

## Task Redundancy Strategy Based on Volunteers' Credibility for Volunteer Thinking Projects

Lesandro Ponciano

Francisco Brasileiro

Guilherme Gadelha

Federal University of Campina Grande  
Department of Computing and Systems

lesandrop@lsd.ufcg.edu.br, fubica@computacao.ufcg.edu.br, guilherme.gadelha@ccc.ufcg.edu.br

### Abstract

In this paper, we propose a task redundancy strategy based on measures of accuracy of volunteers. Simulation results show that our strategy reduces the number of generated task replicas compared to the pessimistic and moderate strategies. It also generates similar or less task replicas compared to the optimistic strategy.

### Introduction

This study focuses on *volunteer thinking projects*, those citizen science systems that gather volunteers willing to contribute executing human computation tasks, e.g., Galaxy Zoo (Lintott, C. J. et al. 2008). This sort of project is constantly recruiting and encouraging volunteers to execute more tasks, its effectiveness relies on *its ability to optimize the contribution provided by each volunteer*.

To achieve a satisfactory quality level in task execution, volunteer thinking projects usually make use of *task redundancy*, i.e., the same task is executed multiple times by different volunteers (Simpson, E. et al. 2013). However, this approach generates a trade-off between redundancy degree and quality. The lower the degree of redundancy the higher the chance of having an unsatisfactory quality level. On the other hand, *if redundancy level is higher than the necessary, the exploitation of volunteers' contribution is not efficient*.

In this work, we analyze variations in volunteers' accuracy over project duration. In the evaluated project, we find that volunteers' accuracy follows a well-defined learning curve, i.e., their accuracy improve in the proportion that they execute more tasks. This indicates that volunteers' accuracy is dynamic throughout their participation in the project. To exploit such behavior in order to optimize task redundancy, we propose a *credibility-based task redundancy strategy* that decides dynamically the redundancy degree, as well as the volunteers to which tasks' replicas will be routed.

### Volunteers' Accuracy Pattern

Between October 2012 and April 2013, we conducted an online experiment that assessed volunteers' accuracy by using a prototype of a volunteer thinking project. In the project, volunteers evaluated the outputs of different algorithms used

to automatically recognize the structure of table in images, so to help in identifying the best algorithm. This is part of an effort to transcribe statistical data contained in old books of the Brazilian Ministry of Finance Library. Each task displays (i) the image of a page from an old book containing a table, and (ii) two options of table recognition that were generated by using different table structure recognition algorithms. The volunteers are asked to select the recognition that best match the table that exists in the book page, or informing that neither is good enough. Snowball recruitment was done by disclosure in students' mailing list of two Brazilian universities: IFPB and UFCG.

We generated 520 different tasks which differ among them only in terms of the book page displayed or options of table recognition. The volunteers were not allowed to execute the same task more than once. In the entire experiment, we collected 3,599 results from 1,098 volunteers, eliminating volunteers who participated in the development of the tasks and those who knew the tasks before the experiment. We already knew the correct results for these tasks, and used this information to analyze volunteers' accuracy as the ratio of the number of tasks executed correctly to the total number of tasks executed. We distinguish between passenger volunteers, those who execute tasks only one day and do not return to execute more tasks, and regular volunteers, those who return at least one more day to execute more tasks after executing the first task in the project.

Figure 1 shows the average volunteers' accuracy rate in the proportion that they execute more tasks, with confidence level of 95%. Each curve stops in the 95-percentile of the volunteers in our experiment, i.e., 95% of the passenger volunteers do not execute more than 19 tasks and 95% of the regular volunteers do not execute more than 30 tasks. Volunteer's accuracy improves in the proportion that he/she executes more tasks, forming a learning curve. It seems to be a consequence of the fact that in this kind of project, volunteers always execute the same type of task. This figure also shows that regular volunteers become slightly more accurate than passenger volunteers.

### Credibility-based Task Redundancy Strategy

Credibility-based task redundancy strategy takes advantage of the volunteers' learning curve in order to dynamically decide the task redundancy degree and the volunteers to which

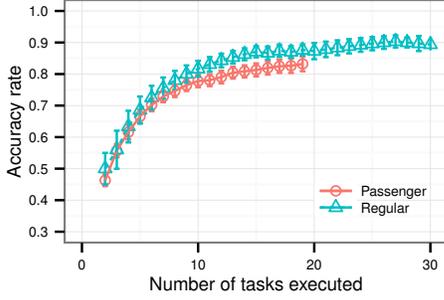


Figure 1: Volunteers' accuracy rate.

