



Optimization of Orchestration of Geocrowdsourcing Activities

Kahina Bessai, François Charoy

► To cite this version:

Kahina Bessai, François Charoy. Optimization of Orchestration of Geocrowdsourcing Activities. Third International Conference on Information Systems for Crisis Response and Management in Mediterranean Countries (ISCRAM-med 2016), Oct 2016, Madrid, Spain. hal-01356053

HAL Id: hal-01356053

<https://inria.hal.science/hal-01356053>

Submitted on 24 Aug 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimization of Orchestration of Geocrowdsourcing Activities

Kahina Bessai, Francois Charoy

University of Lorraine
LORIA-INRIA-UMR 7503
BP 239, Vandoeuvre-les-Nancy, France
`kahina.bessai@loria.fr, francois.charoy@loria.fr`,

Abstract. Geocrowdsourcing has proved to be very valuable in crisis situations. Calling to citizen on the ground or to experts or trained amateurs to help in the mapping of crisis situation is a recognized and valuable practice. However, despite the experience gained from real and dramatic situations, it remains difficult to set up and execute complex processes that require actions of both people on the ground and people on the web, and to understand how to get the best result at the minimal cost in term of users actions. In this paper, we describe a process that can be used to assess a global situation on a map using a combination of services and user operations. We want to understand how best to distribute a limited amount of human actions between different kind of tasks in order to get the most reliable result. Since it is difficult to conduct experimentation, we have decided to use simulation to reach a result that could be applied on the ground. This simulation relies on a geolocalised corpus of tweets. It provides some hints about how to deploy an exercise on the ground that are discussed as a conclusion. In addition, we propose a binary integer programming (BIP) making best use of the available workers.

Keywords: crowdsourcing, GIS, pattern, simulation

1 Introduction

Calling to the crowd to conduct activities out of the reach of individuals or small groups is a very common practice. This kind of initiative is known under the generic name of crowdsourcing, or human computation[1]. It has been applied in catastrophic situations but can also be used in more controlled situations. The goal of this project is to develop and validate a method and a framework in order to get citizen to contribute to contextual city assessment activities at a large scale. More precisely, we want to be able to synthesize information regarding the state of a populated geographic area by coordinating in an optimized way volunteers actions. To achieve that kind of goal, we need to get information from the ground. People must execute actions at some places, mostly collect data. Other participant, connected to the system must then analyse the data to

produce an aggregated result. People could use that framework for area status assessment during a crisis or as a preparedness activity, to measure phenomenon in a city at a defined point in time and to repeat that measure on a regular basis. They could combine it to other kind of sensors to conduct different kind of real time analysis. Real time means that the result could be used to take action as the event unfold. This paper describes an ongoing work where we try to optimize the overall quality of a geographical area status assessment following a predefined process with a limited number of human resources. We have defined a crowdsourcing process and tested it with an existing dataset [2]. This was a first experience to evaluate our framework. Here, we continue this work with a new dataset, with new parameters and with the objective to optimize the resource that we can leverage for such a process in order to get the best possible result as we will explain it in section 4. In section 5 we describe a subset of the results of our experiment and we discuss the best way to find compromise between the coverage of the assessment and its quality.

2 State of the art

People used crowdsourcing in a geographical context for a long time, even before the existence of computers. Bird watchers contribute to science by participating to counting activities periodically to create maps of birds populations and of their evolution. This activity has taken a new dimension with the advent of mobile phones, GPS and the Internet ¹. The construction of maps like OpenStreetMap² is a ten year old project that aims at building accurate maps calling to the crowd to contribute all kind of information. Each of these experiments are done in an ad-hoc way. Humanitarians organisations conduct more organized operations under the umbrella of the Digital Humanitarians [1]. They are able to call to volunteers to geolocalize information during crisis. Other sophisticated research use information produced on social media to make sense of what is going on on a crisis area [3]. These are important steps toward a generalized approach regarding the use of crowdsourcing with a geographical context. The goal of other authors is to allow people to call to the crowd to answer more or less structured queries. In [4], the authors propose a prototype for location-based mobile crowdsourcing consisting of a Web and a mobile client. Through these clients, people from the crowd can search for tasks and submit solutions. Similarly, in [5], the authors design a framework for crowdsourcing location-based queries on the top of Twitter. The Framework relies on Foursquare to find the appropriate people to answer the query. In [6] Kazemi and Shahabi introduce a taxonomy of spatial crowdsourcing and propose three techniques for assigning the maximum number of tasks to the crowd. They also introduce a taxonomy for spatial data. This last work relies on the Gowalla dataset³ for its simulation. Re-

