Cross-domain-aware Worker Selection with Training for Crowdsourced Annotation

Yushi Sun[1], Jiachuan Wang[1], Peng Cheng[2], Libin Zheng[3], Lei Chen[1,4] , Jian Yin[3]

[1]The Hong Kong University of Science and Technology, Hong Kong SAR, China

[2]East China Normal University, Shanghai, China

[3]Sun Yat-sen University, Guangzhou, China

[4]The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China

香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

East China Normal University

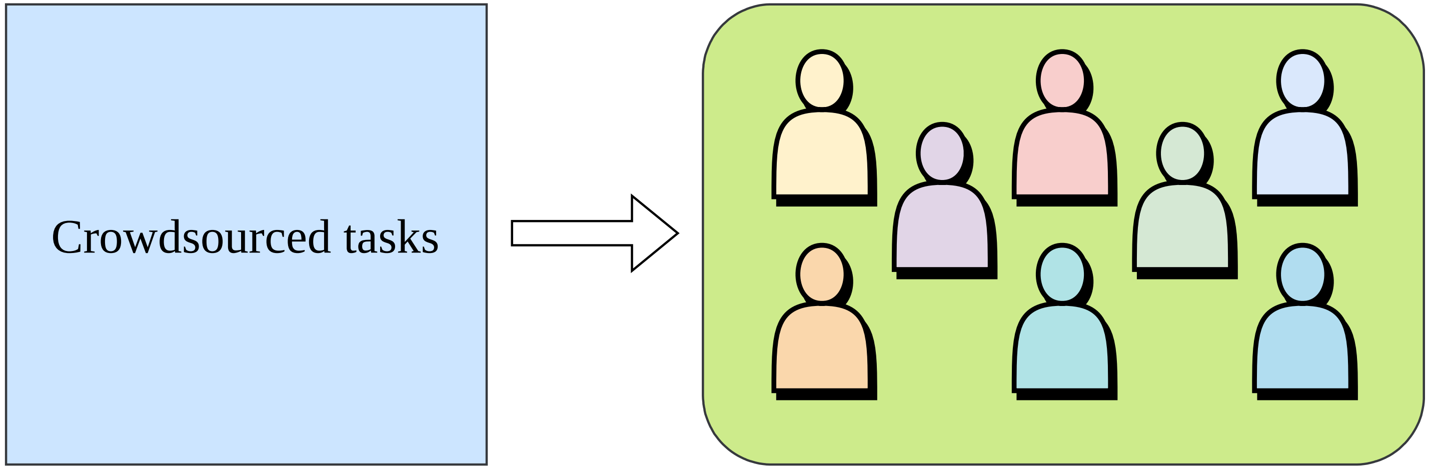
香港科技大學(廣州) THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

