Hyderabad, India
April 7, 2025

End-to-End Learning from Crowdsourced Labels: A Signal Processing perspective

Shahana Ibrahim[1], Panagiotis A. Traganitis[2], Xiao Fu[3], Georgios B. Giannakis[4]

[1]Dept. of ECE, University of Central Florida
[2]Dept. of ECE, Michigan State University
[3]Dept. of ECE, Oregon State University
[4]Dept. of ECE, University of Minnesota

Acknowledgements: NSF 2128593, 2212318, 2312546, 2007836

Relevant links

Learning From Crowdsourced Noisy Labels: A Signal Processing Perspective

https://arxiv.org/abs/2407.06902

LINK TO SLIDES

https://tinyurl.com/crowdslides

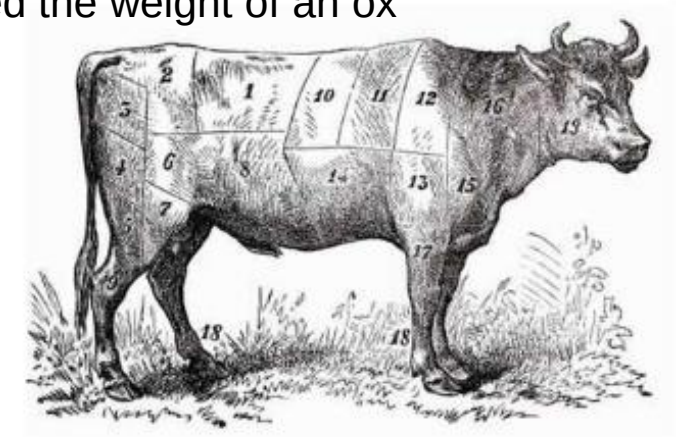
https://github.com/shahana-ibrahim/Learning-from-Crowdsourced-Noisy-Labels

Outline

- Motivation and problem statement
- Part I: Combining crowdsourced labels
- Part II: End-to-end (E2E) learning with crowdsourced labels
- Part III: Other aspects of crowdsourcing
- Conclusions and open issues

The wisdom of crowds

- **The parable of the ox** (Sir Francis Galton, 1906)
  - 787 people guessed the weight of an ox



  - Average crowd guess: **1,197 pounds** - True weight: **1,198 pounds!**

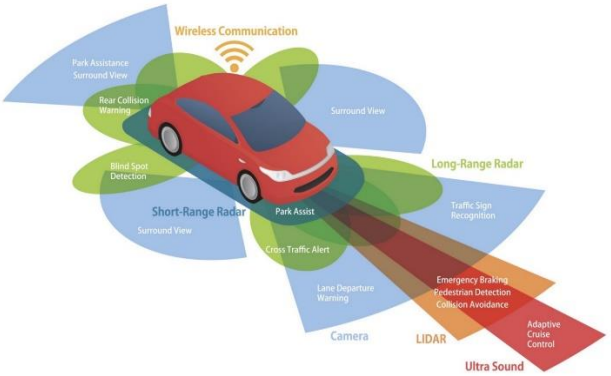
- **Who wants to be a millionaire –** Ask the audience

Q. Can we harness this wisdom in a principled way?

J. Surowiecki, *The Wisdom of Crowds*, Anchor Books, ISBN 978-0-385-72170-7, 2005.

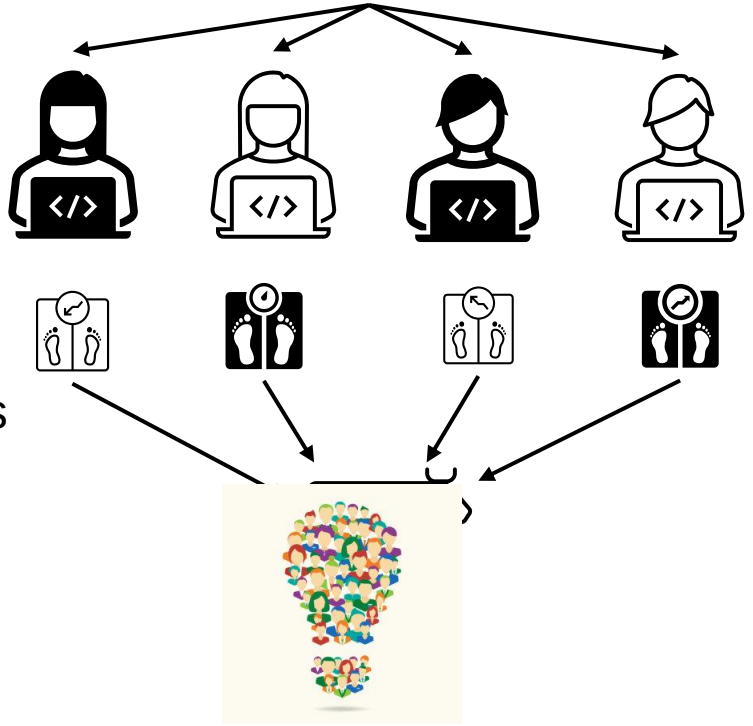
Fusing data versus fusing decisions

- Data fusion is costly [Waltz-Llinas'90, Mitchell'07]



- Distributed (generally noisy) decisions [Tsitsiklis'89]

- **Crowdsourcing vs hierarchical learning**
  - Can benefit from the wisdom of humans
  - Fuses (un) reliable decisions from annotators



- **Redundancy** is key to discover truth and assess annotator reliability!

Crowdsourcing in the learning context

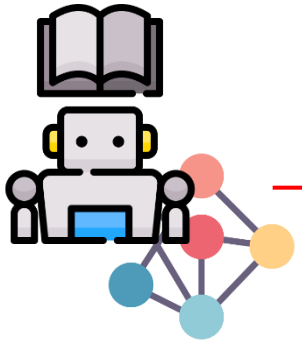
- Machine learning pipeline

Data $\boldsymbol{x}$

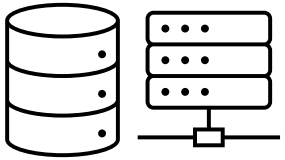
{dog, cat, gecko, … }

Labels y

$y = f(\boldsymbol{x})$



Model training



Machine learning model

No labels, noisy labels or very few labels

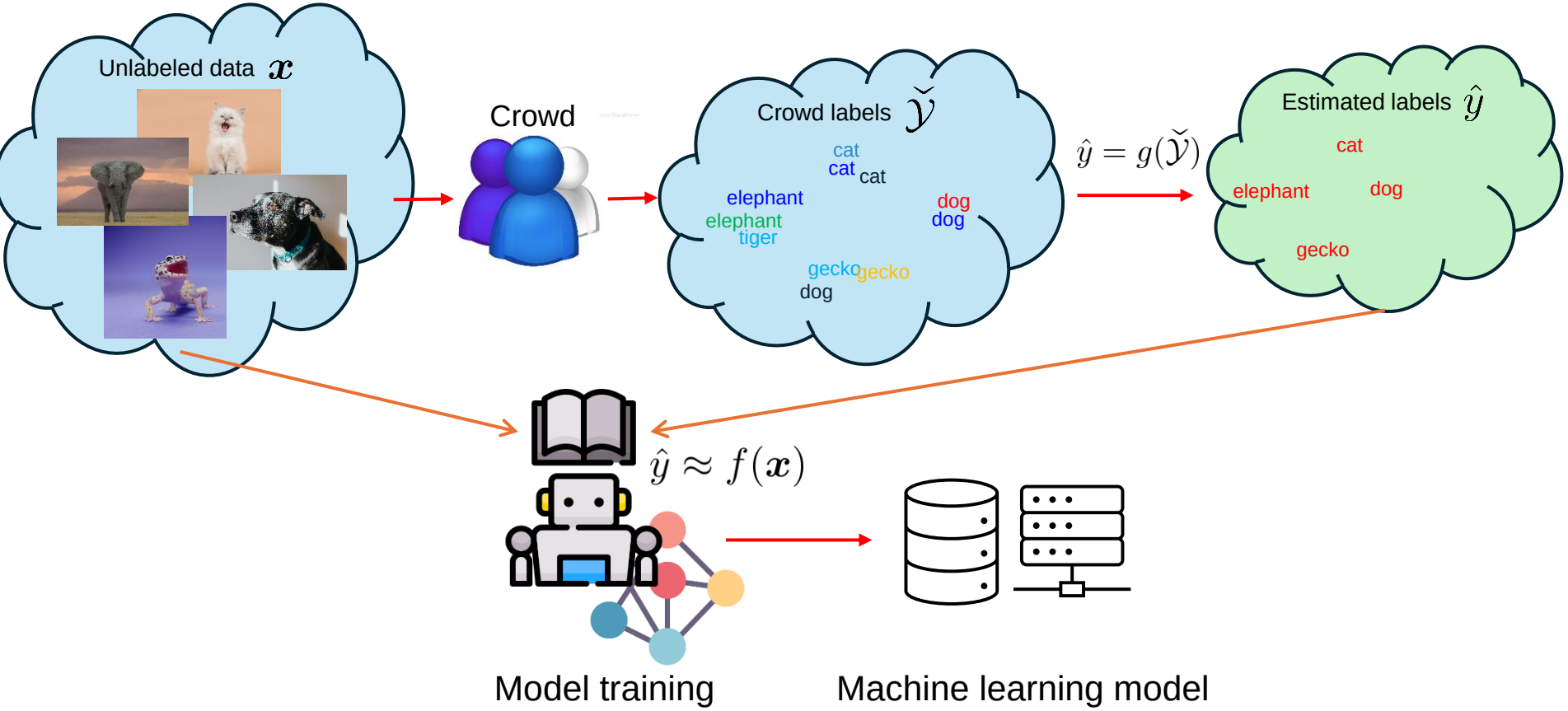
Crowdsourcing encompasses weakly supervised, semi-supervised, self-supervised, transfer learning … and relies on ensembles of (un)reliable learners

- Two major paradigms to **learn from crowds**

Two-stage learning with denoised labels

S1. Given $\breve{\mathcal{Y}} = \{\breve{y}_n^{(m)}\}_{n,m=1}^{N,M}$ learn $g : \hat{y} = g(\breve{\mathcal{Y}})$ — Learns denoiser / fusion rule

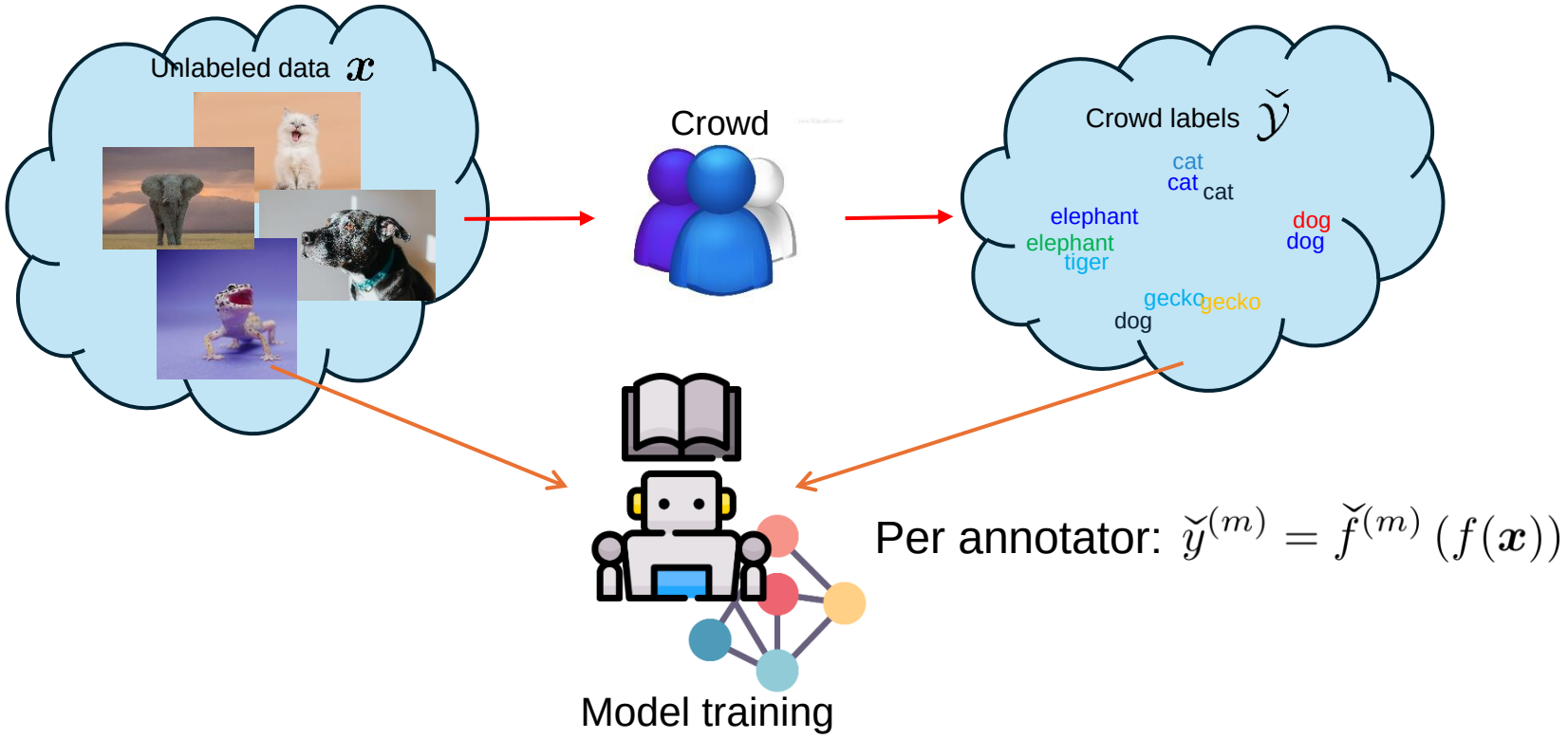
S2. Given $\{\boldsymbol{x}_n, \hat{y}_n\}_{n=1}^{N}$ learn $\hat{f} : \hat{y}_n = \hat{f}(\boldsymbol{x}_n)$ — Learns downstream task



Single stage learning with noisy labels

- End-to-end (E2E) learning: Given $\{\boldsymbol{x}_n, \breve{y}_n^{(m)}\}_{n,m=1}^{N,M}$ learn $\breve{f}^{(m)}$, $\hat{f}$ jointly such that

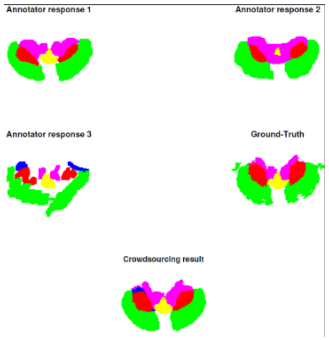
$$\breve{y} = \breve{f}^{(m)}(\hat{f}(\boldsymbol{x}))$$



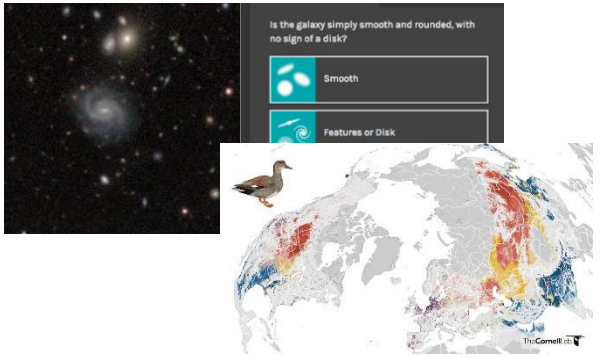
- **Testing phase**: Given $\boldsymbol{x}_{\text{test}}$ infer $\hat{y}_{\text{test}} = \hat{f}(\boldsymbol{x}_{\text{test}}) \approx f(\boldsymbol{x}_{\text{test}})$

Crowdsourcing across disciplines

- Medical image analysis



- Citizen Science



- Crowdsensing in 5G and 6G

- Facilitated via commercial services

amazon mechanical turk

DATA ANNOTATION TECH

Prolific

Crowdsourcing: “Hidden” catalyst behind AI tools

Send us a Tip! | Shop | Subscribe

GIZMODO

The Future Is Here

We may earn a comm

Search

HOME LATEST NEWS GADGETS SCIENCE EARTHER IO9 AI SPACE EN ESPAÑOL VIDEO

ARTIFICIAL INTELLIGENCE

ChatGPT Is Powered by Human Contractors Getting Paid $15 Per Hour

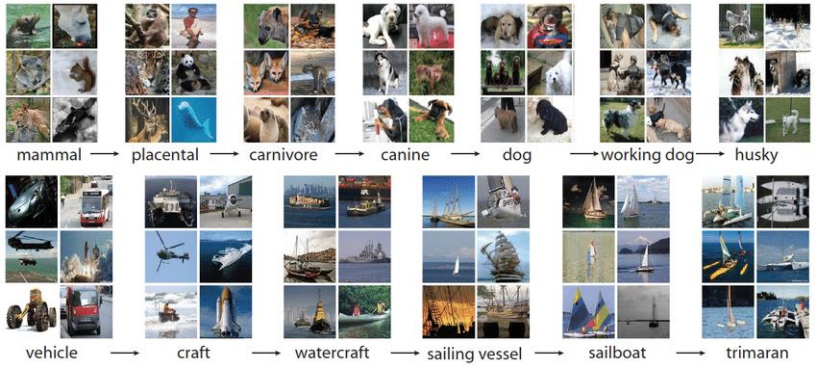
The well known chatbot is automated, but that automation is guided by low-paid human workers labelling data.

- Ukrainians used cell phone app to spot deadly Russian drones and missiles in Crimea

Opportunities and challenges

❑ Opportunities

- From unlabeled to labeled datasets
- No expert supervision needed
- Low-cost and efficient learning
- LLM alignment



❑ Challenges

- Lack of ground-truth labels
- Annotators can be unreliable and even adversarial
- Sparsity of responses per annotator

❑ This tutorial:

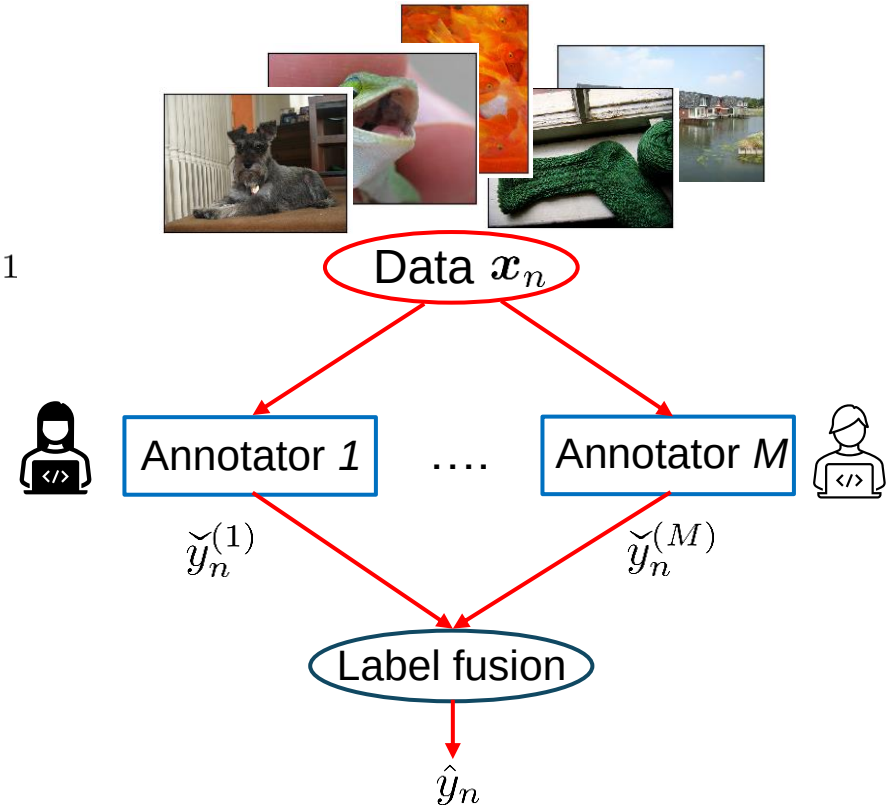
- Crowdsourcing from signal processing, and machine learning points of view
- Statistical learning framework

Crowdsourcing for classification

- N data $\{(\boldsymbol{x}_n, y_n)\}_{n=1}^N$, K classes
- M annotators observe **subsets** of $\{\boldsymbol{x}_n\}_{n=1}^N$
- Annotators provide noisy labels

$$\{\breve{y}_n^{(m)}\}_{n,m=1}^{N,M}$$

Goal: Given $\{\breve{y}_n^{(m)}\}_{n,m=1}^{N,M}$, find $\{\hat{y}_n\}_{n=1}^N$



Q1. Which annotators are reliable? **Q2.** How can we combine (un)reliable answers?

- Inference of labels is sought in a "noisy" unsupervised setting!

Outline

- Motivation and problem statement
- Part I: Combining crowdsourced labels
  - Majority voting and annotator models
  - Estimating model parameters via Expectation-Maximization (EM)
  - Moment matching methods
  - Identifiability
- Part II: End-to-end (E2E) learning with crowdsourced labels
- Part III: Other aspects of crowdsourcing
- Conclusions and open issues

Majority voting or averaging

- Simplest solution to crowdsourcing
  - Per datum *n*, pick most common answer

$$\hat{y}_n = \underset{k \in \{1,\ldots,K\}}{\arg\max} \sum_{m=1}^{M} \mathbb{I}[\check{y}_n^{(m)} = k]$$

Caveat: Majority voting (MV) assumes all annotators are "the same" and "independent"

Q. Can we do better than MV? **A.** Yes, by modeling annotator behavior!

- Simple extension: **weighted** MV

$$\hat{y}_n = \underset{k \in \{1,\ldots,K\}}{\arg\max} \sum_{m=1}^{M} w^{(m)} \mathbb{I}[\check{y}_n^{(m)} = k]$$

Q. How to find weights?

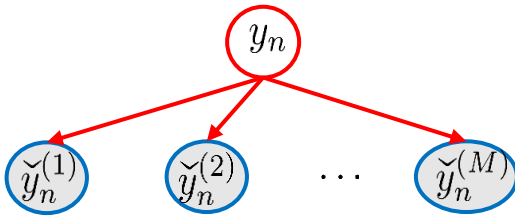
Annotator model with independent responses

- Dawid-Skene (DS) data model $\{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N} \overset{\text{i.i.d.}}{\sim} \mathcal{P}$
- Maximum a posteriori (MAP) fusion of annotator labels $\breve{\mathcal{Y}}$

$$\hat{y}_n = \underset{k\in\{1,\ldots,K\}}{\arg\max}\ \Pr(y_n = k|\breve{\mathcal{Y}}) = \underset{k\in\{1,\ldots,K\}}{\arg\max}\ \Pr\left(\breve{y}_n^{(1)} = k_1, \ldots, \breve{y}_n^{(M)} = k_M | y_n = k\right) \Pr(y_n = k)$$

(as1) Given ground-truth label y_n , annotator responses $\{\breve{y}_n^{(m)}\}_{m=1}^{M}$ are independent

$$\Pr\left(\breve{y}_n^{(1)} = k_1, \ldots, \breve{y}_n^{(M)} = k_M | y_n = k\right) = \prod_{m=1}^{M} \Pr\left(\breve{y}_n^{(m)} = k_m | y_n = k\right)$$



$$\hat{y}_n = \underset{k\in\{1,\ldots,K\}}{\arg\max}\ \log\Pr(y_n = k) + \sum_{m=1}^{M}\sum_{k'=1}^{K} \log\Pr\left(\breve{y}_n^{(m)} = k' | y_n = k\right) \mathbb{I}[\breve{y}_n^{(m)} = k']$$

- Performance

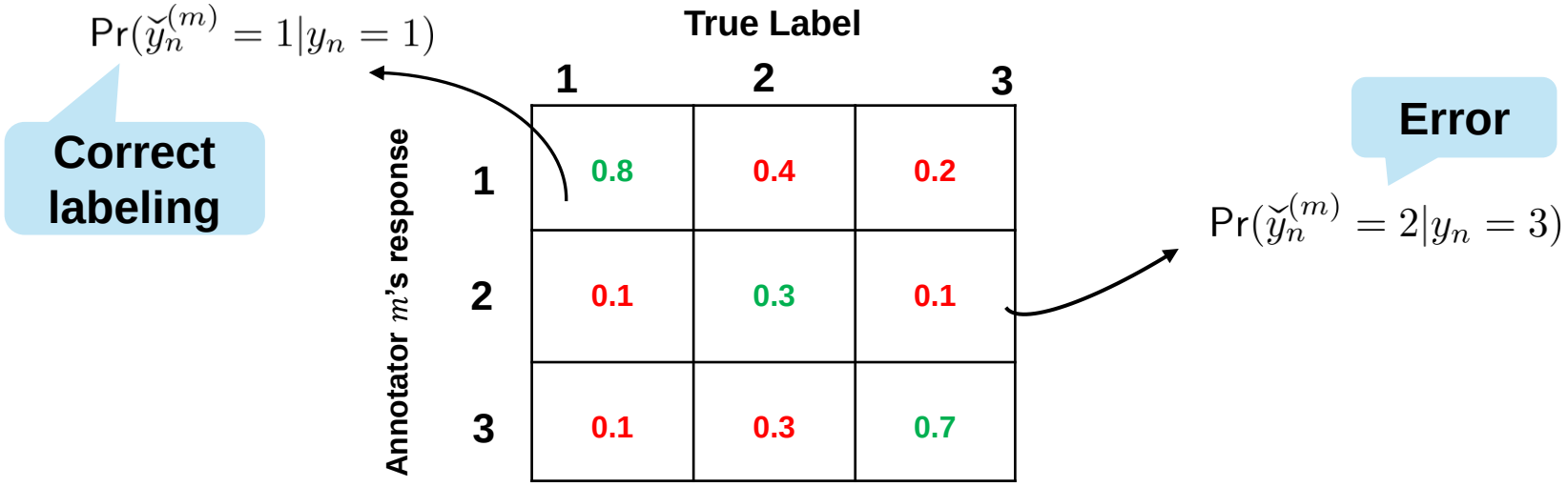
Theorem. Under **(as1)** there are constants α,β > 0 so that the error probability of the MAP classifier satisfies

$$\mathcal{P}_e \leq \alpha e^{-M\beta}$$

A. P. Dawid and A. M. Skene, “Maximum likelihood estimation of observer error-rates using the EM algorithm,” *Applied Stats.*,1979.
P. A. Traganitis, A. Pagès-Zamora, and G. B. Giannakis, "Blind Multiclass Ensemble Classification," *IEEE Transactions on Signal Processing*, vol. 66, no. 18, pp. 4737-4752, September 2018.

