A fuzzy expert system to Trust-Based Access Control in crowdsourcing environments

Olusegun Folorunso *, Olusegun Afeez Mustapha

Department of Computer Science, Federal University of Agriculture, Abeokuta, Nigeria

Received 13 January 2014; revised 20 June 2014; accepted 3 July 2014
Available online 18 July 2014

Keywords
Trust-based access; Control; Fuzzy-expert systems; Crowdsourcing environment; Quality assurance; Trust value

Abstract Crowdsourcing has been widely accepted across a broad range of application areas. In crowdsourcing environments, the possibility of performing human computation is characterized with risks due to the openness of their web-based platforms where each crowd worker joins and participates in the process at any time, causing serious effect on the quality of its computation. In this paper, a combination of Trust-Based Access Control (TBAC) strategy and fuzzy-expert systems was used to enhance the quality of human computation in crowdsourcing environment. A TBAC-fuzzy algorithm was developed and implemented using MATLAB 7.6.0 to compute trust value ($T_{value}$), priority value as evaluated by fuzzy inference system (FIS) and finally generate access decision to each crowd-worker. In conclusion, the use of TBAC is feasible in improving quality of human computation in crowdsourcing environments.

© 2014 King Saud University. Production and hosting by Elsevier B.V. All rights reserved.

* Corresponding author.
E-mail address: folorusno@funaab.edu.ng (O. Folorunso).
Peer review under responsibility of King Saud University.
1. Introduction

Crowdsourcing is sometimes referred to as human computation, a methodology that lets humans’ process tasks which are difficult to implement in software. These tasks may include transcription of documents, reviewing of articles or evaluating the quality of ranking algorithms [20]. Crowdsourcing has recently shown enormous potential in solving highly decentralized target localization tasks [13]. In a crowdsourcing system, tasks are distributed to a population of anonymous Internet users for completion. Man-Ching et al. [12] and John et al. [14] defined crowdsourcing as the use of large distributed groups of people to complete micro tasks or to generate information. Because traditional search relevance evaluation requiring expert assessment is a lengthy process [2,6,8,14], crowdsourcing has gained traction as an alternative solution for these types of high volume tasks [1,2,8].

A crowdsourcing site has two groups of users: requesters and workers. The crowdsourcing site exhibits a list of available tasks, associating with reward and time period, that are presented by requesters; and during the period, workers compete to provide the best submission [12]. Meanwhile, a worker selects a task from the task list and completes the task because the worker wants to earn the associated reward. At the end of the period, a subset of submissions is selected, and the corresponding workers are granted the reward by the requesters [12].

Crowdsourcing is now being pushed by large IT companies such as Amazon, Google, or Yahoo!. They have recognized the opportunities behind such mass collaboration systems [14] for both improving their own services as business case. In particular, Amazon focuses on a task-based crowdsourcing platform called Amazon Mechanical Turk (AMT) [20]. Requesters are invited to issue human-intelligence tasks (HITs) requiring a certain qualification to the AMT. These crowdsourcers post mostly simple tasks that, however, require human capabilities. In particular, 50% of tasks are processed at a cost of $0.10 and less, most of the tasks are usually also offered in chunks to multiple AMT workers [12,20].

Meanwhile, the major problem with current crowdsourcing environments as described in [5,26] is the lack of manageability as a result of the openness of Web based platforms, where anybody can join and participate which make quality assurance challenging.

From literature, Quality Assurance (QA) is seen as a major determinant of e-satisfaction in most business-to-consumer (B2C) and crowdsourcing environments inclusive. QA is further designed into a theoretical framework or a model. These models are Rayport and Jaworski 7Cs Framework [19], DeLone and McLean Electronic Commerce Model [7], The ISO 9126 Quality Model (ISO/IEC 9126, 2011), SERVQUAL [17], WebQual 4.0 Model [3], Palmer’s Model [16] and Stefani and Xenos Quality Model [22]. These models were drawn from the fields of information systems, marketing, human–computer interaction and design and the researchers “focus has been on online consumers” perspective [4].
QA factors of online transactions as aggregated from QA models is usually expressed by performance characteristics such as response time, throughput, predictability, availability, scalability, stability, usability, reliability, efficiency, functionality, speed of download (bit rate), ease of finding information, timeliness, confidentiality and transaction integrity. In this case, we looked into confidentiality and transaction integrity by focusing on Trust-Based Access Control (TBAC).

