Robust Deep Learning from Crowds with Belief Propagation

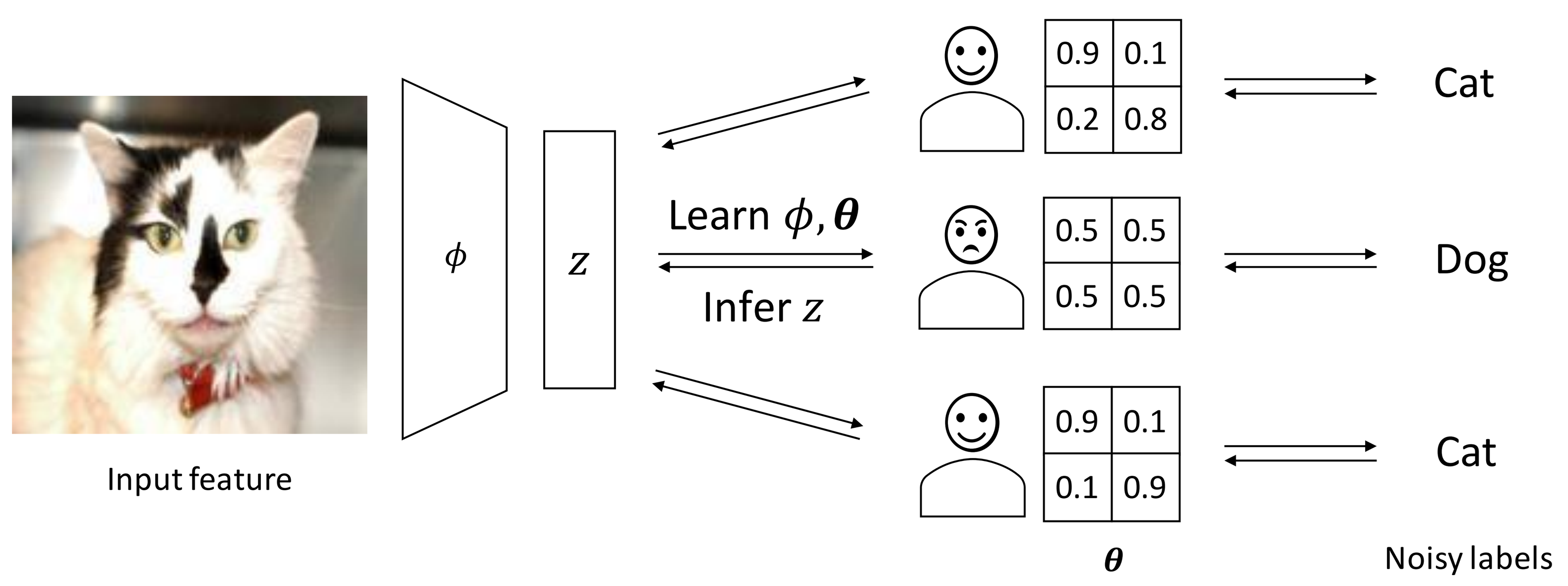
Hoyoung Kim*, Seunghyuk Cho*, Dongwoo Kim, Jungseul Ok

Pohang University of Science and Technology, *equal contribution

1. Introduction

Learning from Crowds with Confusion Matrix

In crowdsourcing systems, we can obtain noisy datasets from workers. We infer true labels $\boldsymbol{z}$ and learn a classifier ϕ by modeling workers with confusion matrix $\boldsymbol{\theta}$.



Crowds from Prior Distribution

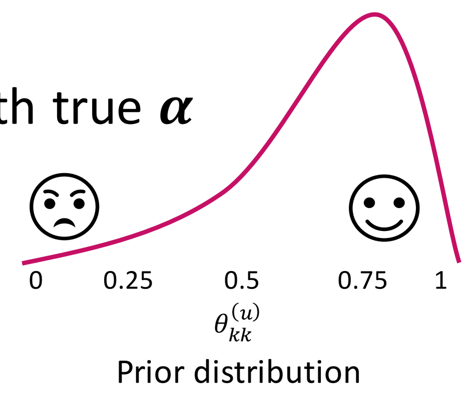
Assuming worker u follows prior distribution parameterized by $\boldsymbol{\alpha}$, confusion matrix $\boldsymbol{\theta}^{(u)} \in [0,1]^{K\times K}$ denoting average ability for K classes is drawn from worker prior.

Importance of worker prior $\boldsymbol{\alpha}$

Inference performance depends on $\boldsymbol{\alpha}$ + Optimal inference with true $\boldsymbol{\alpha}$

Unstable learning of confusion matrix (CM)

Initialization sensitivity + Sparsity in crowds' labels



→ We i) integrate the existing methods using CM through mean-field and ii) propose a robust framework through belief-propagation.