Parametric model

- MAP classifier needs $\boldsymbol{d}(k) := \Pr(y_n = k)$ and $\boldsymbol{A}_m(k', k) := \Pr\left(\breve{y}_n^{(m)} = k' | y_n = k\right)$
  - Confusion matrices $\boldsymbol{A}_m \in \mathbb{R}^{K \times K}, \quad m = 1, \ldots, M \qquad \boldsymbol{A}_m := [\boldsymbol{a}_{m,1}, \ldots, \boldsymbol{a}_{m,K}]$



  - Ideal annotator $\boldsymbol{A}_m = \boldsymbol{I}$
- Model parameters $\{\boldsymbol{A}_m\}_{m=1}^M, \boldsymbol{d}$ must be estimated from annotator responses $\breve{\mathcal{Y}}$
- $\boldsymbol{d}$ lies on a *K-1* simplex; and likewise $\{\boldsymbol{a}_{m,k}\}_{k=1,m=1}^{K,M}$ lie on *K-1* simplexes
  - Number of parameters: $\boldsymbol{d} : K - 1$ and $\boldsymbol{A}_m : K(K-1)$ Total: $\mathcal{O}(MK^2)$

Alternative models

- For *K=2, we* need to estimate 2 parameters per annotator *m*
  - Sensitivity $\boldsymbol{A}_m(1,1) = \Pr(\breve{y}_n^{(m)} = 1 | y_n = 1)$
  - Specificity $\boldsymbol{A}_m(2,2) = \Pr(\breve{y}_n^{(m)} = 2 | y_n = 2)$

- Parsimonious models of $\{\boldsymbol{A}_m\}_{m=1}^M$
  - One parameter (one-coin) or homogeneous D-S (one unknown per annotator *m*)

$$\boldsymbol{A}_m = \left(w^{(m)} - \frac{1 - w^{(m)}}{K-1} \right) \mathbf{I} + \frac{1 - w^{(m)}}{K-1} \mathbf{1}\mathbf{1}^\top, \quad 0 \leq w^{(m)} \leq 1$$

  - Confusion vector $\boldsymbol{w}^{(m)} := [w_1^{(m)} \dots w_K^{(m)}]^\top \quad w_k^{(m)} := \Pr(\breve{y}_n^{(m)} = k | y_n = k)$

$$\boldsymbol{A}_m(k,k') = \begin{cases} w_k^{(m)} & k = k' \\ \frac{1 - w_k^{(m)}}{K-1} & k \neq k' \end{cases}$$

  - Significant reduction in the number of parameters $\mathcal{O}(M)$ and $\mathcal{O}(MK)$, respectively
  - More parsimonious however, implies less expressive modeling

- Item difficulty can be incorporated [Whitehill et al'09]

A.T. Nguyen, B. C. Wallace, J. J. Li, A. Nenkova, and M. Lease, "Aggregating and predicting sequence labels from crowd annotations," *Proc. of Annual Meeting Association for Computational Linguistics*, 2017.

Outline

- Motivation and problem statement
- Part I: Combining crowdsourced labels
  - Majority voting and annotator models
  - Estimating model parameters via Expectation-Maximization (EM)
  - Moment matching methods
  - Identifiability
- Part II: End-to-end (E2E) learning with crowdsourced labels
- Part III: Other aspects of crowdsourcing
- Conclusions and open issues

Estimating model parameters via EM

- Popular for maximum likelihood estimation of parametric mixture models [Dempster-Laird-Rubin'77]
  - Observations $\breve{\mathcal{Y}}$
  - N latent true labels $\boldsymbol{y}$
  - Wanted parameters $\boldsymbol{\psi} := [\{\boldsymbol{A}_m\}_{m=1}^M, \boldsymbol{d}]$

- Wish to maximize $L(\boldsymbol{\psi}) = \log \Pr(\breve{\mathcal{Y}}; \boldsymbol{\psi}) = \log \left(\sum_{\boldsymbol{y}} \Pr(\boldsymbol{y}, \breve{\mathcal{Y}}; \boldsymbol{\psi}) \right)$ Typically intractable
  - Use a tractable surrogate pdf q to maximize an expected lower bound with a two-step iteration

$$\mathbf{E}_{q(\boldsymbol{y})} \left[\log \Pr \left(\boldsymbol{y}, \breve{\mathcal{Y}}; \boldsymbol{\psi} \right) \right] = \log \Pr(\breve{\mathcal{Y}}; \boldsymbol{\psi}) - D_{KL} \left(q(\boldsymbol{y}) || \Pr(\boldsymbol{y} | \breve{\mathcal{Y}}; \boldsymbol{\psi}) \right)$$

S1. Expectation (E-)step $Q(\boldsymbol{\psi}; \boldsymbol{\psi}^i) = \mathbb{E}_{\boldsymbol{y}|\breve{\mathcal{Y}};\boldsymbol{\psi}^i} [\log \Pr(\boldsymbol{y}, \breve{\mathcal{Y}}; \boldsymbol{\psi})]$

Latent variables updated given observed ones and parameters

S2. Maximization (M-)step $\boldsymbol{\psi}^{i+1} = \arg\max_{\boldsymbol{\psi}} Q(\boldsymbol{\psi}; \boldsymbol{\psi}^i)$

Parameters updated given estimated latent variables

- Nondecreasing sequence of $L(\boldsymbol{\psi})$'s converges at least to a stationary point – Initialization is critical

A.P. Dempster, N.M. Laird and D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, *Series B*, vol. 39, pp. 1–38, 1977.

EM for crowdsourcing

S0. Initialize $\psi^0 := [\boldsymbol{d}^0, \boldsymbol{A}_1^0, \ldots, \boldsymbol{A}_M^0]$

S1. E-step

$$Q(\boldsymbol{\psi}; \boldsymbol{\psi}^i) = \mathbb{E}_{\mathcal{Y} \sim \mathsf{Pr}(\mathcal{Y}|\breve{\mathcal{Y}}; \boldsymbol{\psi}^t)}[\log \mathsf{Pr}(\breve{\mathcal{Y}}, \mathcal{Y}; \boldsymbol{\psi})] = \sum_{n=1}^{N} \sum_{k=1}^{K} q(y_n = k; \boldsymbol{\psi}^i) \log \mathsf{Pr}(y_n = k, \breve{y}_n^{(1)}, \ldots, \breve{y}_n^{(M)}; \boldsymbol{\psi})$$

$$q(y_n = k; \boldsymbol{\psi}^i) := \mathsf{Pr}(y_n = k | \breve{y}_n^{(1)}, \ldots, \breve{y}_n^{(M)}; \boldsymbol{\psi}^t) = \frac{\exp\left(\log \boldsymbol{d}^i(k) + \sum_{m=1}^{M} \sum_{k'=1}^{K} \log \boldsymbol{A}_m^i(k', k) \mathbb{I}[\breve{y}_n^m = k']\right)}{Z}$$

Bayes rule

S2. M-step

$$\boldsymbol{\psi}^{i+1} = \arg\max_{\boldsymbol{\psi}} Q(\boldsymbol{\psi}; \boldsymbol{\psi}^i) \Rightarrow$$

$$\boldsymbol{A}_m^{i+1}(k', k) = \frac{\sum_{n=1}^{N} q(y_n = k; \boldsymbol{\psi}^i) \mathbb{I}[\breve{y}_n^m = k']}{\sum_{k''=1}^{K} \sum_{n=1}^{N} q(y_n = k; \boldsymbol{\psi}^i) \mathbb{I}[\breve{y}_n^m = k'']} \quad \forall m, k', k$$

$$\boldsymbol{d}^{i+1}(k) = \frac{\sum_{n=1}^{N} q(y_n = k; \boldsymbol{\psi}^i)}{\sum_{k'=1}^{K} \sum_{n=1}^{N} q(y_n = k'; \boldsymbol{\psi}^i)} \quad \forall k$$

Theorem. If *M* and *N* sufficiently large, and EM is **properly initialized** then correct labels y_n are selected w.h.p.

A. P. Dawid and A. M. Skene, "Maximum likelihood estimation of observer error-rates using the EM algorithm," *Applied Statistics*, pp. 20–28, 1979.
Y. Zhang, X. Chen, D. Zhou, and M. I. Jordan, "Spectral methods meet EM: A provably optimal algorithm for crowdsourcing," *Advances in Neural Information Processing Systems*, pp. 1260–68, 2014.

How about prior information?

Bayesian approaches view ***d, A*** as random, and rely on Gibbs sampling, variational inference (VI)

- Iterative solvers guarantee local optimality – global only with reliable initialization
- Both EM and Bayesian approaches can account for a few ground-truth labels, if available

Majority voting (MV)
simple, but reliable?

$$\hat{y}_n = \underset{k \in \{1,\ldots,K\}}{\arg\max} \sum_{m=1}^{M} \mathbb{I}(\breve{y}_n^{(m)} = k)$$

Initializing reliability parameters, but performance?

$$\boldsymbol{d}^0(k) = \frac{\sum_{m=1}^{M} \sum_{n=1}^{N} \mathbb{I}(\breve{y}_n^{(m)} = k)}{N_1 + \cdots + N_M}, \quad k = 1, \ldots, K$$

$$\boldsymbol{A}_m^0(k,' k) = \frac{\sum_{n=1}^{N} \mathbb{I}(\breve{y}_n^{(m)} = k', \hat{y}_n = k)}{N_m}, \quad \forall m, k', k$$

Resemble histogram estimates

- MV offers simple initialization, but how about nonuniformly unreliable annotators?

H.-C. Kim and Z. Ghahramani, "Bayesian classifier combination," *Proc. Intl. Conf. Artif. Intell. Statist.*, pp. 619-627, 2012.

Outline

- Motivation and problem statement
- Part I: Combining crowdsourced labels
  - Majority voting and annotator models
  - Estimating model parameters via Expectation-Maximization (EM)
  - Moment matching methods
  - Identifiability
- Part II: End-to-end (E2E) learning with crowdsourced labels
- Part III: Other aspects of crowdsourcing
- Conclusions and open issues

Estimating parameters from annotator statistics

Conditionally independent annotators

- Pairwise statistics of annotator responses

$$\boldsymbol{R}_{\boldsymbol{m},\boldsymbol{m'}}(k_m, k_{m'}) := \Pr(\breve{y}_n^{(m)} = k_m, \breve{y}_n^{(m')} = k_{m'})$$
$$= \sum_{k=1}^{K} \Pr(y_n = k) \Pr(\breve{y}_n^{(m)} = k_m | y_n = k) \Pr(\breve{y}_n^{(m')} = k_{m'} | y_n = k)$$

$\Pr(\breve{y}_n^{(m)} = k_m | y_n = k) = \boldsymbol{A}_m(k_m, k)$

$\Pr(\breve{y}_n^{(m')} = k_{m'} | y_n = k) = \boldsymbol{A}_{m'}(k_{m'}, k)$

$$\boldsymbol{R}_{m,m'} = \boldsymbol{A}_m \boldsymbol{D} \boldsymbol{A}_{m'}^{\top}, \boldsymbol{D} = \mathrm{diag}(\boldsymbol{d})$$

- Triplet statistics of annotator responses

$$\underline{\boldsymbol{T}}_{m,m',m''}(k_m, k_{m'}, k_{m''}) := \Pr(\breve{y}_n^{(m)} = k_m, \breve{y}_n^{(m')} = k_{m'}, \breve{y}_n^{(m'')} = k_{m''})$$
$$= \sum_{k=1}^{K} \Pr(y_n = k) \Pr(\breve{y}_n^{(m)} = k_m | y_n = k) \Pr(\breve{y}_n^{(m')} = k_{m'} | y_n = k) \Pr(\breve{y}_n^{(m'')} = k_{m''} | y_n = k)$$

$$\underline{\boldsymbol{T}}_{m,m'm''} = \sum_{k=1}^{K} \boldsymbol{d}(k) \boldsymbol{A}_m(:,k) \circ \boldsymbol{A}_{m'}(:,k) \circ \boldsymbol{A}_{m''}(:,k) = [[\boldsymbol{A}_m \boldsymbol{D}, \boldsymbol{A}_{m'}, \boldsymbol{A}_{m''}]]$$

PARAFAC/CPD tensor

$\{\boldsymbol{A}_m\}_{m=1}^{M}, \boldsymbol{d}$ recoverable from statistics of annotator responses via **moment matching**

P. A. Traganitis, A. Pagès-Zamora, and G. B. Giannakis, "Blind multiclass ensemble classification," IEEE Trans. Signal Process., vol. 66, no. 18, pp. 4737–4752, 2018

Estimating annotator statistics

- Estimates can be readily computed from $\breve{y}$
  - *M* first-order moments

$$\hat{\boldsymbol{\mu}}_m(k) = \frac{1}{N_m} \sum_{n=1}^{N} \mathbb{I}[\breve{y}_n^{(m)} = k]$$

N_m: no. of responses from annotator *m*

  - $\binom{M}{2}$ second-order moments

$$\hat{\boldsymbol{R}}_{m,m'}(k_m, k_{m'}) = \frac{1}{N_{m,m'}} \sum_{n=1}^{N} \mathbb{I}[\breve{y}_n^{(m)} = k_m, \breve{y}_n^{(m')} = k_{m'}]$$

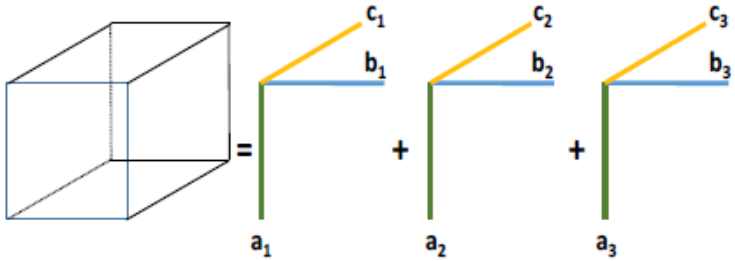
  - $\binom{M}{3}$ third-order moments

$$\hat{\underline{\boldsymbol{T}}}_{m,m',m''}(k_m, k_{m'}, k_{m''}) = \frac{1}{N_{m,m',m''}} \sum_{n=1}^{N} \mathbb{I}[\breve{y}_n^{(m)} = k_m, \breve{y}_n^{(m')} = k_{m'}, \breve{y}_n^{(m'')} = k_{m''}]$$

  - Sparse annotator responses
    - Some moments not available
    - Heteroskedastic noise

Canonical Polyadic Decomposition in one slide

- Data tensor $\underline{\boldsymbol{T}} : I \times J \times N$
- PARAFAC/CPD Tensor model: $\underline{\boldsymbol{T}} = \sum_{f=1}^{F} \boldsymbol{a}_f \circ \boldsymbol{b}_f \circ \boldsymbol{c}_f$



$\mathbf{A} := [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_F] : I \times F \quad \mathbf{B} := [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_F] : J \times F \quad \mathbf{C} := [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_F] : N \times F$

- $\mathbf{A}, \mathbf{B}, \mathbf{C}$ essentially unique
  - Recoverable up to scaling and common permutation ambiguities

$$\hat{\mathbf{A}} = \mathbf{A}\boldsymbol{\Pi}\boldsymbol{\Lambda}_1, \hat{\mathbf{B}} = \mathbf{B}\boldsymbol{\Pi}\boldsymbol{\Lambda}_2, \hat{\mathbf{C}} = \mathbf{C}\boldsymbol{\Pi}\boldsymbol{\Lambda}_3, \quad \boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_2\boldsymbol{\Lambda}_3 = \mathbf{I}$$

$\boldsymbol{\Pi}$: permutation matrix

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \|\underline{\boldsymbol{T}} - \sum_{f=1}^{F} \boldsymbol{a}_f \circ \boldsymbol{b}_f \circ \boldsymbol{c}_f\|_F^2$$

NP-hard!

- Solved via alternating optimization / Gradient descent (GD/SGD) etc.
- Off-the-shelf solvers: *N-way toolbox, Tensorlab, AO-ADMM, SPLATT*

N. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," IEEE Trans. Signal Process., vol. 65, no. 13, pp. 3551–3582, Jul. 2017.

Moment matching with third-order statistics

Available triplet moments

Coupled CPD

$$\text{find } \{\boldsymbol{A}_m\}_{m=1}^{M}, \boldsymbol{d}$$

$$\text{s.t. } \hat{\underline{\boldsymbol{T}}}_{m,m',m''} = [[\boldsymbol{A}_m \boldsymbol{D}, \boldsymbol{A}_m, \boldsymbol{A}_{m''}]], \ (m,m',m'') \in \boldsymbol{\Omega},$$

$$\boldsymbol{A}_m \geq \boldsymbol{0}, \boldsymbol{1}^\top \boldsymbol{A}_m = \boldsymbol{1}^\top, \ \boldsymbol{1}^\top \boldsymbol{d} = 1, \boldsymbol{d} \geq \boldsymbol{0}.$$

Sample triplet moments

- Relaxation

$$\min \sum_{(m,m',m'')\in\Omega} \|\hat{\underline{\boldsymbol{T}}}_{m,m',m''} - [[\boldsymbol{A}_m \boldsymbol{D}, \boldsymbol{A}_m, \boldsymbol{A}_{m''}]]\|_F^2$$

$$\text{s.t. } \boldsymbol{A}_m \geq \boldsymbol{0}, \boldsymbol{1}^\top \boldsymbol{A}_m = \boldsymbol{1}^\top, \ \boldsymbol{1}^\top \boldsymbol{d} = 1, \boldsymbol{d} \geq \boldsymbol{0}.$$

- First- and second-order moments can be included

Theorem. If solution $\mathcal{S}^*$ relies on ensemble moments, and solution $\mathcal{S}^N$ on sample moments, then as $N \to \infty$

$$\mathcal{D}(\mathcal{S}^*, \mathcal{S}^N) \to 0 \quad \text{almost surely.}$$

➢ Reduced complexity via orthogonal tensor decompositions [Zhang et al'16]

P. A. Traganitis, A. Pagès-Zamora, and G. B. Giannakis, "Blind multiclass ensemble classification," IEEE Trans. Signal Process., vol. 66, no. 18, pp. 4737–4752, 2018

Confusion matrix identifiability

Q. Can we identify $\{\boldsymbol{A}_m\}_{m=1}^M, \boldsymbol{d}$ from $\underline{\boldsymbol{T}}_{m,m',m''}, \quad (m,m',m'') \in \boldsymbol{\Omega}$?

- Coupled tensor factorization inherits CPD identifiability guarantees
  - Recovery of $\{\boldsymbol{A}_m\}_{m=1}^M, \boldsymbol{d}$ up to common permutation ambiguity
  - ≥3 full rank $\boldsymbol{A}_m$'s required
  - Annotators have to **co-label** "sufficient" samples
- Additional assumption required to resolve permutation ambiguity

(as) Most annotators are "better than random"

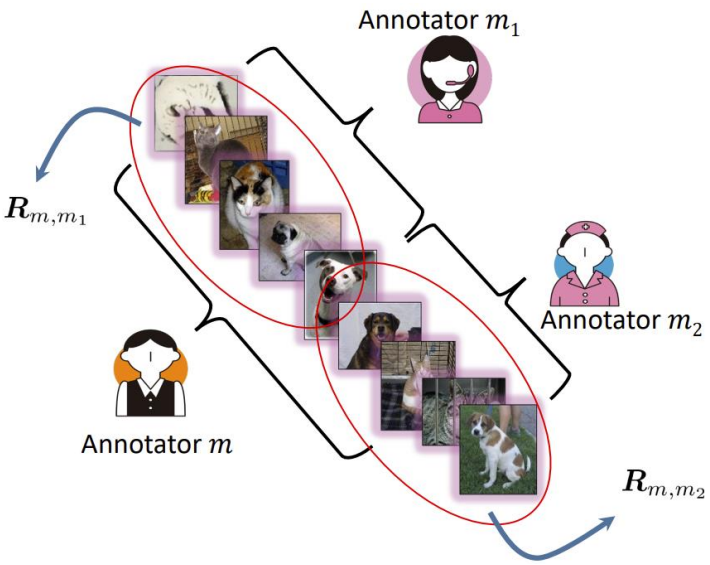
Diagonally dominant $\boldsymbol{A}_m$

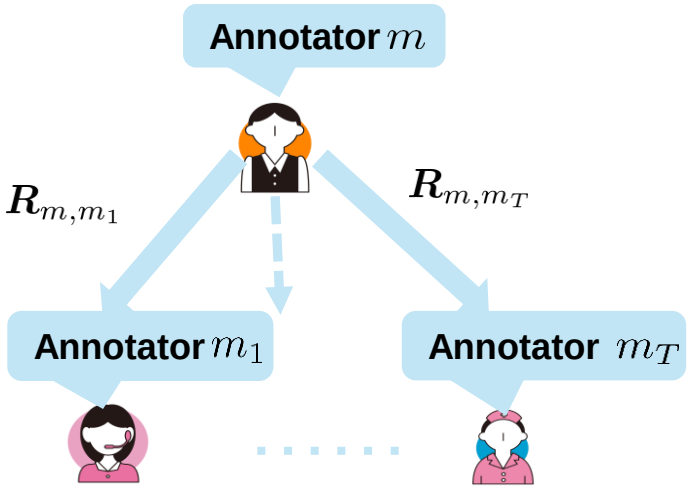
- **Caveats**
  - Complexity increases with *M*
  - Accurate estimation of third order moments requires a lot of data

P. A. Traganitis, A. Pagès-Zamora, and G. B. Giannakis, "Blind multiclass ensemble classification," IEEE Trans. Signal Process., vol. 66, no. 18, pp. 4737–4752, 2018

Learning from Pairwise Statistics

Assume that annotator m co-labels with annotators $m_1, \ldots, m_T$





$$\begin{aligned} \boldsymbol{Z}_m &\triangleq \left[\boldsymbol{R}_{m,m_1}, \boldsymbol{R}_{m,m_2}, \ldots, \boldsymbol{R}_{m,m_T}\right] \\ &= \left[\boldsymbol{A}_m \boldsymbol{D} \boldsymbol{A}_{m_1}^\top, \ldots, \boldsymbol{A}_m \boldsymbol{D} \boldsymbol{A}_{m_T}^\top\right] \\ &= \underbrace{\boldsymbol{A}_m \boldsymbol{D}}_{\boldsymbol{W}_m} \underbrace{\left[\boldsymbol{A}_{m_1}^\top, \ldots, \boldsymbol{A}_{m_T}^\top\right]}_{\boldsymbol{H}_m^\top} \end{aligned}$$

Stack all available pairwise statistics

$$\boldsymbol{Z}_m = \boldsymbol{W}_m \boldsymbol{H}_m^\top \ , \ \forall m$$

$\boldsymbol{Z}_m$ observed, but $\boldsymbol{W}_m$ and $\boldsymbol{H}_m$ are unknown

S. Ibrahim, X. Fu, N. Kargas, and K. Huang, "Crowdsourcing via pairwise co-occurrences: Identifiability and algorithms," NeurIPS, 2019

Nonnegative Matrix Factorization (NMF) model

$$\boldsymbol{Z}_m = \boldsymbol{W}_m \boldsymbol{H}_m^\top \ , \forall m$$

$$\boldsymbol{Z}_m = \boldsymbol{W}_m \boldsymbol{Q} \boldsymbol{Q}^{-1} \boldsymbol{H}_m^\top$$

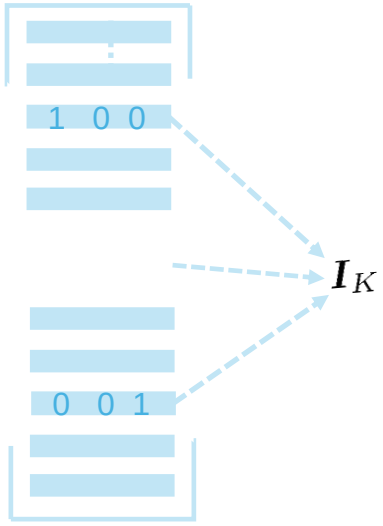
$\boldsymbol{W}_m^\star$ $\quad$ $\boldsymbol{H}_m^{\star\top}$

- Generally not unique
- Identifiability can be established under some conditions
  - Separability
  - Sufficiently scattered condition (SSC)

identifiability challenge

Separability

$$\boldsymbol{H}_m = \begin{bmatrix} \boldsymbol{A}_{m_1} \\ \vdots \\ \boldsymbol{A}_{m_t} \\ \vdots \\ \boldsymbol{A}_{m_T} \end{bmatrix}$$



Suppose there exists indices Λ such that

$$\boldsymbol{H}_m(\Lambda, :) = \boldsymbol{I}_K$$

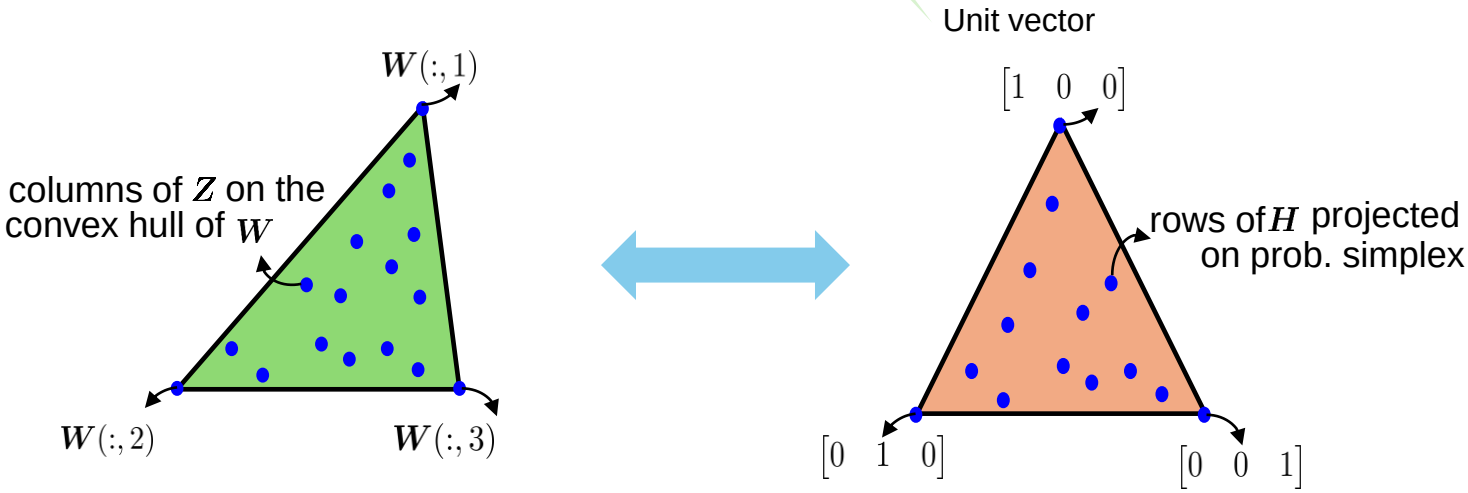
Identity Matrix

X. Fu, K. Huang, N. D. Sidiropoulos, and W.-K. Ma, “Nonnegative matrix factorization for signal and data analytics: Identifiability, algorithms, and applications,” IEEE Signal Process. Mag., vol. 36, no. 2, pp. 59–80, March 2019.

