Proceedings of the 2014 Winter Simulation Conference
A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.

An AGENT-BASED MODEL FOR CROWDSOURCING SYSTEMS

Guangyu Zou

Dalian University of Technology
Panjin, Liaoning 124221, CHINA

Alvaro Gil
Marina Tharayil

PARC
Webster, NY 14580, USA

ABSTRACT

Crowdsourcing is a complex system composed of many interactive distributed agents whom we have little information about. Agent-based modeling (ABM) is a natural way to study complex systems since they share common properties, such as the global behavior emerging on the basis of local interactions between elements. Although significant attention has been given to dynamics of crowdsourcing systems, relatively little is known about how workers react to varying configurations of tasks. In addition, existing ABMs for crowdsourcing systems are theoretical, and not based on data from real crowdsourcing platforms. The focus of this paper is on capturing the relationships among properties of tasks, characteristics of workers, and performance metrics via an ABM. This approach is validated by running experiments on Amazon Mechanical Turk (AMT).

1 INTRODUCTION

Crowdsourcing is defined as the act of outsourcing work to a large distributed group of independent, mostly anonymous workers over a technology mediated platform (Estells-Arolas and de Guevara 2012). As shown in Figure 1, crowdsourcing systems are composed of four major components: Requester, Worker, Task, and Environment. The requester submits tasks to be solved on crowdsourcing platforms where workers choose tasks that they are interested in and are able to finish. The submissions of workers are evaluated by the requester who determines whether or not to pay workers based on the quality of the submissions. The environment defines the rule governing the whole process of submission, evaluation, and rewards etc. Also the environment defines how workers communicate with each other and how workers communicate with the requester, because some crowdsourcing platforms, such as open innovation, allow and encourage collaborations between workers.

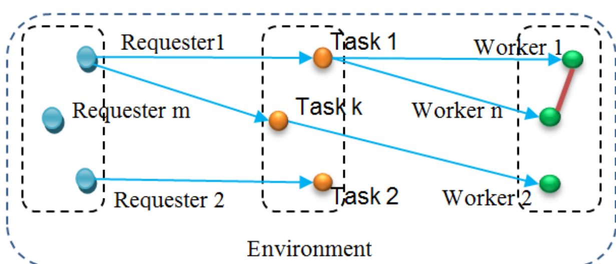


Figure 1: Network of Crowdsourcing Systems.

Usually, the crowdsourcing platforms are hard to predict with respect to various factors such as availability of workers, background of workers, and tendency of workers to attempt the tasks. Crowdsourcing is a complex system composed of many interactive distributed agents, based on which the global pattern emerges. It is difficult to understand the dynamics of a crowdsourcing system without an effective modeling approach.

To learn to employ a crowdsourcing platform for accomplishing business functions most effectively and efficiently, there is a need for a solution that facilitates simulation of the crowdsourcing platforms. Agent based modeling (ABM) is an ideal way to study complex systems because an ABM consists of a large number of autonomous entities. Even a simple ABM can exhibit complex behavior patterns and provide valuable information about the dynamics of the real-world system (Bonabeau 2002). Therefore, an ABM is introduced in the paper to simulate the dynamics of crowdsourcing systems from the perspective of complex systems.

The underlying dynamics of crowdsourcing systems have been studied by researchers from different perspectives, sometimes using different names such as Open Innovation (Chesbrough 2003) or Swarm Intelligence (Krause, Ruxton, and Krause 2010). (Yuen, King, and Leung 2011) conducted a survey about various crowdsourcing systems, and classified them based on four categories: application, algorithm, performance, and dataset. (Martin, Hanrahan, and O'Neill 2014) conducted an ethnomethodological study of Turker Nation forum showing how AMT users reason about and carry out their work activities. (Rajan, Bhattacharya, Celis, Chander, Dasgupta, and Karanam 2013) proposed a novel framework (CrowdControl) for scheduling a batch of tasks on a crowd platform which aims at simultaneously learning crowd performance and optimizing the performance achieved on an input batch of tasks. (Bernstein, Karger, Miller, and Brandt 2012) used queueing theory to analyze realtime crowdsourcing in order for requesters to minimize their cost subject to performance requirements. (Huang, Zhang, Parkes, Gajos, and Chen 2010) used statistic analysis to predict the rate and quality of work based on observations of output to various designs. (Faridani 2012) derived a statistical model for the estimation of completion time. (Mason and Watts 2010) studied the relationship between financial incentives and the performance of workers at the aggregate level. To the best of our knowledge, there is no existing research that studies the participation of workers, the completion time and the quality of tasks by combining properties of tasks and attributes of workers together.

There are three metrics that a requester pays the most attention: cost, completion time, and quality of tasks. In this paper, the models of completion time and quality of tasks are developed by considering properties of requester, task, and worker together as the behavior rules of workers in the ABM for a crowdsourcing platform. This differentiates from (Faridani 2012, Mason and Watts 2010) since it not only includes the impacts of properties of tasks in the completion time, but also takes into account how attributes of workers and properties of tasks affect the quality of tasks. Using agent simulation, we study the dynamics of AMT and observe the performance metrics (completion time and accuracy) to determine relationships between properties of tasks, characteristics of workers, and performance metrics.

