A Hybrid Simulation Model for Crowdsourced Software Development

RAZIEH SAREMI, Stevens Institute of Technology, Hoboken, NJ 07030, USA, rlotfali@stevens.edu

Crowdsourcing as a new emerging software development method contains crowdsourced mini-tasks as demand and online workers as suppliers. The major counter-argument in such systems is that suppliers are volunteers and are not bound by any contract, also, the size of available suppliers varies wieldy throughout the day. Such uncertainty about the receiving service may cause inefficiency and task failure. This research presents a hybrid simulation model to address the risk of task failure in competitive crowdsourcing platforms. The simulation model is composed of three components: the discrete event simulation which represents the task life cycle, the agent-based simulation which illustrates the crowd workers' decision-making process and the systems dynamic simulation which displays the platform.

KEYWORDS

Crowdsourcing, Simulation, Task, Agent-Based, Discrete Event, Systems Dynamic

1 INTRODUCTION

The crowdsourced software development (CSD) platform can be viewed as a demand and supply market for project tasks and workers [1]. While tasks are demands, workers are responding as suppliers. However, preparing presentable demands in a crowdsourcing platform is very important, understanding the suppliers' sensitivity and performance to the arrival demand and rate of failure is crucial. To date most of the developed methods focused on the static aspect of the tasks and neglect the dynamic aspects of it. This issue leads us to propose a hybrid simulation model to reduce task failure ratio in crowdsourced platform.

Simulation techniques have been demonstrated to be a powerful method in modeling and understanding NP-hard problems in software engineering processes. Systems dynamics models have been used in simulating of lifecycle processing, defect detection techniques, business cases in software processes as well as software investment analysis [2]. While, discrete event models used to identify entities, events and detail activities in a system and model the relationship among them [3]. Also, Agent based models used to simulate certain social aspect in software development such as interaction between developers [2]. To the best of our knowledge, there is no existing study using a hybrid simulation model of CSD based on combination of all three different simulation methods. We believe providing a hybrid simulation for CSD is beneficial for both crowd project managers and crowd workers to optimize their own utility model.

 ACM ISBN 978-1-4503-5733-3/18/05. https://doi.org/10.1145/3195863.3195866

2 HYBRID SIMULATION MODEL

In this part we present a hybrid simulation model combination of systems dynamic, discrete event and agent-based model. Figure 1 illustrates the overall view of the hybrid model, composed of three levels.

Micro Level: Crowdsourced projects integrates online and unknown workers elements in to the design. Appling agent based (AB) method to simulate crowd-workers' behavior individually provides the option of observing diversity of attributes among them. Crowd-workers are represented as agents who are having one or more of the following characteristics: 1- identifiable with a set of rules that directing their behavior, 2- autonomous agent that can act independently in the environment and have control over their actions, 3- situated workers that work in the same environment and interacting with each other, 4- flexible agent that can adapt their behavior to be a better fit to the environment [4].

Agents' arrival to the platform is following non-homogenous Poisson distribution [5]. Agents are assigned to unique IDs upon creation in the simulation. To address the rest of the characteristics, each agent has a utility factor [5] that define their behavior. Therefore, each agent has an internal decision-making state consisting of two components of: "Registering" for a task, "Submitting" the task. Agent decision making is related to the information that agent receives from the agent's community and social environment, as well as another agents' choice of competing on a task [6]. Agent's final decision to submit a task and level of submission impacts on the agent's profile. The agent may decide to submit the registered task. Task submission decision follows random numbers greater than 0.7. By submitting a new task, agents' attributes of reliability factor will be updated. According to empirical data submitted tasks which can grant score greater than 75% are considered qualified and the submitter agent is reported as winner or runner-up. Therefore, qualified submission is determined by assigning random number greater than 0.75. When, the submission passed peer-review, agent's attribute of rate and trust-ability factor will be updated. Moreover, the new score will be reported in the agent's profile. Decisions from simulated agents will determine task progress.

Meso Level: In this model, tasks are defined as a set of discrete events (DE) that has start and end. A set of tasks performs in a sequence to achieve the overall goal and project success. Time between a task start and end is called task duration and it is the estimated execution time for an agent to perform a task. According to our empirical analysis task duration is following triangle distribution with the maximum 30 days, minimum 1 days and the