¹ <http://www.nytimes.com/2013/08/20/science/earth/crowdsourcing-for-the-birds.html>

² <http://www.openstreetmap.org>

³ <https://snap.stanford.edu/data/loc-gowalla.html>

garding, the selection of the k -top results several approaches have been proposed which can be classified in four categories: (i) heuristics-based approaches [7][8][9][10], (ii) machine-learning methods [11], (iii) extensions of heap-based methods [12] [13] [14] and (iv) combining rating and comparison based-algorithms [15] [16].

Here, we consider the state of a global area that we want to assess as accurately as possible in a repeatable way. We want to support different kinds of scenarios that include a data collection phase on the ground. We want to define a model that could be experimented in a real setting with the best chances to provide a result of the best quality considering a limited number of resources. We have already done a first simulation of our model, but with the only dataset used by researchers to our knowledge for this purpose, the Gowalla dataset ⁴. It provides set of checkins at accurate places These checkins are used by researchers to simulate crowdsourced activities and to assess their models [2] but these checkins are just dots on a very large map (California) where we would like to assess an area on a city size area. In the following section, we get into the details of this problem.

3 The crowdsourcing pattern

The goal of the crowdsourcing activity that we are considering is to create the most accurate assessment map on a delimited geographical area. It could be for instance a map of the level of damage of a city after a earthquake or a map of flooded area and its evolution. It could also be a map assessing the level of cleanliness of a city or any other criteria that could be computed by aggregating measures from different points on the map. The quality of the result depends on the coverage of the assessment i.e. the percentage of the map covered at the end of the process with the correct result. We consider time bounded crowdsourcing campaign since we want to repeat it in order to produce of view on the evolution of the situation. These are two supplementary constraints. In order to avoid the traditional pitfalls of requiring complex tasks from crowdworkers, that are then difficult to evaluate, we have defined a process following the idea of the find-fix-verify pattern [17]. It is based on three simple steps : data collection, data selection and data assessment. The data collection is done on the ground by people. In our case, data would be geolocalised pictures but it could also be text description, sounds, movie clips or any kind of data that is difficult to process automatically. For the data selection phase, the map is divided in square areas of fixed size. We group the data by area and ask the workers to select the most valuable one area by area. Each worker has to vote for one picture on an area regarding a given criteria. Figure 1 is an example of this task

This result in a smaller set of photos that will be assessed regarding this criteria by another set of workers. We only consider the best pictures that have been selected from the previous step as shown on figure 2 In the end we obtain a

⁴ <https://snap.stanford.edu/data/loc-gowalla.html>

Crowdsourcing Task Example



Select the best picture to represent the catastrophic situation at that location

Fig. 1. Example of a selection task

map where each area is graded according to the opinion of the workers. As said before, we limit in time the execution of this process and thus, we assume that we can count only on a limited number of actions done by workers. The question is to decide how to execute this process in order to get the best coverage with the best quality of the assessment. This will depend on different factors like the quality of the data that are collected, the quality of the worker and the way the process is executed in order to distribute the actions between the different activities.

Since it would be very difficult to conduct experimentation in real setting with actual worker to test all the possible combination, we propose to conduct some simulation that combine a geolocalised corpus and simulated workers.

4 The problem formulation of the orchestration of geocrowdsourcing activities

In this section, we propose a formal model for the tasks assignment in crowdsourcing.. Moreover, we design an efficient approach in such way that the overall quality of the the selection and the assessment process evaluation of a given tasks be maximal. To give a formal model of the considered problem here, let the following variables :

- Z the possible geographical zones with z its cardinality.
- W the available workers with w its cardinality.
- w_i the skill of the worker i .
- q_j the minimum quality required by geographical zone j .
- x_{ij}^k represents the assignment of worker i to geographical zone j for the step k of the evaluation process, where $k \in \{1, 2\}$ represents respectively the selection step and the assessment phase. The value of this binary variable is defined as follows:

$$x_{ij}^k = \begin{cases} 1 & \text{if worker } i \text{ is affected to geographical zone } j \text{ for phase } k \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Recall that the process tasks matching is done in two steps. For instance, if the process consist to classify images in crisis context, the first step consists to