Trust exists in our daily life and thus can be used as a mechanism to make access decisions in networks [27]. Trust models have also been proposed to control anonymity and uncertainty [9,10]. Several novel access control models based on trust have been proposed in recent years. TrustBAC extends the conventional Role Based Access Control (RBAC) with the notion of trust level [25] in which users are assigned to trust levels instead of the roles based on user identities, behavior history, recommendation, etc. The role that is assigned to a user changes along with the change of the trust level of the user.

Trust management systems have been introduced in order to create a coherent framework to deal with trust [11]. Since the importance of building trust models has become vital for the development of some nowadays computer systems the way this trust is derived, that is, the metrics also becomes crucial. Metrics become very important for the deployment of trust management systems as the way of quantifying trust [11]. The simplest way to define a trust metric is by using a discrete model where an entity can be either “trusted” or “not trusted”. This can also be expressed by using numerical values such as 1 for trusted and 0 for not trusted [11].

However, trust management models cannot address security issues at fullest [18]. There is also need of explicit trust model which address trust access control for crowdsourcing environment. Access control mechanism based on trust calculations using fuzzy approach is presented in [24] where access feedback is used for access control.

This paper proposes the fuzzy expert systems approach for TBAC which is necessary in capturing the trust calculations and crowd worker identity of context based trust relationship in crowdsourcing environments as inputs and access decision as output to the system.

The rest of the paper is organized as follows. In Section 2, we discuss the relevant work in the areas of trust and crowd-sourcing, Section 3 is on trust and access model, fuzzy-expert systems design, and CrowdTBAC-fuzzy framework, Section 4 presents results and discussion; and Section 5 concludes the work.

2. Related work

This section presents discusses some related works that are relevant to our research ideas and purpose stated earlier.

Sara et al. [21] presented a model for users’ reputation in Wikipedia. In their research they first parse and mine the entire English Wikipedia history pages in
order to extract detailed user edit patterns and statistics. The pattern and statistics were then used to derive three computational models of a user’s reputation. Finally, these models were validated using ground-truth Wikipedia data associated with vandals and administrators and when used as a classifier, the best model produces an area under the Receiver Operating Characteristics (ROC) curve of 0.98. An automated system was developed to estimate the reputation $R_i(t)$ of a Wikipedia user $i$ at time $t$ based on its past behavior. The reputation index $R_i(t)$ was assumed to be positive and scaled between 0 and 1 and, for the moment, is assumed to be loosely interpretable as the probability that $i$ produces high-quality content. A step was taken toward this long term goal by developing several computational models of $R_i(t)$ and testing them, in the form of classifiers, on the available “ground-truth” associated with Wikipedia-known administrators and vandals. Their approach, which is fairly standard in machine learning applications, requires some explanations. It was reasonable to assume that there exists a true reputation function that is scaled between 0 and 1 and grows monotonically from the user with the lowest reputation to the user with the highest reputation.

The first of the three models that was proposed in their paper measures the reputation of a user as tokens are inserted by the user. In the work, it was discovered that a users’ reputation decreases when its earlier tokens inserted were deleted by another user. The second reputation model measures the time interval between insertion and deletion of content. While the third model was developed as a variation of the second model. (Other works are summarized in Supplementary Table 1.)

3. TBAC-fuzzy framework design

This section explains the Trust-Based Access Control model; trust metrics, fuzzy-expert systems design, TBAC-fuzzy framework and a crowd-sourcing scenario. In this case we first consider a scenario.