Outline

- Background and Motivation
- Definitions
- Methodology
- Experiments
- Summary

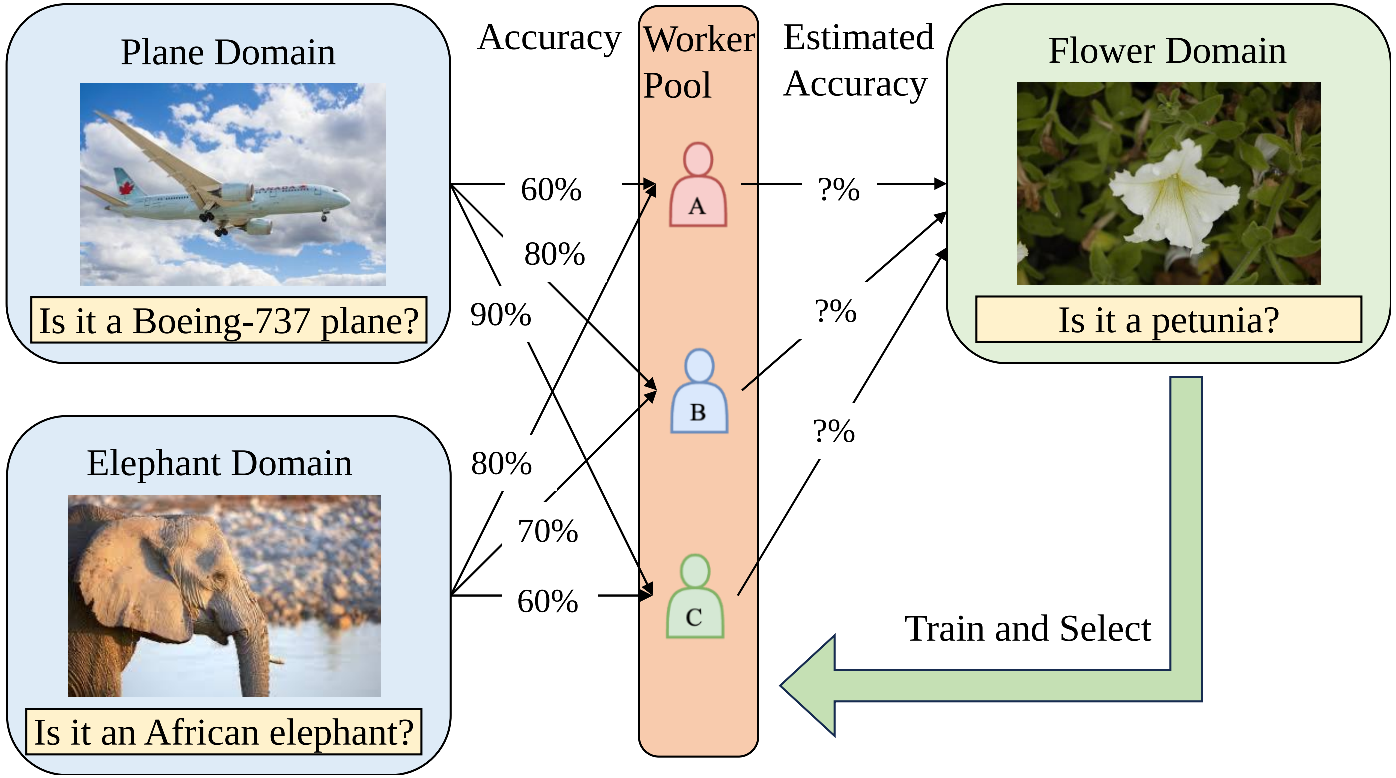
Background and Motivation

- Crowdsourcing is preferable for obtaining **high-quality** annotations for **large-scale** datasets.
- **Worker Selection is important** in Crowdsourcing.
- How to design an allocation scheme to select high-performance crowd workers remains a challenge.



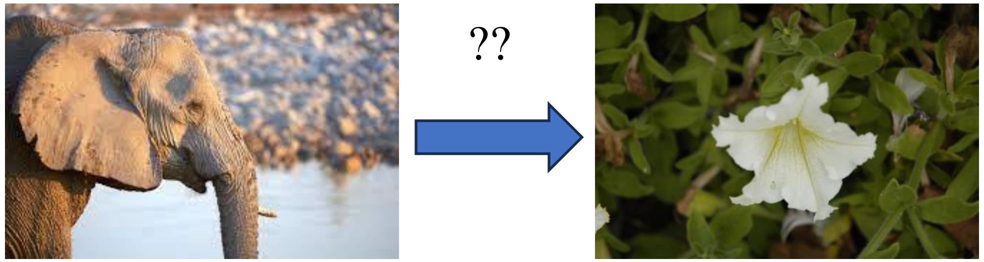
Background and Motivation

- The **answering history of workers** (prior domain knowledge) can help select high-quality workers when **annotating a new domain** (target domain task).



Background and Motivation

- However, there are two challenges:
  - Difficulty in accurately estimating the **correlation between domains** with a limited budget.
  - Difficulty in estimating the **workers' dynamic knowledge change** during the question-answering worker training process.



Outline

- Background and Motivation
- Definitions
- Methodology
- Experiments
- Summary

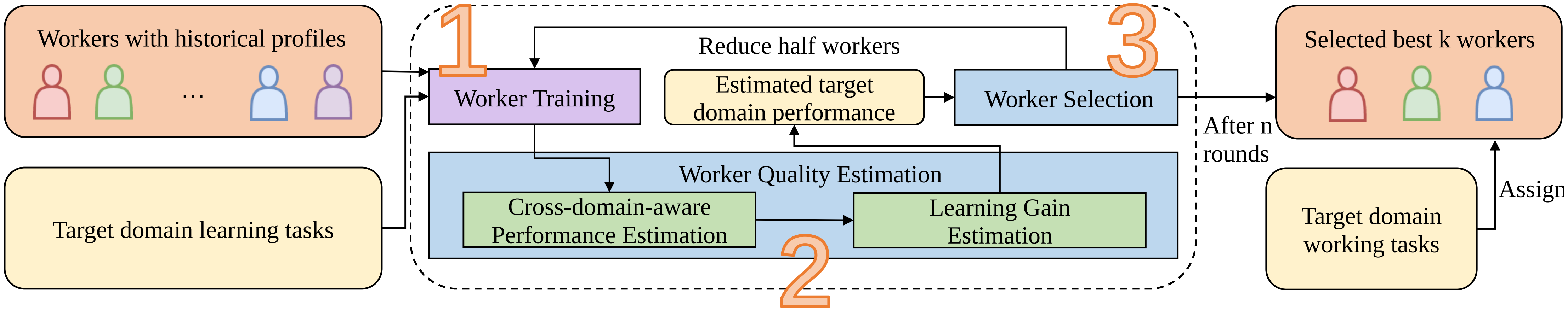
Definitions

- Cross-domain-aware worker selection with training:
  - Given target domain tasks $T = \{T_l, T_w\}$, the total budget B, and worker pool W with each worker w_i's historical profile h_i.
  - Cross-domain-aware worker selection with training problem is to 1) assign no more than B tasks to |W| workers **for training** and 2) **select top k workers** with the highest possible annotation accuracy on working tasks T_w.