According to you what is the status of this area

- ☐ Normal
- ☐ Damaged
- ☐ Destroyed

Fig. 2. Example of an assessment task

select the best picture to represent a given situation and the second steps is related to the assessment of the selected pictures during the first step. We aim to use fairly the available workers to cover maximum geographical zone while the overall quality of each considered zone is maximum on one hand. In order to avoid the quality degradation of a given zone, we have established a quality threshold that must be achieved. The problem can be defined as follows:

$$\forall j \in \{1, \dots, z\}, \max Q_j = \sum_{i \in W} (x_{ij}^1 w_i + x_{ij}^2 w_i) \quad (2)$$

Under the following constraints :

$$\forall i \in \{1, \dots, w\}, \sum_{j=1}^z (x_{ij}^1 + x_{ij}^2) = 1 \quad (3)$$

$$\forall j \in \{1, \dots, z\}, \sum_{i=1}^w (x_{ij}^1 + x_{ij}^2) w_i \geq q_j^s \quad (4)$$

$$\forall i \in \{1, \dots, w\}, j \in \{1, \dots, z\}, (x_{ij}^1, x_{ij}^2) \in \{0, 1\}^2 \quad (5)$$

Equation 2 represents the overall quality of the process tasks selection and assessment of a given location. It is important to note that the considered problem is multi-objectives in the sense that an objective function denoted Q_j is associated for each location. Equation 4 represents the fact that each worker should be assigned to a single geographical zone. In addition, a given worker participate

The coverage is obtained by splitting the space in cells of equal size. We use the size of the cell to decide of the accuracy of the measure. We see on figure 4 the distribution of tweets on the map is not even. Some cells receive no action while others receive several. Figure 5 shows a detail that could be an intermediate

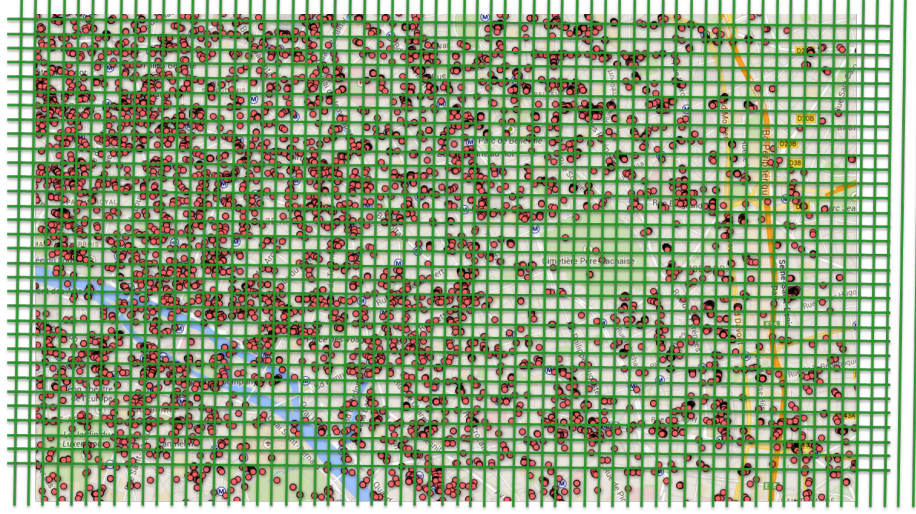


Fig. 4. The grid

result. Cells can be painted based on the assessment. Obviously, we won't get result for all cells since some of them have no data. To simulate the process, we use tweets for geolocalised action and we estimate a number of crowdsourcing actions that will be executed to complete the job. We estimate that if we have N tweets for a given period of time, we can count on $2N$ crowdsourcing actions that we can distribute between selection and assessment. We also execute some simulation with a varying number of worker (in our case from 20000 to 100000). As mentioned previously, the process evaluation consists in two steps: (i) the selection stage, where the user has to decide what, in his opinion is the more accurate data to reflect a situation (see Figure 1) and (ii) the assessment phase, where the user grade the situation on a predefined scale based on the data that is provided to him (see Figure 2). Of course, in a real situation, the description of the purpose of the selection and of the assessment would be of paramount importance but since we are simulating the actions, we consider that the users will have a good understanding of the context. However, the available workers do not provide the same quality. In the following, we describe the considered two stages of our model.

