International Journal of Engineering & Technology, 7 (4.19) (2018) 447-453

SPC

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Performance Measurement towards Crowd-Workers Reliability in Crowd Computing using Bayesian Probability Model

Masnida Hussin[1], Nur Aliya Rozlan[2]

Faculty of Computer Science and Information Technology, University Putra Malaysia
**Corresponding Author Email: [1]masnida@upm.edu.my*

Abstract

Crowd computing becomes increasingly popular nowadays to Internet users. It has been chosen as distributed processing mechanism to solve computing problems which offer less policy but better business. Some of advantages in using the crowd computing are significant in time and cost used. In the crowd computing, the users' tasks are completed by crowd workers. These workers have different skill, knowledge and style in finishing the task that allocated to them. Satisfaction on the completing the tasks is really challenging to measure in the crowdsourcing due to its dynamic environment. Furthermore, it is raised when behaviour or commitments of crowd-workers in providing services are started to query, either it can be trusted or not. In this work, a Bayesian probability model utilized to assess workers reliability in crowd computing is proposed. Specifically, we formulate trust factor by using the Bayesian model for indicating the reliability of the available workers in crowdsourcing platform. The process of prediction and hypothesis of workers' commitment are identified to relate with the crowd-sourced computing system. We designed the significant behavioural factors to measure the workers performance towards user satisfaction. We then developed an automation measurement system that used to verify Bayesian formulation towards workers performance. The web-based automation system is able to identify the workers' reliability values according to the input/response from the users. Optimistically, by using Bayesian probability model provides guidelines for designing trustworthy system in crowd sourcing platform.

Keywords**:** *Crowd computing; Worker Reliability; Bayesian Probability Model; Crowd sourcing;*

1. Introduction

Crowdsourcing systems, in which tasks are electronically distributed to numerous "information piece-workers", have emerged as an effective paradigm for solving of computing problems in several domains. The crowdsourcing platform becomes popular because people believe this approach can be a medium for human-powered problem solving. In computer science context, the crowdsourcing is combination word between "crowd" and "source" that means tasks/works accomplished by crowd of people via Internet. Hence, the "crowd" and "source" is specifically can be seen from two different perspectives i.e. the users and workers. From the users viewpoint, their tasks are been completed by the crowd workers in the form of crowd computing. Currently the crowd computing is being used for processing in distributed way with real time tasks such as image classification, video annotation, form data entry, optical character recognition, translation, recommendation, and proofreading [1-3].

Crowd computing is involved three components namely users, workers and crowdsourcing system. The users may issue task on the crowdsourcing systems through the website with precise requirements/deadline and some reward. Once workers read and understand the given task and committed to complete it, they will bid to get the task and execute it diligently. The quality of worker's finish task is significant to the users. In crowd computing, there is no employee-employer relationship between users and worker like the traditional organization or company.

The worker is anonymous where the users have no idea who they deal with but they expect to get the accurate result within expected time. Meanwhile, the crowdsourcing system presents as policy management where it received the users' requests, find the suitable workers and give payment to the worker who has completed the task [4]. The popular crowdsourcing platform includes Amazon Mechanical Turk [5], Crowd Flower [6], and Zen Crowd [7] .

The crowdsourcing resources or workers used and been hired to solve users' problems and able to process large scale of data [2, 8]. It can be a medium for the workers to earn reward or profit. In the using of crowd computing for completing the tasks, it involved several objectives for example to maintain good reputation of workers for making the crowd users efficiently interact and complimentary the crowdsourcing system. Every user's task basically needed to be executed by the workers and the payment is given according to the result of execution. The results can be expressed in successful execution time, resource utilization and minimum waiting time. However, it is not an easy to determine the correctness of the result. The crowd computing needs to verify and cross-check each result after the workers return the tasks. If the results able to meet users' requirements, it then considered the workers have improved their processing reputation. It looks simple scheduling roles, but in reality it is hard to determine the accuracy of services in crowdsourcing [6].

The reliability challenge in crowdsourcing is important due to every task that has been delivered back to the crowd users must be in good quality in regards Service Level Agreement (SLA) [4, 9]. However, there are many factors that affect reliability towards the workers performance. Specifically, the workers that completed the tasks are varies in regards their attitude, knowledge, region and capability. The distances and network medium that used by the workers, and scheduling strategy are also another issue that needs

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

to take into account. From the users' perspectives, it is expected that once the tasks are been assigned to the workers, they should get the results as per agreed in SLA. The task requirements, on the other hand, are varies and dynamic in terms of characteristics and features. Therefore, the workers' reliability becomes a critical issue in crowdsourcing for satisfying the crowd-users.

