Conceptualizing Multidimensional Reputation Evaluation Model for Crowdsourcing Participants: Blockchain Design Perspective

Research-In-Progress

Muhammad Nauman Shahid
Department of Information Systems and Analytics,
National University of Singapore
nauman@comp.nus.edu.sg

Abstract

Given the weak mechanisms to display real reputation status of crowdsourcing participants (MTurkers), the existing reputation evaluation models of crowdsourcing platforms have several problems arising from centralized and simple evaluation systems and poor differentiation ability. We conceptualize a multidimensional reputation evaluation using concepts of Blockchain and collect indexes of crowdsourcing participants from the perspective of rewards and punishments based on four dimensions: initial reputation dimension, transaction dimension, evaluation dimension and fraud penalty dimension. Our findings based on the empirical analysis of data of participants from the "Mechanical Turk" platform verifies the comprehensiveness, dynamicity, and accuracy than the existing models of a crowdsourcing platform.

Keywords: Blockchain, Crowdsourcing, Evaluation Model

1. Introduction

The popularization and advancement of modern internet technology have enabled real-time information sharing and unbounded crowd interaction. While some crowds use multimedia and gaming in their free time to create value, others use it transforms their wisdom, knowledge, ability, and experience into monetary benefits. Many web crowdsourcing platforms such as ClickWorker, CrowdFlower, Toluna, and Amazon Mechanical Turk (MTurk) have emerged in the last decade. Organizations outsource their tasks to the crowd workers on these platforms and use their wisdom to solve problems in a short time. MTurk has evolved as one of the famous crowdsourcing platforms with crowdsourcing participants exceeding 0.5 million and the annual turnover exceeds 30 million US dollars.

Crowdsourcing is an important form of mass entrepreneurship and innovation-driven development due to the virtual nature of the internet and asymmetric information sharing between the participants. However, malicious actors manipulate the network and often commit transaction frauds worth millions of dollars more likely leading to the "lemon phenomenon" and effect the growth of the market. At present, many irregularities exist in the trading mechanism and despite corrective interventions, the effect is insignificant and online reputation problem has increasingly become a hindrance in the healthy development of MTurk. The work submitted by the crowd workers (specifically MTurkers on MTurk) is an intangible commodity and it is impractical to detect the development process and replace the submitted work. Therefore, it is necessary to establish a reliable reputation evaluation mechanism to ensure work quality.

The development of blockchain by Nakamoto (2008) has disrupted many areas including the way digital trust can be established. Blockchain provides a way to record and share temper-resistant information because the entries are immutable, transparent, and searchable, that makes it easy to look for a transaction. The immutability of record and automatic execution of computer algorithms — smart contract — on pre-agreed terms make it ideal for digital ecosystems. In reputation systems, it can ensure information integrity, trustworthiness, and a common standard for value creation. In addition, as a pre-cursor, a conceptual design of reputation evaluation inspired by existing reputation evaluation systems may be established to lay the foundation for future implementation. Therefore, this paper establishes a multi-dimensional reputation evaluation system at the participant's level to guide further research and

development in crowdsourcing context. In the next section, we present a brief overview of related work followed by our proposed model, conclusions, and future work.

2. Related Work

2.1 General Commentary

The present research in e-commerce focuses on tangible commodities, while the research on intangible commodities is less. In particular, fraud prevention (e.g., Dai et al. 2016) and work quality (e.g., Xie et al. 2016) are key areas of inquiry in extant research. These researchers use game theory and aspects of cognitive theories to posit reputation evaluation schemas to prevent malicious attitude and false data reporting in mobile crowdsourcing.

Ruan et al. (2015) developed a model to distinguish between malicious and honest participants by implementing several features in the code of the proposed system. Bhattacharjee et al. (2017) use an algorithm that uses a regression model to determine the honesty of an actor in the system. Yan et al. (2017) proposed a reputation model based on the work completion time whereas, Liu et al. (2012) combine objective measurement and subjective measurement in a coherent evaluation model.

Several researchers focus on the improvement of quality and balance of system efficiency and transaction cost to be the best determinants of a reputation model. For example, Rui et al. (2016) use game theory to develop a reputation model for crowdsourcing participants. Ma et al. (2016) use game theory-driven incentives mechanism to encourage authentic and credible information sharing. In sum, extant research has a rich theoretical basis for evaluating the reputation of crowdsourcing participants. However, the research on fraud prevention lacks to touch upon the overall reputation status of crowdsourcing participants.

