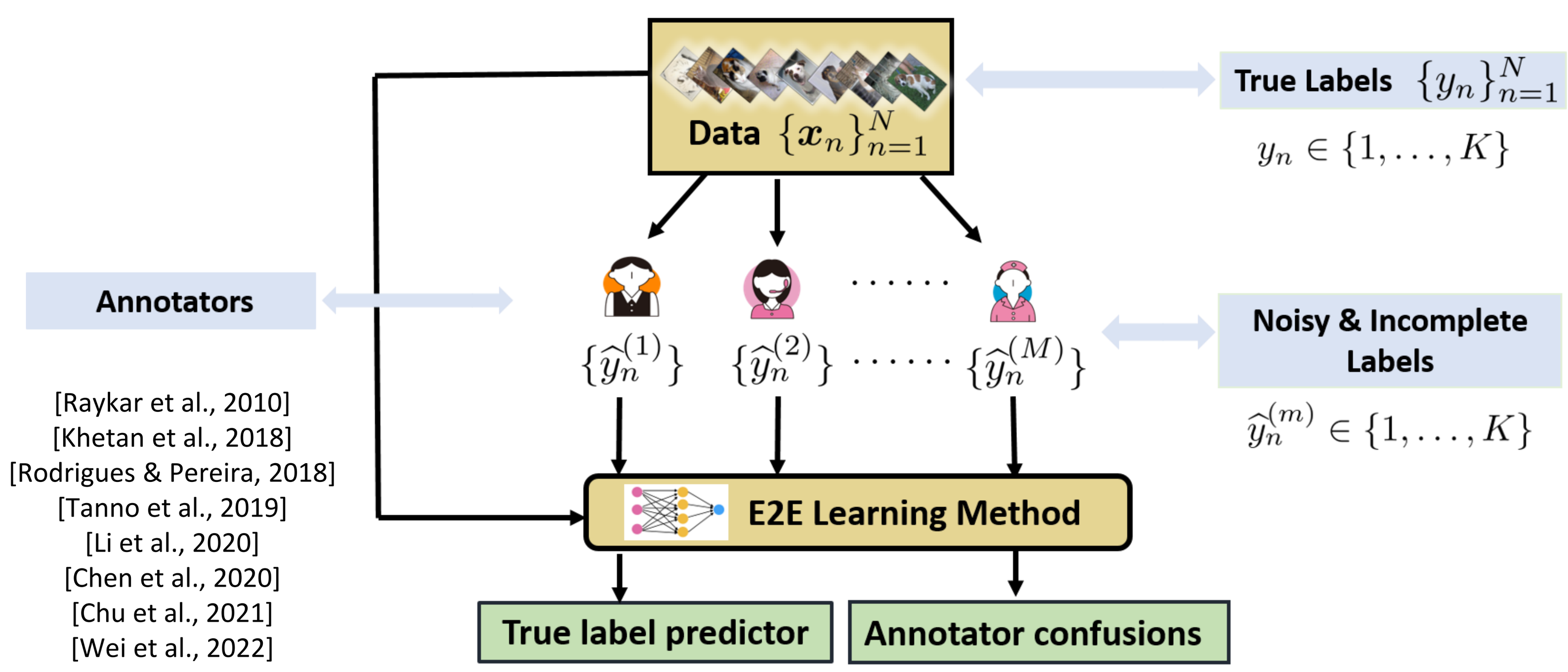
Deep Learning From Crowdsourced Labels: Coupled Cross-Entropy Minimization, Identifiability, and Regularization

Shahana Ibrahim,
Tri Nguyen, Xiao Fu

Oregon State University

ICLR

End-To-End Learning for Crowdsourced Labels



Noisy Label Generation Model

$$\Pr(\widehat{y}_n^{(m)} = k|\boldsymbol{x}_n) = \sum_{k'=1}^K \Pr(\widehat{y}_n^{(m)} = k|y_n = k')\Pr(y_n = k'|\boldsymbol{x}_n)$$