An effective measurement technique is required to estimate the workers' reliability in order to execute the users' tasks. This work developed reliability measurement towards workers that participant in crowdsourcing by using probability model. It aims to evaluate the level of trustworthy in crowd computing. Specifically, we studied about the factors that affected the workers' performance and how the factors able to measure the reliability of the worker. The remainder of this paper is organized as follows. Section 2 describes related work on reliability in crowdsourcing and other works that used Bayesian in crowd-sourced computing. Section 3 details the models used in the paper. Our Bayesian model for worker reliability is presented in Section 4. The automation system for online measurement is presented in Section 5. Finally, Section 6 concludes the paper.

2. Literature Review

Generally, the crowd computing system comes into the picture when the crowdsourcing system received the task from the task provider. Once the crowdsourcing received the task, it will break the task into a smaller task and invites the crowd workers to participate in completing the task [4]. The crowd computing system is then assigned the sub task to the crowd worker. Once the task is completed, the crowdsourcing is responsible to collect the tasks [1]. The system is verified and validated whether the task meets the users' requirements. If the task workers unable to complete the tasks or the tasks are not completely done, then crowdsourcing reassign the tasks to other crowd workers. In the other hand, if the completion of tasks satisfied and met the users' requirements, the system then begin to process the payment/reward for the crowd worker [9]. The task is then integrated and return to task provider.

The reliability of the crowdsourcing is the popular topic today among the crowdsourcing users. The authors in [8] mentioned that the obstacle in the crowdsourcing platform lies in the problem of information veracity. For example, the individual workers might not be reliable due to several factors. Once the crowd users put the tasks to the crowdsourcing platforms, the users do not know who is solving the work, either the workers had a good or bad criteria. Such criterion is important due to it can reflect directly on quality of results/outcomes. The authors in [3] discussed about distributing tasks in the crowdsourcing platform. They highlighted that the crowd users should investigated the workers by giving several questions related to topic-specific expertise and attitude of workers to measure the level of their commitment. It is one of the approaches in order to get the reliable information from the workers. The topical-influence perspective that discussed in [10] is the close factor to indicate the level of commitment in the participant workers. The crowd users may not make observations independently, but instead, they may be influenced by others when making decisions. The existence of user influence may greatly affect the truth discovery procedure stated by [8] [2]. They also claimed that if only have the requirement to identify the reliability of crowdsourcing, it is still not enough to prove it. The authors in [10] proposed scaling technique for evaluating data quality and worker performance in crowdsourcing. Meanwhile, the authors in [7] measured the reliability of the crowdsourcing by using the combination manual matching. Their algorithms are related to the performance of the workers in delivering the outcome to users. The solution systematizes and automatizes through manual matching by dynamically creating micro matching tasks, and publishing them on a popular crowdsourcing platform. The authors in [11] addressed the problem of poor human workers reliability due to redundancy in completing the tasks. It happen when the platform randomly assigning the same task to multiple workers. They proposed the crowdsourcing applications Boolean Crowd technique to avoid such redundancy happened. It is due to the task redundancy will lead to poor quality data.

The probability model is a suitable method to understand and describe the information in detail. It facilitates step by step procedures from the random scenarios towards the most suitable answer [12]. Due to heterogeneous criteria of crowd workers in the crowdsourcing platform, it is required to assess them using concept of hypotheses. It is due to the workers is dynamically joined and leaved the platform that makes it hard to accurately determine their behaviours. Bayesian rules are the best probability model for verifying and analysing dynamic and big data such as in the crowdsourcing [12,13]. Such rules can be used to predict the behaviour of the workers then identify prospect value to ensure the reliability in the crowdsourcing. The authors in [3] used the probability measurement to guide and make the justification for finding the correct/reliable workers in crowdsourcing.

In [14], they proposed the community-based Bayesian label aggregation model for identifying appropriate worker to be matched with the tasks. It assumed that crowd workers belong to a certain community and conform to the identified types. Each type represents a group of workers with similar confusion matrices. Such matrices have the key latent features are (i) the confusion matrix of each community, (ii) the community membership of each user, and (iii) the aggregated label of each item. Even though they do not highlight the reliability issue, the label aggregation model from Bayesian can be enhanced to meet certain trust feature. The author in [2] presented the incremental Bayesian model to accommodate the effective crowd size and data quality for consensus labelling. They poses two unaddressed challenges for consensus models of crowdsourcing are (i) the large number of potential species makes classification difficult, and (ii) limited crowd availability and the lack of relevant skills among the general public. Their incremental Bayesian model minimizes crowd size by re-evaluating the quality of the consensus label. Some of factor that they considered are the cost, scarce responses and feedback in a timely fashion. The authors in [13] presented the information-theoretic approach for matching and scheduling the tasks into the right agents (workers) using a Bayesian method. In the selection criteria, they proposed the automated algorithm for selecting a cohort of agents (workers) to complete informative tasks, hiring new members of the cohort and identifying those members whose services are no longer needed. Such selection factors improve the accuracy of scheduling decisions while requiring fewer responses from the crowd. In this work, we utilized Bayesian rules for determining value trust of available workers according to several possible factors from the users' responses.