The rest of the article is organized as follows. In section 2, we present the relevant research of modeling in crowdsourcing platforms. Section 3 describes the methodology to build an ABM on crowdsourcing based on experimental data. Section 4 shows how to apply this methodology to develop a practical ABM where experiments on AMT are conducted in order to collect data to analyze the dynamics of AMT. We show how to derive behavior rules for crowdsourcing workers and build the ABM accordingly. The verification and validation (V&V) of the created models are conducted by comparing the simulation outputs to the real data. Finally, in section 5, we conclude by summarizing our findings.

2 LITERATURE REVIEW OF MODELING AND SIMULATION ON CROWDSOURCING

Crowdsourcing is a complex system composed of many interactive distributed agents, based on which the global pattern emerges. An effective modeling approach is needed to understand the dynamics of a crowdsourcing system. In (Scekic, Dorn, and Dustdar 2013), three incentive schema are evaluated in crowdsourcing environments using an ABM. In (Khazankin, Schall, and Dustdar 2012), a simulation model is built to predict the quality of service (QOS) by assigning tasks to workers. (Chu, Chen, Liu, and Zao 2011) built and tested a mobility model of crowdsourcing users who update disaster situation information. Crowdsourcing and Open Innovation are two coined terms, since they describe a form of collective intelligence that is enabled by new technologies, particularly Internet connectivity. Some ABMs have been built to mimic the open innovation that takes place both in universities (Bcheler and Sieg 2011) or in virtual scientific communities (Zou and Yilmaz 2011). A general ABM is built in (Bcheler, Fchslin, Pfeifer, and Sieg 2010) (Bcheler, Lonigro, Fchslin, and Pfeifer 2011) to model crowdsourcing as a complex biological system, where attributes of requesters, tasks, and workers are investigated. (Zou, Tharayil, Gil, Chander, and Celis 2013) build a ABM to study worker's availability, capability, and service time by collecting data from MobileWorks. Our work complements and departs from the previous citations by (i) taking into account the impact of additional properties of tasks on quality of tasks, (ii) capturing the behavior rules of workers interacting with the environment, and (iii) considering online workers processing tasks available in crowdsourcing platforms. Therefore, this paper aims at exploring the underlying behavioral rules of online workers on crowdsourcing platforms by conducting design of experiments, and building the generic simulation model of crowdsourcing platforms as a test bed, establishing a foundation in theory and in practice for further research on the dynamics of crowdsourcing platforms.

3 METHODOLOGY TO BUILD AN ABM ON CROWDSOURCING

The dynamics of crowdsourcing systems are based on the activities of crowdsourcing workers and the interactions between them. To better understand crowdsourcing systems, it is necessary to understand the behavior of workers, that is, how workers respond to different configurations of tasks and then how well and how quick workers process tasks. In order to derive the behavior rules of workers, we need to identify the properties of workers and tasks, and then design experiments to collect the data. Finally, according to the behavior rules of workers derived from the data, an ABM is built and validated using a testing set. The process that was used to build an ABM is listed as follows:

1. Identify the inputs and outputs of the platform to be modeled: The inputs are the variables controlled by the requester to observe how the workers perform (outputs) when these inputs are changed.
2. Design and run experiments: The requester decides what type of experimental design will be used to derive models of the crowdsourcing platform.
3. Selection of the best distribution to fit the output data collected from the experiment: For each input combination, we look at all the output values and determine if the distribution is unimodal or multi-modal. Once this is decided we proceed to select the type of distribution to be used across all the input combinations.
4. Derive the parameters of the selected distribution.
5. Derive a model that captures the relationship between the inputs and the parameters of selected distributions.
6. Build an ABM based on the model derived in Step 5.
7. Verify and validate the ABM by comparing the simulated output to the real data.

The following sections describe the application of the methodology to build an ABM on crowdsourcing.

4 BUILDING AN ABM OF AMT

Without loss of generality, AMT is chosen as the target crowdsourcing system to build an ABM since AMT is a contemporarily popular crowdsourcing platform used for microwork (i.e., tasks that can be divided into workloads). In this section we show how to apply the methodology described in Section 3 to a parking lot video where workers are asked to monitor and record cars entering/exiting the lot.

4.1 Inputs and Outputs of the Experiments

In order to understand the dynamics of AMT, we use Design of Experiments (Schmidt and Launsby 2005) to collect performance data of tasks finished by workers. AMT is a crowdsourcing Internet marketplace that enables individuals or businesses (known as Requesters) to coordinate the use of human intelligence to perform tasks that are hard to process by computers. The goal is to figure out what and how factors affect the behavior of workers which in turn determine the overall performance on AMT. We design HITs (Human Intelligence Task) on AMT to show workers a parking lot video and ask the viewers to record when a car enters/exits the parking lot. Figure 2 is the snapshot of a HIT on AMT, which belongs to the category of video analytics. We vary multiple inputs in the system, and measure their impact on the outputs, such as accuracy, acceptance rate, and completion time. In this paper, we interchangeably use the words outputs and performance metrics.

