Abstract
We demonstrate the following four facilities provided by the system CrowdCleaner: (1) an error-monitor to find out which items (e.g., submission date, price of real estate, etc.) are wrong versions according to the reports from the crowds, which belongs to a passive crowdsourcing strategy; (2) a task-manager to allocate the tasks to human workers intelligently; (3) a smart-decision-maker to identify which answer from the crowds is correct with active crowdsourcing methods; and (4) a whom-to-ask-finder to discover which users (or human workers) should be the most credible according to their answer records.

System Framework

Crucial Modules
- **Error Monitor**: It discovers the new errors of multi-version data and evaluates whether each reported error is valuable. Then, the error-monitor module decides which reported errors are the actual errors, or the spam reports.
- **Task-manager**: It assigns the questions to human workers based on the submitted errors from the error-monitor module.
- **Smart-decision-maker**: It employs the entropy-based decision strategy to determine whether the answers of human workers are consistent. Thus, each expected repaired result is actually considered as a discrete random variable.
- **Whom-to-ask-finder**: It finds some credible human workers instead of experts.

Technical Background
- **Entropy-based decision strategy**: From the frequencies of different suggestions, the possibility of each suggestion $x_i (1 \leq i \leq n)$ is denoted $Pr(x_i)$. Formally, we define the entropy of an expected repaired result $X$ as

$$H(X) = -\sum_{i=1}^{n} Pr(x_i) \log Pr(x_i)$$

When the diversity is too large, we further use the submodularity of entropy to clean the uncertainty of spam suggestions.
- **Whom-to-ask strategy**: a group of credible workers $CW_n = \{cw_1, cw_2, ..., cw_n\} \subseteq W$ with size $n$, where each $cw_i$ is associated with an confidence $c_i$, and $W$ is the set of all human workers. Thus, the group confidence of credible workers is

$$GC(CW_n) = Pr(|C| \geq \frac{n}{2}) = Pr(|C| \geq \frac{n+1}{2}) = \sum_{k=\lceil\frac{n}{2}\rceil}^{n} \prod_{i=1}^{k} c_i \prod_{j \in A^k} (1 - c_j)$$

the group confidence is used to measure which human workers are credible.

Demo Interface

Acknowledgements
This work is supported in part by the Hong Kong RGC Project MHKUST602/12, National Grand Fundamental Research 973 Program of China under Grant 2012-CB316200, Microsoft Research Asia Gift Grant and Google Faculty Award 2013.