The literature based on the quality of tasks and their completion encourages individual participation to improve the reputation rather than developing techniques to screen systematically and technically the malicious participants. The intangible nature of crowdsourcing widens the gap in research to monitor the quality and only provides the ability to validate artificially the model by simulation and experimentation. Blockchain provides a fruitful avenue to develop incentives mechanism together with enhanced trustworthiness, punishment for malicious actors based on given consensus mechanism and prevents fraud. Therefore, we construct a four-dimensional – initial reputation, transaction, evaluation and fraud penalty – blockchain-based reputation evaluation model for crowdsourcing participants. We collected data from Amazon's Mechanical Turk platform to investigate empirically the reputation evaluation performance of BMREM model and existing models of a crowdsourcing platform to verify the validity of BMREM model.

2.2 Reputation Model of MTurk

The current model of Amazon's Mechanical Turk uses simple accumulative reputation evaluation model (MTurk SAREM) to determine overall reputation. It considers two factors – 1) transaction amount, 2) transaction evaluation. It has poor ability to distinguish reputation status of participants and feedback the overall reputation of the participants. Mathematically, SAREM is illustrated in equation 1. R_p represents

$$R_p = \sum_{i=1}^{n} r_i . P_i(tr, p) \qquad \text{…Eq. 1}$$

In the equation, the value R_P represents the reputation score of crowdsourcing participant p, r_i represents the transaction evaluation coefficient of participant p given by task requester tr in i^{th} transaction, $i \in (1, 2, \ldots n)$.

There are five main problems in this model. First, new participants are difficult to get tasks due to non-consideration of initial reputation. Second, the credibility of task requester is difficult to determine from the participants perspective. Third, the activity of all participants varies in different times and recent activity is more credible and may differently affect the reputation coefficient, therefore, the time factor is important but neglected. Fourth, the difference between tangible and intangible commodities is often neglected in crowdsourcing platforms that are a form of e-commerce

service. Fifth, no concrete mechanism in place to stop malicious acts such as transaction frauds and unavailability of punishment mechanism.

3. Proposed Model

We collect the transaction data of MTurk from an approved resource and examine the patterns of data, enhance the index features of it, and evaluate it based on its descriptive ability. Therefore, developing on the existing research (e.g., Huang and Chen 2019) and exploring the transactional characteristics of intangible service commodities (such as crowdsourcing), we propose Blockchain Multidimensional Reputation Evaluation Model (BMREM) for crowdsourcing participants. The model is conceptualized hypothetically based on MTurk's data given the reputation system is conceptualized on a blockchain ecosystem. We consider four index dimensions to discuss the model.

3.1 Blockchain Multidimensional Reputation Evaluation Index System

3.1.1 Initial Reputation Dimension: The initial reputation dimension prevents the participants from doing fraudulent transactions thence increasing the probability for new MTurkers to get tasks. This dimension is characterized by four first-level indexes: identity authentication, trustworthy guard, transaction margin, and integrity and credibility of store information. In sum, these four dimensions are refined into ten first-level indexes as shown in Table 1.

3.1.2 Transaction Dimension: The transaction dimension is divided into three first-level indexes: transaction amount, transaction time, and transaction node. The crowdsourcing participants adopt good faith due to the inherent trust enabled by blockchain transactions. However, repeated transaction nodes that supplement both the initial reputation dimension and the transaction dimension are highly credible.

3.1.3 Evaluation Dimension: The evaluation dimension has three indexes: transaction evaluation, evaluator credibility, and evaluation factors. Transaction evaluation means, the interpretation of quality by the task requester after the final submission of work by a participant; the evaluator credibility reflects the credibility of task requestor; theory assumes that the evaluation given by high-credibility task requester is more credible. Combining with the characteristics of intangible commodity transactions, the task completion quality can be determined from the smart contract of after-work quality.