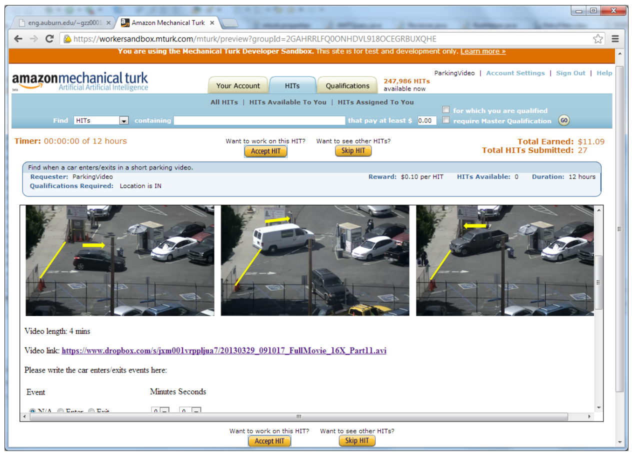


Figure 2: Snapshot of a HIT on AMT

Table 1 lists all the inputs and associated descriptions which are used in the following experiments. These are job design choices such as video speed, time to post, and features offered by AMT platform in their task design options, such as HIT lifetime or qualification of the worker.

Table 1: Inputs of Experiments

Inputs	Description
Incentive (Reward)	The money paid per task.
Video Speed	The speed to play the parking video, such as (16X, 32X)
Video Duration	How long the video is.
Weekday	1 means the task posted during weekdays. -1 means the task posted during weekends.
TimeofDay	8 means morning from 8 AM to 12 PM. 20 means evening from 8 PM to 12 AM.
Country	1 means US. -1 means India.
Approval Rate	The percentage of a worker's submission approved by requesters in the worker's career.
HIT Life Time	How long a HIT is available to workers.
Assignment Duration	Time allowed to finish a task after the task is accepted by a worker.

Table 2 lists all the output metrics which are used in the following experiments to demonstrate the performance of AMT. Since we have the gold data (known answers) for the videos posted in this experiment, we can compare results with gold data to obtain accuracy metrics.

4.2 Experiment Design

We use Design of Experiments (Schmidt and Launsby 2005) to collect the training and testing experimental data set. The training set is obtained by carrying out a screening design that will be used for two purposes: The first one is to create the training set that will be used to build a preliminary model and the second one is to identify the key input variables that will be used for collecting the testing set. We use the screening design in an unconventional way since this is used to derive the behavior rules of workers. The goal of

Table 2: Outputs of Experiments

Outputs	Description
Accuracy	Indicates how close the worker's submission is from the answer.
Acceptance	Binary value means whether or not a task is taken by a worker.
Accepting Time	The time period between when the task is posted and when it is accepted by a worker.
Working Time	The time period between when a worker accepts a task and when the worker submits it.
Completion time	The time period between when the task is posted and when it is submitted by a worker. It equals the sum of accepting time and working time.

the screening test is to figure out the impact the 9 inputs have on performance of AMT. In a screening experiment performed on AMT, 9 different inputs were varied simultaneously and outputs were measured. To measure the performance, we use the following metrics: accuracy, HIT acceptance/rejection, accepting time, working time, and completion time. Table 3 lists the experiment settings of the screening test and the corresponding results. There are 30 replications for each setting, i.e., 30 HITs were posted with identical settings for each of the 12 rows in Table 3. In some settings, not all tasks were accepted by workers. The averages shown below were computed using the completed subset.

Table 3: Settings of Screening Test and Results

	Experiment Input Settings									Outputs	
Row #	Reward	Video speed factor	Video duration (min)	Weekday	Country US	Minimum approval rate	Time of day US	HIT lifetime (hr)	Assignment duration (hr)	Mean Accuracy	Mean Completion Time (min)
1	0.05	16	5	-1	-1	0	8	3	1	0.825	120
2	0.05	16	5	-1	-1	90	20	72	24	0.866	447
3	0.05	16	10	1	1	0	8	3	24	0.717	110
4	0.05	32	5	1	1	0	20	72	1	0.939	2152
5	0.05	32	10	-1	1	90	8	72	1	0.941	2064
6	0.05	32	10	1	-1	90	20	3	24	0.964	197
7	0.5	16	10	1	-1	0	20	72	1	0.922	150
8	0.5	16	10	-1	1	90	20	3	1	0.918	51
9	0.5	16	5	1	1	90	8	72	24	0.94	137
10	0.5	32	10	-1	-1	0	8	72	24	0.861	1259
11	0.5	32	5	1	-1	90	8	3	1	0.839	54
12	0.5	32	5	-1	1	0	20	3	24	0.973	37

The linear regression coefficients of the 9 inputs are shown in Table 4, where the highlighted coefficients in bold show the variables that have significant effect on the outputs. Five out of the nine inputs are chosen to conduct a fractional-factorial design based on the impacts on the output metrics. The selected 5 inputs are incentive, video speed, country, HIT lifetime, and assignment duration. We use this design to create a testing set data that will be used to validate the model created earlier. This is not a conventional use of this design either; however, it serves the purpose of probing the experimental space to capture the interactions of the key input variables selected by the screening design.