3. The Models

In our system model, we aggregate there main entities in the crowdsourcing environment are workers, crowd computing (as the platform) and tasks/sources (Figure 1). The probabilistic model will be added inside crowd computing to formulate the reliability level of workers. Task scheduling strategy is not been considered in this work; where only information about previous processing/performance are used for evaluating process.

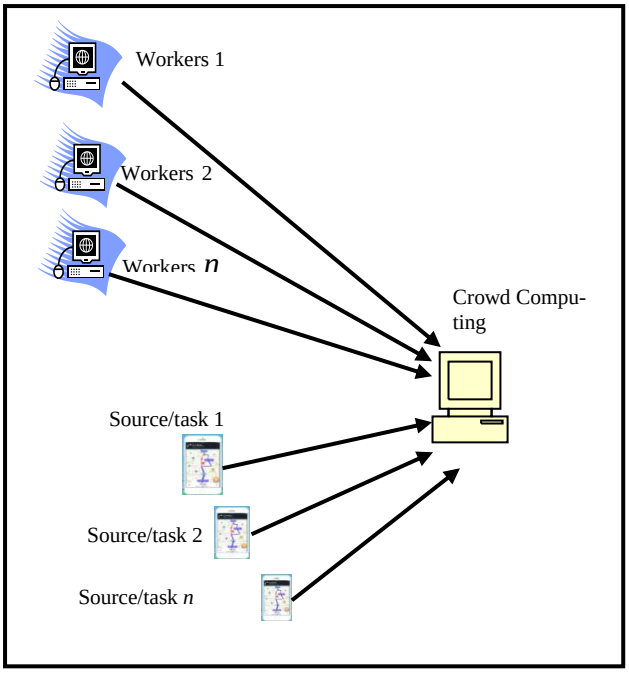


Fig. 1: System Model

Basically, probability model is the mathematical representation from random situation that associated with an event. It describes on how to understand inference about hypotheses from source data. Meanwhile, the Bayesian theorem (Figure 2) represents formation of probability according to the both hypothesis and actual events. The constant A means hypothesis that we make from the actual data, while B representing the actual data. The P(A|B) means to find the combination between the hypothesis and the actual data. The P(B|A) is conditional probability that occurs when certain data is been analysed. P(A) is the probability of data that required to find; where it refers the reliable behaviour in this work. The P(B) refers to all the actual data combination. The generation between actual data and expression will produce a probability value. By using the hypothesis from actual events and make a good prediction from that, it able to give an accurate result. Hence, we employed the Bayes model to determine the reliable value for the participant workers in the crowdsourcing according to the behavioural factors where it represents the hypotheses in our study.

In this work, we utilized the Bayes rules as the probability model to measure reliability in workers. Specifically, the Bayes rules is used to collect and describe the probability of worker's behaviour from the random events to achieve actual condition of the event [10, 15]. There are two significant processes in defining the workers behaviours in our work are hypothesis and prediction that can be formulated as a set of inference. The Bayes rules are suitable to solve a large and complex data like crowdsourcing for finding trust value of workers.

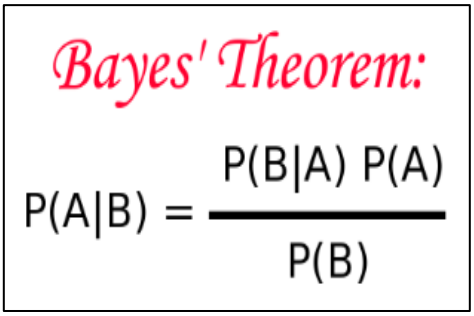


Fig. 2: Bayesian theorem

Specifically, we identified several factors that based crowd computing condition. It is assumed that each worker that participant in the platform is required to give their commitment in completing the tasks. That means the workers cannot leave the platform before they finish executing the tasks. The workers behaviours/factors play an important role in the crowdsourcing platform because worker comes from different side of connection where many probability events can happen. So it is effective to find either the workers in crowdsourcing platform are reliable or not by using factor as element of event. The choosing of the factors for find the workers' reliability is based of study in the crowdsourcing nowadays [2, 8, 9]. In this work, we have been finalise four factors of worker which is the requirement for supporting the condition of event.

