Improving the Quality of Crowdsourced Image Labeling via Label Similarity

Yi-Li Fang, Hai-Long Sun, Peng-Peng Chen, and Ting Deng

Fang YL, Sun HL, Chen PP et al. Improving the quality of crowdsourced image labeling via label similarity. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 32(5): 877–889 Sept. 2017. DOI 10.1007/s11390-017-1770-7

Problem

Parallel approach

Sequential Approach

- **Task Allocation:**

(1) Approach 1: Parallel

Method: Each task is simultaneously published to multiple workers who independently complete the task. In addition, each worker is not permitted to observe the results of others.

Weakness : The noisy and dispersed labels generated in this way always make it hard to infer the actual labels.

(2) Approach 2: Sequential

Method: Tasks are successively distributed to multiple workers who can observe prior versions of the labels.

Weakness: The incorrect labels obtained previously often mislead the successive workers. This way makes labels more likely to be incorrect.

- **Result Inference:**

When characterizing the performance of workers, existing methods treat each incorrect label equally, which is unfair to estimate worker ability.

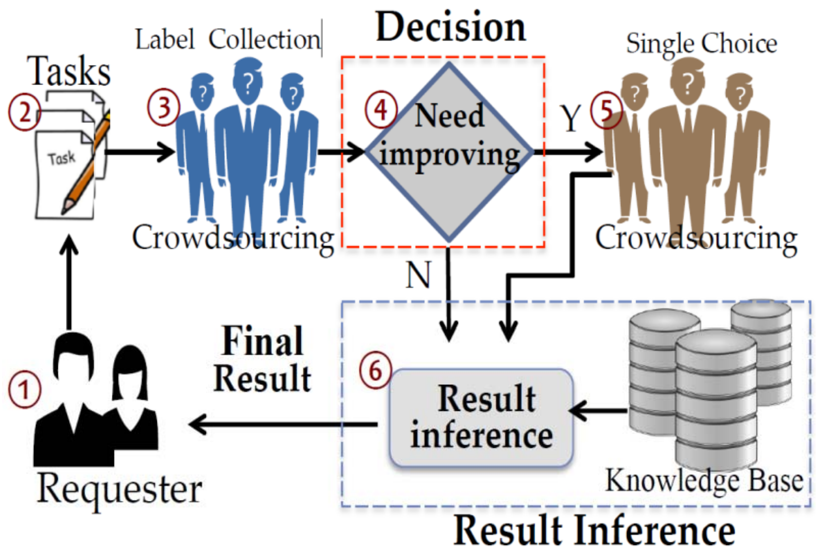
Our Workflow

***Step* 1:** A requester publishes tasks to a crowdsourcing platform, e.g. Crowdflower.

***Step* 2:** Tasks are assigned to workers by using the task allocation method in our crowdsourcing workflow;

***Step* 3:** In label collection, every worker provides a label to describe the object in the corresponding image.

***Step* 4:** This step concerns an either-or decision about im- -plementing the second-round processing in step 5 or the result inference in step 6.

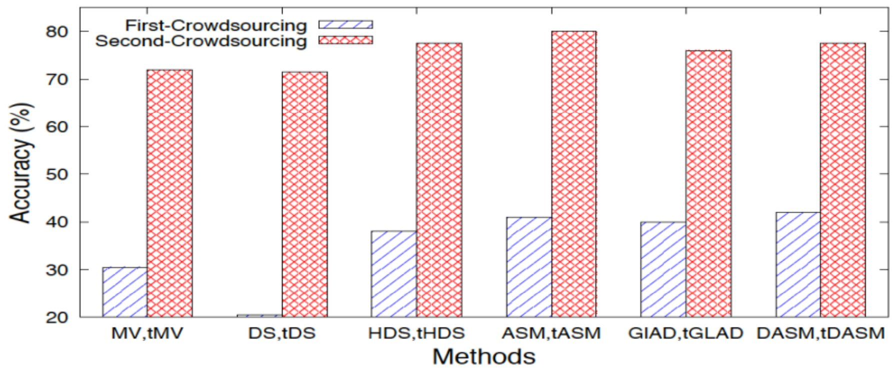


***Step* 5:** Each task is designed as a single choice task, the options of which are collected from step 3. Every worker is asked to choose the best label for each image and then these labels are submitted to step 6.

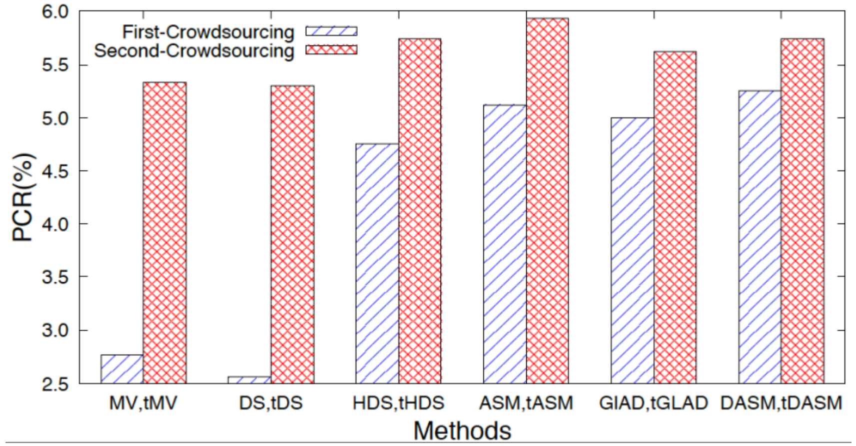
***Step* 6**: Our model involving the similarity of all the labels is applied to infer the final results of all the tasks and these results are submitted to the requester.

Results

- **Accuracy**
  - In terms of accuracy, the two-round allocation can generate higher-quality results.
  - Our two result inference methods (tASM and tDASM) also outperform other four methods.
- **Benefit** (PCR=Accuracy/Cost)
  - As for PCR, two-round allocation always outperforms one-round allocation.
  - Our methods (tASM and tDASM) outperform the other four.



Accuracy varying with different workflows



PCR varying with different workflows

Summary of Our Contributions

- We propose a two-round crowdsourcing workflow to process image labeling, and employ a decision algorithm to decide whether or not to generate the second-round crowdsourcing, which can minimize the number of tasks reprocessed by the second crowdsourcing.
- We incorporate label similarity into result inference and propose two novel inference methods.
- We conducted both real experiments on real crowdsourcing platforms and simulations for our method.