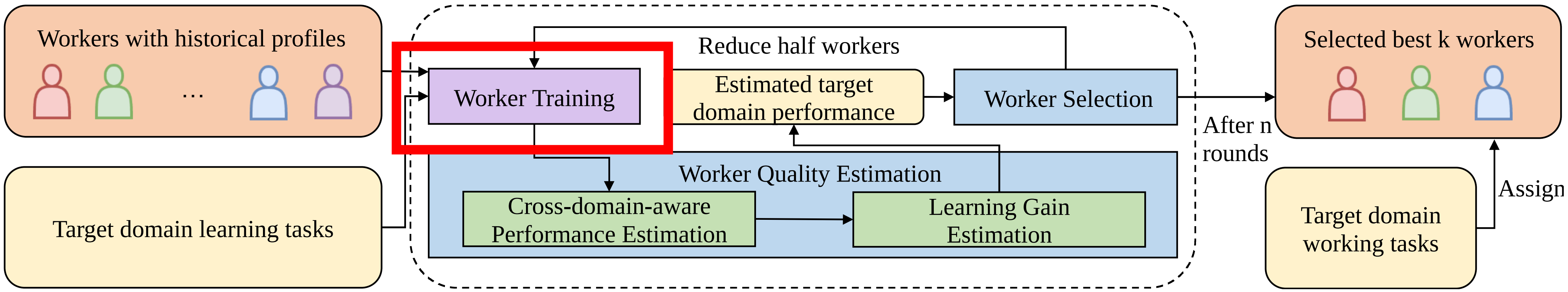
Outline

- Background and Motivation
- Definitions
- Methodology
- Experiments
- Summary

Methodology

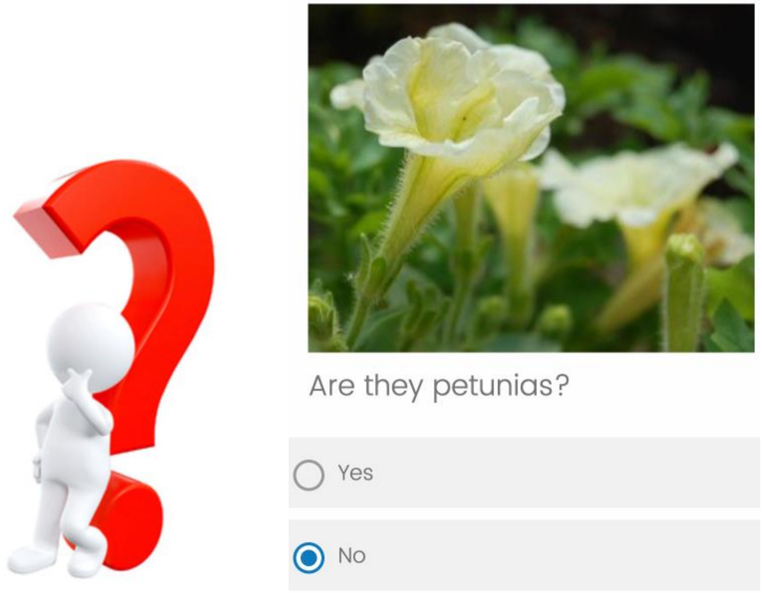


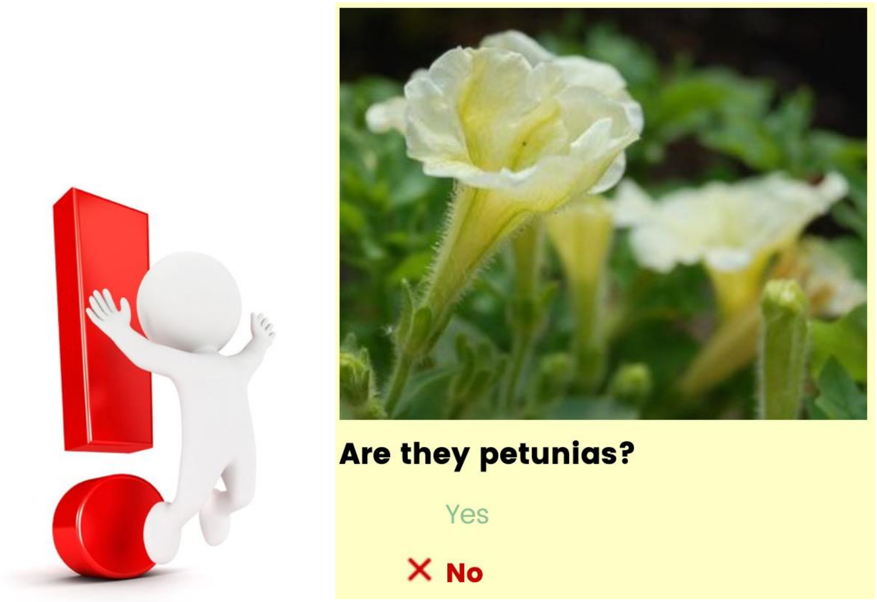
Methodology



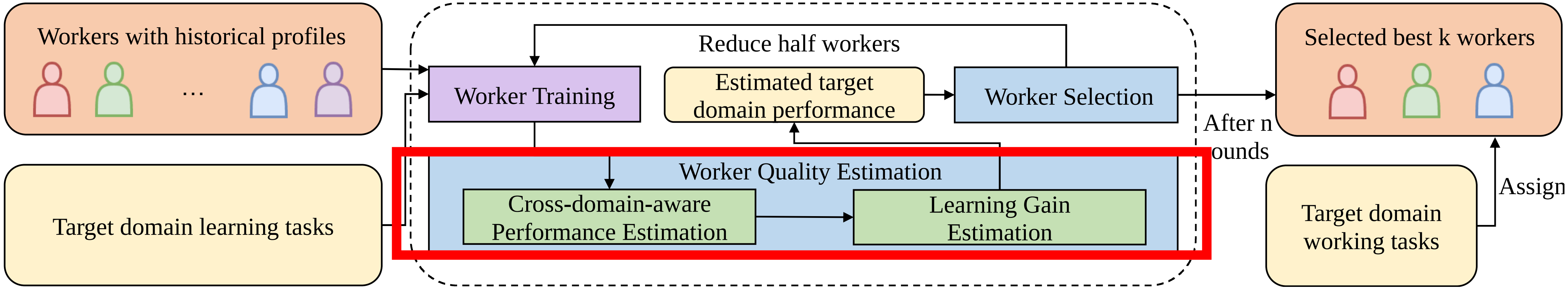
Methodology – Worker Training

- Worker training is treated as an **"Answer and learn"** process for workers.