Prob. Of mth annotator's response given x_n: $\boldsymbol{p}_n^{(m)}$

Confusion Matrix: $\boldsymbol{A}_m$

True label predictor: $\boldsymbol{f}(\boldsymbol{x}_n)$

$$\boldsymbol{p}_n^{(m)} = \boldsymbol{A}_m \boldsymbol{f}(\boldsymbol{x}_n), \forall m, n$$

$$\widehat{y}_n^{(m)} \sim \text{categorical}(\boldsymbol{p}_n^{(m)})$$

Goal : learn $\boldsymbol{f}$ and $\boldsymbol{A}_m, \forall m$

Assumption: Annotator confusions are not data dependent

Coupled Cross Entropy Minimization (CCEM)

The most popular E2E learning criterion [Rodrigues & Pereira, 2018]

$$\underset{\boldsymbol{f}, \{\boldsymbol{A}_m\}}{\text{minimize}} \ \frac{1}{|\mathcal{S}|} \sum_{(m,n)\in\mathcal{S}} \text{CE}(\boldsymbol{A}_m \boldsymbol{f}(\boldsymbol{x}_n), \widehat{y}_n^{(m)})$$
$$\text{subject to} \ \ \boldsymbol{f} \in \mathcal{F}, \ \boldsymbol{A}_m \in \mathcal{A}, \ \forall m.$$

[Rodrigues & Pereira, 2018]
[Tanno et al., 2019]
[Li et al., 2020]
[Chen et al., 2020]
[Chu et al., 2021]
[Wei et al., 2022]

Indices of observed labels

Neural network function class

Constrained set $\{\boldsymbol{A} \in \mathbb{R}^{K\times K} | \boldsymbol{A} \geq \boldsymbol{0}, \boldsymbol{1}^\top \boldsymbol{A} = \boldsymbol{1}^\top\}$

Can CCEM learn the true label predictor and the true annotator confusions?

Analysis Result

CCEM correctly learns the true confusions and the true classifier,

under the assumptions

1. ***Anchor point condition*** → For each class, there is a data sample belonging to that class with prob. close to 1
2. ***Class expert condition*** → For each class, there is an expert which can predict that class correctly with prob. close to 1

Often, its hard to hold the two conditions together

Proposed Learning Criteria

✓ ***Anchor point condition***
✗ ***Class expert condition***
→ **If we have more data, but no experts to label**

GeoCrowdNet (F)

$$\underset{\boldsymbol{f}, \{\boldsymbol{A}_m\}}{\text{minimize}} \ \frac{1}{|\mathcal{S}|} \sum_{(m,n)\in\mathcal{S}} \text{CE}(\boldsymbol{A}_m \boldsymbol{f}(\boldsymbol{x}_n), \widehat{y}_n^{(m)}) \boxed{- \lambda \log\det \boldsymbol{F}\boldsymbol{F}^\top}$$
$$\text{subject to} \ \ \boldsymbol{f} \in \mathcal{F}, \ \boldsymbol{A}_m \in \mathcal{A}, \ \forall m.$$

Regularization term

✗ ***Anchor point condition***
✓ ***Class expert condition***
→ **If we have less data, but there are experts to label**

GeoCrowdNet (W)

$$\underset{\boldsymbol{f}, \{\boldsymbol{A}_m\}}{\text{minimize}} \ \frac{1}{|\mathcal{S}|} \sum_{(m,n)\in\mathcal{S}} \text{CE}(\boldsymbol{A}_m \boldsymbol{f}(\boldsymbol{x}_n), \widehat{y}_n^{(m)}) - \boxed{\lambda \log\det \boldsymbol{W}^\top\boldsymbol{W}}$$
$$\text{subject to} \ \ \boldsymbol{f} \in \mathcal{F}, \ \boldsymbol{A}_m \in \mathcal{A}, \ \forall m.$$

Regularization term

$$\underbrace{\begin{bmatrix} \boldsymbol{p}_1^{(1)} & \dots & \boldsymbol{p}_N^{(1)} \\ \vdots & \ddots & \vdots \\ \boldsymbol{p}_1^{(M)} & \dots & \boldsymbol{p}_N^{(M)} \end{bmatrix}}_{\boldsymbol{P}} = \underbrace{\begin{bmatrix} \boldsymbol{A}_1 \\ \vdots \\ \boldsymbol{A}_M \end{bmatrix}}_{\boldsymbol{W}} \underbrace{[\boldsymbol{f}(\boldsymbol{x}_1) \ \ \dots \ \ \boldsymbol{f}(\boldsymbol{x}_N)]}_{\boldsymbol{F}}$$