4.3 Fractional Factorial Experiment

Table 5 lists the experiment settings of the 5 inputs and the corresponding outputs. The remaining inputs of the screening design have been set to the following values in this design: Video duration (5 mins), Weekday (1), Minimum Approval rate (0%), Time of day US (8-20). There are 25 replications for each setting.

Table 4: Regression Table of Screening Test

	Accuracy	Acceptance	Accepting Time	Completion Time
Incentive	**0.03265**	**0.13333**	**-19842.5**	**-17270.3**
Video speed factor	0.00557	-0.02778	**22236.7**	**24649.9**
Video duration (min)	-0.02088	**-0.12778**	2447.25	4683.40
Weekday	-0.02107	**-0.12778**	-3233.10	-5633.48
Country US	**0.03382**	0.0000000	**13425.5**	**11188.9**
Minimum approval rate	-0.00172	0.00556	-1499.08	-3966.67
Time of day US	0.01628	-0.02778	-290.80	-2643.05
HIT lifetime (hr)	**0.03630**	**0.16667**	**25205.1**	**27444.4**
Assignment duration (hr)	-0.02195	**-0.16111**	**-13649.6**	**-11269.0**

Table 5: Settings of Fractional Factorial Test

	Experiment Settings					Results	
Row	Incentive	Video speed	Country (US)	HIT lifetime (hr)	Assignment duration (hr)	Mean Accuracy	Mean Completion Time (min)
1	0.1	16	-1	3	12	0.812	48
2	0.1	16	-1	36	1	0.842	76
3	0.1	16	1	3	1	0.999	14
4	0.1	16	1	36	12	0.927	233
5	0.1	32	-1	3	1	0.696	92
6	0.1	32	-1	36	12	0.993	302
7	0.1	32	1	3	12	0.885	101
8	0.1	32	1	36	1	0.863	274
9	1	16	-1	3	1	0.927	45
10	1	16	-1	36	12	0.971	141
11	1	16	1	3	12	0.924	66
12	1	16	1	36	1	0.954	37
13	1	32	-1	3	12	0.924	138
14	1	32	-1	36	1	0.905	70
15	1	32	1	3	1	0.973	28
16	1	32	1	36	12	0.947	94

In the next section, the behavior rules of crowdsourcing workers are derived based on the data collected. Then an ABM is developed to implement the behavior rules of workers.

4.4 An ABM of Crowdsourcing

Agent based modeling (ABM) is an ideal way to study complex systems because even a simple ABM can exhibit complex behavior patterns and provide valuable information about the dynamics of the real-world system (Bonabeau 2002). We develop an ABM for a crowdsourcing platform that is composed of three components: requester, task, and worker. Some environmental factors are built into the structure of the system model. In our case of posting parking video on AMT, the behavior of the requester is just submitting tasks at a given hour. Task has properties such as reward, assignment duration, lifetime etc. but no activities. The main effort in our building an ABM for a crowdsourcing platform is to accurately capture the behavior rules of crowdsourcing workers. As shown in Figure 3, a worker firstly searches for a task that he/she is interested in. After such a task is found, the worker makes a decision whether or not to accept it. If the task is accepted, the worker starts working on the task. Otherwise, the worker keeps searching for another task. After the worker finishes the task, the worker submits it for evaluation by the requester. These four actions of a worker corresponds to four performance metrics respectively, that is, accepting time, acceptance, working time, and accuracy. We have also observed cases where the workers accept multiple tasks and work on them at a later time. This behavior is not captured in this simplified model.

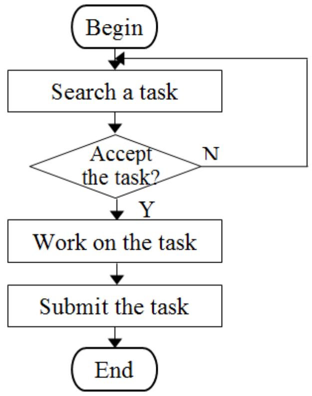


Figure 3: The Behavior of a Worker

Figure 4 presents the box plot of accuracy across the settings of the screening test in section 4.2. The x-axis corresponds to the 12 settings in the screening test as shown in Table 3. The question is how to build the relationship between the inputs of each setting and the accuracy. One way to capture the relationship between the inputs and the outputs is via a linear regression. If a linear regression is used, then the output will be a single point as shown in Figure 5(a). This figure shows the histogram of the first setting in Figure 4 and the red point is the output of the regression model. One of the drawbacks of using linear regression is that this cannot capture the variety of accuracy within a single setting.