Methodology

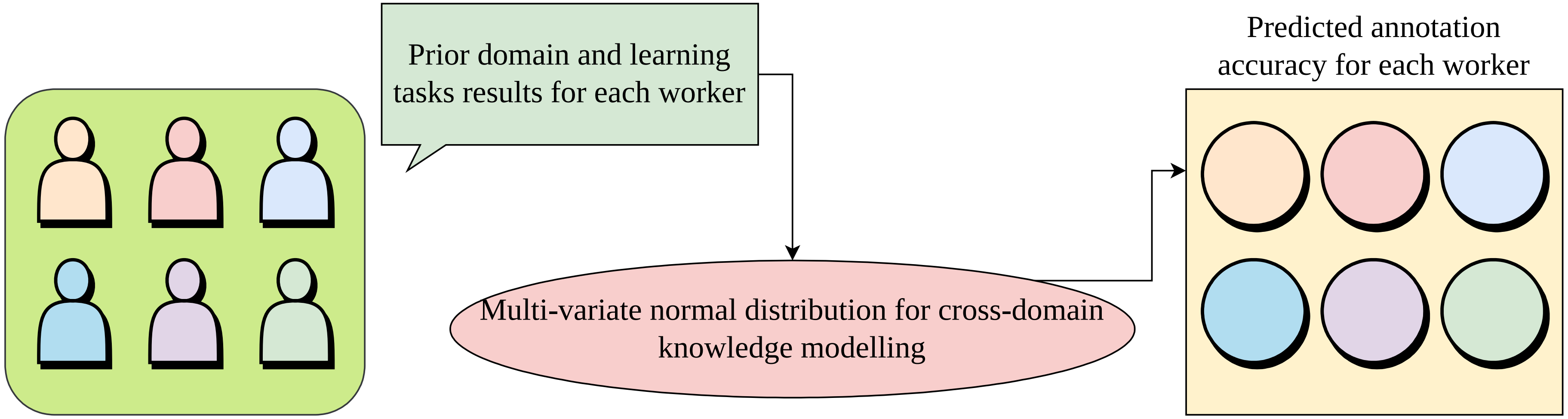


Methodology – Worker Quality Estimation

- We consider two factors in estimating workers' quality:
- **Cross-domain correlation** – Cross-domain-aware Performance Estimation (CPE)
- **Worker learning gain** – Learning Gain Estimation (LGE)

Methodology - CPE

- Model the **correlation** between workers' **prior knowledge and the target domain** knowledge as a **multivariate normal distribution**.
- Record **the correct and wrong number** of learning tasks for each worker.
- Update the distribution with **maximum likelihood estimation**.
- **Predict the annotation accuracy** of each worker.



Methodology - CPE

- Maximum likelihood estimation:

$$\begin{aligned}\log L &= \sum_{i=1}^{|W_c|} \log P(h_{i,T}|h_i) \\ &= \sum_{i=1}^{|W_c|} \log \int_0^1 h_{i,T}^{C_{i,c}} (1-h_{i,T})^{X_{i,c}} \frac{e^{-\Psi}}{\sqrt{2\pi|\bar{\Sigma}|}} \mathrm{d}h_{i,T} \\ &= \sum_{i=1}^{|W_c|} \Big[\log \int_0^1 h_{i,T}^{C_{i,c}} (1-h_{i,T})^{X_{i,c}} e^{-\Psi} \mathrm{d}h_{i,T} \\ &\quad + \log \frac{1}{\sqrt{2\pi}} - \frac{1}{2}\log|\bar{\Sigma}|\Big],\end{aligned}$$

$$\bar{\mu} = \mu_T + \Sigma_{1\times D}\Sigma_{D\times D}^{-1}(h_i - \mu_{1\sim D}),$$
$$\bar{\Sigma} = \Sigma_{1\times 1} - \Sigma_{1\times D}\Sigma_{D\times D}^{-1}\Sigma_{D\times 1},$$