Separability in NMF

- Separability of $\boldsymbol{H}$: $\boldsymbol{H}$ has a diagonal submatrix [Donoho and Stodden'03]
  - $\exists\ \Lambda = \{\ell_1, \ldots, \ell_K\}$ such that

$$\boldsymbol{Z} = \boldsymbol{W}\boldsymbol{H}^\top; \quad \boldsymbol{H}(\ell_i, :) = \alpha_i \boldsymbol{e}_i,\ \alpha_i > 0,\ i = 1, \ldots, K$$

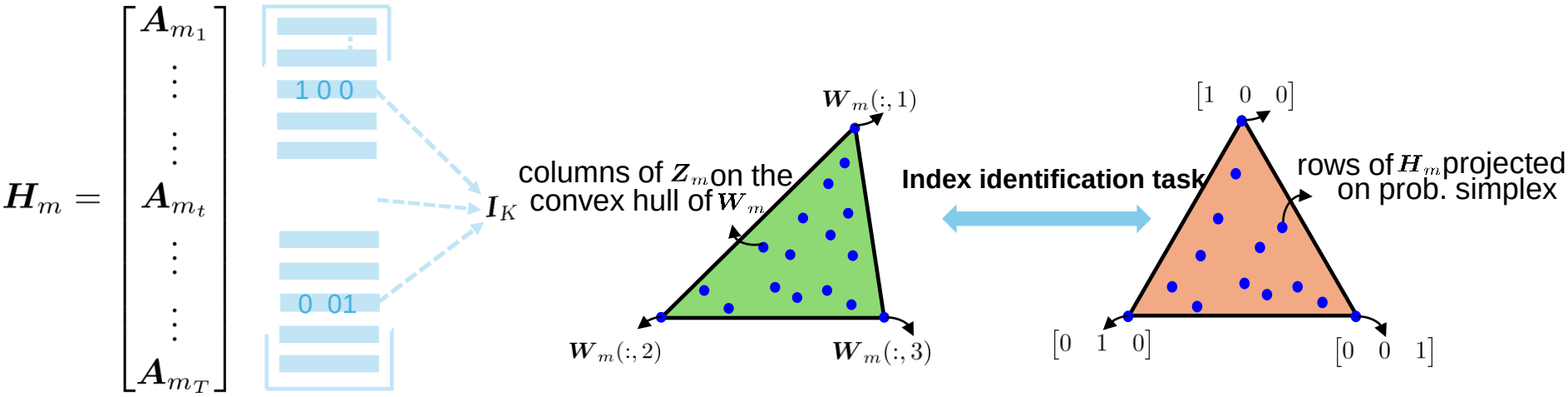


  - Solvable via Gram-Schmidt-like algorithms
    - successive projection algorithm (SPA) [Araújo et al'01, Gillis and Vavasis'14, Fu et al'15]
  - Intuition: **taller** $\boldsymbol{H}$ yields higher chance of separability holding

Identifiability Result

Theorem. If $\boldsymbol{H}_m$ satisfies separability, there exists a greedy algorithm to recover confusion matrices from $\boldsymbol{Z}_m = \boldsymbol{W}_m \boldsymbol{H}_m^\top$.

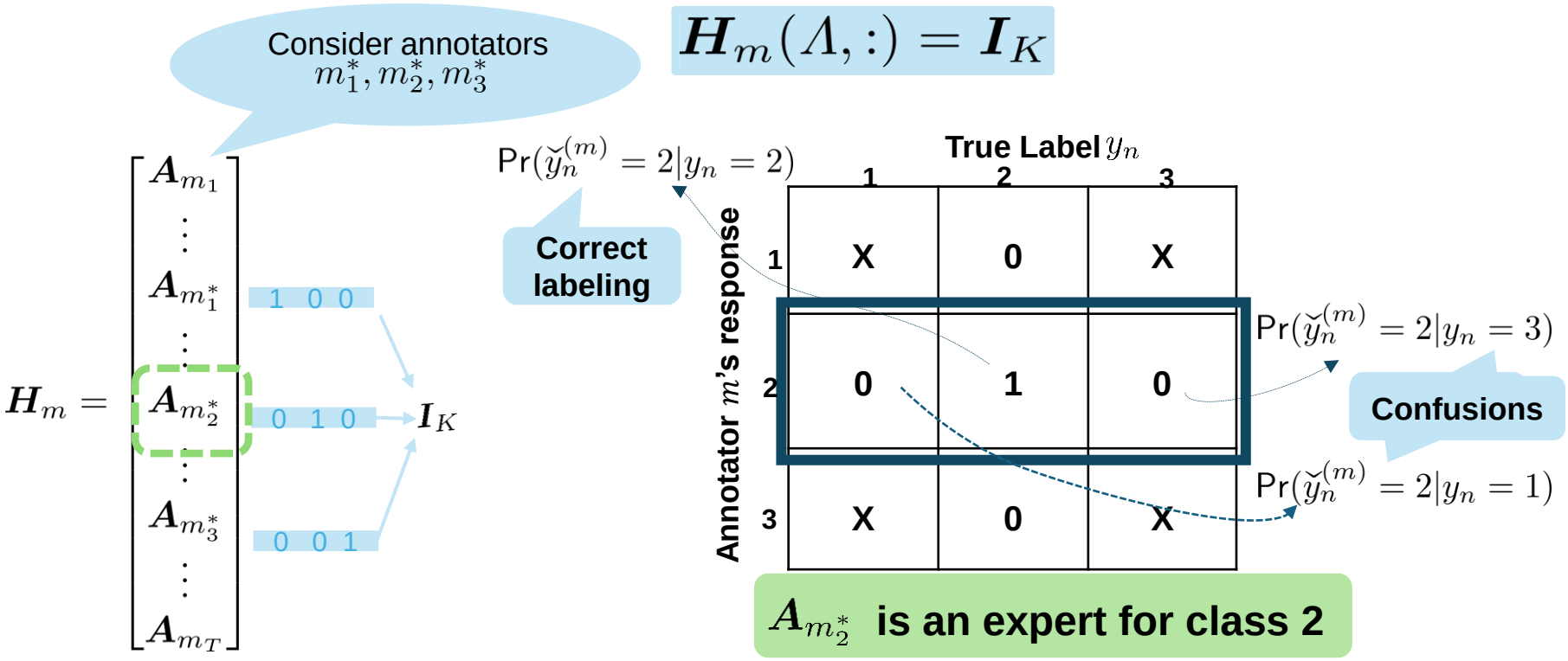
Separability



- Equivalent to estimating vertices of simplex formed by columns of $\boldsymbol{Z}_m$

S. Ibrahim, X. Fu, N. Kargas, and K. Huang, "Crowdsourcing via pairwise co-occurrences: Identifiability and algorithms," NeurIPS, 2019

A Closer Look at Separability



Specialized annotators exist → separability holds

- Approximate specialized annotators exist → Near-separability holds $\boldsymbol{H}_m(\Lambda,:) \approx \boldsymbol{I}_K$
  - Confusion matrix identified using *K* Gram-Schmidt iterations
  - MultiSPA algorithm [Ibrahim et al'19]

S. Ibrahim, X. Fu, N. Kargas, and K. Huang, "Crowdsourcing via pairwise co-occurrences: Identifiability and algorithms," NeurIPS, 2019

Identifiability Result

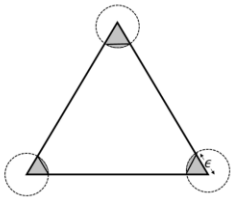
Q. Do more annotators help?

Proposition 1: Let $\rho > 0, \varepsilon > 0$, and assume that the rows of $\overline{\boldsymbol{H}}_m$ (row normalized version of $\boldsymbol{H}_m$) are generated within the $(K-1)$-probability simplex uniformly at random. If the number of annotators satisfies

$$M = \Omega\left(\frac{\varepsilon^{-2(K-1)}}{K}\log\left(\frac{K}{\rho}\right)\right),$$

then, with probability greater than or equal to $1-\rho$, there exist rows of $\overline{\boldsymbol{H}}_m$ indexed by $q_1, \dots q_K$ such that

$$\|\overline{\boldsymbol{H}}_m(q_k,:) - \boldsymbol{e}_k^\top\|_2 \leq \varepsilon, \ k = 1, \dots, K.$$

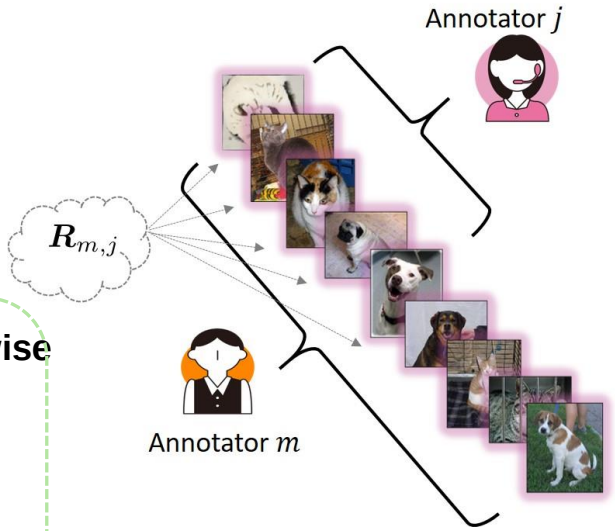


- Larger *M:* more rows in $\boldsymbol{H}_m$
- **Takeaway**: **Large number of annotators assists identifiability**

S. Ibrahim, X. Fu, N. Kargas, and K. Huang, "Crowdsourcing via pairwise co-occurrences: Identifiability and algorithms," NeurIPS, 2019

Enhanced Identifiability Guaranteed Approach

- For some annotators $\boldsymbol{Z}_m = \boldsymbol{W}_m \boldsymbol{H}_m^\top$ may not be identifiable



Coupled NMF

$$\text{find } \{\boldsymbol{A}_m\}_{m=1}^M, \boldsymbol{d}$$
$$\text{s.t. } \hat{\boldsymbol{R}}_{m,m'} = \boldsymbol{A}_m \boldsymbol{D} \boldsymbol{A}_{m'}^\top, \ (m,m') \in \boldsymbol{\Omega},$$
$$\boldsymbol{A}_m \geq \boldsymbol{0}, \boldsymbol{1}^\top \boldsymbol{A}_m = \boldsymbol{1}^\top, \ \boldsymbol{1}^\top \boldsymbol{d} = 1, \boldsymbol{d} \geq \boldsymbol{0}.$$

Available pairwise statistics

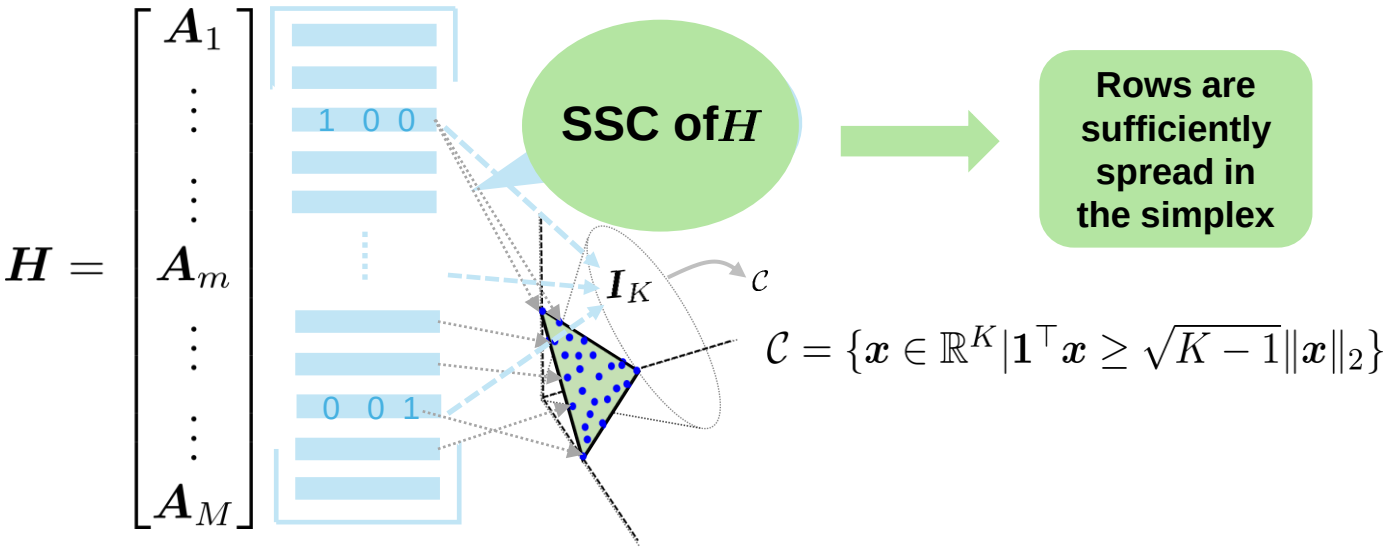
Idea: find a feasible solution that fits whatever data available

- Realized via $\text{minimize} \sum_{m,m' \in \boldsymbol{\Omega}} \text{KL}(\hat{\boldsymbol{R}}_{m,m'} || \boldsymbol{A}_m \boldsymbol{D} \boldsymbol{A}_{m'}^\top)$

Q. Identifiability?

Towards more realistic conditions

- **Sufficiently scattered condition (SSC)** [Huang et al'14]



- SSC on $\boldsymbol{H}$ implies that the rows are "sufficiently spread"
  - A second-order cone $\mathcal{C}$ can be contained inside conic hull spanned by $\boldsymbol{H}$

$$\mathcal{C} \subseteq \text{cone}(\boldsymbol{H}^\top)$$

- **Implications for NMF $\boldsymbol{R} = \boldsymbol{W}\boldsymbol{H}^\top$:**
  - If both factors satisfy SSC, model identifiable up to permutation
  - Relaxed condition compared to separability; more rows help

K. Huang, N. Sidiropoulos, and A. Swami, "Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition," IEEE Trans. Signal Process., 2014

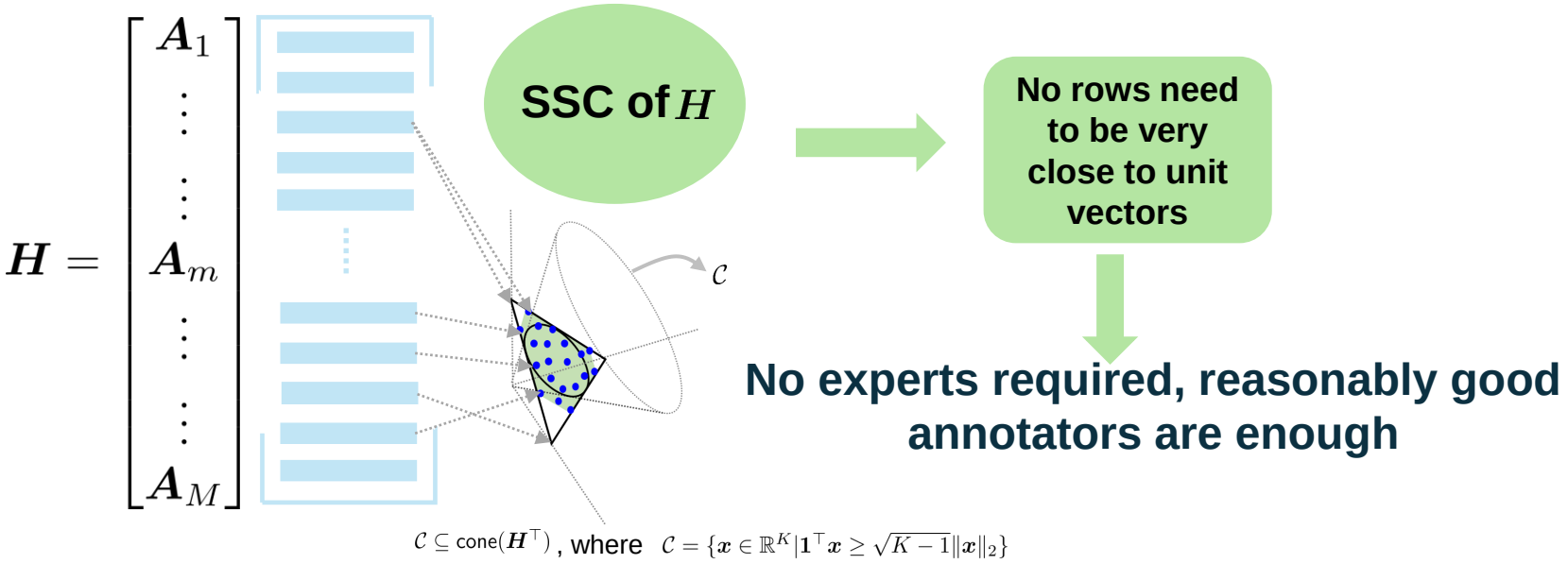
Identifiability via the SSC

Q. How does SSC improve the identifiability conditions?

Idea:

$$\begin{bmatrix} \boldsymbol{R}_{1,3} & \boldsymbol{R}_{1,4} \\ \boldsymbol{R}_{2,3} & \boldsymbol{R}_{2,4} \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \end{bmatrix}}_{\boldsymbol{H}^{(1)}} \boldsymbol{D} \underbrace{[\boldsymbol{A}_3^\top, \boldsymbol{A}_4^\top]}_{\boldsymbol{H}^{(2)\top}}$$

- If both matrices satisfy SSC, model is identifiable
- If $\boldsymbol{R}_{2,5} = \boldsymbol{A}_2 \boldsymbol{D} \boldsymbol{A}_5^\top$ is available, $\boldsymbol{A}_5$ identifiable via least squares ($\mathrm{rank}(\boldsymbol{A}_2) = K$)



K. Huang, N. Sidiropoulos, and A. Swami, "Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition," IEEE Trans. Signal Process., 2014

Towards More Realistic Conditions

Proposition

$$\begin{bmatrix} \boldsymbol{R}_{1,1} & \dots & \boldsymbol{R}_{1,M} \\ \vdots & \ddots & \vdots \\ \boldsymbol{R}_{M,1} & \dots & \boldsymbol{R}_{M,M} \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{A}_1 \\ \vdots \\ \boldsymbol{A}_M \end{bmatrix}}_{\boldsymbol{H}} \boldsymbol{D} \underbrace{[\boldsymbol{A}_1^\top, \dots, \boldsymbol{A}_M^\top]}_{\boldsymbol{H}^\top}$$

Even if there are no class experts, we can learn the model parameters via

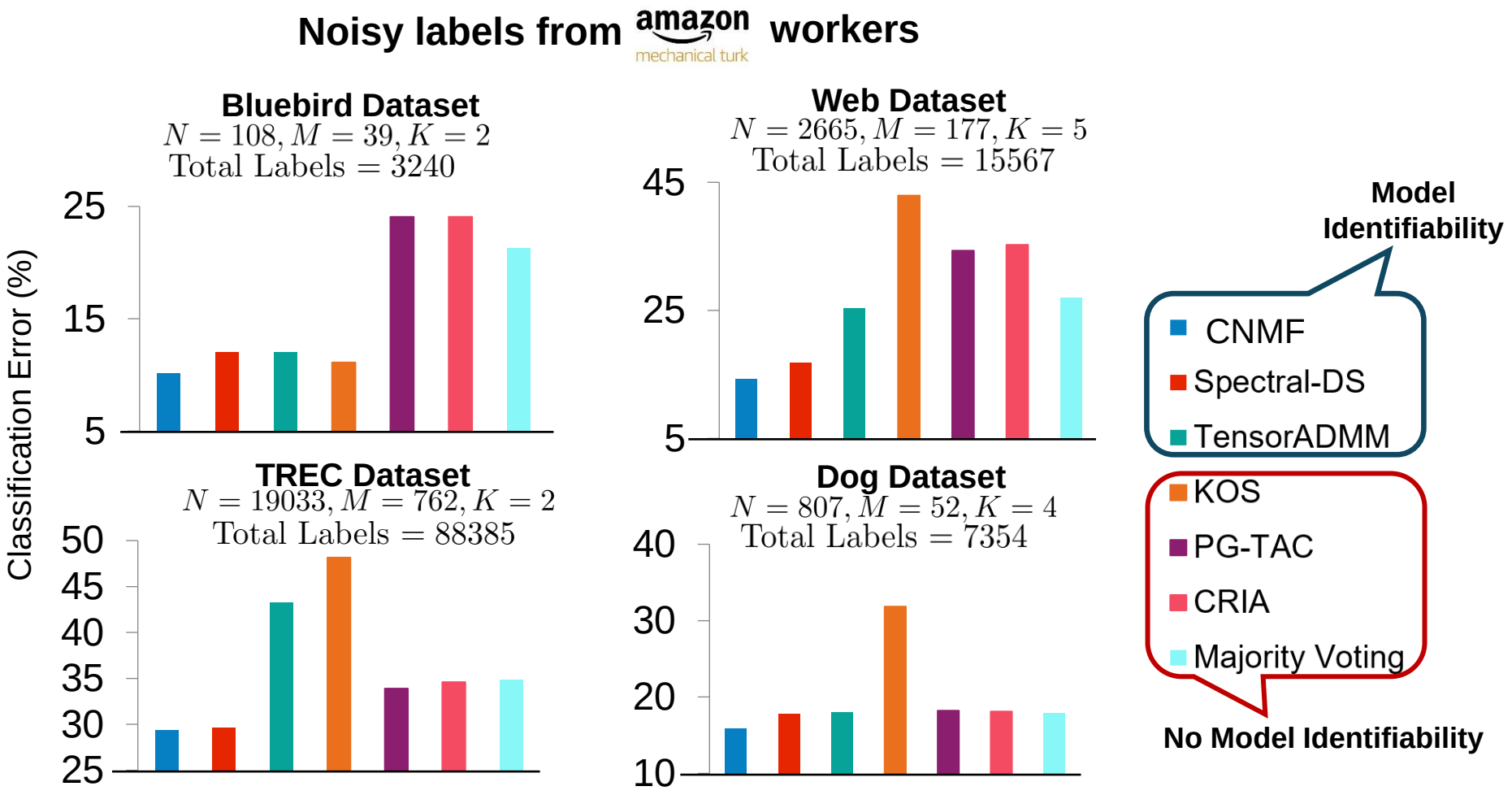
$$\begin{aligned} &\text{find } \{\boldsymbol{A}_m\}_{m=1}^M, \boldsymbol{d} \\ &\text{s.t. } \hat{\boldsymbol{R}}_{m,m'} = \boldsymbol{A}_m \boldsymbol{D} \boldsymbol{A}_{m'}^\top, \ (m,m') \in \boldsymbol{\Omega}, \\ &\quad \boldsymbol{A}_m \geq \boldsymbol{0}, \boldsymbol{1}^\top \boldsymbol{A}_m = \boldsymbol{1}^\top, \ \boldsymbol{1}^\top \boldsymbol{d} = 1, \boldsymbol{d} \geq \boldsymbol{0}. \end{aligned}$$