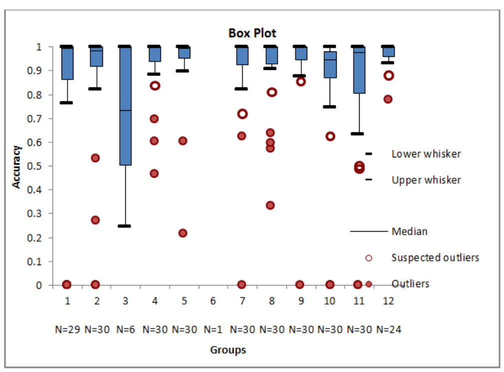
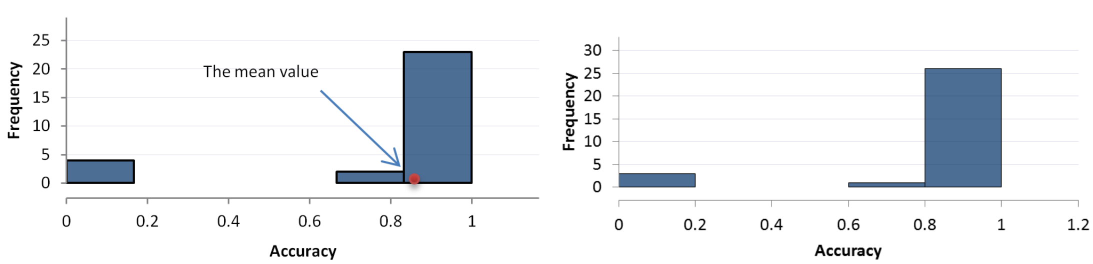


Figure 4: Accuracy in the Screening Test



(a) (b)

Figure 5: Histogram of (a) Real and (b) Simulated Accuracy of the First Setting in the Screening Test

In order to capture the variation of accuracy within a single setting, we use the third item mentioned in Section 3. First, we use a distribution to fit the data and check if that distribution is unimodal or multimodal. After examining multiple distributions, we noticed that the unimodal Lognormal distribution better fits the accuracy. For the first setting of the screening test, Lognormal(-6.72, 4.07) is selected, since the P value of χ^2 test is lower than 0.005. Lognormal distribution has two parameters, μ and σ. Once all the two parameters of the Lognormal distribution are determined for each input setting listed in the experiment, then a linear regression model is built between the inputs and the parameters of the selected distribution (Gil and Zou 2014). Since there are two parameters per input, two linear regression models will be built for μ and σ respectively. The statistical model that captures workers' accuracy is shown in Figure 6, where the R^2 value is 0.81. Under this approach, the output accuracy given for the first setting of the screening test is shown in Figure 5(b). By comparing Figure 5(b) to Figure 5(a), we are confident that the model of accuracy is reasonable capable of reflecting how the accuracy of workers' submissions changes along with the settings of tasks.

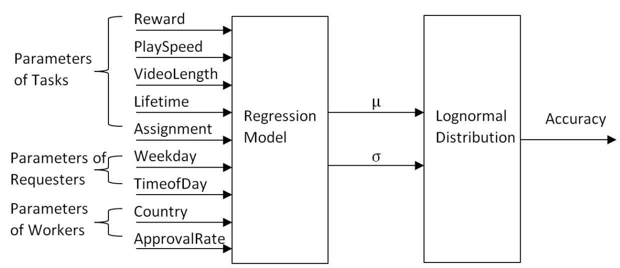


Figure 6: A Behavior Model of Accuracy

Table 6 lists the linear regression coefficients for accuracy, where the coefficients after calibration will be described in Section 4.5. The same approach applies to accepting time and working time, while a variant is applied to acceptance. Since acceptance is a binary number (1 means acceptance, 0 means decline), the distribution parameters (μ and σ) in Figure 6 are replaced by a threshold. Then if a random number drawn from a uniform distribution, U(0,1), is greater than the threshold, the worker accepts the task. Otherwise, the worker declines the task. The threshold needs to be fine-tuned so that the simulated acceptance matches the real data. After the threshold is determined, a similar linear regression model is built between the inputs and the threshold. Thus, the behavior rules of workers are built based on the experimental data collected on AMT.

Table 6: Coefficients of Behavior Model of Accuracy

	Before Calibration		After Calibration	
Factors	Accuracy μ	Accuracy σ	Accuracy μ	Accuracy σ
Const	-8.710	8.969	-8.736	8.301
Incentive	-1.580	0.54765	-1.673	0.530
Video speed factor	0.03999	-0.09460	0.036	-0.092
Video duration	0.31065	-0.10372	0.355	-0.095
Weekday	0.13592	0.06203	0.163	0.063
Country US	-0.48186	0.52922	-0.532	0.403
Approval rate min	0.00338	-0.01212	0.003	-0.010
Time of day US	-0.06305	-0.08359	-0.068	-0.086
HIT lifetime	-0.00748	-0.00358	-0.007	-0.004
Assignment duration	0.05992	-0.01301	0.055	-0.017

Figure 7 is a snapshot of the simulation model built for the AMT platform. The left hand side blue point represents a requester who posts tasks on AMT. The middle layer represents tasks that are taken by workers on the right hand side. Since it is a snapshot when the simulation model is running, there are some empty spaces in the line of tasks. The empty spots mean those tasks have been finished by workers at that point in the simulation. Workers follow the behavior rules derived in this section. Workers are green when they are idle, and turn red when they are busy. The model is developed using RePast Symphony.