2. Related Work

Inference and Learning with Worker Prior

- Q. Liu et al., Variational Inference for Crowdsourcing. NeurIPS, 12.
- O. Jungseul et al., Optimality of Belief Propagation for Crowdsourced Classification. ICML, 16.
- R. Tanno et al., Learning from Noisy Labels by Regularized Estimation of Annotator Confusion. CVPR, 19.
- L. Shao-Yuan et al., Crowdsourcing Aggregation with Deep Bayesian Learning. Sci China Inf Sci , 21.

Belief Propagation with Deep Learning

- Z. Zhang et al. Factor Graph Neural Networks. NeurIPS, 20.
- J. Kuck et al., Belief Propagation Neural Networks. NeurIPS, 20.
- Satorras and Welling, Neural Enhanced Belief Propagation on Factor Graphs. AISTATS, 21.

3. Alternating Inference and Learning

While not converged **do**

Clip the output of a classifier to regulate overfitting in learning

Infer true labels $\boldsymbol{z}$ using clipped output and a variational distribution

Learn a classifier ϕ with maximizing the ELBO

Obtain $\boldsymbol{z}$ and ϕ

Alternating inference and learning with different variational approaches is proposed.

Deep Mean-Field: unified view of previous methods

Variational distribution $q(\boldsymbol{z}, \boldsymbol{\theta}) = q(\boldsymbol{z})q(\boldsymbol{\theta})$ follows mean-field approximation. Equipping with Dirichlet prior Dir($\boldsymbol{\alpha}$), existing methods are special cases of deepMF.

Methods	Worker prior $\boldsymbol{\alpha}$	$q(\boldsymbol{\theta})$	Hyperparameters
CL (Rodrigues and Pereira, 18)	$\boldsymbol{\alpha} = 1$	Dirac delta	Many
Trace (Tanno et al., 19)	$\alpha_{kk} < 1, \alpha_{kk'} = 1\ (k' \neq k)$	Dirac delta	Many
BayesDGC (Li et al., 21)	Not predetermined	Dirichlet	Many
deepMF	Not predetermined	Dirichlet	Few
deepBP	Not predetermined	No	Few

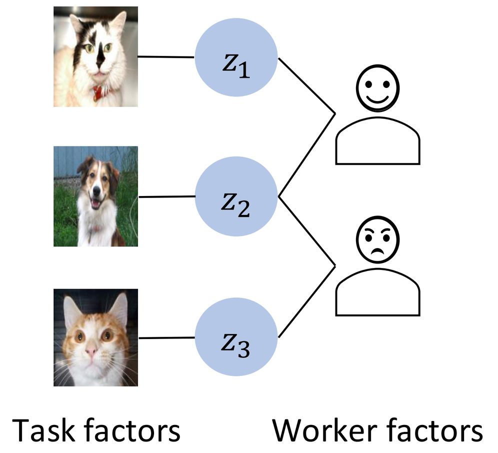
Deep Belief-Propagation: robust Bayesian framework

Variational distribution $q(\boldsymbol{z})$ is obtained from belief-propagation.

Operation of belief-propagation

The likelihood is expressed as factor graph which contains tasks' and workers' factors.

After updating the messages, we aggregate belief for $q(\boldsymbol{z})$.



Fast belief-propagation message update

The message update required $O(2^r)$ is lowered to $O(r \cdot S)$, where r and S is the number of tasks per worker and samples in Monte-Carlo.

Robust on canonical scenarios

DeepBP is robust against i) true prior, ii) feature overfitting and iii) extreme workers.

4. Robustness Analysis

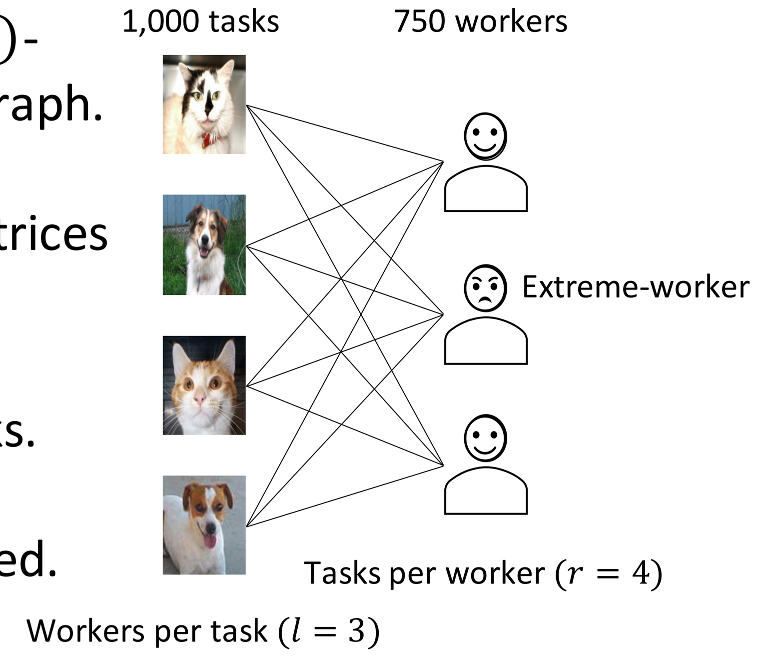
Sparse Crowdsourcing System

Assuming 1,000 tasks and 750 workers, $(l = 3, r = 4)$-regular bipartite graph is generated for assignment graph.