Available pairwise statistics

- Missing blocks handled via imputation (matrix completion)
- Symmetric NMF problem
- Faster, convergence guaranteed algorithm

S. Ibrahim and X. Fu. "Crowdsourcing via annotator co-occurrence imputation and provable symmetric nonnegative matrix factorization", ICML, 2021

Empirical Results



Methods guaranteeing model identifiability consistently perform well

Other Algorithms

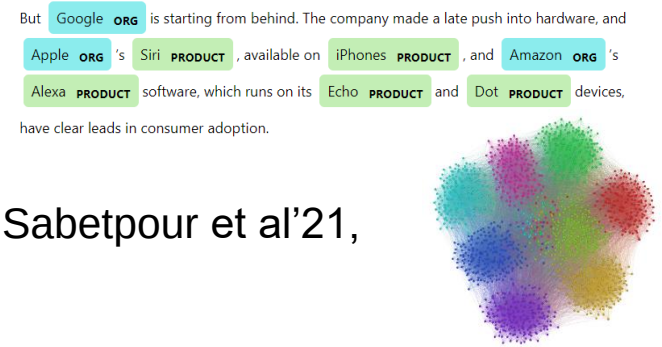
- For *K=2*, second-order moments (inter-annotator covariance) are enough!
  - Sensitivity and specificity estimated via eigendecomposition [Parisi et al'14, Jaffe et al'14]
- Second-order moments sufficient for one-coin model
  [Karger et al'11, Ghosh et al'11, Dalvi et al'13, Ma et al'18]

Generalization. Constrained optimization with $\boldsymbol{w} = [w_1 \dots w_M]$ capturing annotator reliability

$$\min_{\boldsymbol{y},\mathbf{w}} g(\boldsymbol{y},\mathbf{w}) = \sum_{m=1}^{M} w_m \sum_{n=1}^{N} d(y_n, \check{y}_n^{(m)})$$
$$\text{subject to} \quad r(\mathbf{w}) = 1$$

- *d:* loss; e.g., $d(y_n, \check{y}_n^{(m)}) = \mathbb{I}(y_n = \check{y}_n^{(m)})$
  r(***w***): regularization; e.g., for sparsity
- alternating minimization

- Minimax conditional entropy [Zhou et al'12]
- Maximum margin MV [Tian-Zhu'15] - resembles SVM
- Data dependencies can be captured
  - Sequential data [Nguyen et al'17,Simpon-Gurevych'19, Sabetpour et al'21, Traganitis-Giannakis'22, Marinnan et al'24]
  - Graph data [Traganitis-Giannakis'22]

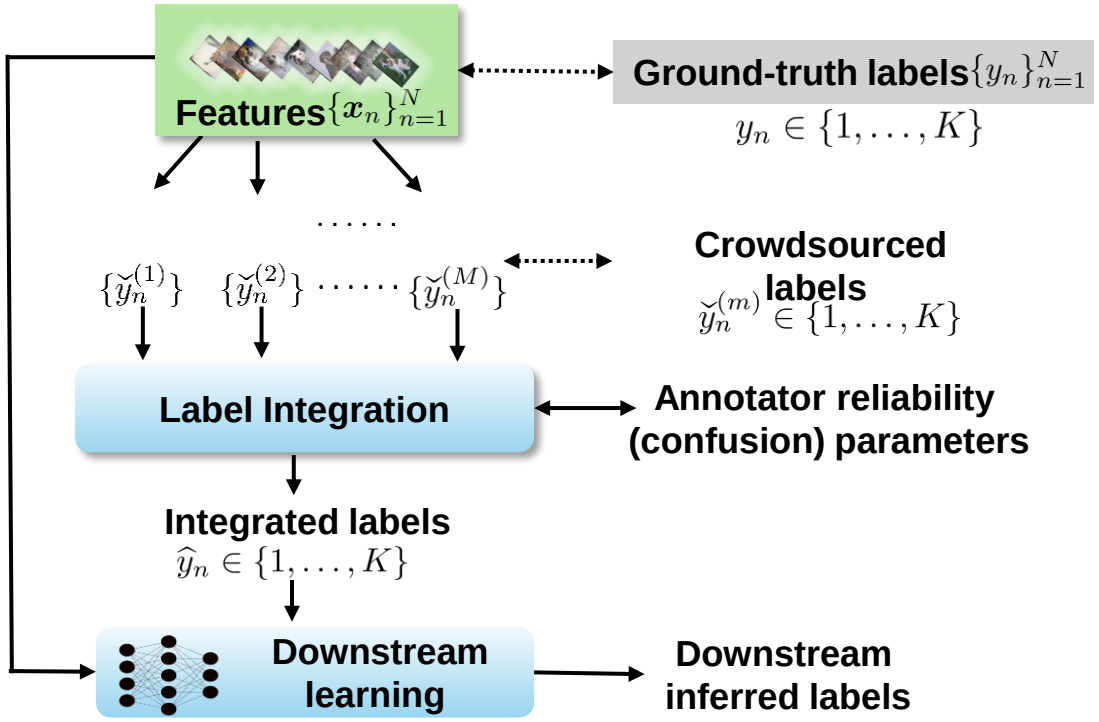


Two-stage Approach: Key Takeaways

- **DS model:** core model for label aggregation
  - Several variants available
  - Identifiability extensively studied
- Model identifiability often key performance indicator
- Sample complexity is crucial when selecting algorithm
- Moment-based methods
  - ✓ **Label efficiency**
  - ✓ **Computational efficiency**
  - ✓ **Model identifiability**

Challenges

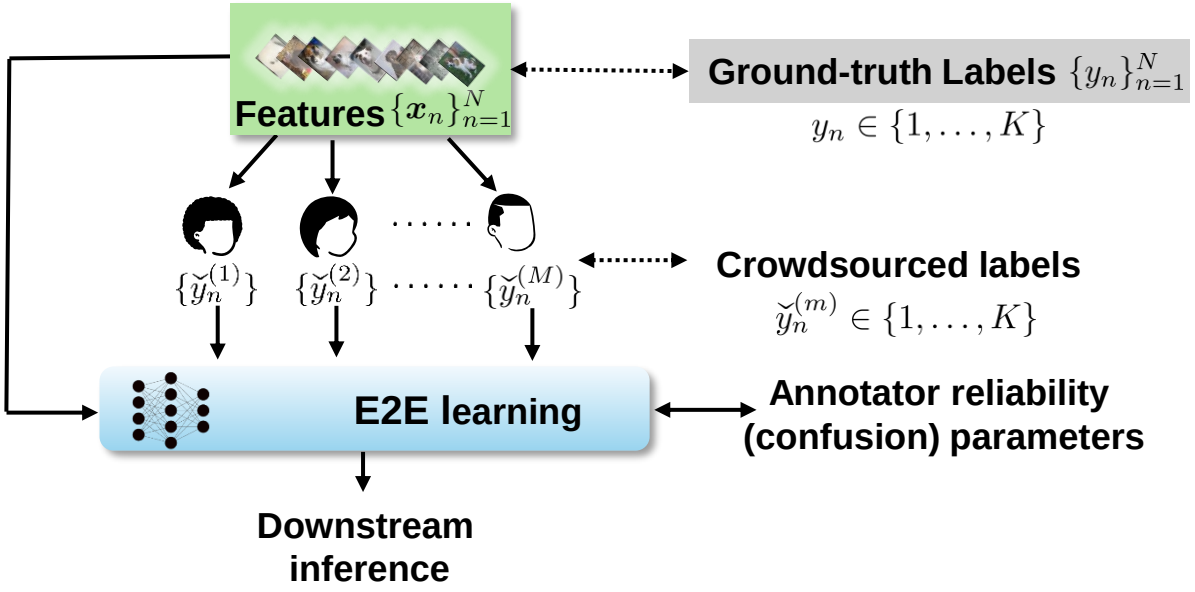
- **Ignore data features**
- **Error propagation**
- **Conditional independence**



Outline

- Motivation and problem statement
- Part I: Combining crowdsourced labels
- Part II: End-to-end (E2E) learning with crowdsourced labels
  - E2E learning with EM
  - Deep learning-based E2E crowdsourcing
  - Instance-dependent E2E crowdsourcing
- Part III: Other aspects of crowdsourcing
- Conclusions and open issues

End-To-End (E2E) Learning



[Raykar et al'10]
[Rodrigues & Pereira'18]
[Khetan et al'18]
[Tanno et al'19]
[Li et al'20]
[Chen et al'20]
[Chu et al'21]
[Wei et al'22]

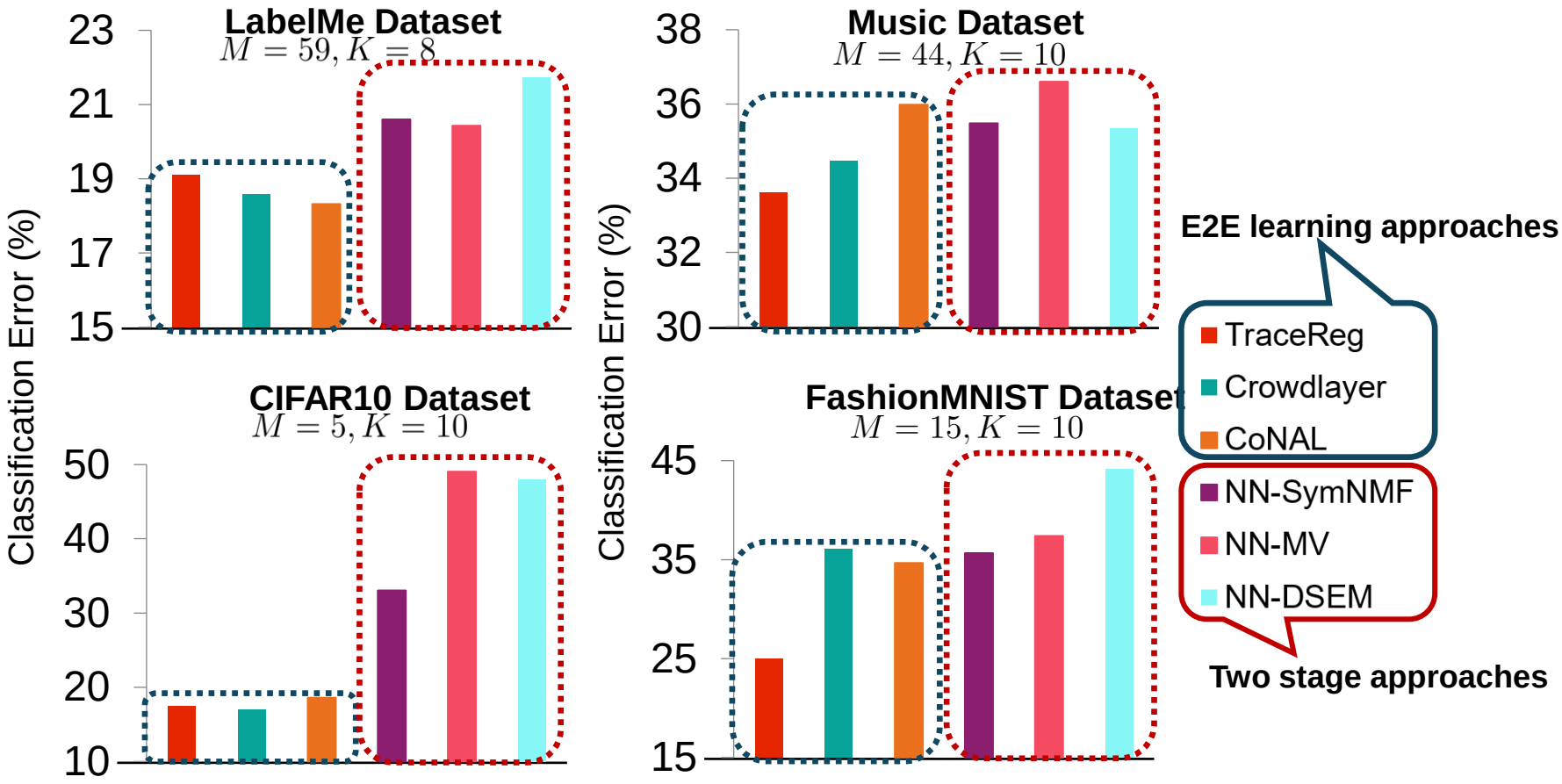
- ❑ Joint label integration and downstream learning
- ➢ For classification, E2E systems aim at

$$\widehat{\boldsymbol{f}}_{\boldsymbol{\theta}} \leftarrow \arg\min_{\boldsymbol{\theta},\boldsymbol{\eta}} \ \ell(\{\boldsymbol{x}_n\}, \{\breve{y}_n^{(m)}\}, \boldsymbol{\theta}, \boldsymbol{\eta})$$

additional model parameters, e.g., annotator confusion matrices

- ➢ The goal of the E2E learning is to learn $\widehat{\boldsymbol{f}}_{\boldsymbol{\theta}}$ corresponds to the optimal classifier

End-To-End (E2E) Learning



- E2E methods in general outperform two-stage approaches
- Key research directions:
  - Understanding the performance of E2E crowdsourcing
  - Enhancing performance with principled design

E2E crowdsourcing via maximum likelihood

- Given dataset $\mathcal{D} := \{\boldsymbol{x}_n, \{\breve{y}_n^{(m)}\}_{m=1}^M\}_{n=1}^N$, consider

$$\Pr(\mathcal{D}) = \prod_{n=1}^{N} \Pr(\boldsymbol{x}_n, \breve{y}_n^{(1)}, \ldots, \breve{y}_n^{(M)}) = \prod_{n=1}^{N} \sum_{y_n=1}^{K} \Pr(y_n|\boldsymbol{x}_n) \prod_{i=1}^{M} \Pr(\breve{y}_n^{(m)}|y_n, \boldsymbol{x}_n)$$

Classifier

$[\boldsymbol{f}(\boldsymbol{x}_n)]_k = \Pr(y_n = k|\boldsymbol{x}_n)$

Confusion Matrix

$[\boldsymbol{A}_m]_{k,k'} = \Pr(\breve{y}_n^{(m)} = k|y_n = k')$

Common assumption:
Confusion matrices are not data dependent

[Li et al'21, Rodrigues & Pereira'18, Tanno et al'19]

$$= \prod_{n=1}^{N} \sum_{y_n=1}^{K} \Pr(y_n|\boldsymbol{x}_n) \prod_{m=1}^{M} \Pr(\breve{y}_n^{(m)}|y_n)$$

$\Pr(y_n|\boldsymbol{x}_n) \rightarrow \boldsymbol{f}(\boldsymbol{x}_n)$

$\Pr(\breve{y}_n^{(m)}|y_n) \rightarrow \boldsymbol{A}_m$

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

EM for E2E crowdsourcing

$$\Pr(\mathcal{D};\boldsymbol{\psi}) = \prod_{n=1}^{N}\sum_{y_n=1}^{K}[\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_n)]_{y_n}\prod_{m=1}^{M}\boldsymbol{A}_m(\breve{y}_n^{(m)}, y_n)$$

Learnable parameters $\boldsymbol{\psi} = [\boldsymbol{A}_1, \ldots, \boldsymbol{A}_M, \boldsymbol{\theta}]$

Classifier parameterized by $\boldsymbol{\theta}$

❑ E-Step

Expected value of complete log-likelihood

$$Q(\boldsymbol{\psi};\boldsymbol{\psi}^t) = \mathbb{E}_{\mathcal{Y}\sim\Pr(\mathcal{Y};\mathcal{D},\boldsymbol{\psi}^t)}[\log\Pr(\mathcal{D},\mathcal{Y};\psi)] = \sum_{n=1}^{N}\sum_{k=1}^{K} q(y_n = k;\boldsymbol{\psi}^t)\log\Pr(\boldsymbol{x}_n, y_n = k, \breve{y}_n^{(1)}, \ldots, \breve{y}_n^{(M)};\boldsymbol{\psi})$$

❑ M-Step

Estimates $\boldsymbol{\psi}$ by maximizing $Q(\boldsymbol{\psi};\boldsymbol{\psi}^t)$

$$\boldsymbol{A}_m^{t+1}(k',k) = \frac{\sum_{n=1}^{N} q(y_n = k;\boldsymbol{\psi}^t)\mathbb{I}[\breve{y}_n^m = k']}{\sum_{k''=1}^{K}\sum_{n=1}^{N} q(y_n = k;\boldsymbol{\psi}^t)\mathbb{I}[\breve{y}_n^m = k'']}$$

$$\boldsymbol{\theta}^{t+1} = \arg\max_{\boldsymbol{\theta}}\ Q(\boldsymbol{\theta};(\boldsymbol{\theta}^t, \{\boldsymbol{A}_m^{t+1}\})).$$

- Similar to EM for DS model
- Several classifier variants
  - logistic regression [Raykar et al'10]
  - neural network [Rodrigues & Pereira'18]
  - conditional random field [Rodrigues et al'14]

Deep learning with Crowd Layer

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

$\boldsymbol{p}_n^{(m)}$

Annotator's confusion

$\boldsymbol{A}_m$

Label predictor

$\boldsymbol{f}(\boldsymbol{x}_n)$

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

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

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

- Model parameters estimated via

$$\underset{\boldsymbol{f_\theta}, \{\boldsymbol{A}_m\}}{\text{minimize}} \frac{1}{|\mathcal{S}|} \sum_{(m,n)\in\mathcal{S}} \text{CE}(\boldsymbol{A}_m \boldsymbol{f_\theta}(\boldsymbol{x}_n), \tilde{y}_n^{(m)})$$

$$\text{subject to} \quad \boldsymbol{f_\theta} \in \mathcal{F}, \ \boldsymbol{A}_m \in \mathcal{A}, \ \forall m.$$

Indices of observed labels

Neural network function class

Constrained set

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

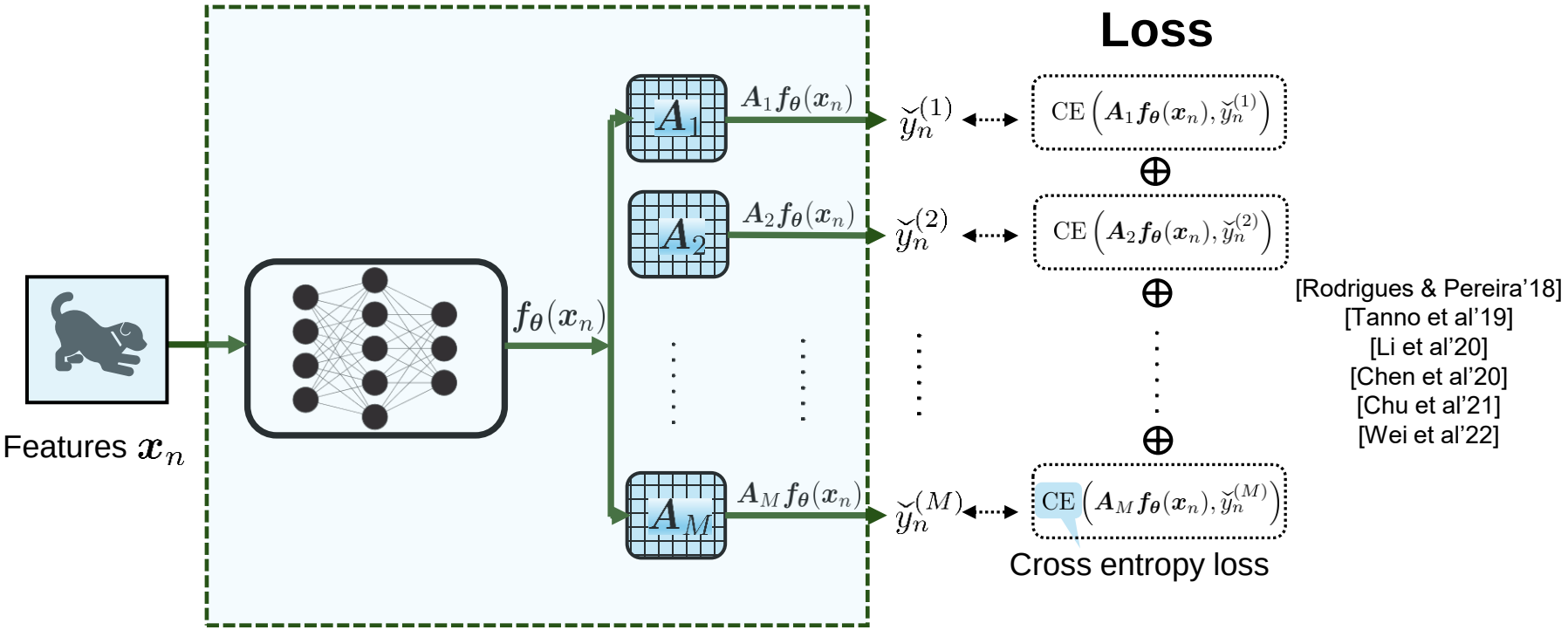
F. Rodrigues and F. Pereira, "Deep learning from crowds," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, no. 1, 2018

Coupled Cross Entropy Minimization (CCEM)

❑ E2E learning criterion

$$\underset{\boldsymbol{f_\theta},\{\boldsymbol{A}_m\}}{\text{minimize}} \quad \frac{1}{|\mathcal{S}|} \sum_{(m,n)\in\mathcal{S}} \mathsf{CE}(\boldsymbol{A}_m \boldsymbol{f_\theta}(\boldsymbol{x}_n), \breve{y}_n^{(m)})$$

$$\text{subject to} \quad \boldsymbol{f_\theta} \in \mathcal{F}, \ \boldsymbol{A}_m \in \mathcal{A}, \ \forall m.$$



[Rodrigues & Pereira'18]
[Tanno et al'19]
[Li et al'20]
[Chen et al'20]
[Chu et al'21]
[Wei et al'22]

Identifiability in E2E crowdsourcing

Q. Identifiability in E2E crowdsourcing?

- When $N \to \infty$, CCEM returns $\boldsymbol{p}_n^{(m)} = \boldsymbol{A}_m \boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_n)$

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

$$\boldsymbol{p}_n^{(m)} = \boldsymbol{A}_m \boldsymbol{Q} \boldsymbol{Q}^{-1} \boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_n)$$
$\boldsymbol{A}_m^\star$ $\boldsymbol{f}^\star(\boldsymbol{x}_n)$

Non-unique decomposition

- There may exist many possible nonsingular matrices s.t. $\boldsymbol{p}_n^{(m)} = \boldsymbol{A}_m \boldsymbol{Q} \boldsymbol{Q}^{-1} \boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_n)$

Q. How can we understand performance of CCEM?

Q. Identifiability under CCEM?

Identifiability for CCEM

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

CCEM

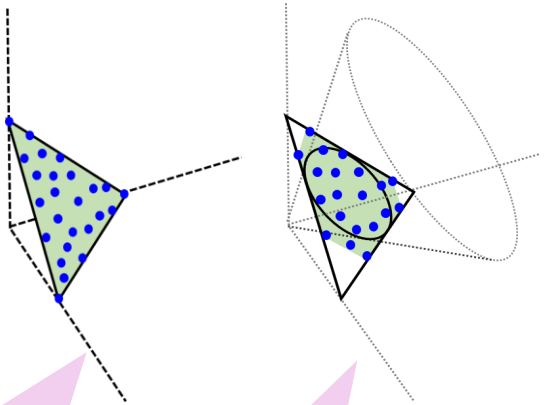
- If both latent factors satisfy ***separability condition or (SSC)***, NMF model is essentially unique up to a permutation matrix

$$\|\widehat{\boldsymbol{A}}_m\boldsymbol{\Pi} - \boldsymbol{A}_m\| \to 0, \quad \|\boldsymbol{\Pi}^\top \widehat{\boldsymbol{f}}_{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x})\| \to 0,$$

- Separability on $\boldsymbol{W}$: class experts for each class
- Separability on $\boldsymbol{F}$: existence of anchor points for each class $\{\boldsymbol{x}_{n_1}, \ldots, \boldsymbol{x}_{n_K}\}$ s.t.

$$\boldsymbol{f}(\boldsymbol{x}_{n_k}) = \boldsymbol{e}_k \Leftrightarrow \Pr(y_{n_k} = k|\boldsymbol{x}_{n_k}) = 1$$

- SSC relaxes these conditions



S. Ibrahim, T. Nguyen, and X. Fu. ``Deep Learning From Crowdsourced Labels: Coupled Cross-Entropy Minimization, Identifiability, and Regularization'', ICLR, 2023.

