ICML 2021

Adversarial Multiclass Learning under Weak Supervision with Performance Guarantees

A. Mazzetto*, C. Cousins*, D. Sam, S. Bach, E. Upfal
cyrus_cousins@brown.edu

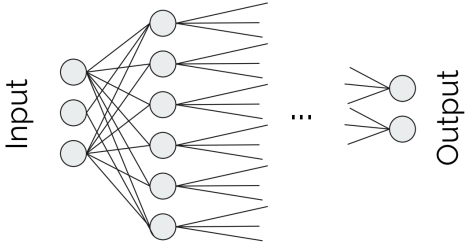
Brown University

18-24th July 2021

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu

Setting the Scene

Supervised Setting



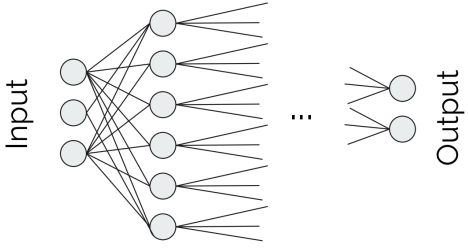
Large labeled set $\boldsymbol{X}_L, \boldsymbol{Y}_L$

Setting the Scene

ICML 2021
Adversarial Multiclass Learning under Weak Supervision with Performance Guarantees

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu

Supervised Setting



Large labeled set $\boldsymbol{X}_L, \boldsymbol{Y}_L$

Semi-Supervised Setting

Small labeled set $\boldsymbol{X}_L, \boldsymbol{Y}_L$ +
Large unlabeled set $\boldsymbol{X}_U$

Examples of Weak Labelers

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu

Examples of Weak Labelers

ICML 2021
Adversarial Multiclass Learning under Weak Supervision with Performance Guarantees

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu

Learn from Related Tasks:

 vs. 

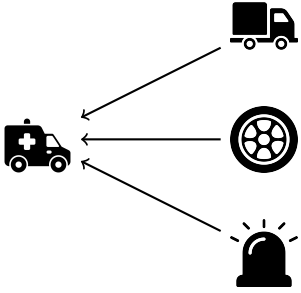
Examples of Weak Labelers

ICML 2021
Adversarial Multiclass Learning under Weak Supervision with Performance Guarantees

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu

Learn from Related Tasks:

 vs. 

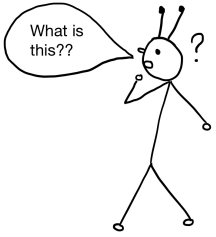


Images made by Those Icons (wheel, siren), fjstudio (ambulance), Freepik (truck) from flaticon.com

Contribution

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu

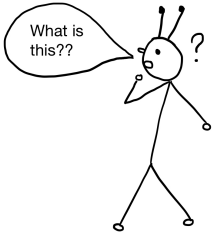
A dversarial
M ulti
C lass
L earning



A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu

Contribution

A dversarial
M ulti
C lass
L earning

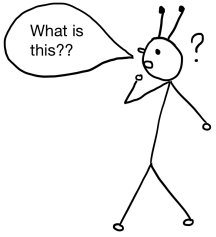


- **Formal framework** for *adversarial learning* from weak labelers
- **Optimization convergence guarantees** for the adversarial learning
- **Generalization bounds**

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu

Contribution

A dversarial
M ulti
C lass
L earning



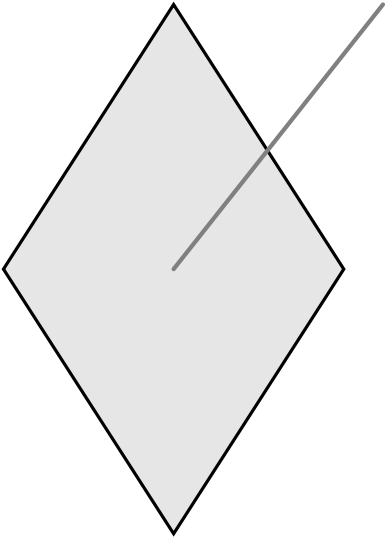
- **Formal framework** for *adversarial learning* from weak labelers
- **Optimization convergence guarantees** for the adversarial learning
- **Generalization bounds**

Relevant previous work:

1. A general framework for adversarial label learning. Arachie et al. JMLR 2021.
2. Optimal binary classifier aggregation for general losses. Balsubramani et al. NeurIPS 2016.

Feasibility of Labeling

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu

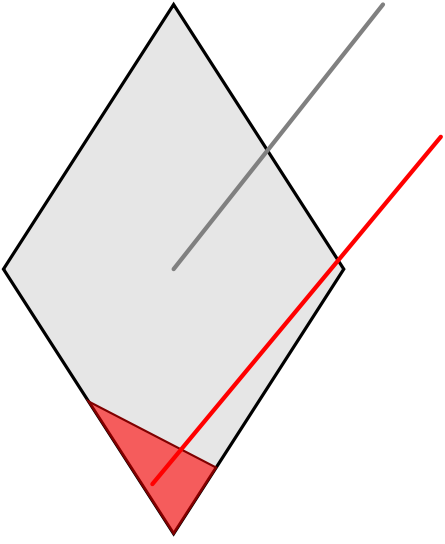


Labeling Space (of the unlabeled data) $\boldsymbol{X}_U$:
$\{\boldsymbol{Y} : \boldsymbol{Y}$ is a labeling of $\boldsymbol{X}_U\}$

Feasibility of Labeling

ICML 2021
Adversarial Multiclass Learning under Weak Supervision with Performance Guarantees