Table 1. Blockchain Multidimensional Reputation Evaluation Index System for Crowdsourcing Participants

Dimension	First level index	Second level index
Initial reputation dimension	Identity authentication	Node-device authentication, Node-user authentication
	Trustworthy guard	Smart contract for previous completions, Smart contract for originality, Smart contract for after-sale guarantee
	Transaction margin	Margin payment
	Integrity and credibility of store information	Case presentation, Sale service introduction
Transaction dimension	Transaction amount	-
	Transaction time	-
	Transaction node	Node block identifier
Evaluation dimension	Transaction evaluation	-
	Evaluator credibility	-
	Evaluation factors	Smart contract for after-work quality
Penalty dimension	Fraud penalty factor	-

Table 2. Parameter interpretation in BMREM model

Parameter	Interpretation
$R_t(p)$	Reputation score of crowdsourcing participant p at t time
$R_{t-1}(p)$	Reputation score of participant p at t-1 time
$R_{t-1}(tr)$	Reputation score of task requester tr at t-1 time
$I(p)$	Initial reputation score
α_{a_j}	Weight of the a_jth index in initial reputation

β	Initial reputation adjustment parameter
$d(t_v, p)$	Time discounting equation
$P(p, tr)$	The transaction amount between crowdsourcing participant p and task requester tr
$f(p, tr)$	Transaction evaluation score is given by task requester tr
$\rho[R(tr)]$	Reputation of task requester tr
$K_{m_i}(p, tr)$	A score of evaluation factor Si given by task requester tr
ω_{s_i}	Weight of the evaluation factor Si
$\mu(K_{m_i}, \omega_{m_i})$	A comprehensive score of evaluation factors
ϑ	Time discounted adjustment parameter
$N(tr)$	Task requester collection
$\varphi_t(p)$	Fraud penalty equation

3.1.4 Fraud Penalty Dimension: The fraud penalty dimension is negatively related to the credibility of crowdsourcing participants. The fraud coefficient is directly proportional to the fraud penalty factor coefficient and inversely proportional to the reputation of crowdsourcing participants.

3.2 The Blockchain Multidimensional Reputation Evaluation Model

In light of the aforementioned analysis, we construct definitions and present the model equation.

Def. 1: While p represents crowdsourcing participant, $Rt\,(p)$ denotes the reputation score of a crowdsourcing participant at time t.

Def. 2: The term $N\,(tr)$ is task requester collection transact with crowdsourcing participant p, if the task requester tr deals with crowdsourcing participant p, then $tr \in N\,(p)$, $f\,(p, tr)$ is the transaction evaluation of crowdsourcing participant p given by task requester tr after the transaction.

Def. 3: The set M as a set of evaluation factors in the evaluation dimension, $M = m1, m2, m3 \ldots mn$, M includes n evaluation factors, $M_i \in m1, m2, m3, \ldots mn$, $K(tr, p)$ is the score of evaluation factors of crowdsourcing participant p given by task publisher tr, $K_{m_i}(tr, p)$ represents the score of the evaluation factor m_i which given by task requester for crowdsourcing participant p, $K(tr, p)$ is an n-dimensional vector.

Def. 4: The term $d\,(t_v, p)$ represents the time discounting equation, t_v shows task requester tr and crowdsourcing participant p traded at t time, $t_v \in (t-1, t)$. The specific model parameters shown in Table 2. From the above definition and model parameters, the BMREM model of crowdsourcing participant p at time t represented as:

$$R_t(p) = \int R_{t-1}\,(p) + d(t_v, p) P(p, tr) f(p, tr) \rho[R(tr)] \mu\left(K_{m_i}, \omega_{m_i}\right) + \varphi_t(p) \qquad N(tr) \neq 0$$

$$R_{t-1}(p)\; N(tr) = 0 \quad \ldots \text{Eq. 2}$$

4. Empirical Validation

The data used in this paper come from a reliable source within Amazon organization. We evaluated not less than 12,734 transaction samples and validated the integrity of data by removing duplicates and redundancies. The cleaning process retained 8,344 samples. Each sample had several indexes categorized by four statistical levels — user, transaction, demographics, and career. The MTurk website uses SAREM to collate the reputation information however; we tested the data to both the proposed model and on SAREM. The main result evaluation parameters used are an incentive and punishment perspective. The empirical validation assumes initial index weights for each parameter in the Eq. 2.