3.1. A crowd-sourcing scenario

Let consider an Internet encyclopedia as a crowd-sourcing environment that allows everyone that has access to the Internet to make contributions or edit articles online. In such environment, workers are classified based on their competences and experiences. The crowd workers can freely access the encyclopedia network but are allowed to make contributions or edit some certain articles based on its trust value acquired during the interaction with the host network. The trust value of each crowd worker is greatly affected by the number of activities performed. The priority level of workers is determined by the rules generated.

The crowd-sourcer specifies a priority value to an existing articles which denotes the sensitive level of that articles which means a worker must have such specified
priority value or more before access can be granted to make contribution to such an article (Supplementary Fig. 1).

3.2. Trust based access control model for crowdsourcing environments

Trust based access control model computes trust quantitatively based on their interaction in a crowdsourcing environment. In order to achieve this, a trust model metric is introduced to evaluate the trust of the workers and then use the trust value to enforce the access control policy with fuzzy expert system.

Our trust based access control model is described as a 3-tuple set model \( \{W_r, T_d, W_t\} \). This describes the set of workers (\( W_r \)), trust degree (\( T_d \)), and workers’ trust degree (\( W_t \)). We further arrived at the following definition:

**Definition 1.** The set of worker (\( W_r \)) is a finite set of group workers. Crowd-workers are classified into; highly skilled, skilled, semiskilled and un-skilled worker.

\[
W_r = \{wr_1, wr_2, \ldots, wr_n\}
\]

**Definition 2.** Trust degree (\( T_d \)) measures the degree of trust of a worker. This value is derived from a trust metric, which is later classified in a linguistic (membership) form.

\[
td \in T_d; \{\text{high, average, low, verylow}\}
\]

**Definition 3.** Worker’s trust degree: is a Cartesian set of a worker to a trust degree.

\[
W_t \subseteq W_r \times T_d;
\]

\[
wt(wr) = td; \{wr \in W_r \text{ and } td \in T_d\}
\]

3.3. Trust metric

An access control policy considers the workers’ trust degree when deciding whether to grant a worker access to perform tasks. The trust value of a worker is determined as the worker visits the host. A trust metric equation from [15] was adapted so as to compute the trust values of every worker from the crowd. The \( T_{\text{value}} \) is the trust value of a worker \( wr \in W_r \) on a visit to the host, \( E_i \) is an events, action or service carried out on the host site by a worker \( W_r \). The value of the weight \( W_i \) is assigned to each event \( E_i \) by the host [15].

\[
T_{\text{value}}(wr) = T_0 + \frac{\sum_{i=0}^{n} W_i E_i}{\sum_{i=0}^{n} W_i} + f(t)
\] (1)
$T_0$ is the initial trust value assigned to a crowd-worker. The initial value has a relation to the evaluation of trust value of a worker. Hence, the crowd-sourcer defines $T_0$ value for a new and existing worker.

The term $f(t)$ is added to the equation to reflect any time-dependent activity (or inactivity) to suggest gain or loss of reliability. The $f(t)$ value represents error in the model when the worker does not have access to the environment due to a network being down or system itself is down.

The trust value of a worker is evaluated based on event performed by the worker. Such an event carries a value and it’s also weighted. An event can be negative if the worker tries to act maliciously; such could be a failure to login or a worker performing brute force attack on the host. The positive value event which varies are the normal event that a worker is expected to carry out on a host.

An anonymous worker is classified as unskilled worker with initial value of 0.0 since the worker is new to the environment and the competence is yet to be ascertained while other workers (semiskilled, skilled, highly skilled) are known in the environment. Anonymous worker can eventually gain recognition to become known in the environment. An existing worker which is known to the environment has a level of skills and can be classified either as skilled worker, semiskilled and highly skilled depending on the worker’s performance history.

