Using Crowdsourcing to Identify the "Best" Item

Kerui Huang and Neoklis Polyzotis
Department of Computer Science
University of California, Santa Cruz

Baskin Engineering UC SANTA CRUZ

Crowdsourcing

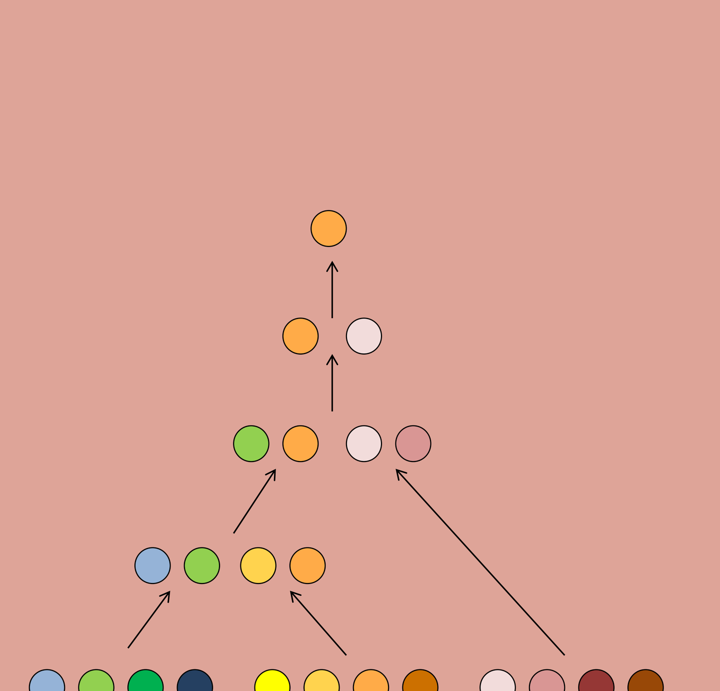
- Humans are simply better than machines at solving several hard problems. Example problems include image categorization, image labeling and fuzzy ranking.
- Crowdsourcing services (e.g., Amazon Mechanical Turk) allow applications to harness human computation at a large scale.

Objective

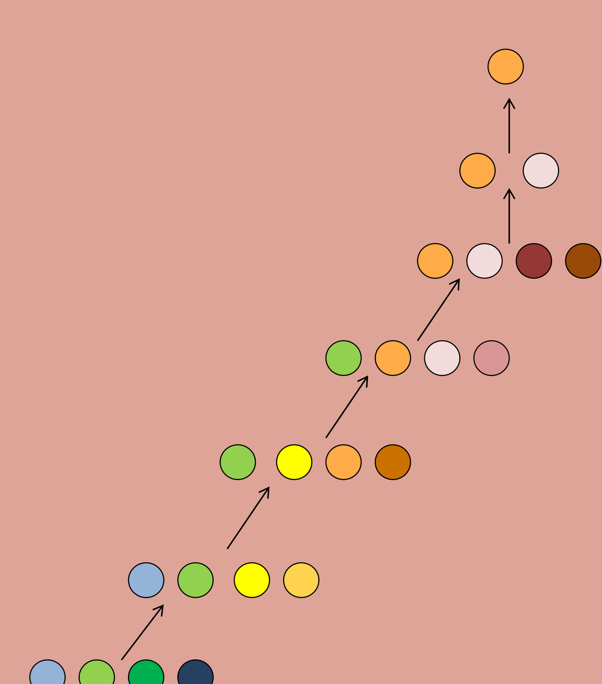
- In this work, we examine the usage of crowdsourcing for a common problem: Identifying the "best" item in a set of items.
- Example applications:
  - Identify the "best" photo out of set of photos for a place or a person.
  - Identify the "best" way to break down a complex task out of several options.
- We investigate two algorithms and their trade-offs in terms of quality, expense and latency.

Algorithms

We have two algorithms for our iterative task: Tournament and Bubble.



The execution of the Tournament algorithm



The execution of the Bubble algorithm

Methodology

- We use a dataset which has 72 blurred images.
- We vary the following parameters: the number of inputs and outputs of a section; the number of sections in a single HIT; the number of assignments of each HIT; how many times the entire algorithm is repeated.
- We use the metric $\sqrt{\sigma^2_{chosen\_image} - \sigma^2_{correct\_image}}$ for judging the quality of the output, where σ is the blur radius in the Gaussian blur function. Since the correct image is not blurred at all, which means $\sigma_{correct\_image} = 0$, the calculation is deducted to σ_{chose_image}.

Two images within different degrees of blur

Table: The parameters in different groups

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10
Number of questions per section	4	8	8	8	8	8	8	8	8	8
Number of answers per section	2	2	2	2	1	4	1	4	2	2
Number of sections per HIT	4	4	1	2	2	2	4	4	4	4
Number of assignments per HIT	3	3	3	3	3	3	3	3	1	5
Times each algorithm is tested	5	5	5	5	5	5	5	5	15	3

