

# CrowdMR: Integrating Crowdsourcing with MapReduce for AI-Hard Problems \*

Jun Chen, Chaokun Wang, and Yiyuan Bai

School of Software, Tsinghua University, Beijing 100084, P.R. China  
chenjun14@mails.thu.edu.cn, chaokun@tsinghua.edu.cn, eldereal@gmail.com

## Abstract

Large-scale distributed computing has made available the resources necessary to solve “AI-hard” problems. As a result, it becomes feasible to automate the processing of such problems, but accuracy is not very high due to the conceptual difficulty of these problems. In this paper, we integrated crowdsourcing with MapReduce to provide a scalable innovative human-machine solution to AI-hard problems, which is called CrowdMR. In CrowdMR, the majority of problem instances are automatically processed by machine while the troublesome instances are redirected to human via crowdsourcing. The results returned from crowdsourcing are validated in the form of CAPTCHA (Completely Automated Public Turing test to Tell Computers and Humans Apart) before adding to the output. An incremental scheduling method was brought forward to combine the results from machine and human in a “pay-as-you-go” way.

## Introduction

Although machine learning does a good job on many problem instances in AI research, there are always instances for which automatic analysis does not provide good answers, e.g. face detection and optical character recognition (OCR). Crowdsourcing alleviates this issue by dissecting a complex problem into pieces of simple questions (often called HITs, *abbr.* Human Intelligence Tasks) which are answered by human in parallel with high accuracy (Howe 2006).

MapReduce (Dean and Ghemawat 2008) provides a model and an architecture to decompose problems in a way that allows them to be solved independently. It offers a software framework to solve data-centric problems with distributed computing and is already applied to many research fields, such as multimedia application and security (Yu et al. 2012).

AI-hard problems (Yampolskiy 2012) are informally defined as difficult problems which need artificial intelligence that matches or exceeds human intelligence to be solved. It includes problems of computer vision or NLP in a common sense. Computer alone is not enough to solve AI-hard problems, and human participation is still needed.

In this paper, we propose CrowdMR — an extended MapReduce model based on crowdsourcing to handle AI-hard problems effectively. Different from pure crowdsourcing solutions, CrowdMR introduces confidence into MapReduce. For a common AI-hard problem like classification and recognition, any instance of that problem will usually be assigned a confidence value by machine learning algorithms. CrowdMR only distributes the low-confidence instances which are the troublesome cases to human as HITs via crowdsourcing. This strategy dramatically reduces the number of HITs that need to be answered by human.

To showcase the usability of CrowdMR, we introduce an example of gender classification using CrowdMR. For the instances whose confidence values are lower than a given threshold  $\alpha$ , CrowdMR distributes HITs in the form of CAPTCHA which automatically discriminates machine and human (Ahn et al. 2008). With gradual completion of HITs, the quality of results is getting better. Due to the flexibility of CrowdMR, we can easily extend it to other AI-hard problems, e.g., word-sense disambiguation and face recognition.

We summarize our major contributions as follows: (1) We integrated crowdsourcing with MapReduce in our CrowdMR model, and the confidence was employed to dissect machine and human computation. (2) An incremental scheduling algorithm based on users’ votes was proposed to deal with the non-real-time property of crowdsourcing. (3) An online gender classification application based on face detection was developed to showcase the usability of CrowdMR.

## Overview of CrowdMR

### Questions in CrowdMR

In CrowdMR, the original AI-hard problem is dissected into pieces of sub-problems. These sub-problems must meet the following criteria: (1) *Clarity*. The answer to sub-problems must be objective and certain. (2) *Simplicity*. Sub-problems must be answerable to all participants, even those without sufficient knowledge in specific fields. (3) *Convenience*. It needs to be convenient throughout the answering process.

### CrowdMR Model

In order to combine human computation with the conventional MapReduce, our CrowdMR model consists of five

\*Corresponding author: Chaokun Wang.

Copyright © 2015, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

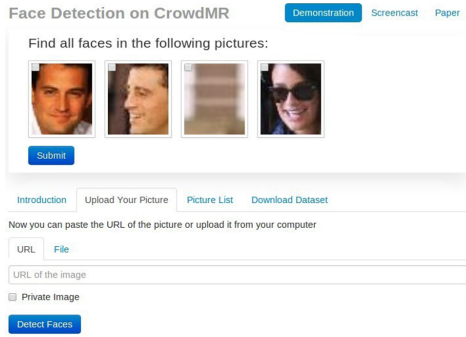


Figure 1: The main page of CrowdMR.

phases: map-auto, map-human, shuffle-auto, shuffle-human and reduce. For each input instance, a confidence is obtained by executing a certain algorithm at the map-auto phase. The instances with high confidence are directly output, while those with low confidence are distributed as HITs in the map-human phase. The pairs output from both map-auto and map-human are combined before sending to grouping phase, where the pairs are clustered using their “Keys” with a certain clustering or classification algorithm (e.g., gender classification). Grouping phase is also composed of shuffle-auto and shuffle-human. The results from shuffle-human via crowdsourcing are combined with those from shuffle-auto before sending to reducer.