### 3.4. Fuzzy expert system design

The fuzzy expert system has 2 input values and 1 output. The fuzzy sets (i.e. linguistic value) for the input variables and the output variable are as follows:

i. Degree of trust denoted as “TrustDegree” is defined with four fuzzy sets \{high, average, low, verylow\}. The “TrustDegree” membership functions are shown in Fig. 4 of the next section. The fuzzy set “VeryLow”, “Low”, “Average” and “High” sets are in trapezoidal form.

ii. Type of crowd worker denoted as “CrowdWorker” with four fuzzy sets (highly skilled worker, skilled worker, semiskilled worker, and unskilled worker).

iii. Priority level denoted as “priority Level” with four fuzzy sets \{high, average, low, verylow\}. The “Priority Level” membership functions are also shown in Fig. 5 where the “VeryLow”, “low”, “average” and “High” sets are in trapezoidal form. The linguistic input variable TrustDegree and CrowdWorker are defined in the Table 2, while Table 3 shows the linguistic output variable Priority Level. The crisp range value was selected based on the degree of membership in the interval [0, 1], where 0 and 1 confirms no membership and full membership respectively.

The membership function is trapezoidal as it is specified by four parameters ($a$, $b$, $c$, $d$) as follows:
Trapezoidal \((x; a, b, c, d)\) = \[
\begin{cases}
0, & x \leq a \\
\frac{x-a}{b-a}, & a \leq x \leq b \\
1, & b \leq x \leq c \\
\frac{d-x}{d-c}, & c \leq x \leq d \\
0, & d \leq a 
\end{cases}
\] \quad (2)

An alternative expression using min and max:

Trapezoidal \((x; a, b, c, d)\) = \[
\max \left( \min \left( \frac{x-a}{b-a}, 1, \frac{d-x}{d-c} \right), 0 \right)
\] \quad (3)

### Table 2

<table>
<thead>
<tr>
<th>Linguistic variable</th>
<th>Fuzzy set (linguistic value)</th>
<th>Crisp range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trust degree</td>
<td>High</td>
<td>(\mu \geq 0.75)</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>(0.5 \leq \mu \leq 0.7)</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>(0.25 \leq \mu \leq 0.6)</td>
</tr>
<tr>
<td></td>
<td>Very low</td>
<td>(\mu \leq 0.35)</td>
</tr>
<tr>
<td>Crowd workers</td>
<td>Highly skilled worker (HS)</td>
<td>(HS \geq 0.75)</td>
</tr>
<tr>
<td></td>
<td>Skilled worker (S)</td>
<td>(0.5 \leq S \leq 0.7)</td>
</tr>
<tr>
<td></td>
<td>Semiskilled worker (SS)</td>
<td>(0.25 \leq SS \leq 0.6)</td>
</tr>
<tr>
<td></td>
<td>Un-skilled worker (US)</td>
<td>(US \leq 0.35)</td>
</tr>
</tbody>
</table>

### Table 3

<table>
<thead>
<tr>
<th>Output</th>
<th>Fuzzy set</th>
<th>Crisp value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Priority_level</td>
<td>High_priority</td>
<td>0.75 above</td>
</tr>
<tr>
<td></td>
<td>Average_priority</td>
<td>0.5–0.75</td>
</tr>
<tr>
<td></td>
<td>Low_priority</td>
<td>0.25–0.5</td>
</tr>
<tr>
<td></td>
<td>Very low_priority</td>
<td>0.0–0.25</td>
</tr>
</tbody>
</table>

#### 3.5. TBAC-fuzzy framework for crowdsourcing environments

The proposed framework comprises of the Trust-Based Access Control and fuzzy expert system. The trust based access control (TBAC) has a model that computes the trust value of workers. The trust metric operates at the preprocessing while the access control module operates at postprocessing module. Supplementary Fig. 2 shows the TBAC and Fuzzy expert system components incorporate into a single framework.

The TBAC-fuzzy expert system can be explained with the following module of the framework.

**Preprocessing module:** The first module of the framework contains the user module and trust metric module of the TBAC. At this preprocessing module,
the user module classifies workers into different classes while the trust metric module computes the trust value of the worker using Eq. (1). The preprocessing module provides a crisp value which serves as the input value to the fuzzy-expert system. These crisp values are forwarded to the fuzzification module and decode to a fuzzy value.

**Fuzzification module:** In the process of fuzzification, membership functions defined on input variables are applied to their actual values so that the degree of trust for each rule premise can be determined. The crisp values are fuzzified into linguistic values.

