Science*, 3, 1, (Sept. 2014), 13. ISSN: 2193-1127. DOI: 10.1140/epjds/s13688-014-0013-1. https://doi.org/10.1140/epjds/s13688-014-0013-1.
[4] Odysseas Tsatalos Stratis Karamanlakis. [n. d.] website. (). Retrieved Sept. 4, 2018 from www.upwork.com.