Model identifiability under CCEM

Theorem 1 [Ibrahim et al., 2023]. Suppose that the Assumptions 1-5 hold true, Assume that each $[\boldsymbol{A}_m^\natural \boldsymbol{f}^\natural(\boldsymbol{x}_n)]_k$ and $[\boldsymbol{A}_m \boldsymbol{f}(\boldsymbol{x}_n)]_k$, $\forall \boldsymbol{A}_m \in \mathcal{A}, \forall \boldsymbol{f} \in \mathcal{F}$ are at least $(1/\beta)$. Also assume that $\sigma_{\max}(\boldsymbol{A}_m^\natural) \leq \sigma$, $\forall m$, for a certain $\sigma > 0$. Then, for any $\alpha > 0$, with probability greater than $1 - {}^{K}/_{N^\alpha}$, any optimal solution $\widehat{\boldsymbol{A}}_m$ and $\widehat{\boldsymbol{f}}$ of the CCEM satisfies the following relations:

$$\min_{\boldsymbol{\Pi}} \|\widehat{\boldsymbol{A}}_m - \boldsymbol{A}_m^\natural \boldsymbol{\Pi}\|_{\mathrm{F}}^2 = K\sigma^2(\eta + \xi_1 + \xi_2), \ \forall m \in [M],$$

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mathcal{D}} \left[\min_{\boldsymbol{\Pi}} \|\widehat{\boldsymbol{f}}(\boldsymbol{x}) - \boldsymbol{\Pi}^\top \boldsymbol{f}^\natural(\boldsymbol{x})\|_2^2\right] = K(\eta + \xi_1 + \xi_2),$$

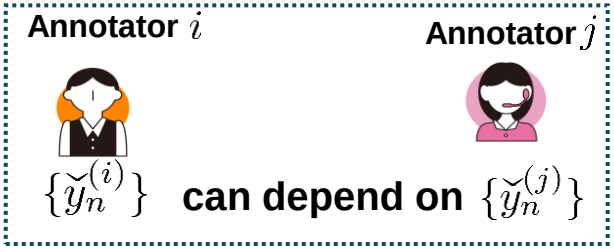
where $\eta^2 = \mathcal{O}\left({}^{\beta MN^\alpha}/_{\sqrt{S}} \left(\sqrt{M \log S} + (\|\boldsymbol{X}\|_{\mathrm{F}} \mathscr{R}_{\mathcal{F}})^{\frac{1}{4}}\right) + \beta\sqrt{K} MN^\alpha \nu\right)$, $\boldsymbol{\Pi} \in \{0,1\}^K$ is a permutation matrix, and $\boldsymbol{X} = [\boldsymbol{x}_{n_1}, \ldots, \boldsymbol{x}_{n_S}]$, $(m_s, n_s) \in \mathcal{S}$, if conditions $\xi_1, \xi_2 \leq {}^{1}/_{K}$, $\nu \leq {}^{1}/_{\beta K^2 M^2 N^\alpha}$, and $S = |\mathcal{S}| = \Omega\left(\beta^2 M^2 N^{2\alpha} K^2 \max\left(M \log S, \sqrt{\|\boldsymbol{X}\|_{\mathrm{F}} \mathscr{R}_{\mathcal{F}}}\right)\right)$ hold.

CCEM correctly learns the true confusions and the true classifier, under the assumptions

1. ***Anchor point condition*** ⇨ Per class, there is a data sample belonging to that class with prob. close to 1
2. ***Class expert condition*** ⇨ Per class, there is an expert

✓ **no conditional independence required**

✓ **finite number of data and missing labels**



Towards more realistic settings

1 ***Anchor point condition*** ⇨ Per class, there is a data sample belonging to that class with prob. close to 1

Reflects diversity of data → **Easier to satisfy for very large datasets**

Natural Images

Social Media Sentiment Data

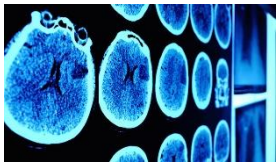
Social Network Data

Labels often obtained from non-experts, e.g., via AMT

amazon mechanical turk

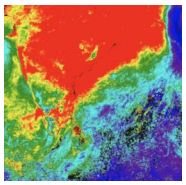
2 ***Class expert condition*** ⇨ At least one expert, per class

Reflects wisdom of the crowd



Medical Data

Ecological Data



Geospatial Data

Requires domain experts

Identifiability by volume minimization

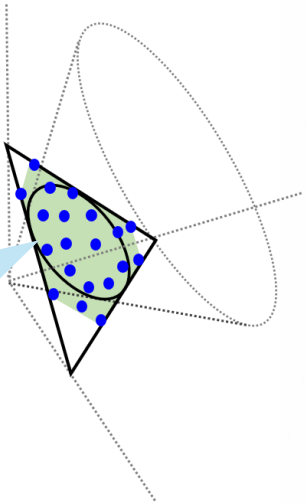
- Identifiability of vanilla CCEM requires conditions on
  - latent factors • ***class experts*** • ***anchor points***
- Conditions on one of the latent factors can be relaxed via regularization

$$\underbrace{\begin{bmatrix} p_1^{(1)} & \cdots & p_N^{(1)} \\ \vdots & \ddots & \vdots \\ p_1^{(M)} & \cdots & p_N^{(M)} \end{bmatrix}}_{P} = \underbrace{\begin{bmatrix} A_1 \\ \vdots \\ A_M \end{bmatrix}}_{W} \underbrace{\begin{bmatrix} f(x_1) & \cdots & f(x_N) \end{bmatrix}}_{F}$$

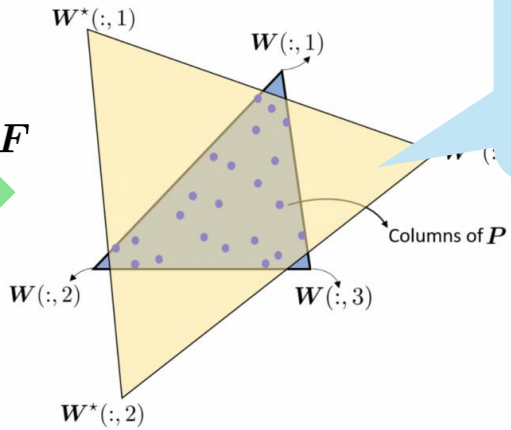
NMF Volume Minimization

If one of the latent factors satisfy the SSC and other is full rank, finding the minimum volume data enclosing simplex guarantees NMF identifiability [Fu et al'15]

***Sufficiently Scattered Condition* (SSC) on F**



$P = WF$



Minimize the volume occupied by conic hull of W

S. Ibrahim, T. Nguyen, and X. Fu. ``Deep Learning From Crowdsourced Labels: Coupled Cross-Entropy Minimization, Identifiability, and Regularization'', ICLR, 2023.

Enhanced identifiability under CCEM

Regularized CCEM:

If F satisfies the SSC
(we have more data, but may have *no experts* to label)

amazon mechanical turk

GeoCrowdNet (F)

$$\underset{\boldsymbol{f_\theta},\{\boldsymbol{A}_m\}}{\text{minimize}} \ \frac{1}{|\mathcal{S}|}\sum_{(m,n)\in\mathcal{S}} \mathsf{CE}(\boldsymbol{A}_m \boldsymbol{f_\theta}(\boldsymbol{x}_n), \breve{y}_n^{(m)}) - \lambda \log\det \boldsymbol{F}\boldsymbol{F}^\top$$

$$\text{subject to} \ \ \boldsymbol{f_\theta} \in \mathcal{F}, \ \boldsymbol{A}_m \in \mathcal{A}, \ \forall m.$$

Volume minimization-based regularization

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

S. Ibrahim, T. Nguyen, and X. Fu. ``Deep Learning From Crowdsourced Labels: Coupled Cross-Entropy Minimization, Identifiability, and Regularization'', ICLR, 2023.

Enhanced identifiability under CCEM

Regularized CCEM:

If W satisfies the SSC
(we may have *less data*, but have experts to label)

GeoCrowdNet (W)

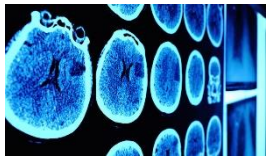
$$\underset{\boldsymbol{f_\theta},\{\boldsymbol{A}_m\}}{\text{minimize}} \ \frac{1}{|\mathcal{S}|}\sum_{(m,n)\in\mathcal{S}} \text{CE}(\boldsymbol{A}_m \boldsymbol{f_\theta}(\boldsymbol{x}_n), \breve{y}_n^{(m)}) - \lambda \log\det \boldsymbol{W}^\top \boldsymbol{W}$$

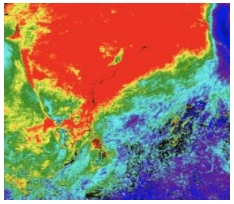
$$\text{subject to} \ \ \boldsymbol{f_\theta} \in \mathcal{F}, \ \boldsymbol{A}_m \in \mathcal{A}, \ \forall m.$$

Volume minimization-based regularization

$$\begin{bmatrix} \boldsymbol{p}_1^{(1)} & \cdots & \boldsymbol{p}_N^{(1)} \\ \vdots & \ddots & \vdots \\ \boldsymbol{p}_1^{(M)} & \cdots & \boldsymbol{p}_N^{(M)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_1 \\ \vdots \\ \boldsymbol{A}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{f}(\boldsymbol{x}_1) & \cdots & \boldsymbol{f}(\boldsymbol{x}_N) \end{bmatrix}$$

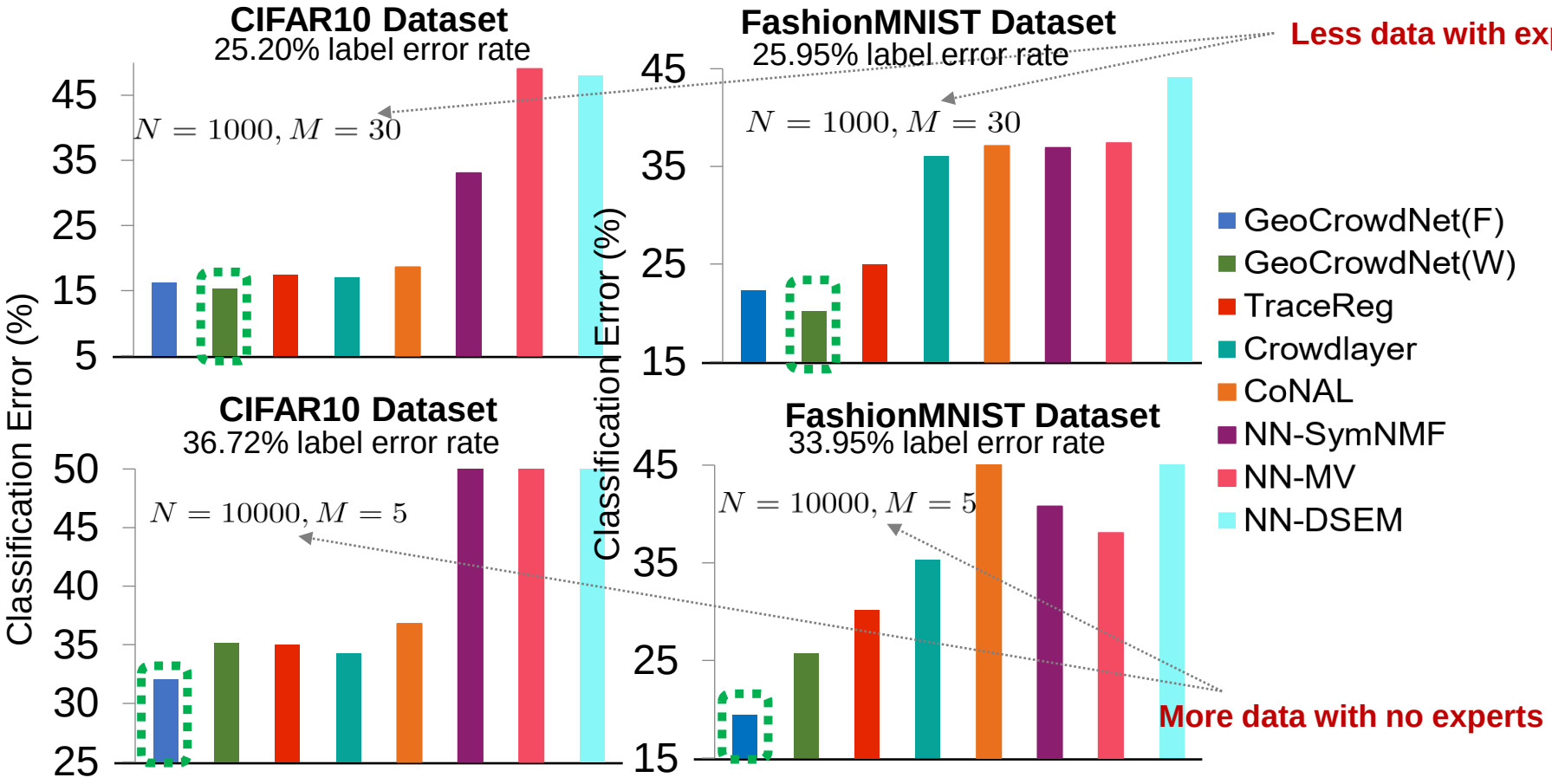
$\boldsymbol{W}$





S. Ibrahim, T. Nguyen, and X. Fu. ``Deep Learning From Crowdsourced Labels: Coupled Cross-Entropy Minimization, Identifiability, and Regularization'', ICLR, 2023.

Empirical results



- Around 57,000 training samples; any annotator only labels 10%.
- The annotators are randomly sampled from a pool of machine classifiers whose accuracy ranges from 15% to 85%

Other E2E crowdsourcing approaches

- **CCEM with Trace Regularization** [Tanno et al'19]

$$\underset{\boldsymbol{f_\theta},\{\boldsymbol{A}_m\}}{\text{minimize}} \ \frac{1}{|\mathcal{S}|}\sum_{(m,n)\in\mathcal{S}} \mathsf{CE}(\boldsymbol{A}_m \boldsymbol{f_\theta}(\boldsymbol{x}_n), \breve{y}_n^{(m)}) + \beta \sum_{m=1}^{M} \text{trace}(\boldsymbol{A}_m)$$

$$\text{subject to} \ \ \boldsymbol{f_\theta} \in \mathcal{F}, \ \boldsymbol{A}_m \in \mathcal{A}, \ \forall m.$$

  - If average of confusion matrices diagonally dominant – identifiable!

- **Agreement-based Model** [Peng et al'19]
  - Maximizes agreement between label predictor and predictor trained on $\breve{\mathcal{Y}}$

$$\underset{\boldsymbol{\theta},\boldsymbol{\phi}}{\text{maximize}} \ \text{MIG}^f(\boldsymbol{f_\theta}, \boldsymbol{g_\phi}; \{\boldsymbol{x}_n\}, \{\breve{\boldsymbol{y}}_n^{(m)}\})$$

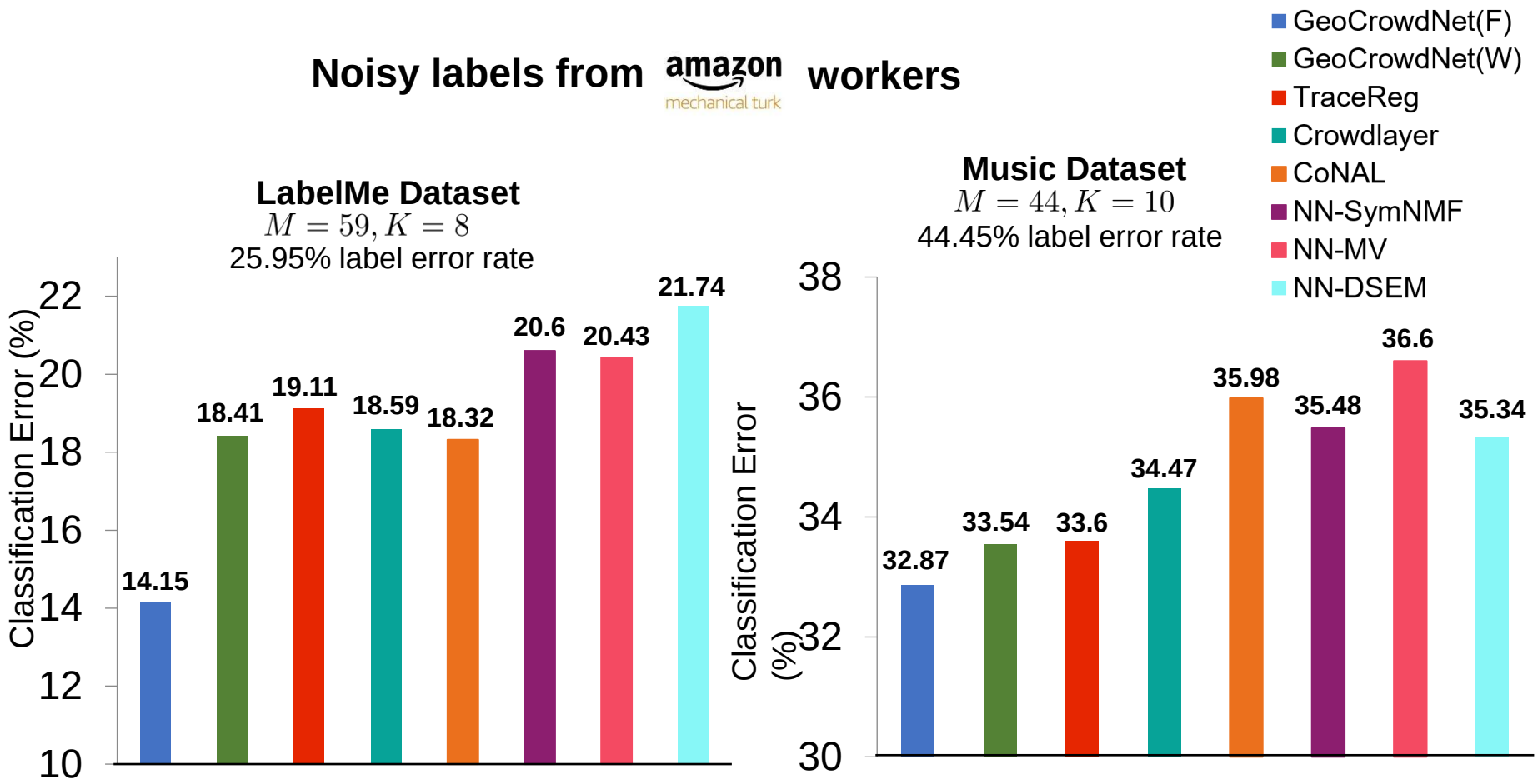
$$\boldsymbol{g_\phi}(\{\breve{\boldsymbol{y}}_n^{(m)}\}) = \text{softmax}\left(\textstyle\sum_{m=1}^{M} \boldsymbol{W}^{(m)} \breve{\boldsymbol{y}}_n^{(m)} + \boldsymbol{b}\right)$$

one hot encoding of annotator labels

Average agreement between annotators

  - Asymptotic optimality under conditional independence

Empirical results



Criterion designed for "no experts case" shows edge in practice

Instance-dependent Label Noise

- Annotator errors can be instance-dependent in practice

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

$\boldsymbol{p}_n^{(m)}$

Data-dependent confusion

$\boldsymbol{T}_m(\boldsymbol{x}_n)$

Label predictor

$\boldsymbol{f}(\boldsymbol{x}_n)$

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

Goal : learn $\boldsymbol{f}(\boldsymbol{x})$ and $\boldsymbol{T}_m(\boldsymbol{x}), \forall m$

- Several approaches for single-annotator scenario
  - Part-dependent label noise [Xia et al'20]
  - Similarity regularization-based [Cheng et al'22]
  - Bayes label-based [Yang et al'22]
  - Sample selection-based [Cheng et al'21, Berthon et al'21]
  - Graphical model-based [Yao et al'21, Wang et al'22]
- Identifiability challenging even with single annotator

Methods for instance-dependent label noise

- Common approach for multiple annotators: **learn two functions**

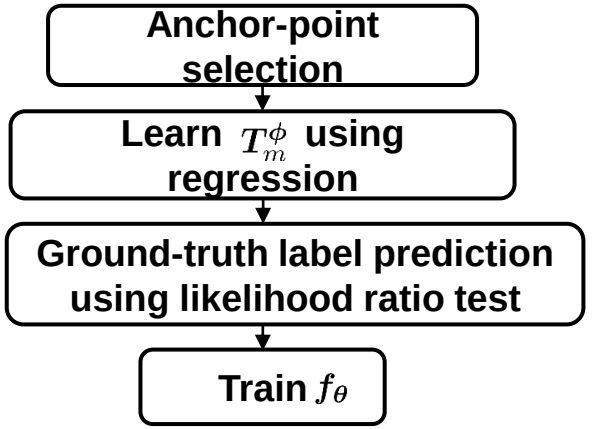
$$\Pr(\breve{y}_n^{(m)} = k|\boldsymbol{x}_n) = \sum_{k'=1}^{K} \Pr(\breve{y}_n^{(m)} = k|y_n = k', \boldsymbol{x}_n) \Pr(y_n = k'|\boldsymbol{x}_n)$$

$\boldsymbol{p}_n^{(m)}$ $\quad \boldsymbol{T}_m^{\boldsymbol{\phi}} : \mathbb{R}^D \to \mathbb{R}^{K \times K}$ $\boldsymbol{T}_m(\boldsymbol{x}_n)$ $\quad \boldsymbol{f}_{\boldsymbol{\theta}} : \mathbb{R}^D \to \mathbb{R}^K$ $\boldsymbol{f}(\boldsymbol{x}_n)$

- [Zhang et al'20] extended trace regularization

$$\underset{\{\boldsymbol{T}_m^{\boldsymbol{\phi}} \in \mathcal{A}\}, \boldsymbol{f}_{\boldsymbol{\theta}} \in \boldsymbol{\Delta}_K}{\text{minimize}} \; -\frac{1}{|\mathcal{S}|} \sum_{(m,n) \in \mathcal{S}} \sum_{k=1}^{K} \mathbb{I}[\breve{y}_n^{(m)} = k] \log[\boldsymbol{T}_m^{\boldsymbol{\phi}}(\boldsymbol{x}_n) \boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_n)]_k + \beta \sum_{m=1}^{M} \sum_{n=1}^{N} \text{trace}(\boldsymbol{T}_m^{\boldsymbol{\phi}}(\boldsymbol{x}_n))$$

- [Guo et al'23] used a multi-stage strategy:
  - Learn $\boldsymbol{T}_m^{\boldsymbol{\phi}}$ using anchor points
  - Train $\boldsymbol{f}_{\boldsymbol{\theta}}$
- Identifiability still remains as a critical challenge



Instance-dependent noise as outliers

- Instance-dependent label noise may happen *sparingly* [Nguyen et al., 2014]

$$\boldsymbol{T}_m(\boldsymbol{x}_n) = \boldsymbol{A}_m + \boldsymbol{E}_m(\boldsymbol{x}_n)$$

Instance-independent confusion matrix

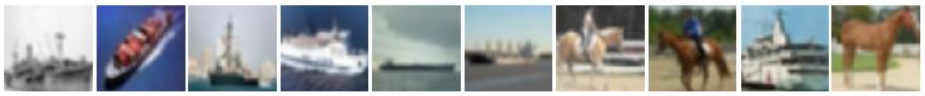
Instance-dependent perturbation matrix

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

$$\boldsymbol{p}_n^{(m)} = \begin{cases} \boldsymbol{A}_m\boldsymbol{f}(\boldsymbol{x}_n) + \boldsymbol{e}_n^{(m)}, & \forall n \in \mathcal{I} \\ \boldsymbol{A}_m\boldsymbol{f}(\boldsymbol{x}_n), & \forall n \in \mathcal{I}^{\mathrm{c}}, \end{cases} \quad \boldsymbol{e}_n^{(m)} = \boldsymbol{E}_m(\boldsymbol{x}_n)\boldsymbol{f}(\boldsymbol{x}_n)$$

Outlier index set with $\boldsymbol{E}_m(\boldsymbol{x}_n) \neq \boldsymbol{0}$

(as) For many data samples $\boldsymbol{E}_m(\boldsymbol{x}_n) = \boldsymbol{0}$



Intuition:
- Nominal images (left) exhibit similar labeling difficulty
- Special/outlier images (right) display a range of labeling challenges

T. Nguyen, S. Ibrahim, and X. Fu. ``Noisy Label Learning with Instance-Dependent Outliers: Identifiability via Crowd Wisdom", NeurIPS, 2024.