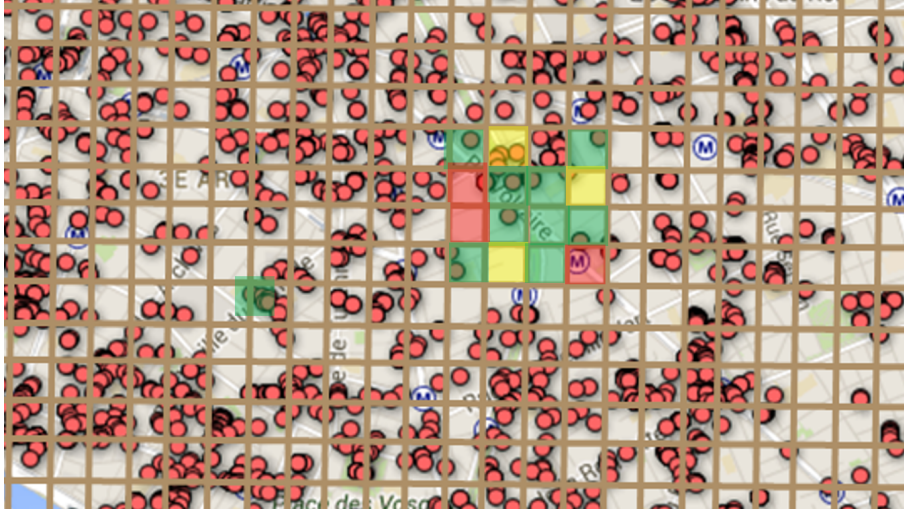


Fig. 5. The expected result

5.1 The selection phase of the proposed model

As its name implies, the selection phase consists to select the representative picture of a given situation. Formally this phase can be described as follows. Without loss of generality, let a set of picture p_1, p_2, \dots, p_m , belonging to the same geographical zone, where their quality are random variables with cumulative distributions functions denoted respectively F_{p_i} (and probability density functions f_{p_i}). Similarly, let a set of workers w_1, w_2, \dots, w_n where their quality are random variables with cumulative function $F(w_i)$. It is previously that if the workers vote of a given picture randomly in a uniform way, the number of votes follows a binomial distribution:

$$P(N_i = k) = C_n^k \left(\frac{1}{m}\right)^k \left(\frac{m-1}{m}\right)^{n-k} \quad (6)$$

Where, $N_i, i \in \{1, \dots, m\}$ are binomial random variables representing the number of votes of photos p_1, p_2, \dots, p_m . One only constraints 3, 4 and 5 of the proposed model are considered, the optimization problem can be formulated as follows:

$$\min_{k \in \{1, \dots, n\}} \mathcal{E}(Q_i) > q_i \quad (7)$$

Where, $\mathcal{E}(Q_i)$ represents the average quality of picture p_i and q_i its minimal required quality.

In other words, the objective consists to determinate the minimal number of workers (or voters) who must choose a given picture so that it will be the most representative one. To study the impact of each considered parameter in the proposed model, we have conducted a simulation with the following method: A

user action is a vote to select the best picture from a set of pictures corresponding to a bucket, based on a calculated quality considering the quality of the pictures and the quality of the worker. Picture quality is calculated with a Gaussian distribution with a mean of 0.8. We consider that the data are of good quality in general. The user task execution depends on the user quality. Again, we use a Gaussian distribution to compute this quality. We will vary the mean regarding the user quality. User quality is an important issue regarding crowdsourcing. We consider that user voluntarily contribute to the task but that they are not experts nor trained. They may provide by results genuinely.

We have defined a value function for the pictures. Let P_i be the picture quality, w be the worker quality. The picture value is drawn as a random number based on the following distribution : $N(P_i, 1 - W)$ The worker votes for the picture with the highest value.

We select the first picture that gets k vote. Vote distribution is random among buckets. We stop to vote for a bucket as soon as a picture reaches k votes.

5.2 The assessment phase model

Regarding the assessment of a picture we use the following model. Photos are graded on a discrete scale (0,1,2). The model assumes that the better the photo is, the more accurate the grading. We are just computing the chance that a user with a given worker quality W gives the right answer.

The correct result is assumed if $P_i * W > Random()$. Else, we chose the resulting vote equi-probably between the two remaining choices.