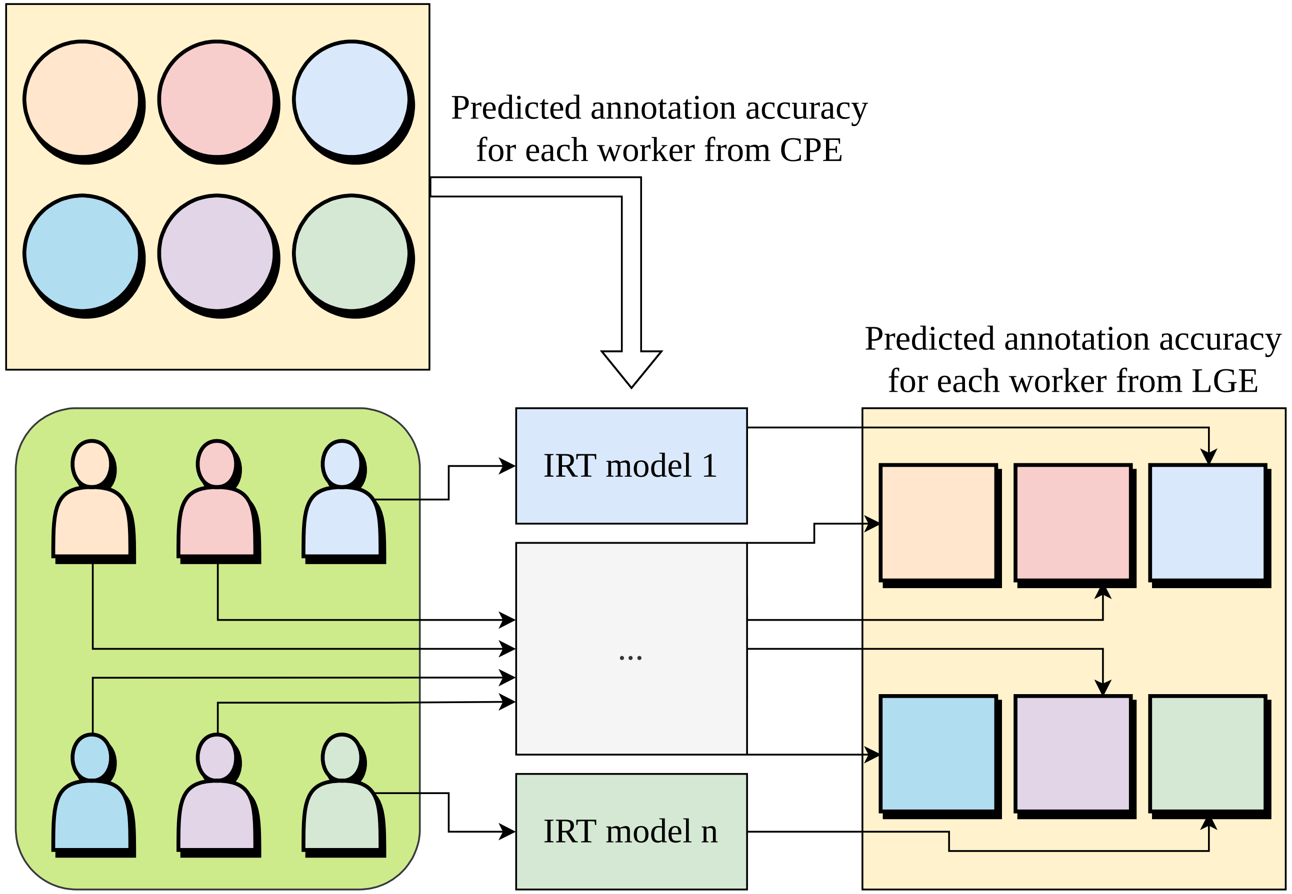
and $\Psi = \frac{(h_{i,T}-\bar{\mu})^{\mathsf{T}}(h_{i,T}-\bar{\mu})}{2\bar{\Sigma}}$.

- Updated annotation accuracy:

$$\begin{aligned}p_{c,i} &= E[h_{i,T}|h_i] \\ &= \int_0^1 h_{i,T} P(h_{i,T}|h_i) \mathrm{d}h_{i,T} \\ &= \int_0^1 h_{i,T} \frac{P(h_i, h_{i,T})}{P(h_i)} \mathrm{d}h_{i,T},\end{aligned}$$

Methodology - LGE

- Adapt the **Item Response Theory (IRT)** model to estimate the learning gain.
  - Compute the IRT scores on the **prior domains**.
  - Compute the IRT scores on the **target domain learning tasks**.
  - Update the learning parameter α_i for each worker based on the **CPE scores and answering history**.
- Predict the estimated scores in the current round.



Methodology - LGE

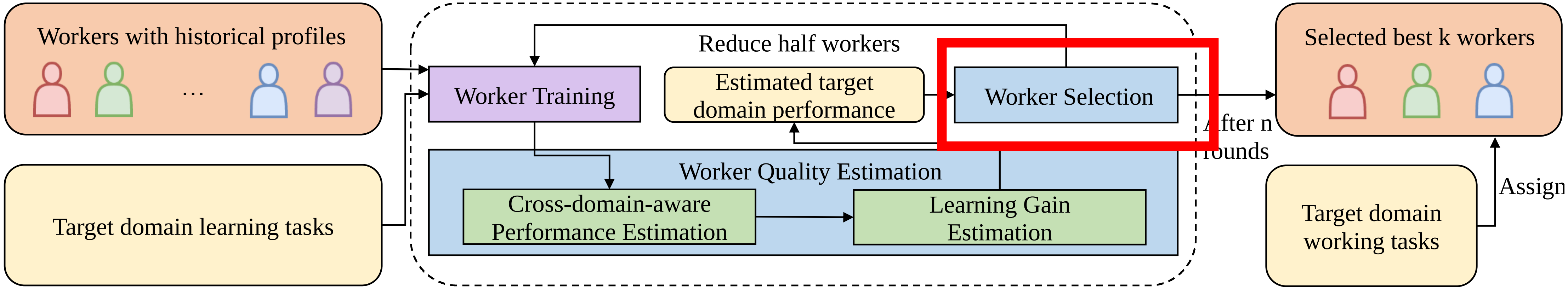
- IRT score:

$$\hat{p}_{j,i,d} = g(\alpha_i, \beta_d, K_j) = \frac{1}{1 + e^{-(\alpha_i \ln(K_j+1) - \beta_d)}}.$$