Outlier-robust E2E crowdsourced learning

- **One annotator insufficient to detect instance-dependent outliers** or learning the ground-truth label classifier [Nguyen et al'24]

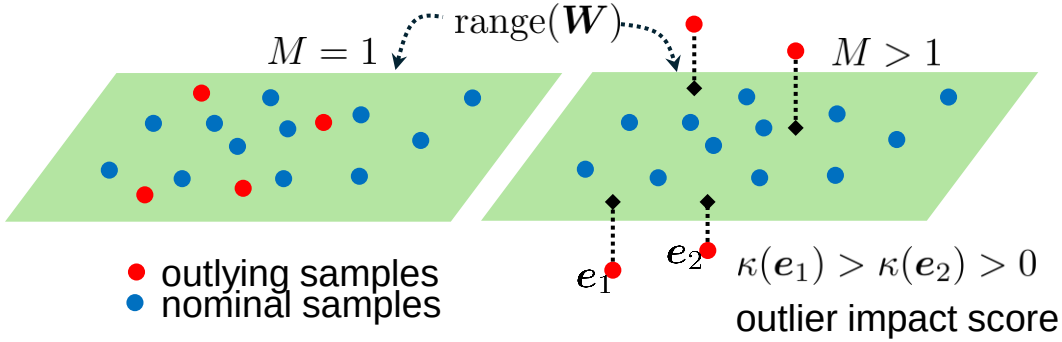
$$\underset{\{\boldsymbol{A}_m\in\mathcal{A}\},\{\boldsymbol{e}_n^{(m)}\in\mathcal{E}\},\boldsymbol{f}\in\mathcal{F}}{\text{minimize}} \ \frac{1}{S}\sum_{(m,n)\in\mathcal{S}} \mathsf{CE}(\boldsymbol{A}_m\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_n)+\boldsymbol{e}_n^{(m)},\breve{y}_n^{(m)}),$$

$$\text{subject to} \ \sum_{n=1}^{N}\mathbb{I}\left\{\sum_{m=1}^{M}\|\boldsymbol{e}_n^{(m)}\|_2>0\right\}\leq C,$$

Outlier no. estimate (instance-dependent label noise samples)

Wisdom of the Crowd helps in instance-dependent label noise learning!!

$$\boldsymbol{W}=[(\boldsymbol{A}_1)^\top,\ldots,(\boldsymbol{A}_M)^\top]^\top, \qquad \boldsymbol{e}_n=[(\boldsymbol{e}_n^{(1)})^\top,\ldots,(\boldsymbol{e}_n^{(M)})^\top]^\top$$



Identifiability under Instance-dependent Label Noise

Theorem 3.5 [Nguyen et al., 2024] (Informal) Let $(\{\widehat{\boldsymbol{A}}_m\}, \{\widehat{\boldsymbol{e}}_n^{(m)}\}, \widehat{\boldsymbol{f}})$ be any optimal solution of (1). The following result holds with probability greater than $1 - 2/S - K/T^{\alpha}$:

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mathcal{D}_{\boldsymbol{x}}}\left[\min_{\boldsymbol{\Pi}} \|\widehat{\boldsymbol{f}}(\boldsymbol{x}) - \boldsymbol{\Pi}^{\top}\boldsymbol{f}(\boldsymbol{x})\|_2^2\right] \leq K(\eta + \xi_1 + \xi_2),$$

$$\min_{\boldsymbol{\Pi}} \|\widehat{\boldsymbol{A}}_m - \boldsymbol{A}_m\boldsymbol{\Pi}\|_{\mathrm{F}}^2 = K\sigma^2(\eta + \xi_1 + \xi_2),\ \forall m,$$

where $\eta^2 = \mathcal{O}\left(\beta M T^{\alpha}/\sqrt{S}\left(\sqrt{M\log S} + (\|\boldsymbol{X}\|_2\mathscr{R}_{\mathcal{F}})^{0.25}\right)\right)$, $\boldsymbol{\Pi}$ a permutation matrix, $\boldsymbol{X}$ is defined as before, and $T = N - |\mathcal{I}|$. In addition, we have exact outlier detection, i.e., $\widehat{\mathcal{I}} = \mathcal{I}$

Outlier-robust E2E crowdsourced learning

Robust NMF

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

$$\boldsymbol{P} = \boldsymbol{W}\boldsymbol{F} + \boldsymbol{E},$$

- Identifiability via
  - Anchor point condition
  - Class expert condition
  - Scarce outliers – lying outside the simplex spanned by $\boldsymbol{W}$

Empirical results

Implementation: COINNet

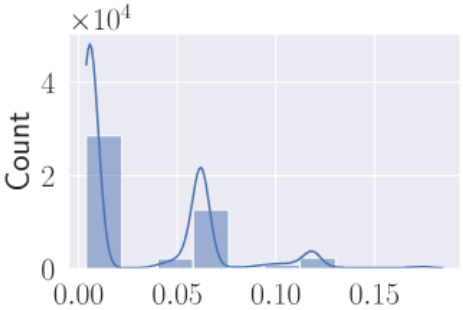
$$\underset{\boldsymbol{A}_m \in \mathcal{A}, \boldsymbol{e}_n^{(m)} \in \mathcal{E}, \boldsymbol{f} \in \mathcal{F}}{\text{minimize}} \quad \mathsf{L}_{\text{ce}} + \mu_1 \mathsf{L}_{\text{outlier}} + \mu_2 \mathsf{L}_{\text{vol}}$$

$$\mathsf{L}_{\text{CE}} = \frac{1}{S} \sum_{(m,n) \in \mathcal{S}} \text{CE}(\boldsymbol{A}_m \boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_n) + \boldsymbol{e}_n^{(m)}, \breve{\boldsymbol{y}}_n^{(m)})$$

$$\mathsf{L}_{\text{outlier}} = \sum_{n=1}^{N} (\sum_{m=1}^{M} \|\boldsymbol{e}_n^{(m)}\|_2^2 + \zeta)^{\frac{p}{2}}$$

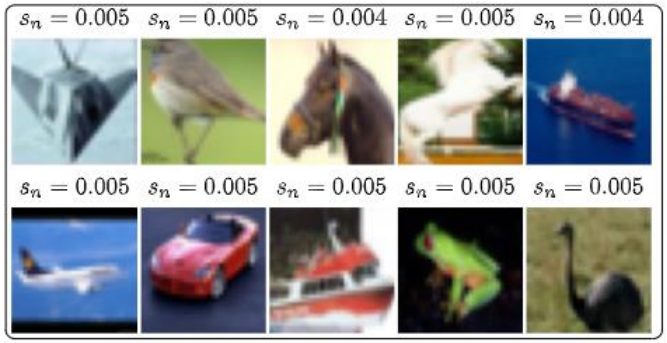
$$\mathsf{L}_{\text{vol}} = -\log\det(\boldsymbol{F}\boldsymbol{F}^\top)$$

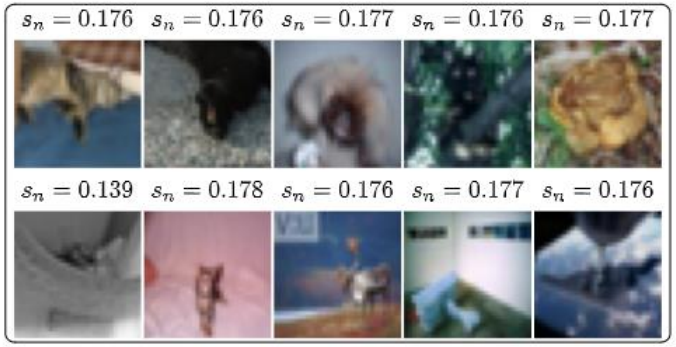
Dataset: CIFAR10-N [Wei, et al., 2022];
$N = 60000, K = 10, M = 3$;
Annotator error rates: 17.23%, 18.12%, 17.64%.



Histogram of $s_n = \sum_{m=1}^{M} \|\widehat{\boldsymbol{e}}_n^{(m)}\|_2^2$

Images with high score values show more instance-dependent confusion characteristics (such as background noise and blurring)

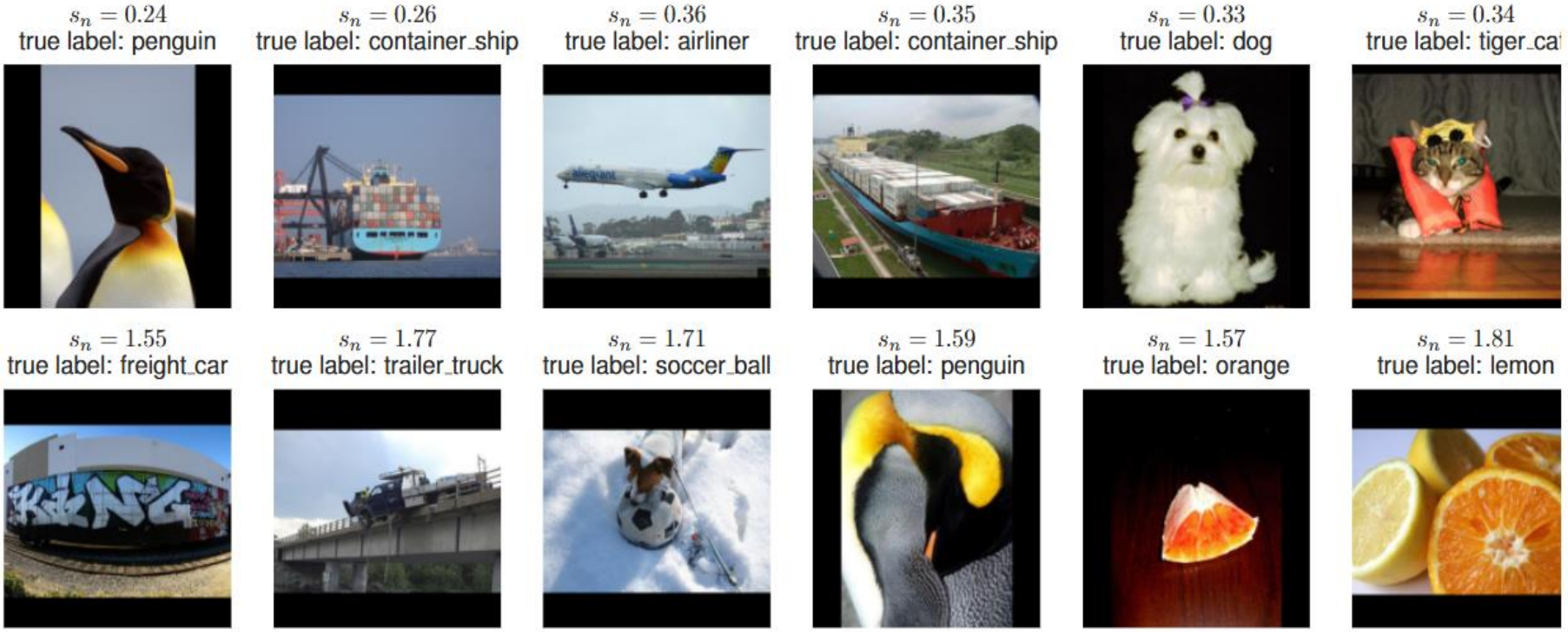




Empirical results

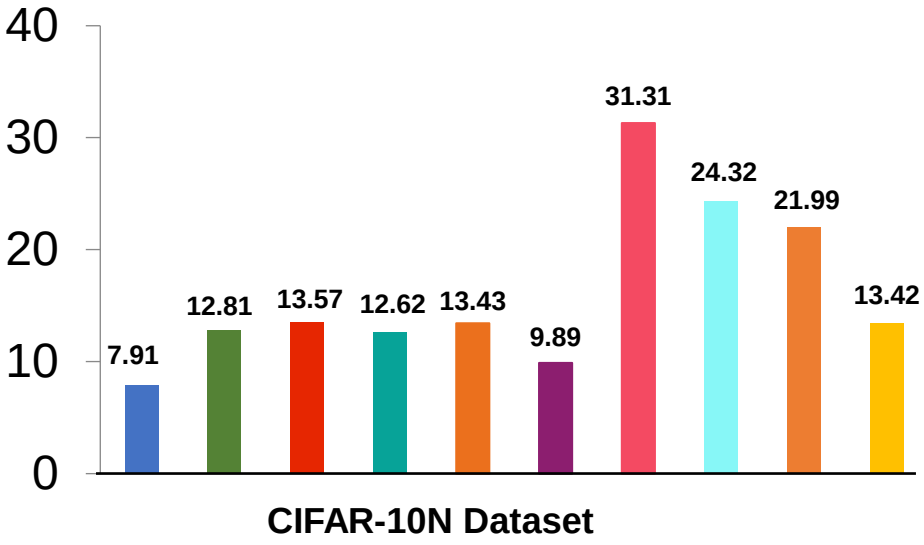
Dataset: ImageNet-15N [Nguyen et al'24]; annotations collected by AMT workers annotating images from ImageNet;
Average annotator error rate: 42.68%.

$$N = 2514, K = 15, M = 100;$$

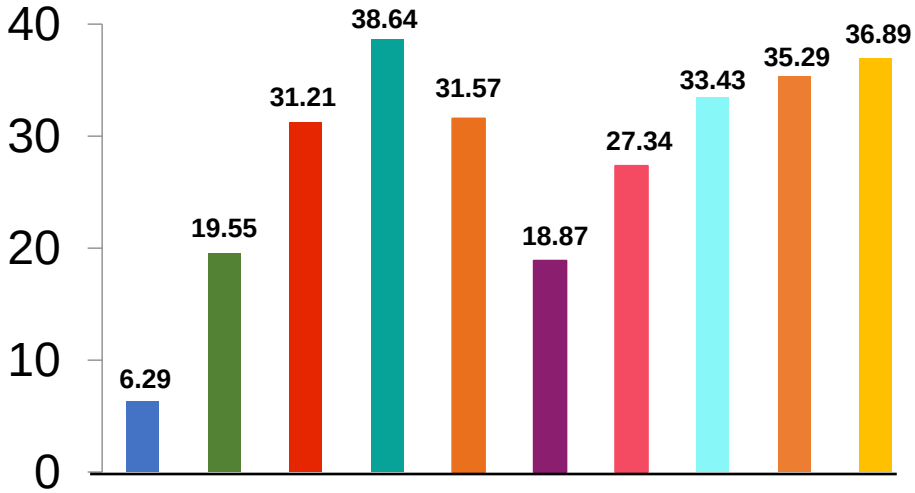


T. Nguyen, S. Ibrahim, and X. Fu. ``Noisy Label Learning with Instance-Dependent Outliers: Identifiability via Crowd Wisdom", NeurIPS, 2024.

Empirical results



CIFAR-10N Dataset



ImageNet-15N Dataset

Noisy labels from amazon mechanical turk workers

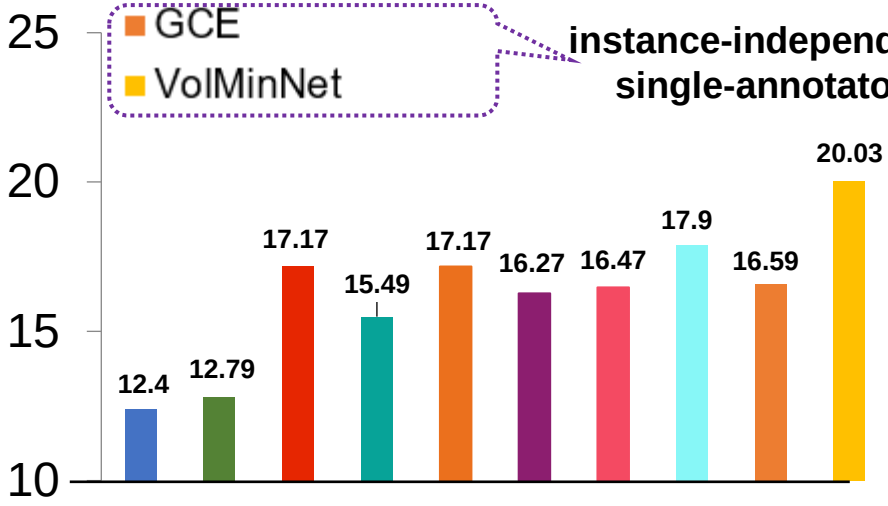
- COINNet
- GeoCrowdNet (F)
- GeoCrowdNet (W)
- CrowdLayer
- TraceReg
- MaxMIG
- MEIDTM
- BLTM
- GCE
- VolMinNet

instance-dependent crowdsourcing

instance-independent crowdsourcing

instance-dependent single-annotator

instance-independent single-annotator



LabelMe Dataset

Outline

- Motivation and problem statement
- Part I: Combining crowdsourced labels
- Part II: End-to-end (E2E) learning with crowdsourced labels
- Part III: Other aspects of crowdsourcing
  - Other types of annotations
  - Adversary aware crowdsourcing
  - Active crowdsourcing
  - Bias and fairness issues
- Conclusions and open issues

Crowdsourced regression

- Ground-truth labels and annotations take continuous real values $y_n, \{\breve{y}_n^{(m)}\} \in \mathbb{R}$
  - e.g., bounding box or timestamp annotations
- Simple aggregation approach: **Averaging** $\widehat{y}_n = \frac{1}{|\mathcal{M}_n|} \sum_{m \in \mathcal{M}_n} \breve{y}_n^{(m)}$
- Gaussian priors for ground-truth labels and annotator responses

$$y_n \sim \mathcal{N}(\mu_n, \sigma_n^2) \qquad \breve{y}_n^{(m)} | y_n = \alpha \sim \mathcal{N}(\alpha, \sigma_m^2)$$

Annotator response noise

  - Solved via Bayesian iterative algorithm [Ok et al'19]
- Moment-based approach [Tenzer et al'22]
  - Sparse + low rank decomposition of annotator covariance matrix

$$[\boldsymbol{C}]_{i,j} = \mathbb{E}[(\breve{y}_n^{(i)} - \mu_i)(\breve{y}_n^{(j)} - \mu_j)], \ \mu_i = \mathbb{E}[\breve{y}_n^{(i)}]$$

- E2E approach – CrowdLayer for regression [Rodrigues et al'18]

Similarity annotations

- Annotators indicate the similarity of data items
  - E2E learning can be formulated as graph clustering

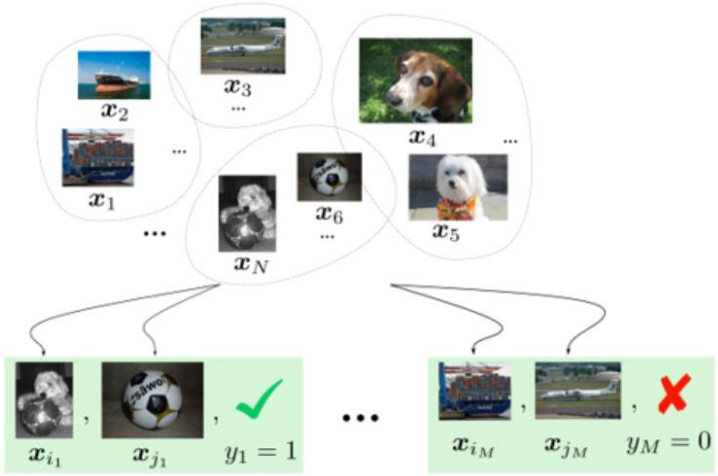
$\boldsymbol{G} \in \{0,1\}^{N \times N}$ Graph adjacency

$$\boldsymbol{G}(i,j) \sim \text{Bernoulli}\left(\boldsymbol{f}(\boldsymbol{x}_i)^\top \boldsymbol{f}(\boldsymbol{x}_j)\right)$$

modeling similarity

$$[\boldsymbol{f}(\boldsymbol{x}_n)]_k = \Pr(y_n = k | \boldsymbol{x}_n)$$

- Requires lower expertise level from annotators



Maximum likelihood Estimation (MLE) **Pairwise annotations**

$$\underset{\boldsymbol{f_\theta} \in \boldsymbol{\Delta}_K}{\text{minimize}} \sum_{(i,j) \in \boldsymbol{\Omega}} \left[\mathbb{I}[\boldsymbol{G}(i,j) = 1] \log\left(\boldsymbol{f_\theta}(\boldsymbol{x}_i)^\top \boldsymbol{f_\theta}(\boldsymbol{x}_j)\right) + \mathbb{I}[\boldsymbol{G}(i,j) = 0] \log\left(1 - \boldsymbol{f_\theta}(\boldsymbol{x}_i)^\top \boldsymbol{f_\theta}(\boldsymbol{x}_j)\right)\right]$$

- Identifiability of $\boldsymbol{f}$ established by treating model as quantized NMF problem
- Similarity annotations easily acquired by the crowd
- Often referred to as ***crowdclustering***

T. Nguyen, S. Ibrahim, and X. Fu "Deep Clustering with Incomplete Noisy Pairwise Annotations: A Geometric Regularization Approach", ICML 2023

Ranking

- Popular model for pairwise preference is the ***Bradley-Terry (BT)*** one:

$$\Pr(\boldsymbol{x}_n \succ \boldsymbol{x}_{n'}) = \frac{e^{s_n}}{e^{s_n} + e^{s_{n'}}},$$

ranking score: s_n

$\boldsymbol{x}_n \succ \boldsymbol{x}_{n'}$ ➡ $\boldsymbol{x}_n$ is preferred over $\boldsymbol{x}_{n'}$

- Annotator-specific preference can be modeled as [Chen et al'13]

$$\Pr(\boldsymbol{x}_n \succ_m \boldsymbol{x}_{n'}) = \Pr(\boldsymbol{x}_n \succ_m \boldsymbol{x}_{n'} | \boldsymbol{x}_n \succ \boldsymbol{x}_{n'}) \Pr(\boldsymbol{x}_n \succ \boldsymbol{x}_{n'}) + \Pr(\boldsymbol{x}_n \succ_m \boldsymbol{x}_{n'} | \boldsymbol{x}_n \prec \boldsymbol{x}_{n'}) \Pr(\boldsymbol{x}_n \prec \boldsymbol{x}_{n'})$$

ω_m $\quad$ $\frac{e^{s_n}}{e^{s_n}+e^{s_{n'}}}$

  - MLE-based optimization to jointly learn both ω_m and s_n
  - In E2E learning, s_n represented using a NN $\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_n)$ [Chhan et al'24]

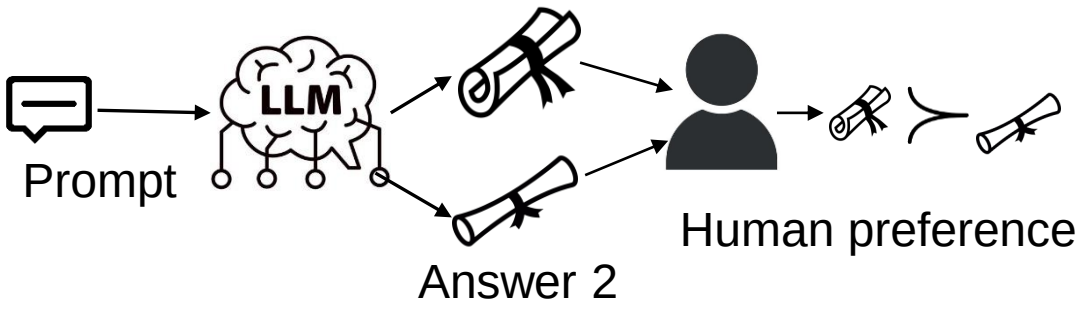
- Ordinal ranking: convert preference order into set of binary labels [Raykar et al'10]

$$[\breve{\boldsymbol{y}}_m^{(m)}]_k = \begin{cases} 1 & \text{if } \breve{y}_n^{(m)} > k \\ 0 & \text{otherwise} \end{cases}, \; k = 1, \ldots, K-1$$

$K-1$ binary labels

Preference annotations

- LLM alignment relies on human preference annotations
  - Reinforcement Learning from Human Feedback (RLHF)
  - Direct Preference Optimization (DPO)



- Single annotator setting: Robust DPO model-preference noise modeled as [Chowdhury et al'24]

$$\epsilon := \Pr(\boldsymbol{x}_n \widehat{\succ} \boldsymbol{x}_{n'} | \boldsymbol{x}_n \succ \boldsymbol{x}_{n'}))$$

Preference flipping rate

- RLHF with **multiple annotators:** Weighted MV [Li et al'24]

➢ Alternative: preferences integrated via mixture of reward models [Chakraborty et al'24]

LLM alignment using crowd feedback remains an active research area with exciting open directions

Outline

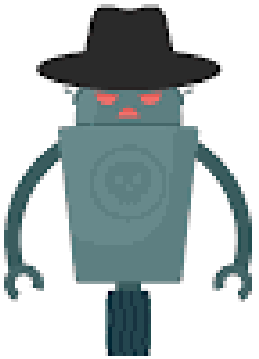
- Motivation and problem statement
- Part I: Combining crowdsourced labels
- Part II: End-to-end (E2E) learning with crowdsourced labels
- Part III: Other aspects of crowdsourcing
  - Other types of annotations
  - Adversary aware crowdsourcing
  - Active crowdsourcing
  - Bias and fairness issues
- Conclusions and open issues

Crowdsourcing under attack

- **Adversarial attacks**
  - Adversaries may hide as legitimate annotators
  - Adversaries manipulate labels to hamper performance or drain resources
  - Adversaries poison crowdsourcing datasets

Q1. Which are the worst adversarial attacks?