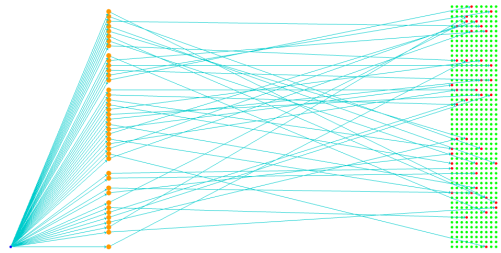


Figure 7: Snapshot of Simulation Model for AMT

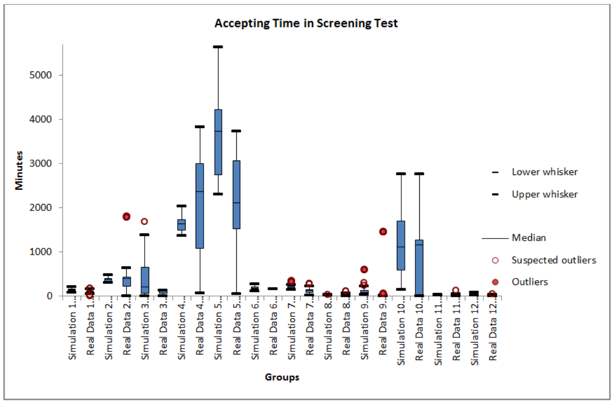
4.5 VERIFICATION AND VALIDATION

The evaluation of a simulation model involves two major activities, one is verification, and the other is validation that includes conceptual and operational validation (Banks 1998). Verification is the process of determining that a computer model, simulation, or federation of models and simulation code and their associated data accurately represent the developer's conceptual description and specifications (Agency 2008). To achieve this goal, unit and integration tests are used. Unit testing involves determining the correctness of the simulation program at the function level. All functions are tested using boundary and error conditions, and the outputs are observed for consistency against expected regularities. Integration testing is the activity of software testing in which individual software modules are combined and tested as a group (Wikipedia 2010). The ABM built in Section 4.4 is verified through both unit test and integration test.

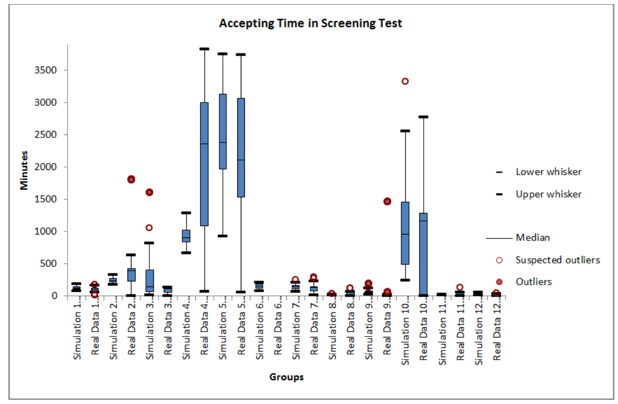
Validation substantiates the accuracy of model's behavior against the system behavior for its intended purpose and domain of applicability. (Rykiel 1996) Without loss of generality, we have developed a framework based on Genetic Algorithms (GA), since GA is capable of searching a large feasible domain. The factors to be tuned include the linear regression coefficients for accuracy, acceptance, accepting time, and working time respectively. After encoding the tuning parameters into chromosomes, the objective of validation is to find the best parameters so that discrepancy between the target metrics and the observed metrics is minimized. This process can be finished automatically. The coefficients for accuracy after calibration are listed in Table 6.

Figure 8(a) displays the comparison of the real data and the simulated accepting time using the coefficients before calibration. The X axis shows 12 pairs of simulated outputs and real data, which correspond to the 12 input settings in the screening test. Figure 8(b) is the comparison of accepting time using the GA-tuned coefficients. By comparing Figure 8(a) to Figure 8(b), we observe that for the 5th setting, the simulation outputs using GA-tuned coefficients are much closer to the real data than the simulation using original coefficients. Similar results are available for the other tests including acceptance, working time, and accuracy.

The simulation model is trained using the data obtained from the screening test. Then it is validated by comparing the results to the data of 5-factorial test, as shown in Figure 9. We observe that the simulation outputs fall into or close to the range of real data. It indicates that the model presented in the paper reflects the dynamics of the AMT platform. Similar things are observed in accepting time, working time, and accuracy, as shown in Figure 9(a), Figure 9(b) and Figure 9(c) respectively. For most settings of the 5-factorial test, simulation outputs are close to the real data except the setting 6 that corresponds to the sixth row of Table 5. It is clear that more efforts are needed for the specific setting. Nevertheless, the model captures the main trends of the AMT platform.