- First factor : Attitude of the worker.
- Second factor: Topical influence by the worker.
- Third factor : The effected by cost and distance
- Fourth factor : Task completion time

The first factor is the attitude of the worker. It indicates either the worker provides correct or fake information of itself. The second factor is topical influence. It refers on how the topic or subject that given to the worker had been influence on the results. It is because the workers capability or knowledge is not the same where some workers find the tasks are easy to solve while others experiencing it difficult. The third factor is cost and distance. The factors refers to the distance of the worker within the crowd source located and how it can be affected the results. These factors needed to be considered due to network connection is explicitly involved in crowd computing. Hence, the distance for transmitted tasks and results are much affected by communication links and signals that might leads to increase waiting time and packet dropped. The final factor is time completion. This time or duration for completing the task execution is determining by the source requirements (i.e., priority, due date). The identified factors aims to assess the workers in terms of successfully completing the users' tasks. By assessing the workers through the factors, we then concluded with the value of trust state. It means on how much considerably the crowd users trusted in the available workers in completing the tasks.

4. Bayesian Model for Worker Reliability

In this section, it describes in details the process of the Bayesian as a probability model. Further, we describe in the details on evaluating the value trust towards the workers using the Bayes rules.

4.1. Designing the Probability Model

There are four main procedures for designing the prediction and hypothesis in the probability model as in Figure 3. In the first procedure, all related factors are identified and collected. In the crowdsourcing there are many factors used to measure the worker performance such as reliability, availability and performance [1, 3]. Therefore, in order to design probability model, the fixed factors needed to be identified for achieving accurate results. Furthermore, the probability element in crowdsourcing such as the number of arrival tasks and criteria of task requirements also is important to study due to the precise assumptions are necessity to identify for better-quality formularization.

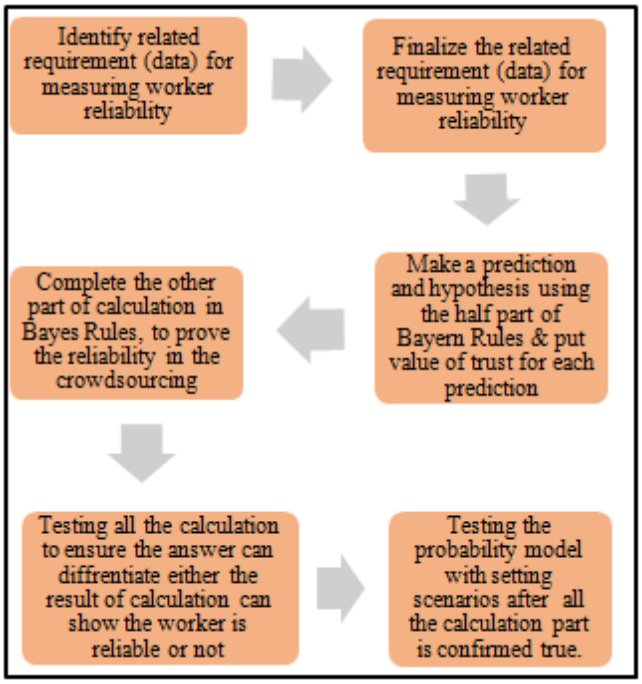


Fig. 2: Steps in designing probability model

In this work, we defined worker reliability based on the four factors that mentioned in Section 3. Basically, all factors aims to meet users' requirements while improving user satisfaction towards crowdsourcing. The input of those factors is assumed gathered after the users received their completed tasks. The available workers in the platform are assumed capable to complete any type of tasks from users. The mapping mechanism used between the tasks and the workers is not considered in this work. The hypothesis and prediction process in designing our probability model is gained from the number of event in the crowdsourcing. Basically, the event can be process from task arrival, task departure and scheduling. However, in this work, we only focus on the feedback event from the users. Such even is received by the platform after the tasks are completed. Due to there are no limitation of giving the feedback towards the workers, it might be a user judged the workers for more than one time. It is also assumed that the feedbacks from one user to another are different and divergent. The probability model using Bayes rules is then tested using automation system. The system aims to test the quality of probability towards worker reliability by exercising the above-mentioned system model (Section 3).

4.2. Assessing the workers

In designing the probability model, it involved several steps of mathematical process, as follows.

4.2.1. First step

In the first step, the requirement (data) or the criteria (factor) that related to the event are collected. We compile all necessity information and data from our reading about crowd-sourced computing environment [1, 2, 5]. In concluded, we identified four factors that most affected to the worker performance. They are according to worker attitude (F1), topical influence (F2), cost and distance (F3) and completion time (F4). Such factors are important in crowdsourcing due to the user satisfaction is the most demanded in the platform. Hence, capability for the workers to meet the deadline and successfully execute the tasks is significant performance in crowd computing.