- Update the learning parameter α_i:

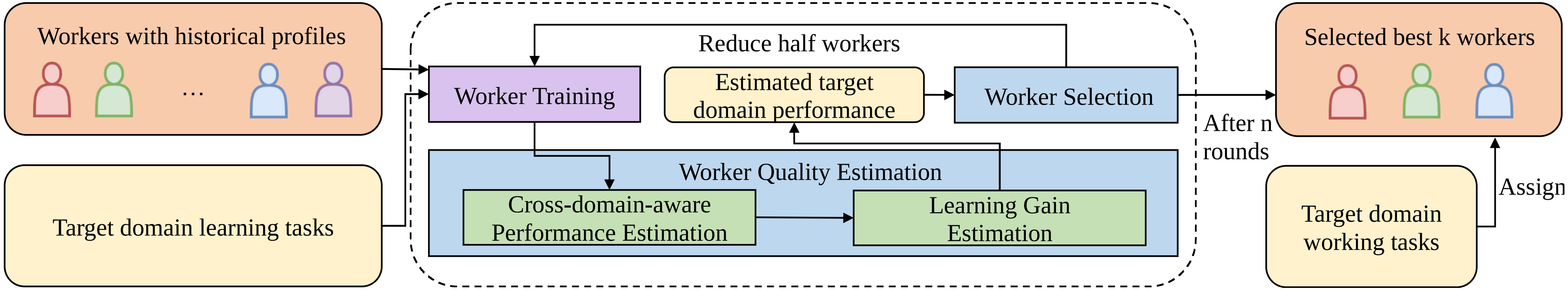
$$\alpha_i = \arg\min_{\alpha_i} \left[\sum_{d=1}^{D} (\hat{p}_{1,i,d} - h_{i,d})^2 + \sum_{j=1}^{c} (\hat{p}_{j-1,i,t} - p_{j,i})^2 \right]$$

Methodology



Methodology – Worker Selection

- Adapt the ME algorithm to select the **top half of the workers** in the current round.
- Error bound: $O(\sqrt{\frac{nk}{B}\ln\frac{1}{\delta_c}})$.



Outline

- Background and Motivation
- Definitions
- Methodology
- Experiments
- Summary

Experiments

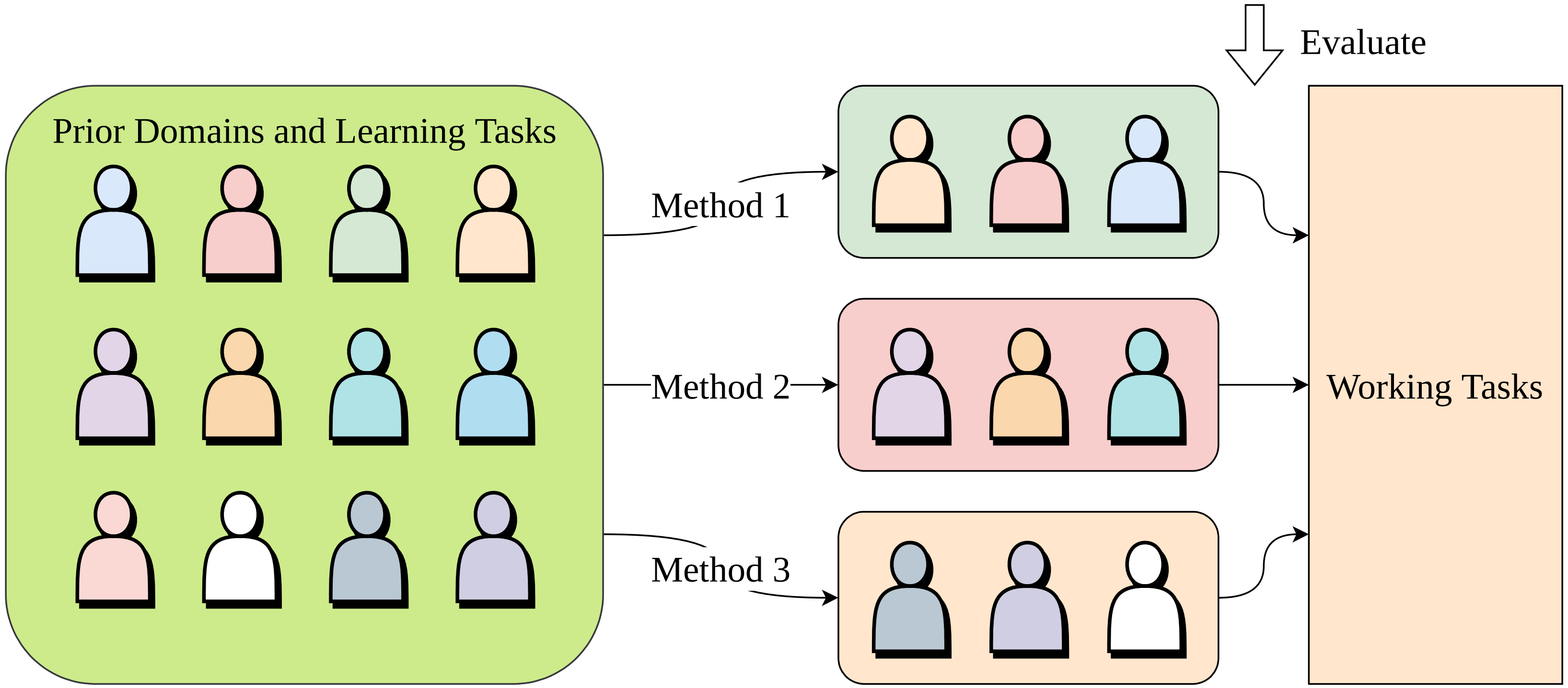
- Datasets:

TABLE II
DATASET STATISTICS

Datasets	$\|W\|$	Q	k	total # of batches	B
RW-1	27	10	7	3	540
RW-2	35	10	9	3	700
S-1	40	20	5	7	2400
S-2	50	20	5	7	3000
S-3	80	20	5	15	6400
S-4	160	20	5	31	16000

Experiments

- Metric: averaged annotation accuracy of the selected top-k workers on the target domain working task.



Experiments

- Baselines: We considered three baselines, Universal Sampling (US), Medium Elimination (ME), and Li et al.
  - US: use the budget for all the workers equally and select the top k workers
  - ME: allocates the budget in rounds and eliminates the workers by half in each round based on the accuracy of the learning tasks
  - Li et al.: compute the correlation between the prior domain historical results with the target domain performance

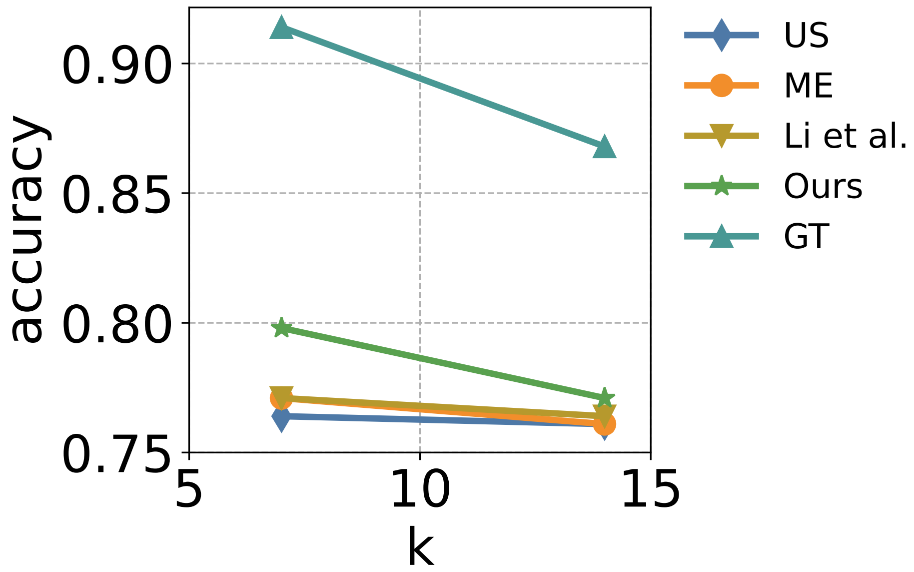
Experiments

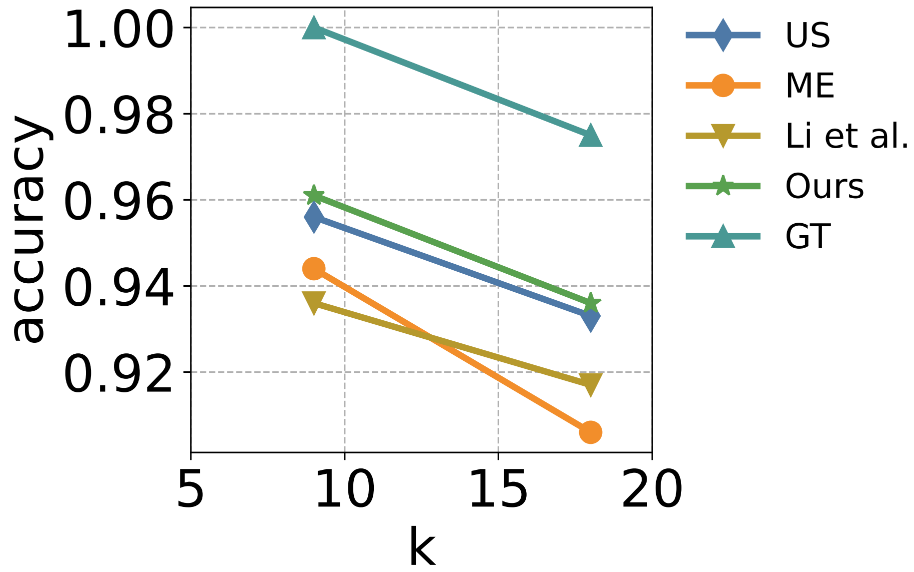
TABLE V
EXPERIMENT RESULTS

	RW-1	RW-2	S-1	S-2	S-3	S-4
US [11], [19]	0.764 (4.5% ↑)	0.956 (0.5% ↑)	0.765 (8.5% ↑)	0.775 (6.8% ↑)	0.815 (4.3% ↑)	0.865 (2.4% ↑)
ME [11], [19]	0.771 (3.5% ↑)	0.944 (1.8% ↑)	0.720 (15.3% ↑)	0.785 (5.5% ↑)	0.795 (6.9% ↑)	0.880 (0.7% ↑)
Li et al. [31]	0.771 (3.5% ↑)	0.936 (2.7% ↑)	0.780 (6.4% ↑)	0.805 (2.9% ↑)	0.845 (0.6% ↑)	0.870 (1.8% ↑)