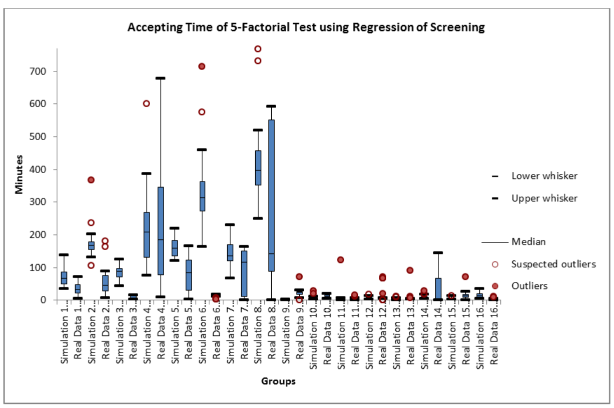


(a)

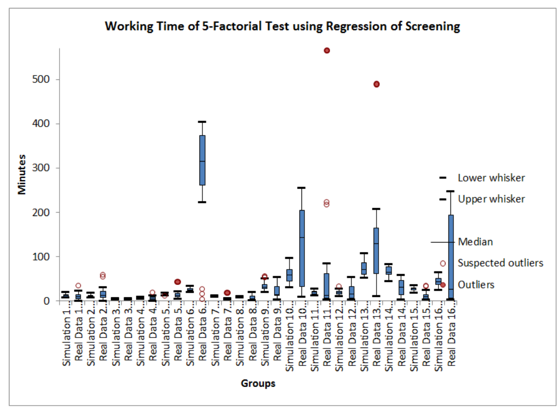


(b)

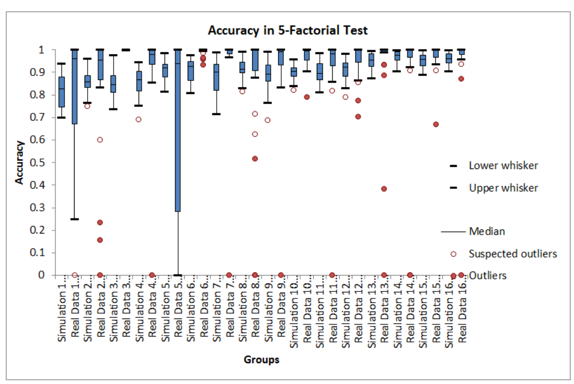
Figure 8: Comparison of Accepting Time Before and After Calibration



(a)



(b)



(c)

Figure 9: Accepting Time, Working Time and Accuracy of 5-Factorial Test using the Coefficients of Screening Test

5 CONCLUSIONS

In this paper, we have shown how to derive behavior rules from workers that can be incorporated into an ABM. We performed experiments on AMT to derive these rules by varying the configuration of tasks in the experiment to gain insights on how crowdsourcing workers respond to these changes. The wide variation on workers' behaviors has been captured using probabilistic models that dynamically change given specific combinations of the tasks. We have also used Design of Experiments to plan the experiments conducted on AMT so that we can identify the main task characteristics in the model that have more influence on the performance metrics. Furthermore, we have used Design of Experiments to define how to collect testing data to validate the ABM such that the number of task characteristics utilized for this set is minimized. All these steps combined result in a way to capture the relationships among properties of tasks, characteristics of workers, and performance metrics via an ABM. This model can be used as a tool for task design, forecasting, and scenario planning. It can also be tuned to capture the behavior of other crowdsourcing platforms, or combination of platforms required to complete jobs in parallel.

ACKNOWLEDGEMENT

The research presented in this paper was conducted when G. Zou worked at Xerox Research Center Webster.

REFERENCES

Agency, M. D. 2008. "Department of Defense Documentation of Verification, Validation & Accreditation (VV&A) for Models and Simulations". Technical report, Missile Defense Agency.

Banks, J. 1998. *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*. Wiley-Interscience.

Bernstein, M. S., D. R. Karger, R. C. Miller, and J. Brandt. 2012. "Analytic Methods for Optimizing Realtime Crowdsourcing". In *Proceedings of Collective Intelligence conference*.

Bonabeau, E. 2002. "Agent-based modeling: Methods and techniques for simulating human systems". *Proceedings of the National Academy of Sciences of the United States of America*.

Bcheler, T., R. Fchslin, R. Pfeifer, and J. Sieg. 2010. "Crowdsourcing, Open Innovation and Collective Intelligence in the scientific method: a research agenda and operational framework". In *Proceedings of Artificial Life XII – Twelfth International Conference on the Synthesis and Simulation of Living Systems, Odense, Denmark*, 679–686.

Bcheler, T., and J. H. Sieg. 2011. "Understanding Science 2.0: Crowdsourcing and Open Innovation in the Scientific Method". *Procedia Computer Science* 7:327329.

Bcheler, T. A., R. Lonigro, R. M. Fchslin, and R. Pfeifer. 2011. "Modeling and Simulating Crowdsourcing as a Complex Biological System: Human Crowds Manifesting Collective Intelligence on the Internet". In *Proceedings of the Eleventh European Conference on the Synthesis and Simulation of Living Systems*, 109 116.

Chesbrough, H. W. 2003. *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Havard business School Press.