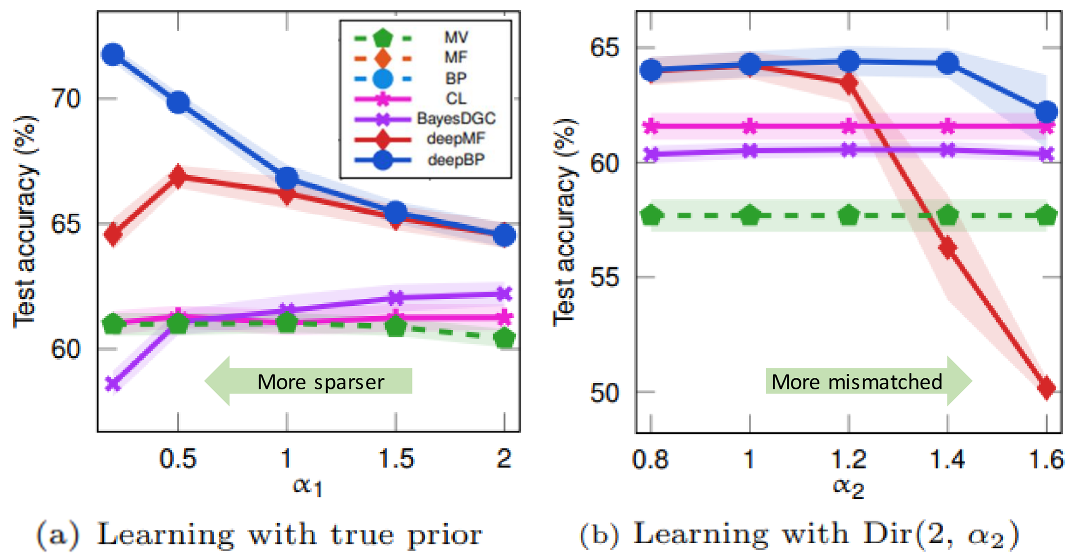
Labels are provided following workers' confusion matrices drawn from true prior.

Extreme-worker labels uniformly across all 1,000 tasks.

Performance of inference $\boldsymbol{z}$ and learning ϕ is measured.



Robustness to Prior and Extreme-worker

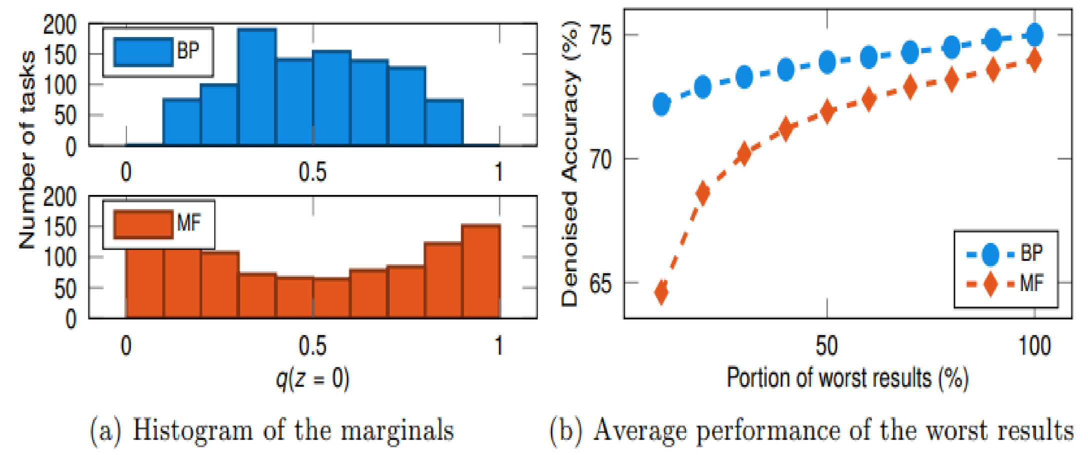


(a) Learning with true prior (b) Learning with Dir(2, α_2)

(a) BP robustly utilize any true prior.

(b) Even with an extreme-worker and mismatched prior, BP is robust.

Overconfidence and Local Minima Issues in Mean-Field



(a) Histogram of the marginals (b) Average performance of the worst results

(a) The marginals of MF are more overconfident than BP's.

(b) MF falls into a local minima more easily than BP.

5. Takeaways

- Previous deep crowdsourcing methods are special cases of deepMF with specific choices of worker prior.
- By the theoretical guarantee on BP for inference, BP-based methods are more robust than MF-based methods in canonical scenarios.