Table 1: Level of Trustworthy

Value of trust	Prediction/Hypothesis
4 & 5	A reliable worker that leads good results, e.g., satisfies the users.
3	A reliable worker, but not all results able to satisfy the users.
1 & 2	A not reliable worker that leads to very weak trust and unable to satisfy the users.

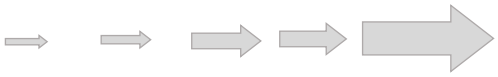
We use user satisfaction as indicator of the level of trust (Table 1) because the users are the main entity in crowdsourcing platform. Note that without users' contribution the crowdsourcing platform will not be used in distributed computing system. Normally, the available workers in crowdsourcing can be trusted and able to complete the users' tasks. However due to the results do not meet users' requirements e.g., missed deadline and error formatting makes the worker is unpopular and unreliable to be chosen.

The users provide feedback according to the service that deliver by the workers to judge the workers performance. Such providing response activity refers the event. There are good events where the users satisfy with the service and poor event when the service not satisfies them. In response to that, we employed the level of trust according to hypothesis that made suitable with the worker behaviour in crowdsourcing (Table 1).

4.2.2. Second step

After list the requirements are identified, the Bayes mathematical formula comes into the picture for processing the prediction and hypothesis from the available requirements. In this step, in order to achieve the reliability there are many probability events needed to execute and manipulate. It is because by using a lot of events, it can help in finding the accurate calculation. Hence, in our testing system we have set many possibilities comments/responses from the users in order to get accurate results. The event is a scenario where the users provide their satisfaction level towards the factors by giving a particular value for every requirement or factor. It means that the higher number of event on the same workers is more reliable response compared than justification from few numbers of users.

Value of trust (system view for end user)	1	2	3	4	5
Value of trust (process in mathematical view)	0.1	0.2	0.3	0.4	0..5



Reliability improvement

Fig. 3: Value of trust

Each condition (factor) is measured using scaling system and each scale consists of certain weightage. Specifically, the scale is designed from 1 to 5 where 1 represents as 0.1 and increased by 0.1 until 0.5 for the scale of 5 (Figure 4). The users will be responded to the worker performance according to the scale in all our four factors. Then, an average of weightage/value of the worker is generated. Such division process (i.e., $\sum TR_e$ / CV) represents the factor value for a worker w. Due to there are merely 4 factors, hence CV is always equal to 4.

Calculation 1in Step 2

a. F_e = F1 + F2 + F3 + F4; F is total value of factor at each event e

b. $T_e = F_e / CV$; the summation of total factor value for a worker w is divided by CV

4.3.3. Third step

In the third step, it is proceed with multiplication and division processes to get a final value of reliability. We show each calculation process in the following figure. Specifically, we measure the probability of worker reliability according to the number of event. As mentioned before this, the event represents the input by the users in terms of response or feedback. There are probabilities that the same users give several responses to the same worker. Therefore, the reliability of workers, *P (T)* towardsresponse from the users, *e* can be formulated as follows.

Calculation 1 in Step 3

a. $P(T) = P(T_{e1})\,P(e) + P(T_{e2})\,P(e) + P(T_{e3})\,P(e) + P(T_{e4})\,P(e) + P(T_{e5})\,P(e) + \ldots\ldots\ldots + P(T_{en})\,P(e)$

where;

P(T) = Probability on how many events has been response to the listed factors for one individual worker

$T_{e1}, \ldots, T_{en}$ = Total factor value in each event; from event 1 until event *n*.

P(e) = Probability number of event.
(For example: if there are 4 events involved in measuring the worker then P(*e*) represents ¼)

By referring the Bayes rule, the constant A and B are reflecting to each other. It means that if there are 80% of users give good response (high factor value) to the worker, than it is assumed 20% of them does not gives favourable feedback. From this indication, the probability P(A|B) refers a comparison made among the events where a factor of trust and not trust is been identified. The average of satisfaction from the good events then is calculated and represents P(B|A). Meanwhile, the P(B) is the total average of satisfaction from all events either they are good or poor events. Details on the formula and its operation are given as below.

Calculation 2 in Step 3

b. $P(A|B) = \dfrac{P(B|A)\,.\,P(A)}{P(B)}$

where;

P (B|A) = The criteria comparison for identifying rate of event that gives good response compared than the negative one.

P(A) = *P(T)*

P(B) = The total of data hypothesis that received at P(B|A) from the reliable and non-reliable event calculation

5. Automation System for Online Assessment

The automation system aims to verify the evaluation process of worker performance through online. Our prototype system required the users to measure the quality of services that provided by the workers. Note that the users are merely allowed to use the measurement system when they are registered as users where they sent the tasks to the platform. Otherwise they are not allowed to judge the workers in the crowd-sourced computing.