5.3 Simulations

We have conducted several simulations using different kind of parameters. The first parameter is the bucket size. We use squares of side length of 50m, 100m and 200m. With smaller bucket, we can expect a more accurate map but a bad coverage. On our corpus (Paris) it represents respectively 50779, 12576, 3120 buckets. The number of vote k required to decide for a result (1,3,5,7,9). Again, the more vote we require, the less result we will get since more actions will be required to take a decision in each area. On the other side, if we require more vote, we may expect more accurate results. Since we consider that we have a limited number of user actions, we can distribute them between selection and assessment, from 50/50 to 80/20. The more action we allocate to selection, the more results we will have to assess. The worker quality is the last parameter that can vary. We assume that workers are doing their best but that may not have always a perfect judgement regarding the task to accomplish. Thus we assign a quality to workers that varies from 0.6 to 0.9 with a gaussian distribution.

In the next section, we describes the result of our experiments based on this model.

6 Results

The simulation has been executed using data collected on the area of Paris. We have collected tweets on this area on periods of 24 hours. The number of geolocalised tweets collected did not vary much from one day to the other, around 60000 tweets per 24 hours. It is very difficult to identify clear patterns in the distribution (cf. figure 3). The coverage in the city center is better than in the

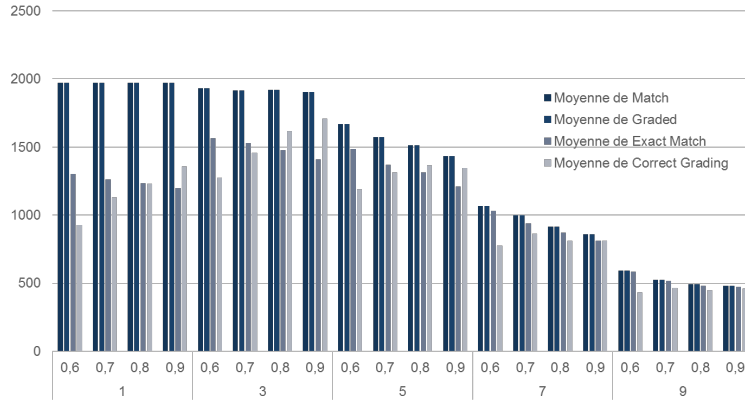


Fig. 6. Number of results depending on the worker quality and the number of votes

suburban areas. The same exercise done in other big cities like New York or London give the same kind of results. There is a high density of tweets in touristic area or business districts. It means that we cannot expect a full coverage of a geographical area with tweets. But since we want to compare the resulting coverage based on different distribution of actions, we are concerned with relative results more than with absolute results.

We have conducted several experiments, varying the parameters. In Figure 6, we vary the number of vote and the worker quality. Half of the actions are used for selection and half for the assessment. Match means that we have been able to select a picture. Exact match means that the best one has been selected by the voters. Graded means that the picture has been graded and Exact grading means that the grading is correct for the bucket.

With one vote to make the selection, the number of action is sufficient to select and grade pictures for every bucket. We see that in this case, and with the parameters that we have selected, the variation of the number of exact match does not vary much with the quality of the worker but the variation of the exact grading is much more important. With 3 votes, the number of results is lower but the number of exact match is higher in the end compared to 1 vote whatever the quality of the worker. With 5 votes, as it could be expected, the number of

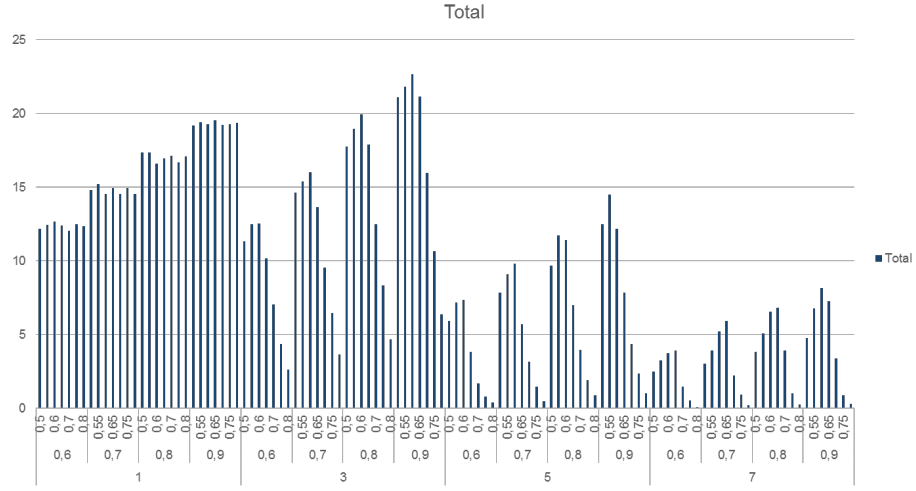


Fig. 7. Number of results depending on the worker quality and the number of votes