Chu, E. T.-H., Y.-L. Chen, J. W. S. Liu, and J. K. Zao. 2011. "Strategies for crowdsourcing for disaster situation information". *Transactions of the Wessex Institute*.

Estells-Arolas, E., and F. G.-L. de Guevara. 2012. "Towards an integrated crowdsourcing definition". *Journal of Information science* 38 (2): 189–200.

Faridani, S. 2012. "Models and Algorithms for Crowdsourcing Discovery". *Chapter 3 of dissertation in Industrial Engineering and Operations Research. University of California, Berkeley*.

Alvaro E. Gil and Guangyu Zou 2014. "Modeling of Crowdsourcing Platforms Using Performance Distributions". Patent Pending. Attorney Docket Number 20131501US01.

Huang, E., H. Zhang, D. C. Parkes, K. Z. Gajos, and Y. Chen. 2010. "Toward automatic task design: a progress report". In *Proceedings of ACM SIGKDD workshop on human computation*, 77–85.

Khazankin, R., D. Schall, and S. Dustdar. 2012. "Predicting QoS in Scheduled Crowdsourcing". *Advanced Information Systems Engineering Lecture Notes in Computer Science* 7328:460–472.

Krause, J., G. D. Ruxton, and S. Krause. 2010. "Swarm intelligence in animals and humans". *Trends in Ecology and Evolution* 25 (1): 28–34.

Martin, D., B. V. Hanrahan, and J. O'Neill. 2014. "Being A Turker". In *Proceedings of CSCW '14*.

Mason, W., and D. Watts. 2010. "Financial incentives and the performance of crowds". *ACM SIGKDD Explorations Newsletter* 11 (2): 100–108.

Rajan, V., S. Bhattacharya, E. Celis, D. Chander, K. Dasgupta, and S. Karanam. 2013. "CrowdControl: An online learning approach for optimal task scheduling in a dynamic crowd platform". In *Proceedings of ICML Workshop: Machine Learning Meets Crowdsourcing*.

Rykiel, J. 1996. "Testing ecological models : the meaning of validation". *Ecological Modeling*.

Scekic, O., C. Dorn, and S. Dustdar. 2013. "Simulation-Based Modeling and Evaluation of Incentive Schemes in Crowdsourcing Environments". In *OTM 2013 Conferences Lecture Notes in Computer Science Volume 8185*, 167–184.

Schmidt, S., and R. Launsby. 2005. *Understanding Industrial Designed Experiments*. Air Academy Press.

Wikipedia 2010. "Integration testing". http://en.wikipedia.org/wiki/Integration_testing.

Yuen, M.-C., I. King, and K.-S. Leung. 2011. "A Survey of Crowdsourcing Systems". In *Proceedings of IEEE International Conference on Social Computing*.

Guangyu Zou and Marina Tharayil and Alvaro E. Gil and Deepthi Chander and Laura Elisa Celis 2013. "System and Method to Analyze and Optimize Crowdsourcing Systems". Patent Pending. Attorney Docket Number 20130880US01.

Zou, G., and L. Yilmaz. 2011. "Dynamics of knowledge creation in global participatory science communities: open innovation communities from a network perspective". *Computational & Mathematical Organization Theory* 17 (1): 35–58.

AUTHOR BIOGRAPHIES

GUANGYU ZOU is Assistant Professor at Dalian University of Technology. Before joining DLUT, he was a research scientist at XRCW. He holds PhD in Computer Science and MS in Industrial Engineering from Auburn University. His research interests are in modeling and simulation, agent-based modeling, and complex adaptive systems. His email address is gyzou@dlut.edu.cn.

ALVARO GIL received his B.S. and M.S. degrees in electrical engineering from Instituto Universitario Politcnico, Barquisimeto, Venezuela, in 1990 and 1998, respectively, and the Ph.D. degree in electrical engineering from The Ohio State University, Columbus, in 2003. From 1990 to 1999, he held engineering positions in the Automation Department at Petrleos de Venezuela (PDVSA) in Maracaibo, Venezuela. From 2002 to 2003 he was a research associate at the Department of Electrical Engineering, The Ohio State University. He worked as a postdoctoral researcher at Ohio State University between 2003 and 2005. He is now a Research Scientist at Xerox Corporation, Webster, NY. His current research interests include cooperative scheduling, anomaly detection via dimensionality reduction, failure diagnostics and prognostics, and multiobjective control. His email address is Alvaro.Gil@xerox.com.

MARINA THARAYIL is a Research Competency Manager of Systems Design and Controls group in the Xerox Innovation Group. She joined XRCW in 2005 as a research scientist working with the centers Advanced Media Handling group where she specialized in mechatronics. Her work with Xerox includes paper path design and control for various Xerox printers. She led the media registration technology development effort for the iGen3 digital press family, which resulted in the improved registration performance in the iGen4 and subsequent models. Her PhD research from the University of Illinois at Urbana Champaign was in Iterative Learning Control (ILC) and Repetitive Control (RC). Her email address is Marina.Tharayil@xerox.com.