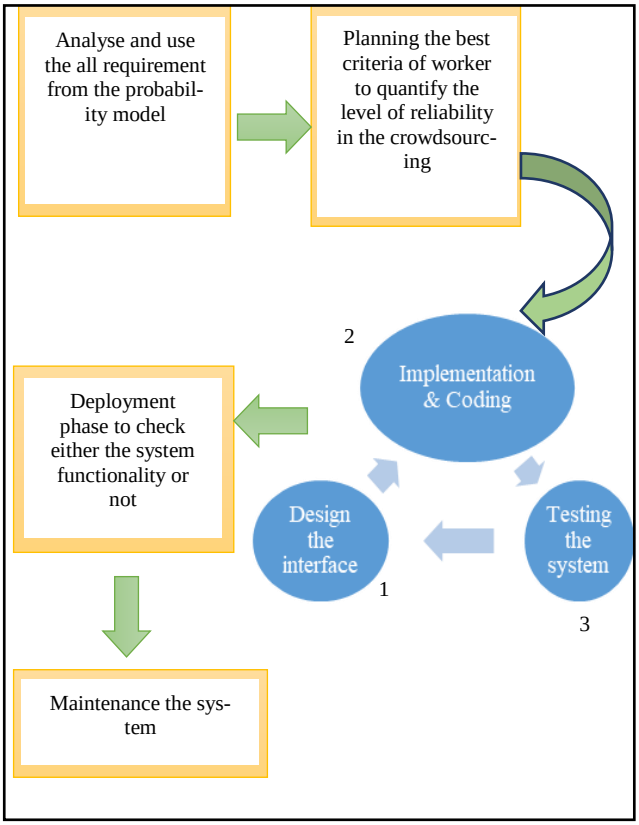


Fig. 4: Agile development model phases

The methodology used in developing the automation system is Agile System Model. The agile model is able to quantify the level of reliability in workers because the model has sufficient function in order to assess the worker reliability before proceed to another phase. Basically, there are five main phases in the agile development model. In the first phase, all requirements from the probability model are identified and classified. It is important to clearly analyze the requirements that suitable for the crowd source for evaluating the worker reliability. The second phase we did in depth planning on requirement selection to be input into the automation system. Due to the automation system aims to be used by the crowd users that demand for easy to use and friendly system, hence the well plan requirements (criteria) in the system must in presentable way.

In the third phase is where the interface design took place. We design a user-friendly interface and the automation system using web-based application (Figure 6). It uses PHP to do calculations that rely on our designed probabilistic model. Our simple web based application leads to ease of use by the crowd users. Specifically, the user will be required entering the value or by using the button clicked for each scale. There are scales given for each criteria that following the four factors; mentioned in Section 3. The users clicked the value for each factor then submitted for the worker's trust value. The value of trust for the worker indicates its reliability. Due to the tasks might be processed by more than one worker, so the users needed to measure all respective workers. The result is then will be displayed when the user submit their worker assessment stage. The process flow of user input shows in the Figure 7.

Table 2: Test Setting

Item	Setting
Number of Worker	20
Number of users	60
Max. worker per user	2 to 10

The fourth phase is the important stage in the development of reliable system where the process of validation of Bayesian prob-

ability model takes place. There is iteration count in confirming the model able to calculate the trust value of the respective workers. It is also fixed and matched process in phase four that regards the process of adopt the Bayesian model in crowdsourcing system. We combine the testing and maintenance phases. We set several workers that executed the tasks owned by the users. The test setting is given in Table II. In the testing phase also we asked other users to pretend as the users and provide their judgment for the workers. It purposely to make sure our system is easy to use and understandable. Based on the automation system, the results of worker reliability are correct and meet the Bayes rules. The worker can accurately assess based on their behavior factors and the trust value can reached through Bayesian probability model. The results also preserve as guidance to the crowd source to quantify the level of worker reliability that will be selected for future involvement in the crowdsourcing.

