

iCrowd: An Adaptive Crowdsourcing Framework

Paper presents new crowdsourcing algorithm. Aim of this new algorithm is increasing the credibility of collected data. Algorithm consists of 2 parts. These are estimating workers' accuracy for different domains and assigning tasks to workers according to calculated accuracy. iCrowd firstly secretly qualify each worker and finds their best accuracy domain. Then it tries to assign most suitable domain tasks to each worker to increase overall result reliance.

First, additional information such as lemmas at the end of the paper and this makes paper more readable. Also, when they mentioned the figures, they put the values in the parenthesis so that, readers don't have to scroll back to the figures. Secondly, iCrowd algorithm is a dynamic which means while the algorithm working, new workers can join or leave the system and this doesn't affect the system results. Task similarity calculations, worker accuracy estimations and task assignments are also dynamic. So that, each of them are calculated while the workers are answering the questions. Besides, algorithm is not domain specific so every company can create their own set of questions and iCrowd can work on it. Moreover, algorithm qualifies the workers, but they are not aware of it. That is a good point, because if they know they are being qualified, that can affect their answering abilities. The other good point is, it can eliminate the workers that aren't good enough to answer questions correctly. It uses threshold for this and eliminating is necessary because it increases the overall data quality.

The algorithm is logically nearly perfect and it's hard to find a weakness in it. But I found a few questionable points. For example, in order to find the task similarities and workers' strongest knowledge area, you need to use lots of questions. Because, each domain consists of a lot of subdomains. For instance, strong knowledge about iPhones doesn't indicate adequate knowledge about iMacs. So, in order to find each users' strength, companies that are crowdsourcing need to prepare at least one questions about each subdomain/area/product. This put a load on the companies because qualification algorithm requires questions with ground truths for accuracy estimation.

In the AMT experiment with *YahooQA* data, the authors only used 110 questions with 6 domains. The second experiment with *ItemCompare* dataset has reasonable number of data but I don't think 110 is a good number for showing the results of the new algorithm. They should have been use more questions and maybe more domains as well. In addition to that, they only show the efficiency/scalability of the same algorithm in the last graph. I would like to see the other algorithms results in the same graph for comparison.

2019-02-10

Baran Kaya, kayab@mcmaster.ca, 400284996