results decreased as well as the number of correct results. It is the percentage of good results compare to the overall result that increase. The more voter we use, the less results we get but of better quality.

In Figure 7 we vary distribution of actions, quality and votes. This graphs shows the coverage of the correct results. It is the % of correct results. Varying the distribution of action is interesting especially if we use several votes for each bucket. The more action we use for selection, the more results we get for grading. We try to understand here how the percentage of correct grading vary based on the variation of these three parameters. Again the best result depends on the quality of the worker ($W=0.9$ and $\text{vote}=3$). But we can observe than if we use one vote and good workers we get a lot of good results. We can also notice that with these parameters, the quality of the result decrease rapidly with the quality of the worker.

In figure 9 we did the same simulation, varying the number of actions from 20000 to 100000. This figure presents the best result depending on the number of action and on the worker quality.

Number of actions	20000	40000	60000	80000	100000
Distribution	0.6	0.7	0.6	0.7	0.7
Vote	3	3	5	7	7
Best Result	13.37	15.95	17.06	18.04	19.4

Fig. 8. Best results by varying the number of actions and votes (worker quality fixed to 0.6)

We are now focusing on the quality according to the workers distributions. Table 8 represents a subset of the result we get. This table represents the best results obtained for all the parameters (distribution, vote and number of actions) for a workers quality (the mean quality of a worker is fixed to 0.6). We can observe that the increase of the correct results by varying these parameters is very slow. The simulation suggests that when we increase the number of actions done by worker, we must increase the number of votes to get better results. By varying workers quality, we notice that the percentage of best results increases. This percentage increases also according to the number of actions. However, as shown in Table 8 and Figure 9, the gain slows above a certain number of actions. This trend is still consolidating for different values of workers quality figure 9(0.7, 0.8 and 0.9).

As we can see in figure 9 it is interesting to note that the gain obtained when we increase the number of actions is less interesting than gain obtained when we have access to better worker . That would suggest that it is better to get access to a smaller number of skilled worker that to a large number of workers. In the results, we also note that when we have access to more action, we obtain a better result when we increase the number of votes.

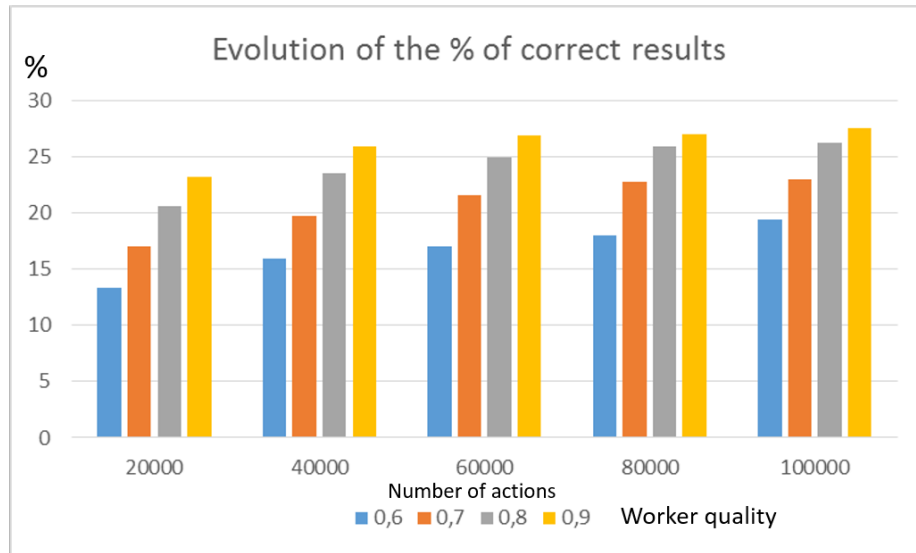


Fig. 9. Evolution of the number of correct results depending on the number of action and on the worker quality