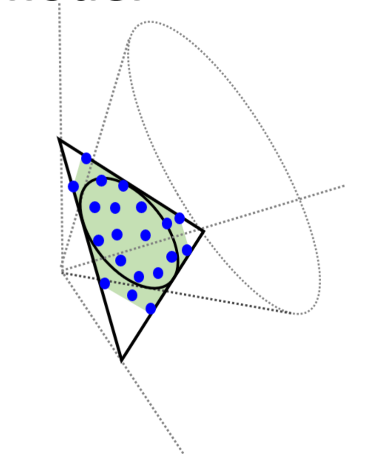
Underlying NMF model

Analysis Result

If $\boldsymbol{F}$ satisfies the SSC (no geometric condition on $\boldsymbol{W}$), the below criterion identifies the true parameters of the model

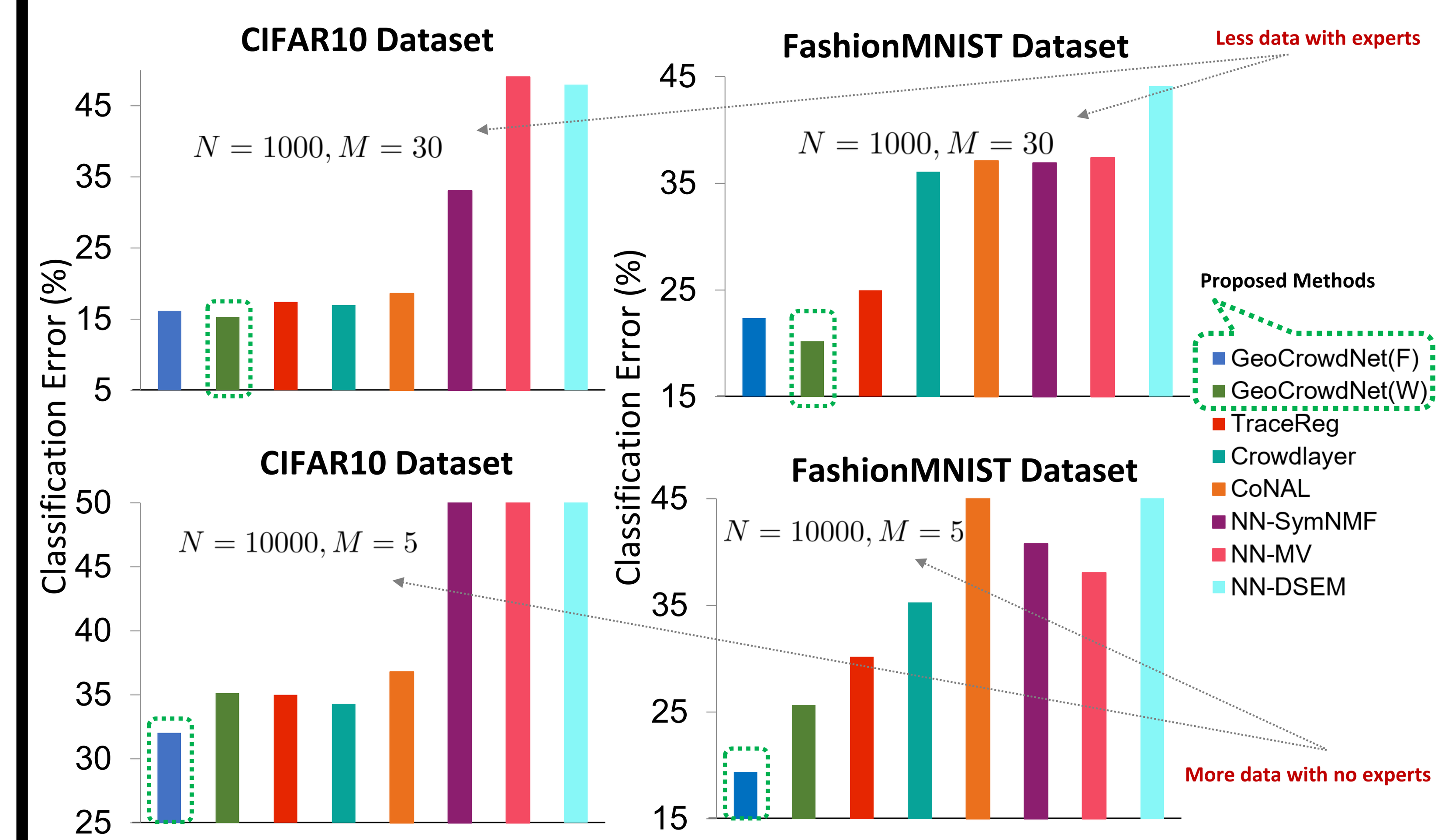
$$\underset{\boldsymbol{W}, \boldsymbol{F}}{\text{maximize}} \ \det(\boldsymbol{F}\boldsymbol{F}^\top)$$
$$\text{s.t.} \ \boldsymbol{P} = \boldsymbol{W}\boldsymbol{F}$$
$$\boldsymbol{F} \geq \boldsymbol{0}, \boldsymbol{1}^\top \boldsymbol{F} = \boldsymbol{1}^\top$$