tasks' replicas will be routed. It is inspired by the credibility-based fault-tolerance technique proposed by Sarmenta (Sarmenta 2001). It is based on four credibility measures: *required result credibility* ( $r$ ), *credibilities of the volunteers*, *credibilities of the result*, and *credibility of a group of results*. The required result credibility ( $r$ ) is the task owner expected level of credibility. Volunteer's credibility is the probability of the volunteer to execute a task correctly. This credibility changes over time based on a learning curve. The credibility of a result is the credibility of the volunteer that generates it. Equal results generated by different volunteers for the same task are joined in the same group of results. The credibility  $C(G_a)$  of a group of results  $G_a$  is computed as the conditional probability of correctness of the group results, given their credibilities and the credibilities of other results in the other groups for the same task, as expressed by Equation 1, where  $g$  is the number of distinct groups of results.

$$C(G_a) = \frac{P(G_a \text{ good}) \prod_{i \neq a} P(G_i \text{ bad})}{\prod_{j=1}^g P(G_j \text{ bad}) + \sum_{j=1}^g P(G_j \text{ good}) \prod_{i \neq j} P(G_i \text{ bad})} \quad (1)$$

Given  $r$  and a task to be executed, our *redundancy strategy* works as follows. The task is routed to an available volunteer that has the highest credibility  $v_i$ . If volunteer credibility is larger than the credibility required ( $v_i \geq r$ ), no redundancy is required and the credibility of the result is equal to the credibility of the volunteer. Otherwise, if  $v_i < r$ , the current result is stored and a task replica is routed to another volunteer. Similar results received are aggregated in the same group, and different results generate distinct groups. This process is repeated until the credibility of a group becomes equal or larger than  $r$ . The task final result is the result in the group with the highest credibility. At each task performed by a volunteer, his/her credibility is improved following a learning curve pattern.

We use simulation to evaluate our redundancy strategy in terms of the number of generated replicas. We simulate a project with  $P = 10,000$  volunteers, being 10% of  $P$  regular volunteers and the rest passenger volunteers. Their accuracy patterns fit the learning curve shown in Figure 1. At each simulation step,  $P$  tasks are executed and the number of tasks executed by the volunteers is incremented, updating their accuracy. Each simulation consists of a pool of  $N$  tasks

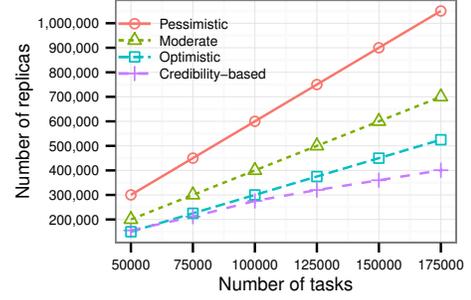


Figure 2: Number of replicas for different number of tasks.

to be executed; we vary  $N$  from 50,000 to 175,000 generating a task-per-volunteer ratio  $N/P = 5, \dots, 18$ , enough to analyze the effect of the volunteers' learning curve. We ran 30 simulations and present results for a statistical confidence level of 95%.

We compare our policy to a *pessimistic*, *moderate*, and an *optimistic* fixed task redundancy strategy. These strategies estimates the minimum redundancy as a function of volunteers' error rate  $f$  and required credibility  $r$  as  $\lceil \log_f(1-r) \rceil$  (Sarmenta 2001). Considering the same sample of volunteers used in the simulation of the credibility-based strategy,  $f$  for the pessimistic, moderate, and optimistic strategies is estimated as 0.45, 0.29, and 0.12, which generates, respectively, the redundancy of 5, 4, and 3 replicas per task. All four strategies consider  $r = 0.99$ .

Figure 2 shows the number of replicas generated by the four strategies. In general, credibility-based strategy reduces the total number of generated replicas compared to the pessimistic and moderate strategies. It also generates similar or less task replicas compared to the optimistic strategy. Note that, in this analysis, credibility-based strategy is also more robust; it guarantees the credibility  $r$  if volunteers' error rate is less than 0.45, while the optimistic strategy only guarantees  $r$  if volunteers' error rate is not larger than 0.12.

## Conclusions and Future Directions

In this work, we have proposed a credibility-based strategy that optimizes tasks' redundancy when volunteers' accuracy patterns fit a learning curve. Naturally, volunteers' accuracy patterns may be different in different types of tasks and not necessarily be a learning curve, so, in the next steps of this work, we will analyze other patterns and evaluate how our credibility-based task redundancy strategy may automatically learn and take advantage of them.

## References

- Lintott, C. J. et al. 2008. Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society* 389.
- Sarmenta, L. 2001. *Volunteer Computing*. PhD Thesis, Massachusetts Institute of Technology, Cambridge, USA.
- Simpson, E. et al. 2013. Dynamic bayesian combination of multiple imperfect classifiers. In *Decision Making and Imperfection*. Springer. 1–35.