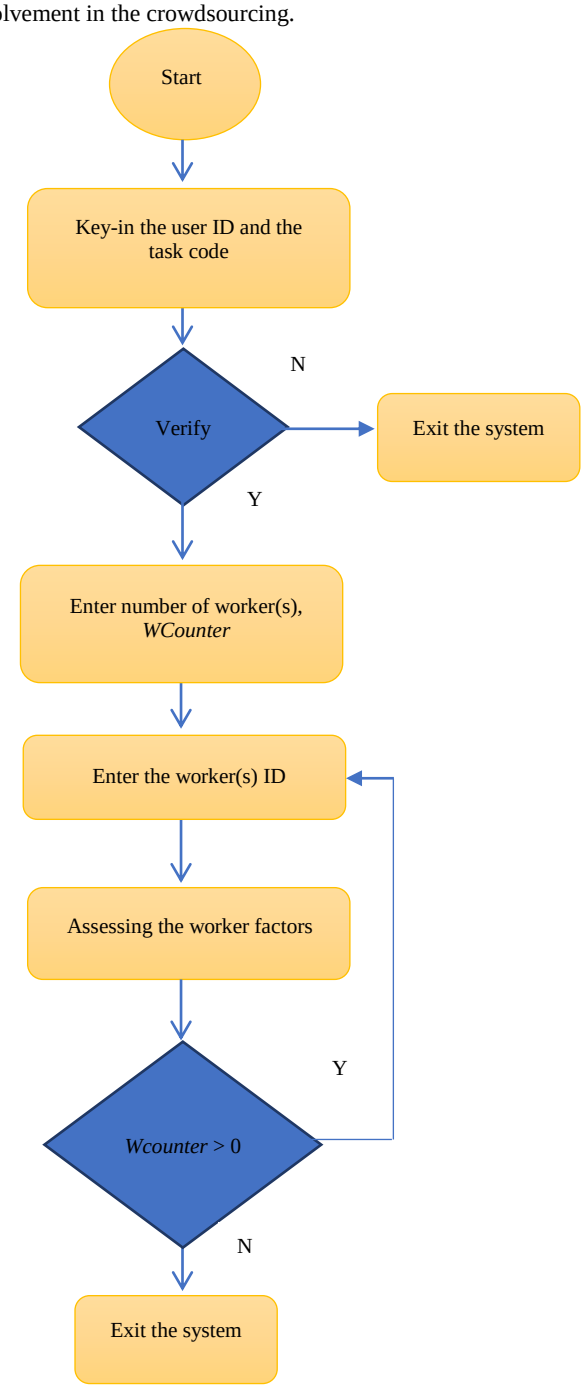


Fig. 5: The flow chart of user input

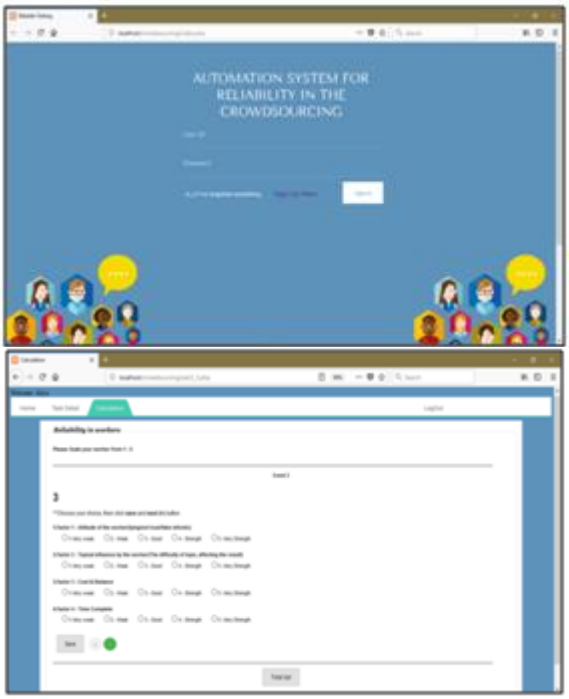


Fig. 6: Sample of user interface

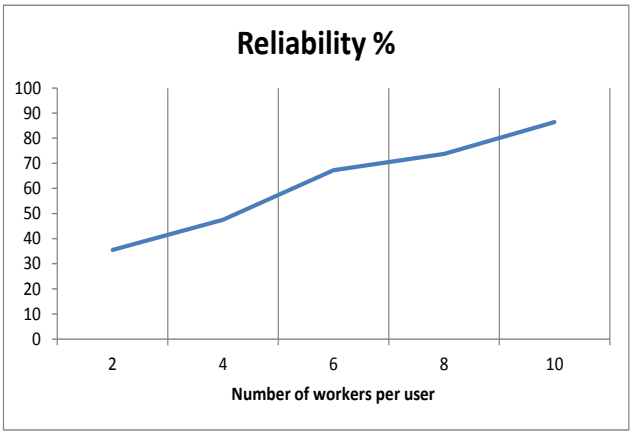


Fig. 7: Reliability level towards the number of workers

Based on the test setting in Table II we analysed the percentage of reliability. The Figure 7 shows the number of workers per user against percentage of reliability. It reveals that increasing in the number of workers for executing the users' tasks improves reliability. Such reliability is calculated based on the average trust value from the Bayesian formulation in Section 4. It also shows in the graph where the exponential growth of reliability towards the number of workers is achieved. That proves that the Bayes rules concept able to measure the reliability in the crowdsourcing.

