DOCS: A Domain-Aware Crowdsourcing System Using Knowledge Bases

Yudian Zheng*, Guoliang Li#, Reynold Cheng*

*University of Hong Kong, #Tsinghua University

Crowdsourcing Workflow

- **Task Assignment**: When a worker comes to the platform, the worker will be assigned to a set of tasks;
- **Truth Inference**: When a worker accomplishes tasks, the platform will collect answers from the worker.

Existing Works Fail in QA Tasks

- Each task is related to different domains
- Select the most suitable tasks for assignment

Build Domain Aware Task Model

- **Entity linking** (map entity to knowledge bases)
- Hierarchical domains in knowledge bases
- Obtain the task model (a vector of distribution)

Build Domain Aware Worker Model

- Use qualification test (like an “exam”)
- Two rules for selecting qualification test
  1. Each selected task should capture a certain domain
  2. The domain distribution of selected tasks should approximate the distribution of all tasks

![KL-divergence formula](image)

Truth Inference

- 1. **Quality for each worker** → **Truth for each task**
- 2. **Truth for each task** → **Quality for each worker**

Task Assignment

- Select the most suitable tasks for assignment

Experiments

- Worker Characteristics on Dataset D_Item
- System Comparisons

![Graph showing system comparisons](image)