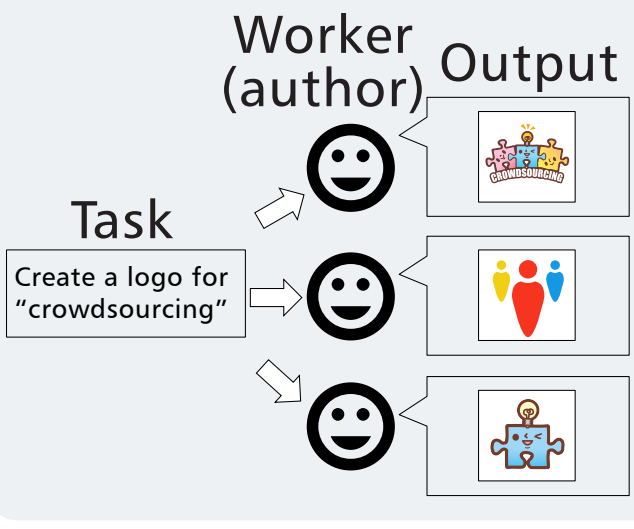
Statistical Quality Estimation for General Crowdsourcing Tasks

Yukino Baba and Hisashi Kashima (Univ. Tokyo)

OVERVIEW

We propose an unsupervised statistical quality estimation method for general crowdsourcing tasks

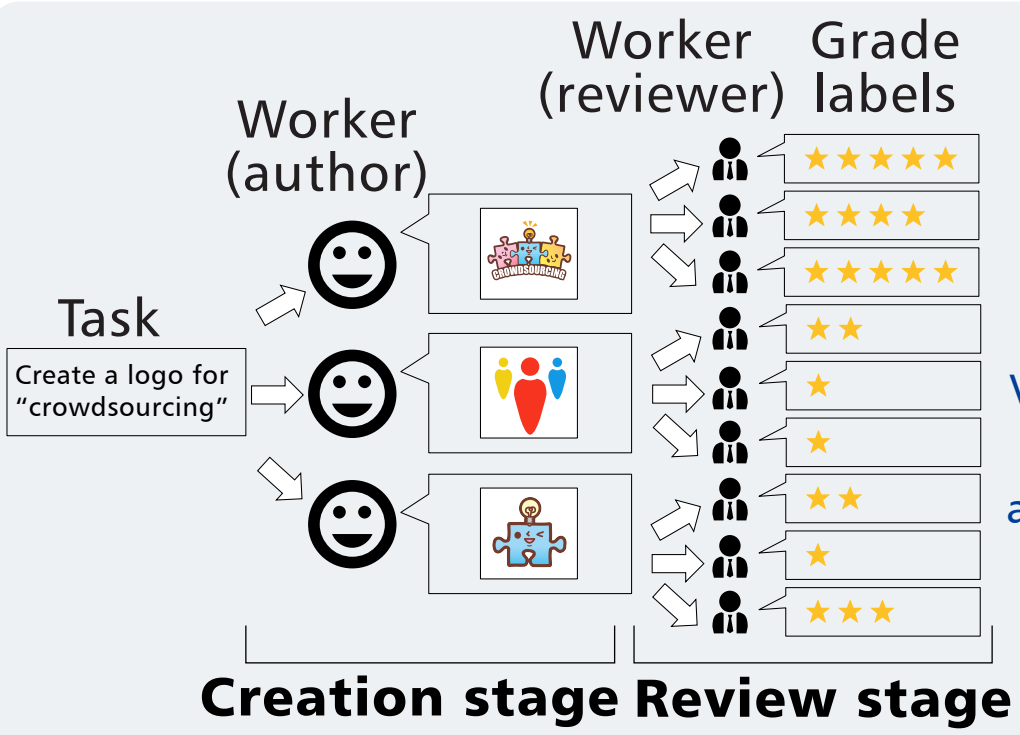


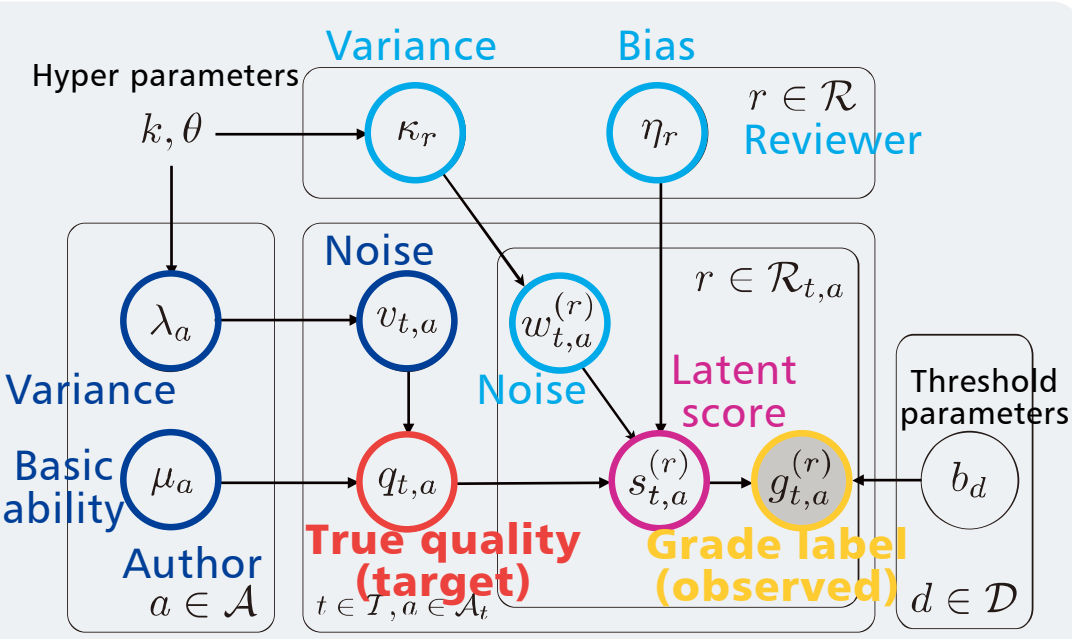
General crowdsourcing tasks:
Tasks with unstructured response format (e.g., logo design and article writing)

It is a common quality control technique on crowdsourcing to assign multiple workers to the same task.
Our method allows you to choose the best one from the multiple worker outputs.

APPROACH

We introduce a **two-stage generative model** consisting the **creation stage** and the **review stage**





Output: (t, a)

$g_{t,a}^{(r)} \in \mathcal{D}, \mathcal{D} = \{1, 2, \cdots, n\}, q_{t,a} \in \mathbb{R}$

RESULTS

Our method outperforms popular vote aggregation methods in terms of the estimation accuracy and performance in finding the best output

Three tasks:
- Logo designing
- Image description
- Language translation

Baselines:
- Averaging of grade labels
- Ordinal label aggregation [Rayker and Yu, '11] (considering only reviewer abilities)