Results

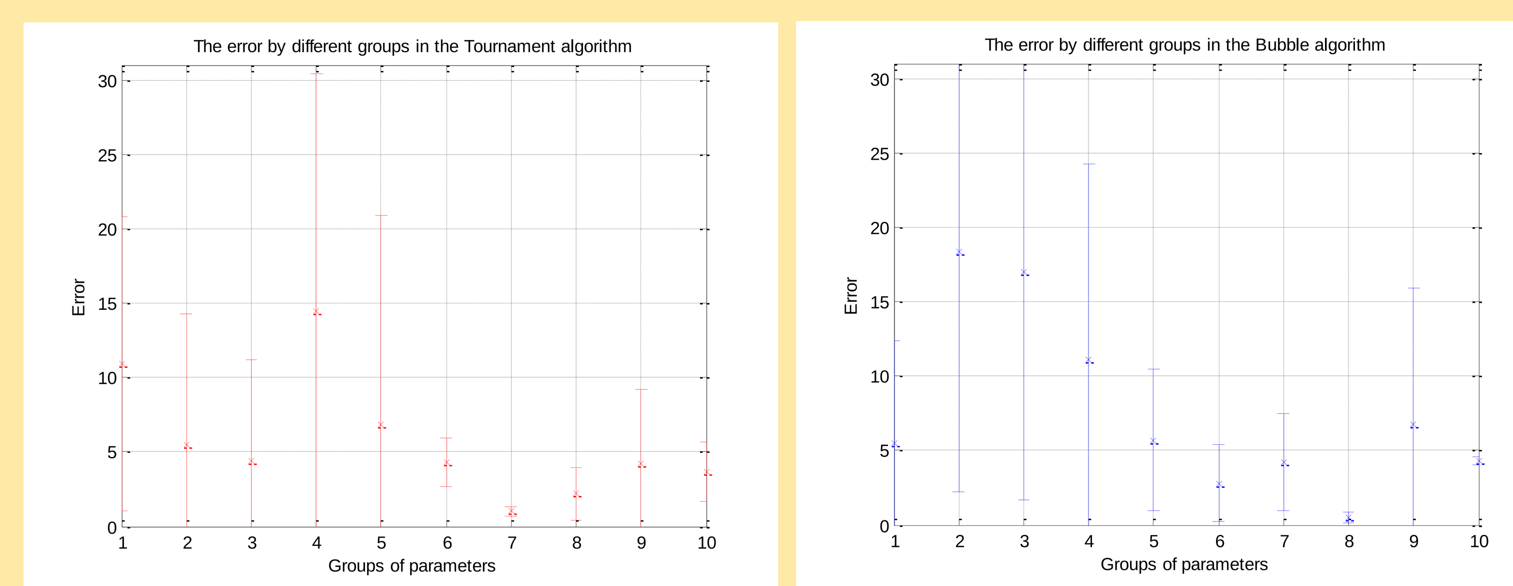


Figure 1. Output error (lower is better) by different groups of parameters in the Tournament and Bubble algorithms. Error is measured as the difference between the output image and the actual best image. Each point corresponds to the average of several experiments.

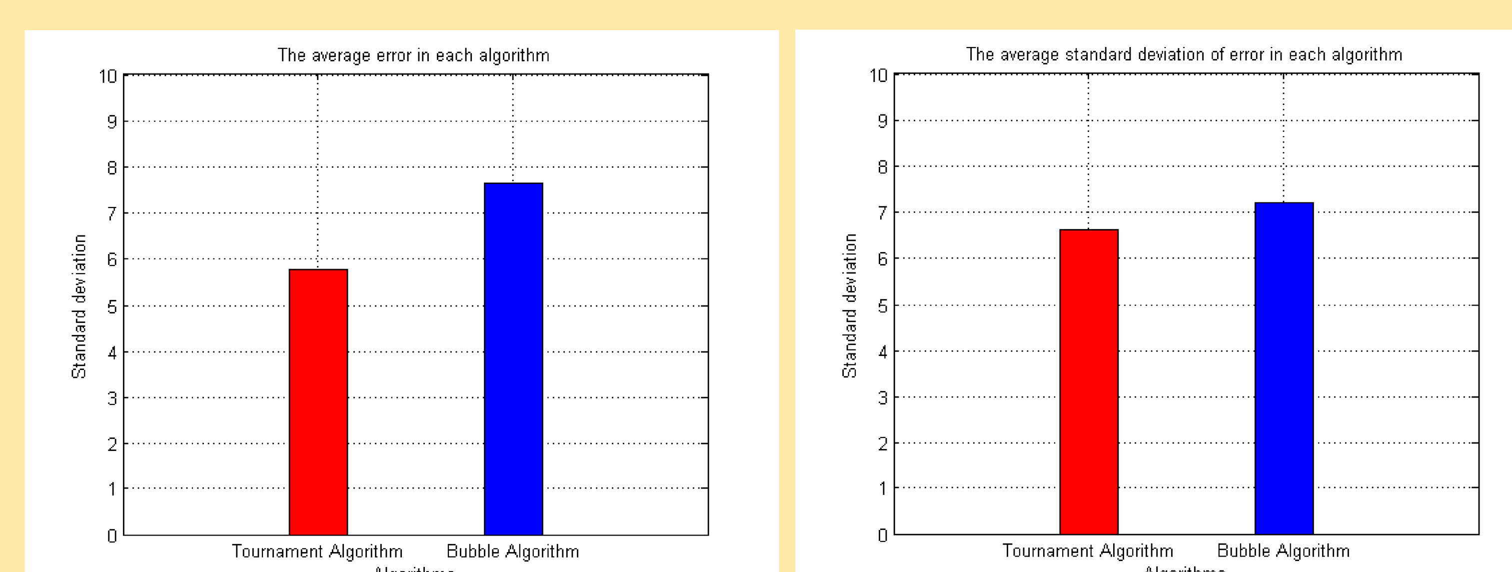


Figure 2. Overall comparison between Tournament and Bubble.

Key Findings:

- Tournament works better than Bubble (but requires more HITs => higher cost).
- Requiring fewer outputs per section is likely to reduce error.
- Repetitions within each HIT work better than repetitions of the whole algorithm.

Conclusions

More and more applications are resorting to human computation to solve hard problems.

Algorithms that effectively combine human and machine computation will be a key building block.

Code and documentation are available at: http://users.soe.ucsc.edu/~khuang/top1doc/

Acknowledgments

Thanks to Paul Heymann (Bionica) for providing the data set of blurred images.