Initial results show that the proposed model performs exceptionally well as compare to SAREM model that lacks evaluation criteria for new participants desiring to establish themselves on MTurk. Task requestors usually choose

MTurkers who are experienced and have completed the most tasks. The integrity status of new participants is measured by indexes shown in Table 1 therefore, mutual trust is reasonably established that benefits both parties.

Table 3 shows the reputation scores of crowdsourcing participants across both models. The participant p_4 has the highest initial reputation score of 956 whereas p_3 and p_8 due to non-compliance to initial authentication dimension, have the lowest initial reputation score of 273 and 322 respectively. It is evident from the perspective of transaction dimension and evaluation dimension, MREM is more effective than SAREM model.

Table 3. Reputation score table for participants samples

Participants	IRS	Total Tran. In USD	Tran. Time	Praise Rt.	Avg. Score of Eval. Fac.	$t_0 - t$	Fraudulent times	MREM score	SAREM score
p_1	859	598	315	0.95	1.00	0	0	13.6	12.9
p_2	698	744	55	0.66	0.96	0	0	9.16	13.6
p_3	273	563	16	0.11	0.71	2	3	0.40	18.5
p_4	956	236	59	0.56	0.95	1	0	14.2	30.9
p_5	799	596	158	0.59	0.95	4	0	13.9	40.9
p_6	568	524	9	0.89	0.96	0	1	31.0	40.2
p_7	668	458	25	0.88	0.99	0	0	23.6	4.55
p_8	322	448	56	0.83	1.00	1	0	29.6	32.6
p_9	858	635	55	0.59	0.70	2	1	30.0	22.3
p_{10}	823	889	129	0.23	0.84	3	0	29.3	10.2

5. Preliminary Findings and Future Work

We proposed a Blockchain Multidimensional Reputation Evaluation Model (BMREM) for a famous crowdsourcing platform. The model regulates the overall participant behavior and ensures fair evaluation of new entrants based on the four dimensions. In the future, we set three main goals for further research. First, illustrate each dimension mathematically and develop an algorithm for simulation. Second, evaluate the generalizability with and without blockchain context on different data sets and compare the robustness in each case. Third, we will perform an analysis of effectiveness from both incentives perspective and fraud detection perspective.

References

Bhattacharjee, S., Ghosh, N., Shah, V. K., and Das, S. K. 2017. "QnQ: A Reputation Model to Secure Mobile Crowdsourcing Applications from Incentive Losses," in *2017 IEEE Conference on Communications and Network Security, CNS 2017*.

Dai, W., Wang, Y., Jin, Q., and Ma, J. 2016. "An Integrated Incentive Framework for Mobile Crowdsourced Sensing," *Tsinghua Science and Technology*.

Huang, Y., and Chen, M. 2019. "Multidimensional Reputation Evaluation Model for Crowdsourcing Participants Based on Big Data," in *2019 International Conference on High Performance Big Data and Intelligent Systems, HPBD and IS 2019*.

Liu, Y., Wu, Y. J., Peng, X., and Zhao, W. Y. 2012. "Research on Prestige Capability Evaluation Mechanism of Witkey Platform," in *Computer Science*.

Ma, X., Ma, J., Li, H., Jiang, Q., and Gao, S. 2016. "RTRC: A Reputation-Based Incentive Game Model for Trustworthy Crowdsourcing Service," *China Communications*.

Nakamoto, S. 2008. "Bitcoin: A Peer-to-Peer Electronic Cash System," *Consulted*.

Ruan, S. ., Wang, X. P., and Xue, X. P. 2015. "Crowd Sourcing Quality Monitoring Based on Evidence Theory Trust Model," in *Computer Application*.

Rui, L. L., Zhang, P., Huang, H. Q., and Qiu, X. S. 2016. "A Crowd Oriented Incentive Mechanism Based on Reputation Value," *Journal of Electronics and Information*.

Xie, H., Lui, J. C. S., and Towsley, D. 2016. "Design and Analysis of Incentive and Reputation Mechanisms for Online Crowdsourcing Systems," *ACM Transactions on Modeling and Performance Evaluation of Computing Systems*.

Yan, J., Ku, S. ., and Yu, S. 2017. "Activity Based Crowdsourcing Worker Reputation Model," in *Computer Application*.