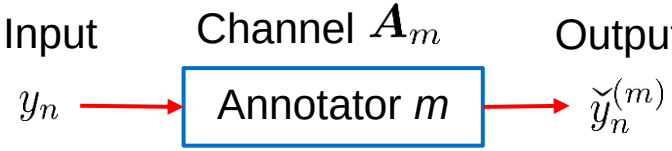
Q2. Can we identify such attacks and discard corresponding annotators?



- Two types of adversaries considered: **spammers** and **colluding adversaries**

Information-theoretic characterization of spammers

Idea. View annotators as independent information-bearing channels



- Per annotator *m*, performance can be benchmarked by channel capacity

$$C^{(m)} := \max_{\boldsymbol{\pi}} I(y_n, \breve{y}_n^{(m)}) \geq 0$$

  - Overall capacity $C = \sum_{m=1}^{M} C^{(m)}$

- Worst spammer annotator: $C^{(m)} = 0$, output not related to input y_n

$$\boldsymbol{A}_m(k', k) := \Pr(\breve{y}_n^{(m)} = k' | y_n = k) = \Pr(\breve{y}_n^{(m)} = k') := s_{k'}^{(m)}$$

Irrelevant to y_n, $\Rightarrow C^{(m)} = 0$

  - Zero-capacity confusion matrix $\boldsymbol{A}_m = \boldsymbol{s}^{(m)} \mathbf{1}^\top$

- Two groups of annotators
  - Spammers $m \in \mathcal{S}$ to be removed
  - Honest $m \in \mathcal{H}$ follow DS annotator model

- Spammer detection while fusing labels via modified EM [Raykar-Yu'12]

P. A. Traganitis and G. B. Giannakis, "Identifying spammers to boost crowdsourced classification," *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, Toronto, Canada, June 2021.

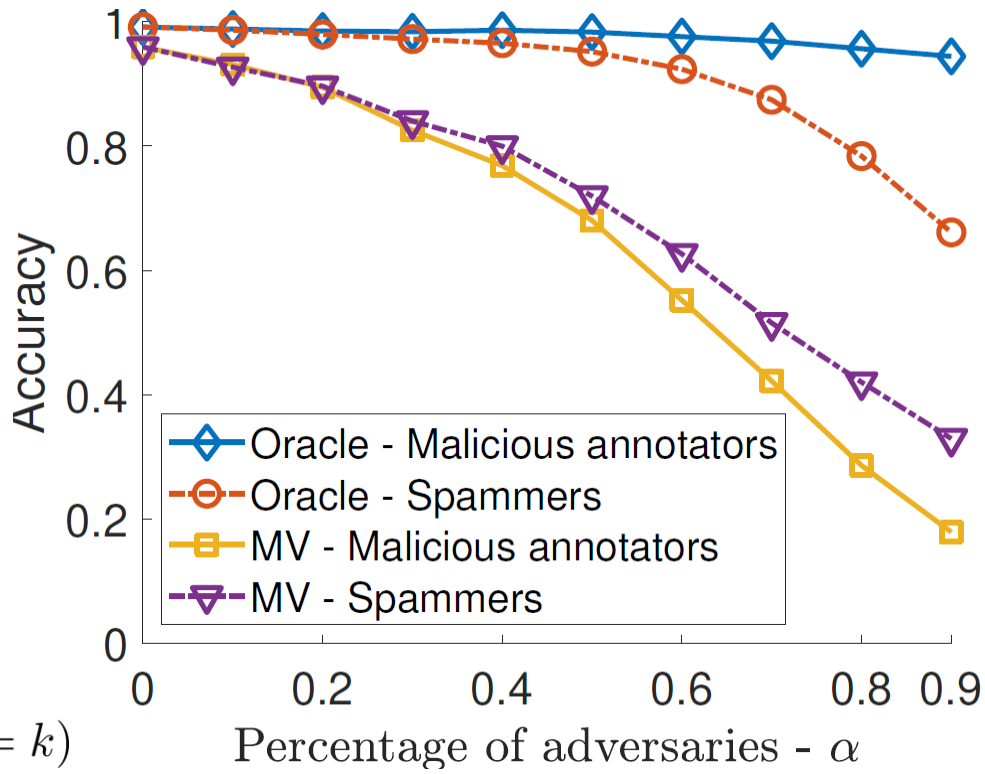
Testing the effect of spammers

- Synthetic dataset, *N=10,000*, *K = 4*, *M* = 20
  - *αM* annotators as adversaries, *(1-α)M* honest
  - Oracle: classifier knows $\{\boldsymbol{A}_m\}_{m=1}^{M}, \boldsymbol{d}$
  - MV: majority voting
- Adversaries
  - Spammers give random responses

$$\Pr(\breve{y}_n^{(m)} = k' | y_n = k) = \Pr(\breve{y}_n^{(m)} = k')$$

  - Malicious annotators $k' \neq k$

$$\Pr(\breve{y}_n^{(m)} = k' | y_n = k) > \Pr(\breve{y}_n^{(m)} = k | y_n = k)$$



- Spammers challenge oracle approaches; we can “flip” malicious responses

Flagging spammers with cross-covariances

- Covariance of annotators $m \in \mathcal{H} \cup \mathcal{S}$ and $m' \in \mathcal{S}$

$$c_{m,m'} := \mathbf{E}[\breve{y}_n^{(m)} \breve{y}_n^{(m')}] - \tilde{\mu}_m \tilde{\mu}_{m'} = \mathbf{E}[\breve{y}_n^{(m)}]\mathbf{E}[\breve{y}_n^{(m')}] - \tilde{\mu}_m \tilde{\mu}_{m'} = 0$$

Can be used to flag spammers

- Sample cross-covariance matrix has **low rank + diagonal** structure

$$\hat{\mathbf{C}} = \mathbf{L} + \mathbf{D} = \mathbf{V}\mathbf{V}^\top + \mathbf{D}, \quad \mathbf{V} \in \mathbb{R}^{M \times K}$$

  - Rows of ***V*** close to **0** indicate spammers

- Spammer identification algorithm

  S1. Recover ***V*** from sample cross-covariance matrix

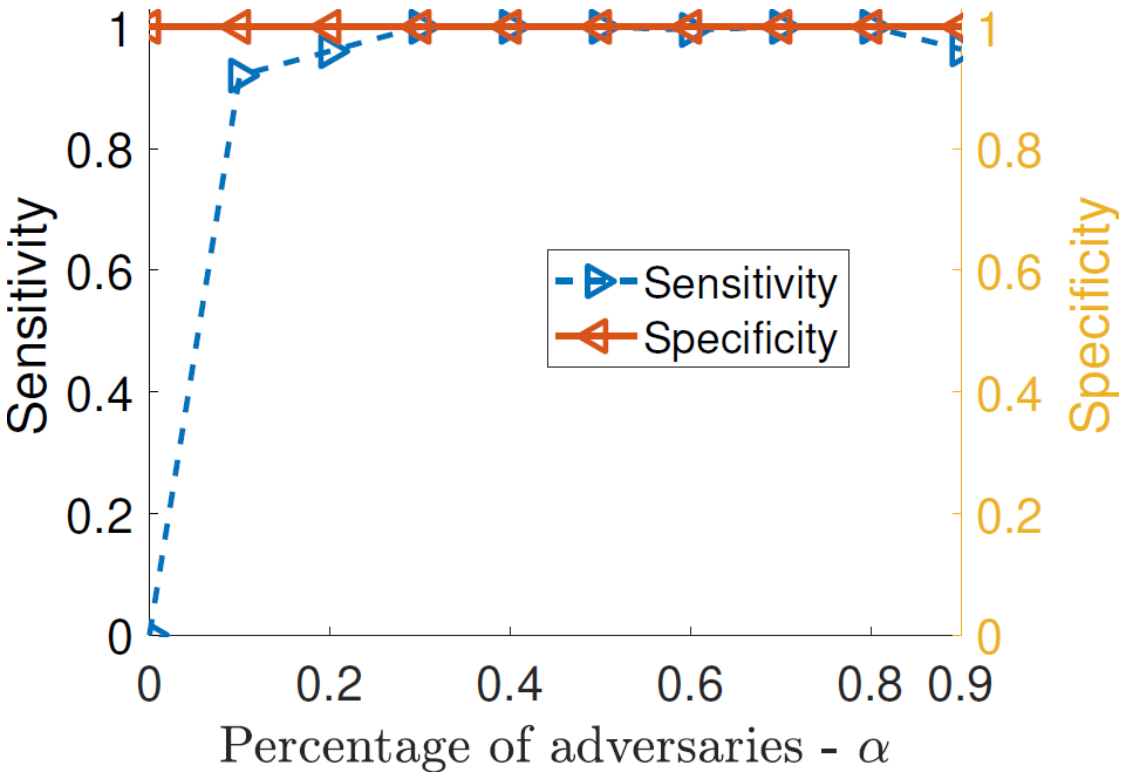
  S2. Cluster rows of ***V*** using e.g., *k*-means

  S3. Identify spammers as the cluster heads with norm closest to 0

P. A. Traganitis and G. B. Giannakis, "Identifying spammers to boost crowdsourced classification," *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, Toronto, Canada, June 2021.

Spammer detection performance

- Synthetic dataset with *N=10,000*, *K = 4*, *M* = 20
  - Figures of merit: Sensitivity (true positive rate) and Specificity (true negative rate)



Bottomline. A few spammers misclassified as honest; all honest classified as honest

Real crowdsourced data

- Bluebird dataset *N=108, K = 2, M=39*
- Dog dataset *N=807, K = 4, M=109*
- Web dataset *N=2,655, K = 5, M=177*

- Annotators deemed spammers were removed from dataset
- **Alg. 1:** Covariance-based spammer identification algorithm

Classification accuracy

Dataset	MV	EM	Alg. 1 + MV	Alg. 1 + EM
Bluebird	0.759	0.88	0.852(**22**)	0.899(**22**)
Dog	0.817	0.834	0.819(**12**)	0.834(**12**)
Web	0.776	0.871	0.841(**158**)	0.91(**158**)

Parentheses indicate number of pruned annotators

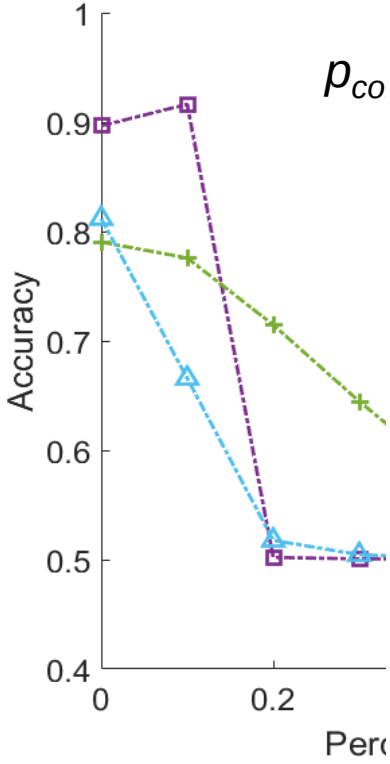
Bottomline. Identifying and removing spammers boosts crowdsourcing performance

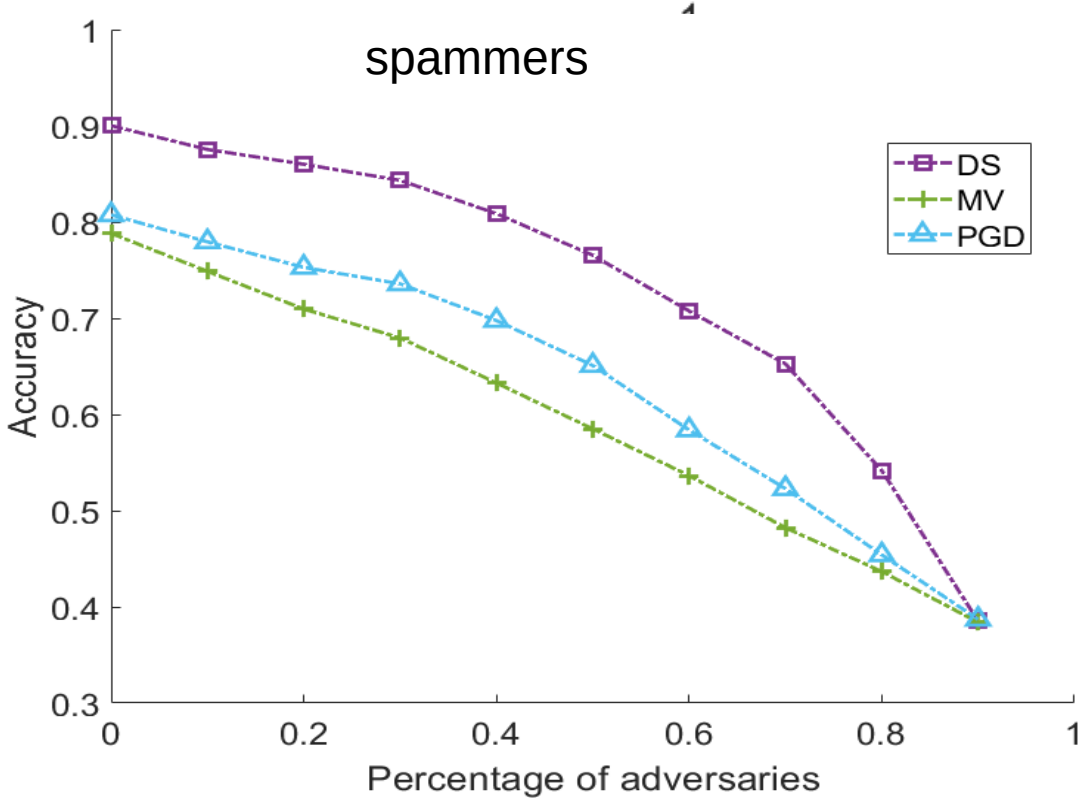
P. Welinder et al, "The multidimensional wisdom of crowds," *Proc. of NIPS*, pp. 2424–2432, 2010.
J. Deng et al, "Imagenet: A large-scale hierarchical image database," *Proc. of IEEE CVPR*, pp. 248–255, 2009.
D. Zhou et al, "Learning from the wisdom of crowds by minimax entropy," *Proc. of NIPS,* pp. 2195–2203, 2015.

Effect of colluding adversaries

Q. What if adversaries cooperate?

- Synthetic dataset with *N = 5,000, M=60, K=3.* Percentage of adversaries = *α*
  - Colluding adversaries provide wrong response w.p. p_{corr}, and ground-truth label w.p. *1* - p_{corr}.
    - PGD [Ma '18]: Moment matching for one-parameter model
  - Colluding adversaries hamper performance more severely than spammers





Colluding adversaries

- Adversaries cooperate to change labels of some data
  - Model mismatch as DS model no longer applies!
- Two groups of annotators
  - Adversaries deviate from presumed model $m \in \mathcal{A},\ \mathrm{card}(\mathcal{A}) = M_{\mathcal{A}}$
  - Honest follow presumed model $m \in \mathcal{H},\ \mathrm{card}(\mathcal{H}) = M_{\mathcal{H}}$

Q. How can colluding adversaries be identified?

- Adversaries alter characteristics associated with the expected model
  - Arbitrary adversaries under one-coin model, not exceeding 50% of annotators [Jagabathula et al'17, Kleindessner-Awasthi'18, Ma-Olshevsky'20]
  - Arbitrary adversaries under DS model [Traganitis-Giannakis'21]
  - Graph-based detection of adversaries and attacked data [Karaaslanli et al'25]

This ICASSP

- E2E learning
  - Honest annotators – corrupted data $\boldsymbol{x}_n$ [Chen et al'22]

Annotator agreement matrix

- Agreement rate of honest annotators *m, m'*

$$\sigma_{m,m'} := \Pr\left(\breve{y}_n^{(m)} = \breve{y}_n^{(m')}\right) = \mathrm{tr}\left(\boldsymbol{A}_m \boldsymbol{D} \boldsymbol{A}_{m'}{}^\top\right) = \boldsymbol{u}^{(m)\top} \boldsymbol{u}^{(m')}$$

$$\boldsymbol{u}^{(m)} := \mathrm{vec}(\boldsymbol{D}^{1/2} \boldsymbol{A}_m{}^\top) : K^2 \times 1$$

$$\sigma_{m,m} := \Pr(\breve{y}_n^{(m)} = \breve{y}_n^{(m)}) = 1$$

- Annotator agreement matrix $[\boldsymbol{\Sigma}]_{m,m'} = \sigma_{m,m'} : M \times M$

(as) $M_{\mathcal{H}} > K^2$

- W.l.o.g. First $M_{\mathcal{H}}$ annotators are honest and the rest adversarial

$$\boldsymbol{\Sigma} = \mathbf{C} + \mathbf{I} = \left[\begin{array}{c|c} \mathbf{C}_{\mathcal{H}} & \mathbf{C}_{\mathcal{H},\mathcal{A}} \\ \hline \mathbf{C}_{\mathcal{A},\mathcal{H}} & \mathbf{C}_{\mathcal{A}} \end{array}\right] + \left[\begin{array}{cc} \mathbf{I}_{\mathcal{H}} & \\ & \mathbf{I}_{\mathcal{A}} \end{array}\right]$$

unknown

$$\mathrm{rank}\left(\begin{bmatrix} \mathbf{C}_{\mathcal{H}} \\ \mathbf{C}_{\mathcal{A},\mathcal{H}} \end{bmatrix}\right) \leq K^2$$

Low rank + Diagonal

P. A. Traganitis and G. B. Giannakis, "Detecting Adversaries in Crowdsourcing," *Proc. of IEEE International Conference on Data Mining*, 2021.

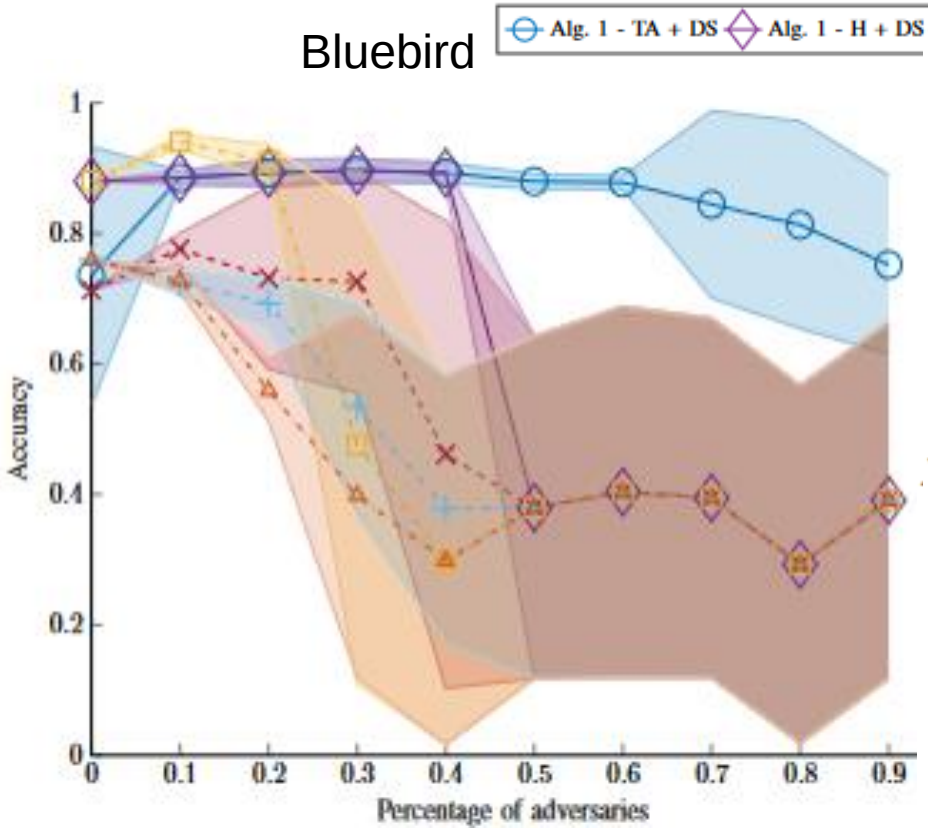
Real data tests

Dataset	Bluebird	Sentence Polarity	Dog	Web
N	108	5,000	807	2,665
M	39	203	109	177
K	2	2	5	5

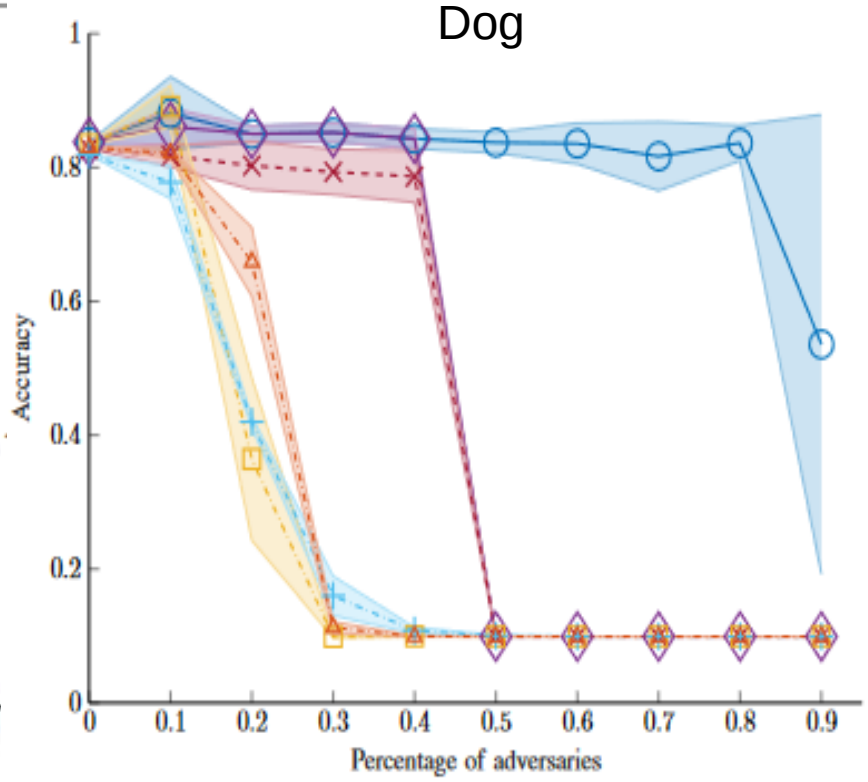
- MMSR: State-of-the-art adversary aware method under one-coin model
- Alg. 1 – TA: [Traganitis-Giannakis'21] with knowledge of one trusted annotator
- Alg. 1 – H: [Traganitis-Giannakis'21] assuming at least 50% of annotators are honest

$p_{corr} = 0.9$

Bluebird



Dog

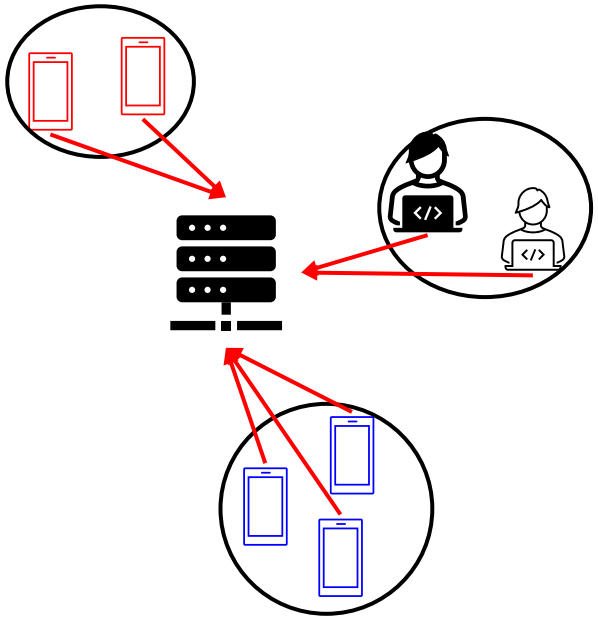


Crowdsourcing with dependent annotators

- So far, annotators were considered **independent**
  - Not always the case: nearby sensors; collaboration; similar devices …

Q1. Can label integration methods be broadened to deal with dependent annotators?

Q2. Can we identify this dependence structure?