Volume occupied by the columns of

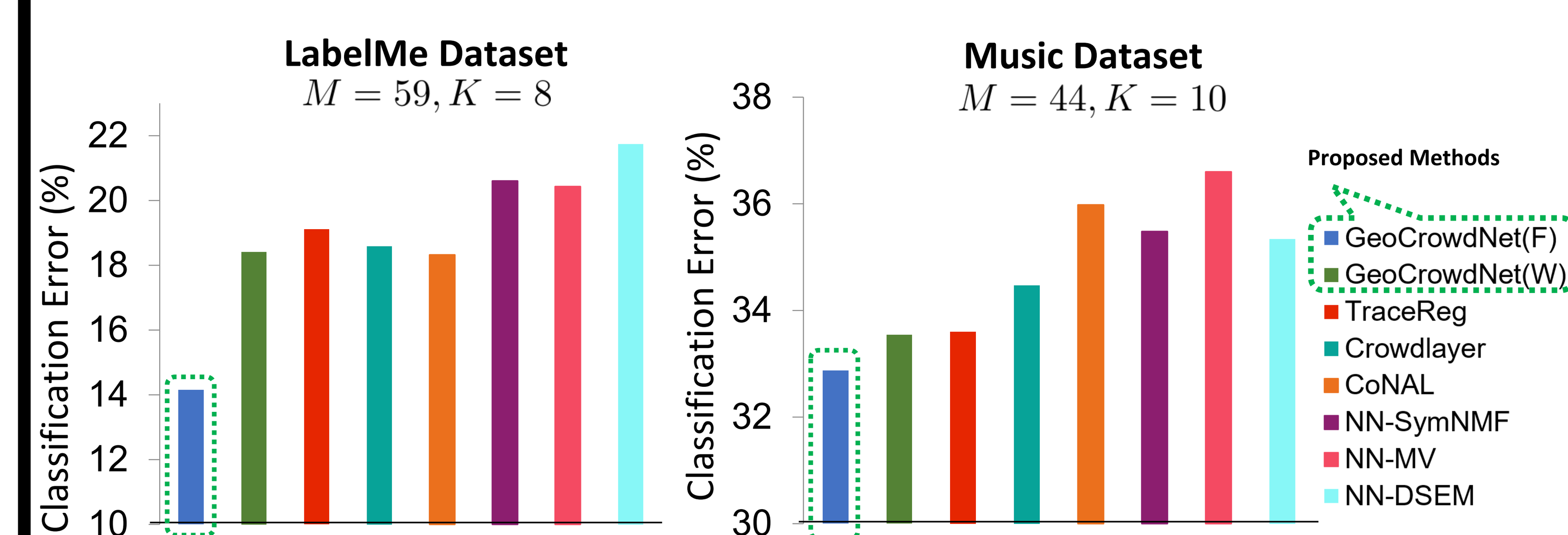


Sufficiently Scattered Condition (SSC)

Empirical Results



Noisy labels from amazon mechanical turk workers



The criterion designed for ``no experts case" shows edge in practice

References

F. Rodrigues and F. Pereira. ``Deep learning from crowds". Proceedings of the AAAI Conference on Artificial Intelligence, 2018.

R. Tanno, A. Saeedi, S. Sankaranarayanan, D. C. Alexander, and N. Silberman. ``Learning from noisy labels by regularized estimation of annotator confusion". IEEE CVPR, 2019.