QASCA: A Quality-Aware Task Assignment System for Crowdsourcing Applications
Yudian Zheng*, Jiannan Wang$, Guoliang Li#, Reynold Cheng*, Jianhua Feng#
#Tsinghua University, *University of Hong Kong, $UC Berkeley

Crowdsourcing
Coordinate a crowd to answer questions that solve computer-hard applications.

Task Assignment
Given n questions, which k questions should be batched in a HIT and assigned to a worker?

Here we have n=4 questions, and a HIT contains k=2 questions.

Evaluation Metric
An application is often associated with an Evaluation Metric.

<table>
<thead>
<tr>
<th>Application</th>
<th>Sentiment Analysis</th>
<th>Entity Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question</td>
<td></td>
<td></td>
</tr>
<tr>
<td><a href="Opinion">I had to wait for six friggin hours in line at the pug store.</a></td>
<td>[iPad 2 == iPad 3rd Gen?](equal or non-equal)</td>
<td></td>
</tr>
<tr>
<td>Evaluation Metric</td>
<td>Accuracy</td>
<td>F-score (&quot;equal&quot; label)</td>
</tr>
</tbody>
</table>

Our Goal
Evaluation Metric ➔ Assignment
When a worker comes, given the evaluation metric, for each set of k questions, we estimate the improvement of quality, if these k questions are answered by the worker, and we select the k questions that can maximize the quality improvement for the worker.

Challenges & Solutions
- [Unknown Ground Truth] With ground truth unknown, how to evaluate the quality of returned results?
  We build a distribution matrix for the n questions’ answers, and evaluate the quality of returned results by computing over the distribution matrix.
- [Expensive Enumeration] Enumerating all possible assignments is exponential.
  We develop linear-time algorithms to compute optimal assignments for Accuracy and F-score, respectively.

QASCA System Architecture

Experiments
We compare with five systems (Baseline, CDAS[1], AskIt[2], MaxMargin and ExpLoss) on real-world datasets.

References