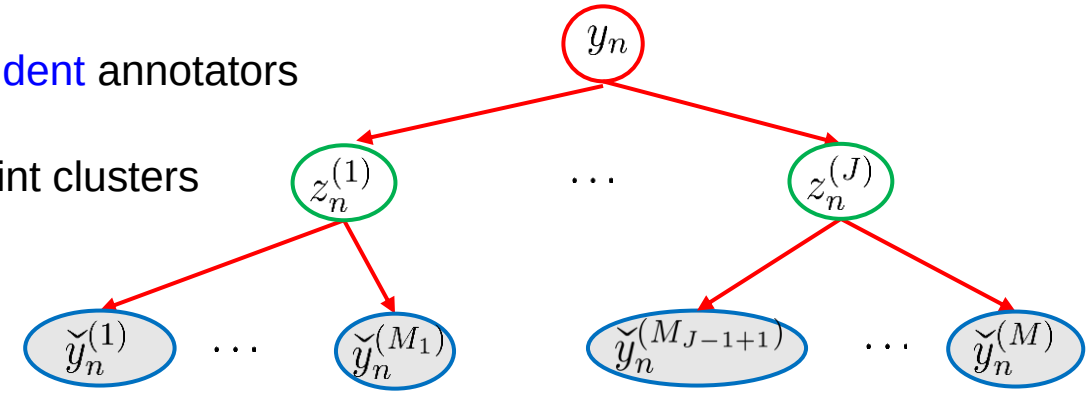


- Two types of approaches:
  - EM and Bayesian methods
  - Moment-based methods

DS model for dependent classifiers

- DS model does not hold for dependent annotators
- Annotators grouped in $J < M$ disjoint clusters
  - Indices collected in $\{\mathcal{C}_j\}_{j=1}^{J}$

$$M_j = |\mathcal{C}_j|$$



Dependencies captured by a hidden variable $z_n^{(j)} \in \{1, \ldots, K\}$ per group

(as) Given $z_n^{(j)}$ annotator responses, $\{\breve{y}_n^{(m)}\}_{m \in \mathcal{C}_j}$ are independent

(as) Given y_n hidden variables $\{z_n^{(j)}\}_{j=1}^{J}$ are independent

- Confusion matrices for annotators and hidden variables

$$\Xi^{(j)}(k, k') = \Pr\left(z_n^{(j)} = k | y_n = k'\right) \qquad \tilde{\boldsymbol{A}}_m(k, k') = \Pr\left(\breve{y}_n^{(m)} = k | z_n^{(j)} = k'\right), m \in \mathcal{C}_j$$

- Hierarchical label fusion

Caveat. Group indices are unknown, and must be estimated!

A. Jaffe, E. Fetaya, B. Nadler, T. Jiang, and Y. Kluger, "Unsupervised ensemble learning with dependent classifiers," *Proc. of AISTATS*, 2016.
H. Chen, B. Chen, and P. K. Varshney, "A new framework for distributed detection with conditionally dependent observations," *IEEE Trans. on Signal Processing*, vol. 60, no. 3, pp. 1409–1419, March 2012.

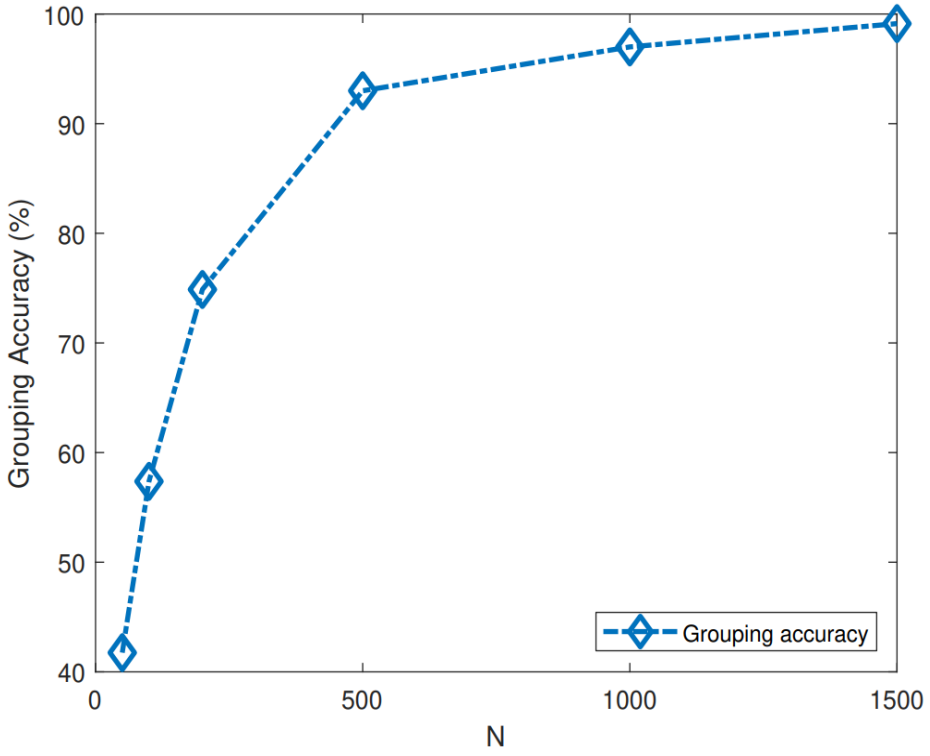
Identifying annotator dependencies

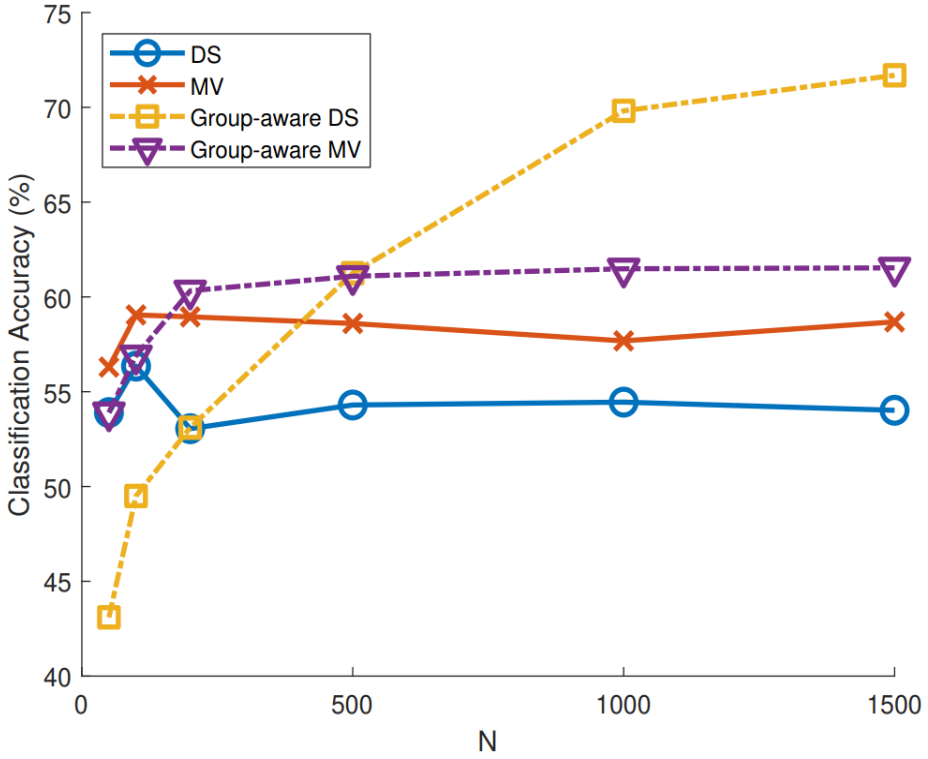
- Spectral methods for binary and general DS model [Jaffe et al'16,Traganitis-Giannakis'22]
  - Annotator clusters identified from second-order statistics
- Worker-cluster model [Imamura et al'18]
  - *J* distinct confusion matrices
- Bayesian methods – joint group and label inference
  - Community Bayesian Classifier Combination [Venanzi et al'14]
  - Bayesian nonparametrics [Moreno et al'15]

P. A. Traganitis, G. B. Giannakis, "Identifying Dependent Annotators in Crowdsourcing," *Proc. Asilomar Conf.*, 2022.

Synthetic tests

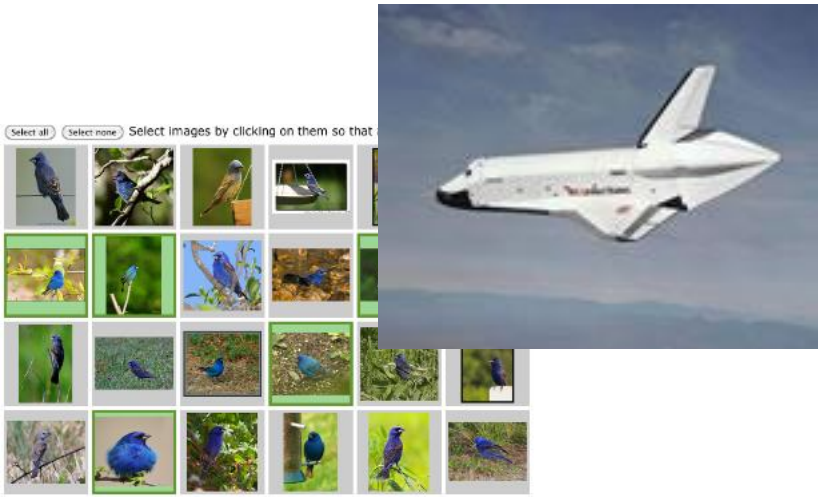
- Synthetic dataset: *N = 5,000, M=40, K=3, J = 5*
- DS: Expectation maximization
  MV: Majority voting
- Group-aware DS (MV): Proposed algorithm + DS (MV)





Real data tests

- Bluebird, *N = 108, M = 39, K = 2*
- Shuttle, *N = 58,000, M = 15, K = 7*
- CoverType, *N = 581,012 , M = 15 , K = 7*



Classification accuracy

Dataset	MV	DS	Group-aware MV	Group-aware DS
Bluebird	75.92%	87.96%	72.22% ***(J = 7)***	91.6% ***(J = 7)***
Shuttle	98.82%	97.54%	98.93% ***(J = 3)***	99.46% ***(J = 3)***
CoverType	75%	48.46%	74.49% ***(J = 4)***	74.09% ***(J = 4)***

P. A. Traganitis, G. B. Giannakis, "Identifying Dependent Annotators in Crowdsourcing," *Proc. Asilomar Conf.*, 2022.

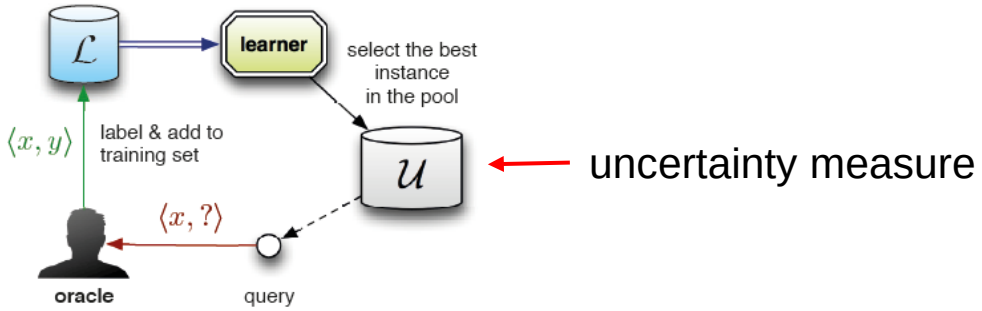
Outline

- Motivation and problem statement
- Part I: Combining crowdsourced labels
- Part II: End-to-end (E2E) learning with crowdsourced labels
- Part III: Other aspects of crowdsourcing
  - Other types of annotations
  - Adversary aware crowdsourcing
  - Active crowdsourcing
  - Bias and fairness issues
- Conclusions and open issues

Learning with label budget constraints

General Goal: For a given budget of queries, maximize accuracy by actively selecting which instances (data) to label ("query").

- Standard machine learning approach: Active Learning (AL)
  - allows for sample (label) complexity reduction
- Data selection – uncertainty sampling
  - Select the most uncertain instance according to the learner's model



  - Crowdsourcing requires choosing which annotators to query

B. Settles, "Active learning," Synthesis Lectures on Artif. Intel. and Machine Learning, vol. 6, no. 1, pp. 1–114, 2012.

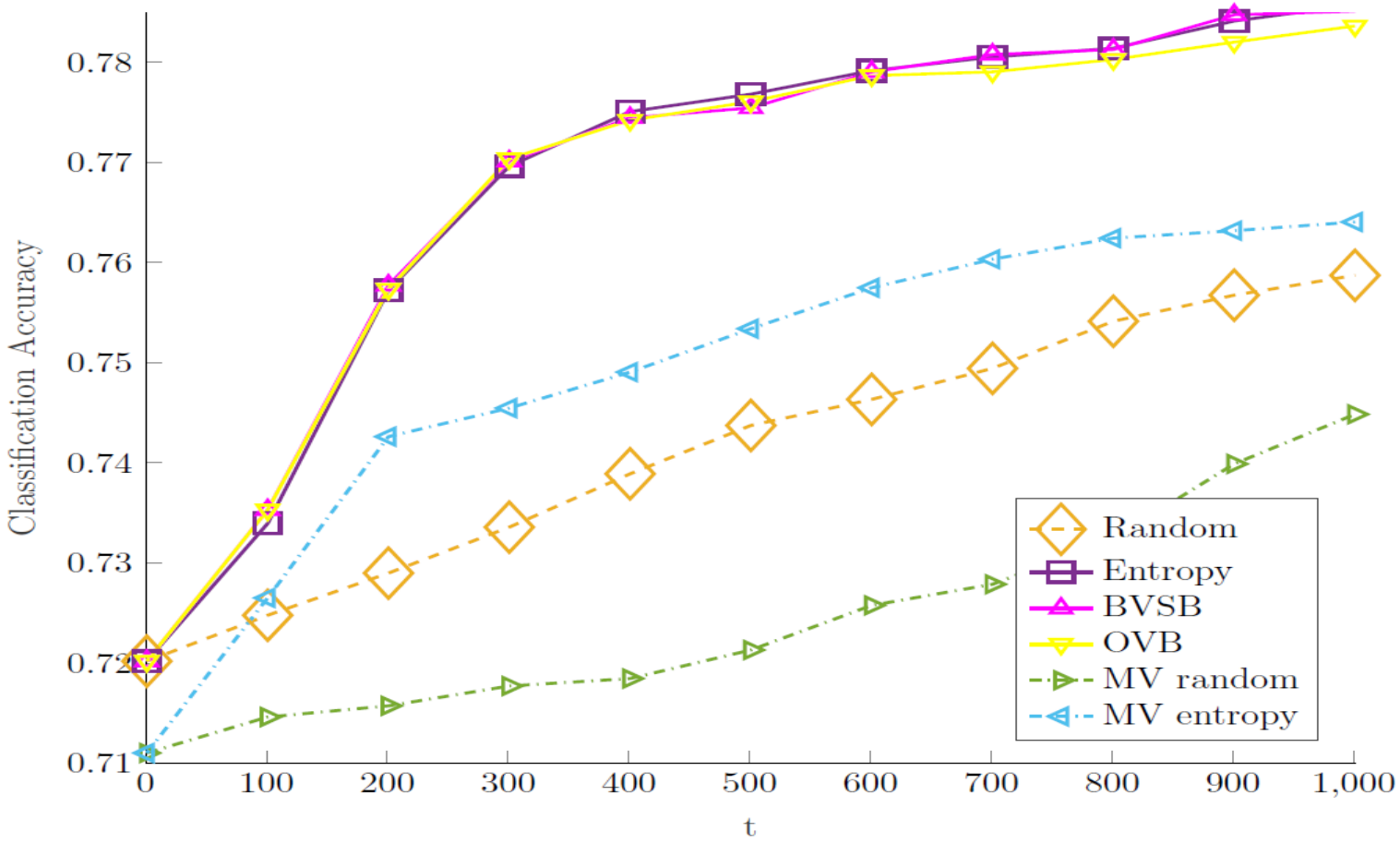
Active crowdsourcing

- Data selection
  - Select the most uncertain instance according to “crowd”
  - Uncertainty quantified by current label posterior $\Pr(y_n|\check{\mathcal{Y}})$
- Annotator selection
  - Select “best” given current annotator reliabilities
  - Confusion matrix estimates may be unreliable – **exploration** required
- Online or streaming algorithms preferable!
  - Parametric & Gaussian Process models [Welinder et al ’10, Yan et al ’11, Rodrigues et al ‘14]
  - AL w/ batch algorithms [Venanzi et al ‘15]
  - Online EM [Traganitis et al ‘20]
  - Multi-armed bandit approach [Rangi-Francescetti‘18]

P. A. Traganitis, D. Berberidis, and G. B. Giannakis, “Active learning with unsupervised ensembles of classifiers,” in Proceedings of the 2020 IEEE ICASSP, pp. 3967–3971, 2020.

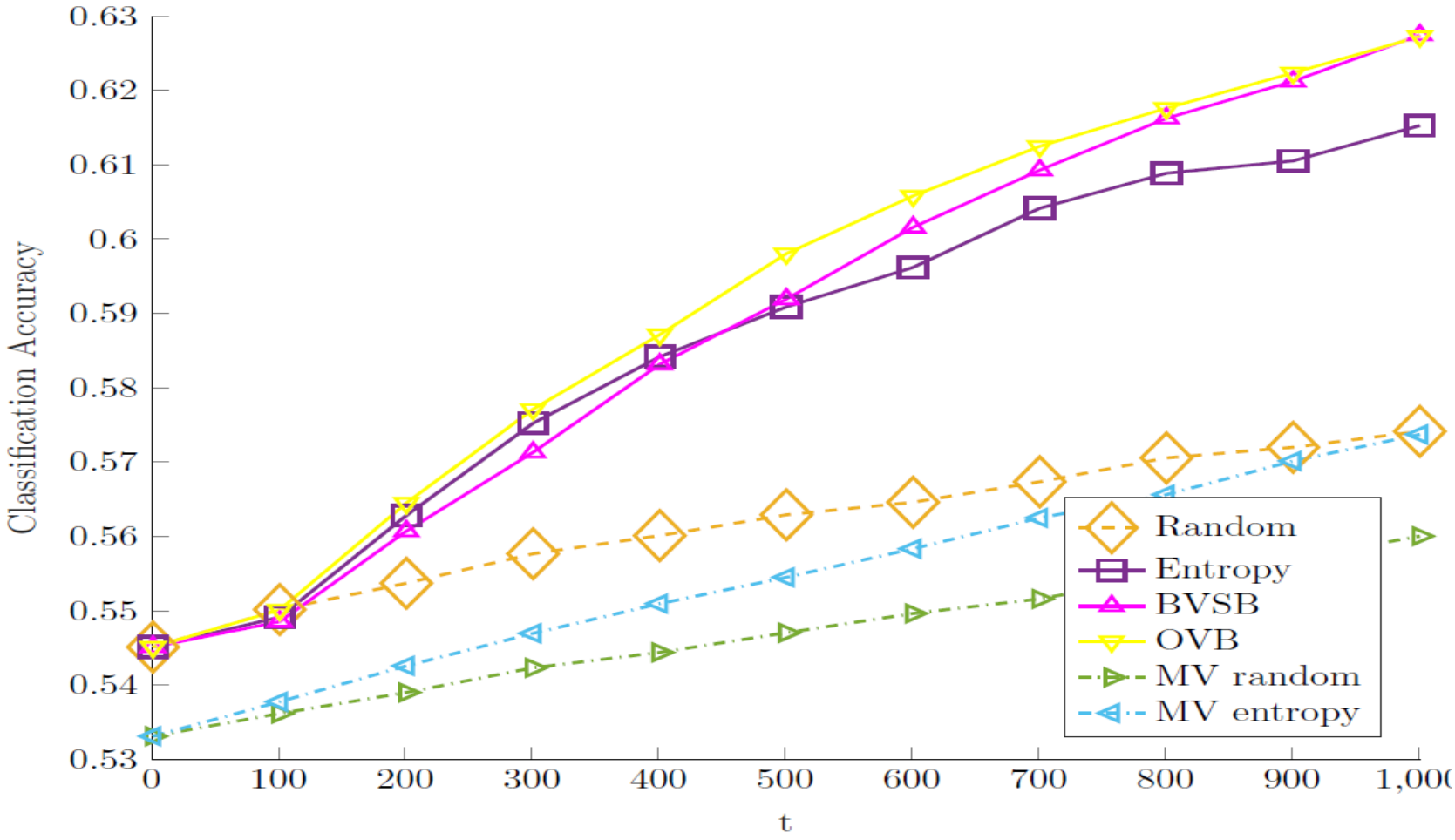
Numerical tests: Dog dataset

- $N = 807, M = 109, K = 4$ $\quad \lambda_t = t^{-0.6}$ $\quad \varepsilon = \begin{cases} 0.5 & t = 1, \ldots, 100 \\ 0.05 & t = 101, \ldots, 1000 \end{cases}$

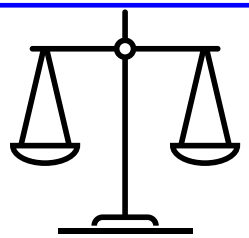


Web dataset

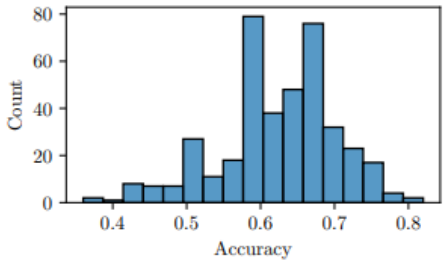
- $N = 2,665, M = 155, K = 5$



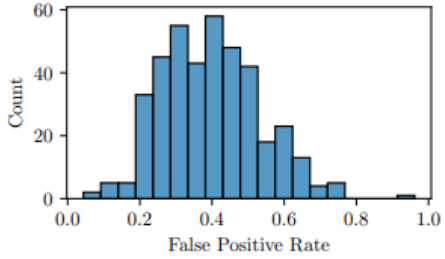
Bias and fairness in crowdsourcing



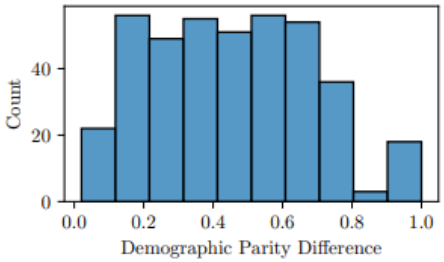
- ML/AI applied in areas w/ societal impact
  - Loans, credit score, penal system, job applications etc.
- Typical performance metrics do not capture bias against sensitive attributes
  - Race, age, affiliation etc.
  - Fairness metrics required – Demographic parity, equal opportunity etc.
- Annotators also exhibit bias



(a) Accuracy



(b) False Positive Rate



(c) Demographic Parity

S. Lazier, S. Thirumuruganathan, and H. Anahideh, “Fairness and bias in truth discovery algorithms: An experimental analysis,” arXiv preprint arXiv:2304.12573, 2023.

Taking bias into account

- Crowdsourcing perpetuates annotator biases
  - Unfair annotators yield unfair ML models

Classifier	TD Algorithm	Delta Accuracy	Delta Demo Par. Diff
Random Forest	Majority Voting	6.04	0.06
Random Forest	Dawid-Skene	7.63	0.04
Random Forest	Learning from Crowd	7.54	0.04
Logistic Regression	Majority Voting	2.55	0.04
Logistic Regression	Dawid-Skene	2.77	0.05
Logistic Regression	Learning from Crowd	2.86	0.05

  - Label integration does not optimize for fairness

- Annotator behavior w.r.t. sensitive data attribute $z_n \in \{0, 1\}$ captured via

$$[\boldsymbol{A}_m(z_n = z)]_{k,k'} = \mathsf{Pr}(\breve{y}_n^{(m)} = k | y_n = k', z_n = z)$$

  - Similar to instance specific confusion matrices
  - Fair annotator assignment [Goel et al'19]

requires ground-truth labels to estimate confusion matrices

Outline

- Motivation and problem statement
- Part I: Combining crowdsourced labels
- Part II: End-to-end (E2E) learning with crowdsourced labels
- Part III: Other aspects of crowdsourcing
- Conclusions and open issues

Conclusions and open issues

Take home 1. Crowdsourcing learns from unreliable information sources

- Also assesses source reliability in learning tasks
- Two major paradigms: Two-stage and E2E
- ML/AI catalyst

Take home 2. Two-stage approaches require only annotator labels

- Available approaches tradeoff accuracy for simplicity
- Workhorse algorithms: EM or VI
- Moment-based algorithms for reliable initialization

Take home 3. E2E approaches simultaneously learn a model and annotator reliabilities

- Inputs: data features and annotator labels
- Improved performance compared to two-stage
- Flexible DL models

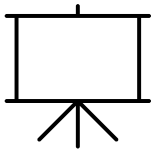
Open issues and future directions

- Application-specific crowdsourcing
  - Unique challenges and design
  - Collaboration of human and machine annotators
- Crowdsourcing in classical data/information fusion
  - Robust learning from multi-view data
  - Game-theoretic crowdsourcing
- Open research directions
  - Crowdsourcing and AI/LLMs
  - Bias and fairness
  - Transfer learning/meta learning

Relevant links

Learning From Crowdsourced Noisy Labels: A Signal Processing Perspective

https://arxiv.org/abs/2407.06902



LINK TO SLIDES

https://tinyurl.com/crowdslides

https://github.com/shahana-ibrahim/Learning-from-Crowdsourced-Noisy-Labels

Thank you!