7 Discussion and Conclusion

The simulation we have done has been repeated several times on the same area with very similar results. This could be anticipated since a lot of data we use for the simulation are synthesized. Overall, based on our model and on the corpus we rely on, we can expect good results with 3 to 7 votes depending on the number of actions. Still, we see that the quality of workers is very important and that it's more efficient to select a reduced number of skilled workers. It means that it could be very valuable either to train the workers to increase their capacity to provide good results or to try to assess worker quality and distribute the work to the best ones first. To do that, we need to get more insight on the way worker behave with the kind of tasks that we plan to ask them to execute. This is the next step of this project, i.e. conduct experiments to compare the execution of these kinds of tasks with domain experts and with ordinary people, with or without training and to evaluate the average quality we can obtain with different kind of task and questions.

References

1. Meier, P.: Digital Humanitarians: How Big Data Is Changing the Face of Humanitarian Response. null edition edn. Routledge, Boca Raton, FL (January 2015)
2. Benouaret, K., Valliyur-Ramalingam, R., Charoy, F.: Answering complex location-based queries with crowdsourcing. In: 9th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing, Austin, TX, USA, October 20-23, 2013. (2013) 438–447
3. Purohit, H., Hampton, A., Bhatt, S., Shalin, V., Sheth, A., Flach, J.: Identifying Seekers and Suppliers in Social Media Communities to Support Crisis Coordination. *Journal of CSCW* (2014)
4. Alt, F., Shirazi, A.S., Schmidt, A., Kramer, U., Nawaz, Z.: Location-based Crowdsourcing: Extending Crowdsourcing to the Real World. In: Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries. NordiCHI '10, New York, NY, USA, ACM (2010) 13–22
5. Bulut, M.F., Yilmaz, Y.S., Demirbas, M.: Crowdsourcing location-based queries. In: PerCom Workshops. (2011) 513–518
6. Kazemi, L., Shahabi, C.: GeoCrowd: enabling query answering with spatial crowdsourcing. In: SIGSPATIAL/GIS. (2012) 189–198
7. Guo, S., Parameswaran, A., Garcia-Molina, H.: So who won? dynamic max discovery with the crowd. Technical report, Stanford University (November 2011)
8. Pomerol, J.C., Barba-Romero, S.: Multicriterion Decision in Management: Principles and Practice. Springer Publishing Company, Incorporated (2012)
9. Adelman, R.M., Winston, A.B.: Sophisticated voting with information for two voting functions. *Journal of Economic Theory* **15**(1) (1977) 145–159
10. Eriksson, B.: Learning to top-k search using pairwise comparisons. In: Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2013, Scottsdale, AZ, USA, April 29 - May 1, 2013. (2013) 265–273
11. Pfeiffer, T., Gao, X.A., Rand, D.G.: Adaptive polling for information aggregation. In: AAAI. (2012)

12. Peng Ye, Doermann, D.: Combining preference and absolute judgements in a crowd-sourced setting (June 2013) ICML'13 workshop: Machine Learning Meets Crowdsourcing.
13. Davidson, S.B., Khanna, S., Milo, T., Roy, S.: Using the crowd for top-k and group-by queries. In: Joint 2013 EDBT/ICDT Conferences, ICDT '13 Proceedings, Genoa, Italy, March 18-22, 2013. (2013) 225–236
14. Feige, U., Raghavan, P., Peleg, D., Upfal, E.: Computing with noisy information. *SIAM J. Comput.* **23**(5) (October 1994) 1001–1018
15. Khan, A.R., Garcia-Molina, H.: Hybrid strategies for finding the max with the crowd: Technical report. Technical report, Stanford University (February 2014)
16. Wauthier, F., Jordan, M., Jojic, N.: Efficient ranking from pairwise comparisons. In Dasgupta, S., Mcallester, D., eds.: Proceedings of the 30th International Conference on Machine Learning (ICML-13). Volume 28., JMLR Workshop and Conference Proceedings (May 2013) 109–117
17. Bernstein, M.S., Little, G., Miller, R.C., Hartmann, B., Ackerman, M.S., Karger, D.R., Crowell, D., Panovich, K.: Soylent: a word processor with a crowd inside. In: UIST. (2010) 313–322