Summary
The authors presented a prototype of a crowdsourcing system to solve complex analytics problems, called CDAS. CDAS can use the power of crowdsourcing to solve problems such as image tagging and natural language processing problems (movie rating). The proposed system relies on breaking down a complex task into smaller jobs for computers and human workers. We assume that not all workers are trustworthy; we might have malicious users who intentionally give inaccuracy answers.

Key ideas in the paper:
- The complex job in Amazon Mechanical Turk is divided into Human Intelligence Tasks (HITs). Each task is assigned to multiple workers to obtain the result.
- Since there are untrustworthy users, the adaptive query model tries to improve the accuracy by first predicting how many workers we need for each task. The next step is waiting for answers from workers and accepting answers until the answers constitute a clear winner or final answer.
- In the first step, given an user’s query, and minimum accuracy, we can compute the minimum number of workers per task using Chernoff bound and the Bayesian theorem.
- In the second step, the probability verification is done by the system as answers come in. The main contribution of the paper is here, where we compute the confidence of each answers based on the partial responses. If the confidence of the answer exceeds the specified threshold, then we stop accepting responses and so we do not have to pay those workers who have not responded. The model minimizes the economic cost where fixed salary and wage for each HIT response is constant for workers and a certain confidence level is achieved. The three proposed stopping criteria methods are MinMax, MinExp, ExpMax; the criteria are mainly based on the minimum expectation of the best/accurate answer vs. the second best.
- The experiment result demonstrates the superiority of the proposed verification method vs. half-voting or majority voting used in current crowdsourcing frameworks.

Positive/Strong Points
- S1 The proposed framework is context-independent/generic. We do not have to consider the context where the queries come from in the verification process.
- S2 The computation of minimum number of workers per task is very fast – the Chernoff bound combined with binary search is a new idea to formally define the number of workers per task.
- S3 The authors showed a very detailed analysis/formal proof of uncertainty in user’s answers – no empirical result is used in building the model (except for some parameter’s settings).
- S4 This method can be done online; and it provides superior performance comparing to pure machine learning methods even when there is one worker employed.

Negative/Weak Points
- W1 The accuracy for this crowdsourcing method is very high due to the fact the user’s accuracy is already very high for high complex tasks and in addition, user’s judgment is also used as the ground truth in many cases.
- W2 The authors mentioned economic costs in the method section, but did not present a formal equation/computation/result and perform a comparison vs. cost using only machine learning techniques.
- W3 The second application: image tagging was not
- W4 We assume the number of workers per task is always more than 1 and we can assign multiple workers to one task (in the prediction method where we compute the minimum number of workers to achieve the desired accuracy). However, in practice, there might be a few available workers and a limited budget; how can we incorporate the partial answers from users? That is whether the system can filter through data and send the most representative questions to users, and based on the responses, the system would learn from users and perform the rest of the queries (in a supervised learning manner).

Research Questions and Points for Discussion
- D1 How can we improve the accuracy estimation while preserving privacy? Sampling rates in the paper is very prone to error and yet does not capture each user’s knowledge in specific domains.
- D2 How can we better utilize the context? The probabilistic approach for verification implicitly assumes independency between answers. We could like to improve the accuracy on numerical answers by modifying the
- D3 One of many assumptions for this framework is that there is only one accurate answer for each query. In the case of multiple correct answers, how can we modify the equations for number of workers and answer confidence.
- D4 See Weakness 4