Considering the non-real-time property of crowdsourcing, we proposed an incremental scheduling algorithm which outputs the machine results first, and then updates the results incrementally as more HITs are answered. This procedure provides a “pay-as-you-go” strategy, which means the result of a job is always accessible after the machine tasks are finished, and users can always get better results as more HITs are completed and merged.

## CrowdMR Application

We show the application of gender classification using our CrowdMR model. The system contains two major parts: (1) The front end uses a graphical way to show all the procedures carried out in CrowdMR. It also provides a Crowdsourcing Task Distribution Interface based on CAPTCHA. (2) The back end executes face detection and gender classification upon the uploaded images. The map phase deals with face detection while the grouping phase performs classification. The main page of our system is shown in Fig. 1.

### Crowdsourcing Task Distribution Interface

Crowdsourcing task distribution interface (CTDI) is the user interface between our system and participants who answer HITs. We distribute crowdsourcing tasks as CAPTCHAs. CAPTCHA can distinguish human from machine to ensure that the answers obtained from the CTDI are reliable. To achieve that goal, several HITs are combined to a single question. The HITs come from 2 groups: (1) Ground-truth questions which are already validated by human. (2) HIT questions which need to be answered by human. For exam-

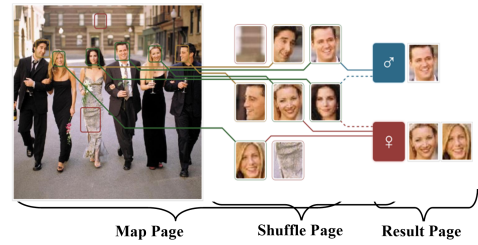


Figure 2: An example of results on different pages.

ple, as shown in Fig. 1, the 1<sup>st</sup>, 3<sup>rd</sup>, 4<sup>th</sup> images are ground-truth questions where the 1<sup>st</sup> and the 4<sup>th</sup> are positive and the 3<sup>rd</sup> is negative, and then the 2<sup>nd</sup> is the real HIT question. If a participant gives correct answers to the 1<sup>st</sup>, the 3<sup>rd</sup> and the 4<sup>th</sup>, her answer to the 2<sup>nd</sup> is also considered reliable.

## Gender Classification

Fig. 2 shows the three major pages in CrowdMR. The **map page** shows the output from mapper. We use different colors to distinguish results with different confidence. Green lines point to results whose confidence values are larger than the threshold  $\alpha$  and are used as positive ground-truth, while yellow lines point to low-confidence results which are used as HIT questions. The unlinked results represent the extremely low-confidence ones and are used as negative ground-truth. The **shuffle page** shows the results of gender classification. The results with higher confidence are shown as solid lines and generate ground-truth questions in the shuffle-human phase. The low confidence results are represented by dashed lines and generate HIT questions. The final gender classification results are shown in the **result page**. More results may be shown here along with more HITs being answered.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (No. 61373023, No. 61170064) and the National High Technology Research and Development Program of China (No. 2013AA013204). We thank Ping Qiu and Yichi Liu for their discussion with the authors.

## References

- Ahn, L.; Maurer, B.; McMillen, C.; Abraham, D.; and Blum, M. 2008. reCAPTCHA: human-based character recognition via web security measures. *Science* 321(5895):1465–1468.
- Dean, J., and Ghemawat, S. 2008. MapReduce: simplified data processing on large clusters. *Communications of the ACM* 51(1):107–113.
- Howe, J. 2006. The rise of crowdsourcing. *Wired magazine* 14(6):1–4.
- Yampolskiy, R. 2012. AI-complete, AI-hard, or AI-easy – classification of problems in AI. In *The 23rd Midwest Artificial Intelligence and Cognitive Science Conference*.
- Yu, Z.; Wang, C.; Thomborson, C.; Wang, J.; Lian, S.; and Vasilakos, A. 2012. Multimedia applications and security in MapReduce: opportunities and challenges. In *Concurrency and Computation: Practice and Experience*, 2083–2101.