**Fuzzy inference engine module:** The inference engine consists of the knowledge base which contains the rules for rating workers (generated rules are available in the supplementary material).

**Defuzzification:** In defuzzification, the fuzzy output set is converted to a crisp number. Some commonly used techniques are the centroid and maximum methods.

- In the **centroid method**, the crisp value of the output variable is computed by finding the variable value of the center of gravity of the membership function for the fuzzy value.
- In the **maximum method**, one of the variable values at which the fuzzy set has its maximum truth value is chosen as the crisp value for the output variable.

**Postprocessing module:** In this module, access right is defined on the workers based on the output result of the fuzzy expert system. The output of the system shows the priority level of the worker. The environment filters the worker based on the output and decides access right to workers that falls within a priority level and thereby improves the quality of the computation. A CrowdTBAC Fuzzy algorithm is described in (Algorithm available in supplementary material).

### 4. Simulation, results and discussion

#### 4.1. Result

Our fuzzy expert system trust model is evaluated using a MATLAB fuzzy logic toolbox version 7.0.6. The fuzzy interference system of our TBAC-fuzzy expert system has two inputs and an output, a mamdani-type inference and a centroid method were used for defuzzification. The implementation process was carried out as shown in Figs. 4–8. Firstly, we defined the number of input and the output on fuzzy interference system editor and define the membership functions of the linguistic variable.

The surface-viewer above reflects the trust value relative to trust and workers that may help us to analyze trust variance. Fig. 8 shows the output surface for
the priority value versus trust value and worker. This priority value output is a factor that controls and enhances the quality assurance of the crowd-sourcing activities.

4.1.1. Simulation

TBAC-fuzzy expert system is simulated in our crowd-sourcing environment scenario. The simulation parameters are shown in Table 5 with the range of values based on assumptions. In this process, 10 workers of different classes (5 unskilled, 2 semiskilled, 2 skilled and 1 highly skilled worker) are considered to make contribution in encyclopedia environment. Within the range of values for event and weight defined below, three distinct values are generated using a java programming language. The values generated are further used to compute the workers’ trust values on the assumption that same events were performed in Table 4.
The initial values ($T_0$) of the workers are considered in the computation to determine their actual trust value. In this case, we defined $T_0$ to be 0.00, 0.01, 0.025, and 0.05 for a new worker (unskilled), semiskilled, skilled and highly skilled workers respectively. The last three workers are the existing workers and are classified based on their past achievements and experiences.

The designed fuzzy rules are simulated using the trust values of crowd-workers with a fuzzy tool box in MATLAB. The simulation graph above revealed that a task with a priority value of 0.37 can only be performed by workers that gain such a value or more. The threshold value of 0.37 is been defined by the crowd-sourcer.
for basic trust level. The threshold/user-defined ($Dval = 0.37$) value represented by the red line as shown in Fig. 9 gives the crowd-source an insight into the worker that is allowed to performed a task. Any worker below the threshold level will be denied access on the task.

It was observed furthermore that not all the unskilled workers can be allowed to carry out a task. Individual workers’ priority level is determined by fuzzy rules