Ours	**0.798**	**0.961**	**0.830**	**0.828**	**0.850**	**0.886**
Ground Truth	0.914	1.000	0.885	0.875	0.915	0.975

Experiments

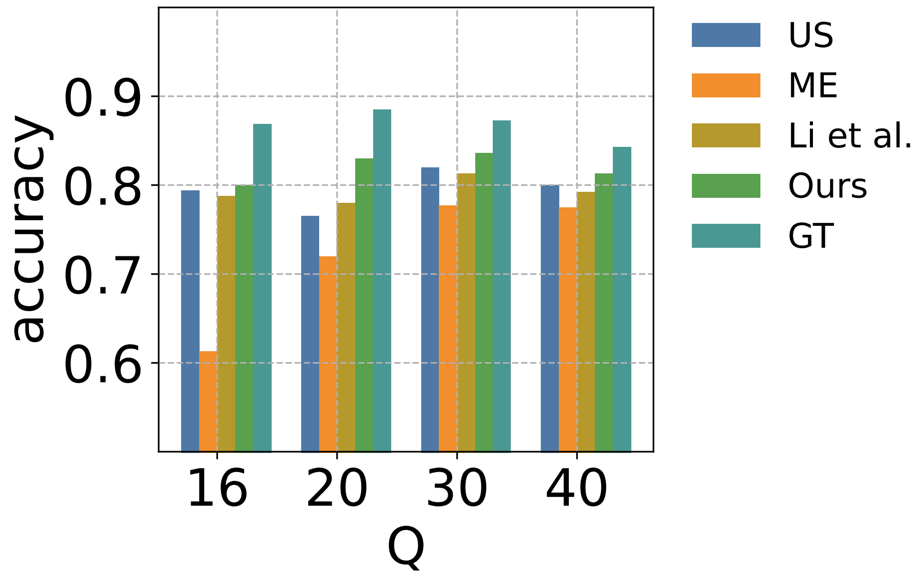
- Stability over the parameter k (number of desired workers)

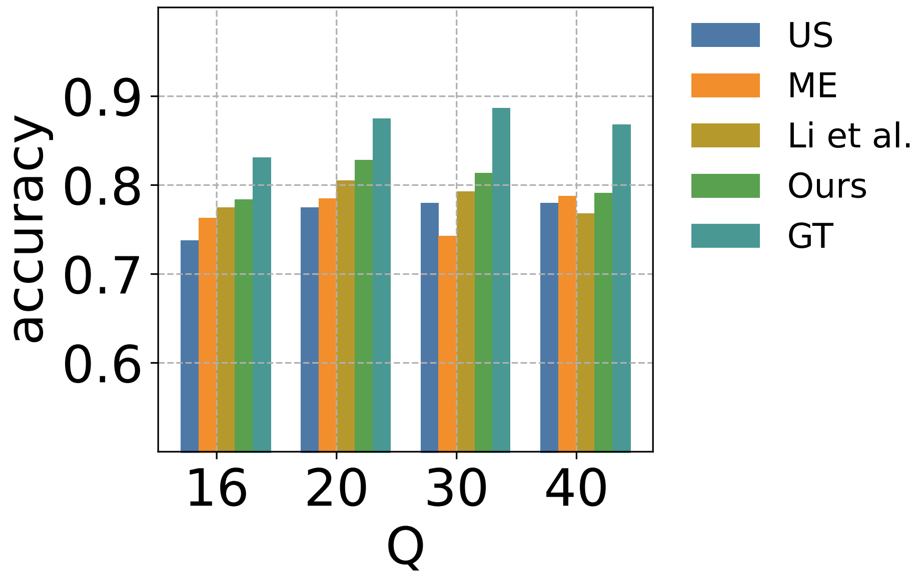




Experiments

- Stability over the parameter Q (number of learning tasks per batch)





Outline

- Background and Motivation
- Definitions
- Methodology
- Experiments
- Summary

Summary

- We incorporate the **cross-domain knowledge** information and propose a novel **Median Elimination-based** worker selection with training algorithm to find high-quality workers.
- We comprehensively consider the **learning gain** of workers during the learning task worker training process over the new domain to get a better estimate of the **dynamic change** in worker quality.
- We collect **two novel cross-domain worker selection datasets** for the crowdsourcing research community to study the problem of cross-domain worker selection with training.
- We conduct **extensive experiments** on real-world and synthesized datasets to evaluate the performance of our proposed method comprehensively.

Thank you