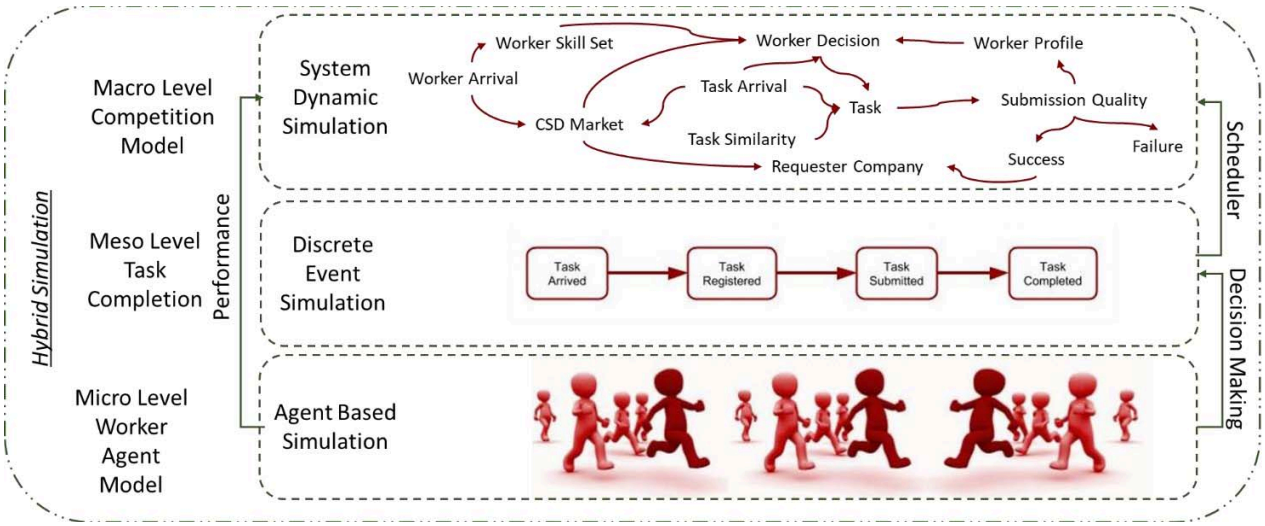

Fig 1: Overview of Hybrid Simulation Model

mode of 16 days.

Task arrival in the model, is an event that follows Poisson distribution [7]. According to our previous analysis, arrival tasks will attract agents to work on them with the rate of 70%. If the agent decides to submit the task, it will move to the submit state, and go under peer- review process.

If the submitted task can grant the quality score of 0.75 and above, it will be recorded as successful task, otherwise, task will be recorded as a failure.

Macro Level: While agent-based model is responsible for individual agent's behavior and discrete event manages task sequential, systems dynamic (SD) shows the interactions among systems parameters and feedbacks with in the platform. This model contains 8 variables including Task, Agent Decision, Submissions' Quality, Task Similarity, Worker Profile, Worker Skill-set and different available Crowdsourced markets. The SD model represents the causal loops among different levels of the platform.

Changing in agents' attributes not only illustrates agents' performance in the platform, but also represents platform trust-ability to the requestor company. In order to measure platform trust-ability, the model provides platform failure ratio in any given time. Task Failure Ratio is the ratio of tasks with zero submissions or not passing peer review.

3 SIMULATION INITIAL RESULTS

The presented simulation model is created based on Topcoder [6], one of the largest competitive platform, workflow. In such platforms task failure ratio is the main metric to measure effectiveness of a scheduling method. Therefore, it is important to understand if the proposed model positively impacts task failure ratio in task arrival framework. The hybrid model is designed to perform with maximum pool of 100 available tasks and 52000 available workers, with averages task arrival of 13 tasks per day and average worker arrival of 1000 workers per day.

As illustrates in Figure 2, modeling platform by SD model will provide failure ratio of 12% on average, adding the DE model on top or it will reduce the failure ratio to 11%. While adding AB model to the combination of SD and DE will reduce the task failure ratio to 8.4%. More details about the proposed hybrid simulation model is reported in [7].

In future work we are considering developing a facilitating tool based on the presented hybrid simulation technique to support decisions during task scheduling phase, e.g. CSD process simulator.

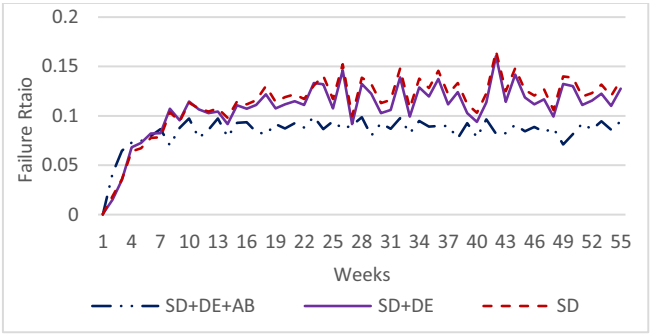

Fig 2: Failure Ratio in the proposed model

REFERENCES

[1] Difallah.D.E, Demartini.G and Cudré-Mauroux.P, "Scheduling Human Intelligence Tasks in Multi-Tenant CrowdPowered Systems," in ACM, 2016.

[2] Chao Gao, He Zhang, and Shu Jiang. 2015. Constructing hybrid software process simulation models. In Proceedings of the 2015 International Conference on Software and System Process, ICSSP 2015, Tallinn, Estonia, August 24 - 26, 2015. 157–166. https://doi.org/10.1145/2785592.2785610C. M.

[3] A. M. Law and W. D. Kelton. Simulation Modeling and Analysis. McGraw-Hill, 2nd edition, 1991.

[4] S. Faradani, B. Hartmann, and P.G. Ipeirotis, "What's the Right Price? Pricing Tasks for Finishing on Time", In Proc. Human Computation, 2011.

[5] N. Archak, "Money, glory and cheap talk: Analyzing strategic behavior of contestants in simultaneous crowdsourcing contests on Topcoder.com," in 19th international conference on World wideweb, 2010.

[6] https://www.topcoder.com/

[7] https://web.stevens.edu/softwareanalytics/?page_id=12