6. Conclusion

Bayes rules, is a probability model that best to find result in dynamic computing environment such as crowd-sourced computing. The reliability in the crowdsourcing platform is important due to the performance of the crowd-sourced computing is subjected to user satisfaction. Hence, by providing better services further improves the system reliability**.** In order to quantify the worker behaviours, several factors need to be concerned. In our work, there are 4 factors that chosen for contributing in achieving the worker reliability in the crowdsourcing. We thoroughly investigated the utility of constant A and B from the Bayes theorem to be matched into our crowd-sourced computing scenario. Based on the results from the automation evaluation system that developed in this work,

the trust value of worker can successfully achieved using Bayes rules. Furthermore, the automation evaluation system becomes a medium for the crowd users to provide feedback towards the participants workers in the system. This work is provides initial analysis of our study and in future, we will consider more factors toward worker behaviours in more complex and large system.

Acknowledgment

This work is supported by a Research Management Center (RMC) University Putra Malaysia under GeranPutra/Feb/2018.

References

[1] S.Tranquillini, F.Daniel, P.Kucherbaev, and F.Casati, "Modeling, enacting, and integrating custom crowdsourcing processes. ," ACM Transactions on the Web (TWEB), vol. 9, 2015.

[2] A. Siddharthan, C. Lambin, A.-M. Robinson, N. Sharma, R. Comont, E. O'mahony, C. Mellish, Ren, and V. D. Wal, "Crowdsourcing Without a Crowd: Reliable Online Species Identification Using Bayesian Models to Minimize Crowd Size," ACM Trans. Intell. Syst. Technol., vol. 7, pp. 1-20, 2016.

[3] Q. Liu, J. Peng, and A. T. Ihler, "Variational inference for crowdsourcing," in Advances in neural information processing systems, 2012, pp. 692-700.

[4] Y.Zhao and Q.Zhu, "Evaluation on crowdsourcing research: Current status and future direction. ," Information Systems Frontiers, vol. 16, pp. 417-434, 2014.

[5] A. M. Turk, "Amazon mechanical turk," Retrieved August, vol. 17, p. 2012, 2012.

[6] C. Van Pelt and A. Sorokin, "Designing a scalable crowdsourcing platform," in Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, 2012, pp. 765-766.

[7] G. Demartini, D. E. Difallah, and P. Cudr, "ZenCrowd: leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking," presented at the Proceedings of the 21st international conference on World Wide Web, Lyon, France, 2012.

[8] Q. Li, F. Ma, J. Gao, L. Su, and C. J. Quinn, "Crowdsourcing High Quality Labels with a Tight Budget," presented at the Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, San Francisco, California, USA, 2016.

[9] D.Schall., "Automatic Quality Management in Crowdsourcing [Leading Edge]." IEEE Technology and Society Magazine. , vol. 4, pp. 9-13, 2013.

[10] A. T. Nguyen, M. Halpern, B. C. Wallace, and M. Lease, "Probabilistic modeling for crowdsourcing partially-subjective ratings," in Fourth AAAI Conference on Human Computation and Crowdsourcing, 2016.

[11] L. de Alfaro, V. Polychronopoulos, and M. Shavlovsky, "Reliable aggregation of boolean crowdsourced tasks," in Third AAAI Conference on Human Computation and Crowdsourcing, 2015.

[12] Y. Bachrach, T. Graepel, T. Minka, and J. Guiver, "How to grade a test without knowing the answers---a Bayesian graphical model for adaptive crowdsourcing and aptitude testing," arXiv preprint arXiv:1206.6386, 2012.

[13] S. E. and R. S., "Bayesian Methods for Intelligent Task Assignment in Crowdsourcing Systems," in Decision Making: Uncertainty, Imperfection, Deliberation and Scalability. Studies in Computational Intelligence. vol. 538, K. M. Guy T., Wolpert D. (eds), Ed., ed: Springer, Cham, 2015, pp. 1-32.

[14] M. Venanzi, J. Guiver, G. Kazai, P. Kohli, and M. Shokouhi, "Community-based bayesian aggregation models for crowdsourcing," presented at the Proceedings of the 23rd international conference on World wide web, Seoul, Korea, 2014.

[15] D. Marquez, M. Neil, and N. Fenton, "Improved reliability modeling using Bayesian networks and dynamic discretization," Reliability Engineering & System Safety, vol. 95, pp. 412-425, 2010.