<table>
<thead>
<tr>
<th>Workers crisp value</th>
<th>Trust crisp value</th>
<th>Priority crisp value</th>
<th>$Dval = 0.37$ (user-defined value to grant access)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.3547</td>
<td>0.1271</td>
<td>Denied</td>
</tr>
<tr>
<td>0.25</td>
<td>0.47668</td>
<td>0.1342</td>
<td>Denied</td>
</tr>
<tr>
<td>0.25</td>
<td>0.40471</td>
<td>0.3988</td>
<td>Granted</td>
</tr>
<tr>
<td>0.5</td>
<td>0.53168</td>
<td>0.4864</td>
<td>Granted</td>
</tr>
<tr>
<td>0.5</td>
<td>0.59501</td>
<td>0.6</td>
<td>Granted</td>
</tr>
<tr>
<td>0.25</td>
<td>0.53451</td>
<td>0.2257</td>
<td>Denied</td>
</tr>
<tr>
<td>0.25</td>
<td>0.58843</td>
<td>0.3765</td>
<td>Granted</td>
</tr>
<tr>
<td>0.5</td>
<td>0.60677</td>
<td>0.3987</td>
<td>Granted</td>
</tr>
<tr>
<td>0.5</td>
<td>0.65498</td>
<td>0.6</td>
<td>Granted</td>
</tr>
<tr>
<td>1.0</td>
<td>0.73293</td>
<td>0.7508</td>
<td>Granted</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Data type</th>
<th>Values range</th>
<th>Probability distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Events (E)</td>
<td>Real</td>
<td>0.0–0.5</td>
<td>Uniform</td>
</tr>
<tr>
<td>Weight (W)</td>
<td>Real</td>
<td>0.0–0.1</td>
<td>Uniform</td>
</tr>
<tr>
<td>$T_0$ for highly skilled</td>
<td>Real</td>
<td>0.05</td>
<td>Uniform</td>
</tr>
<tr>
<td>$T_0$ for skilled</td>
<td>Real</td>
<td>0.025</td>
<td>Uniform</td>
</tr>
<tr>
<td>$T_0$ for semiskilled</td>
<td>Real</td>
<td>0.01</td>
<td>Uniform</td>
</tr>
<tr>
<td>$T_0$ for un-skilled</td>
<td>Real</td>
<td>0.0</td>
<td>Uniform</td>
</tr>
</tbody>
</table>
generated using the two inputs (classes of worker and trust value) to determine the output (priority value). The graph also clearly shows that not all two semiskilled workers can be given access to a task of 0.45 priority value that could be defined by the crowd-sourcer.

4.2. Discussion

The defined value chosen by the environment signifies the sensitive level of the activities to be performed. A lower value say 0.2 which could signifies an activities which required worker with little knowledge or experience and this will accept more workers which means 2 workers will be rejected and denied access to make contributions. Whenever a worker visits the host, trust computations were made based on some events performed on the system. The events on the system have a value assigned and also they are weighted.

In the design architecture, trust metric module compute workers trust and transform into a crisp value which is then decoded at the fuzzification module. The surface view in Fig. 7 shows the dependency of the output on the two inputs.

Postprocessing module of the fuzzy expert system grants access to a worker. This module makes the decision based on the output of the fuzzy system. This module filters and classifies the needed worker based on a defined priority value to make contribution or suggestion. The trust metric equation of TBAC (Eq. (1)) was tested with the fuzzy expert system by some generated values. The outcome of the result has it that majority of the workers gained access to perform activities in the crowdsourcing environment while others are denied. Meanwhile, this can be justified for the fact that half of the workers that participated have little
or more experience while the other workers have little or no experience. In this case a task with a priority value 0.37 can be worked on by 70% of the workers as shown in the simulation chart in Fig. 9. This analysis with the simulation chart helps to enhance the quality assurance of the crowdsourcing activities.

5. Conclusions and future work

This paper addressed the issues of trust in crowd-sourcing alongside its incentive and how the quality of response from the crowd-workers would be improved. The paper identified answer ranking as a problem in crowd-sourcing leading to cheating by workers. To look for more efficient and reliable way to solve this problem, a fuzzy expert system and trust based access control were employed as a single framework to enforce access control. A trust based access control fuzzy algorithm was used to compute and assigned trust value to workers. The trust value assigned to crowd-worker serves as one of the inputs to the fuzzy system and access is granted to workers whose priority value is above threshold level based on the fuzzy output. In this case, the crowdsourcing could be assured that only workers with little or more experience could perform certain tasks. This however helps to enhance the quality of response from the crowd. Experiment was conducted on Trust-Based Access Control fuzzy algorithm using systematic generated values on MATLAB. The results indicated that high quality response and service could be obtained from CrowdWorkers.

In the future, effort on mood positivity with desired model would be conducted to determine the performance of workers in a crowdsourcing environment.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.aci.2014.07.001.

References

A fuzzy expert system to Trust-Based Access Control in crowdsourcing environments