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu



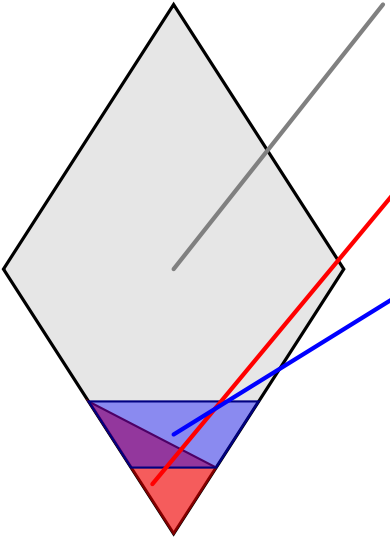
Labeling Space (of the unlabeled data) $\boldsymbol{X}_U$:
$\{\boldsymbol{Y} : \boldsymbol{Y}$ is a labeling of $\boldsymbol{X}_U\}$

Weak-Labeler ϕ_1 **Risk Constraint**:
$\hat{\text{Risk}}(\phi_1; \boldsymbol{X}_U, \boldsymbol{Y}) \approx \hat{\text{Risk}}(\phi_1; \boldsymbol{X}_L, \boldsymbol{Y}_L)$

Feasibility of Labeling

ICML 2021
Adversarial Multiclass Learning under Weak Supervision with Performance Guarantees

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu



Labeling Space (of the unlabeled data) $\boldsymbol{X}_U$:
$\{\boldsymbol{Y} : \boldsymbol{Y}$ is a labeling of $\boldsymbol{X}_U\}$

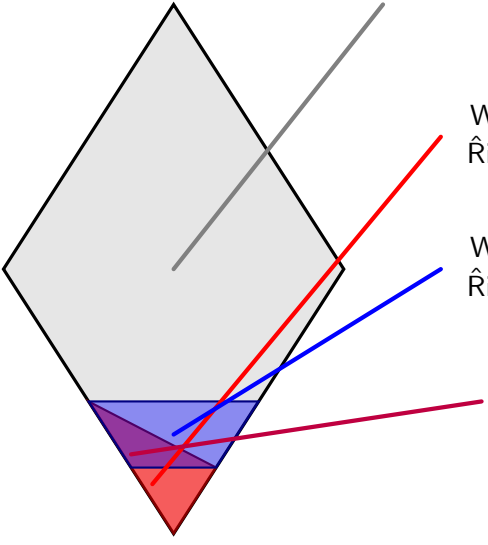
Weak-Labeler ϕ_1 **Risk Constraint**:
$\hat{\text{Risk}}(\phi_1; \boldsymbol{X}_U, \boldsymbol{Y}) \approx \hat{\text{Risk}}(\phi_1; \boldsymbol{X}_L, \boldsymbol{Y}_L)$

Weak-Labeler ϕ_2 **Risk Constraint**:
$\hat{\text{Risk}}(\phi_2; \boldsymbol{X}_U, \boldsymbol{Y}) \approx \hat{\text{Risk}}(\phi_2; \boldsymbol{X}_L, \boldsymbol{Y}_L)$

Feasibility of Labeling

ICML 2021
Adversarial Multiclass Learning under Weak Supervision with Performance Guarantees

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu



Labeling Space (of the unlabeled data) $\boldsymbol{X}_U$:
$\{\boldsymbol{Y} : \boldsymbol{Y}$ is a labeling of $\boldsymbol{X}_U\}$

Weak-Labeler ϕ_1 **Risk Constraint**:
$\hat{\text{Risk}}(\phi_1; \boldsymbol{X}_U, \boldsymbol{Y}) \approx \hat{\text{Risk}}(\phi_1; \boldsymbol{X}_L, \boldsymbol{Y}_L)$

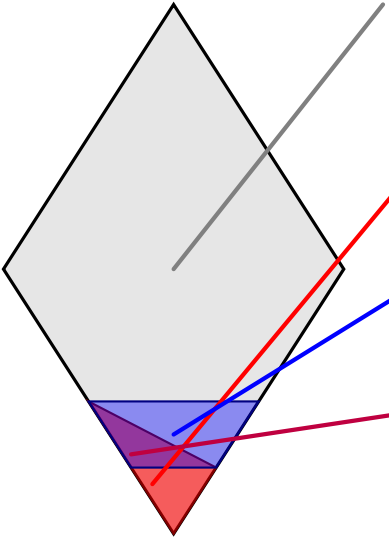
Weak-Labeler ϕ_2 **Risk Constraint**:
$\hat{\text{Risk}}(\phi_2; \boldsymbol{X}_U, \boldsymbol{Y}) \approx \hat{\text{Risk}}(\phi_2; \boldsymbol{X}_L, \boldsymbol{Y}_L)$

Feasible Set
Intersection of all constraints

Feasibility of Labeling

ICML 2021
Adversarial Multiclass Learning under Weak Supervision with Performance Guarantees

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu



Labeling Space (of the unlabeled data) $\boldsymbol{X}_U$:
$\{\boldsymbol{Y} : \boldsymbol{Y}$ is a labeling of $\boldsymbol{X}_U\}$

Weak-Labeler ϕ_1 **Risk Constraint**:
$\hat{\text{Risk}}(\phi_1; \boldsymbol{X}_U, \boldsymbol{Y}) \approx \hat{\text{Risk}}(\phi_1; \boldsymbol{X}_L, \boldsymbol{Y}_L)$

Weak-Labeler ϕ_2 **Risk Constraint**:
$\hat{\text{Risk}}(\phi_2; \boldsymbol{X}_U, \boldsymbol{Y}) \approx \hat{\text{Risk}}(\phi_2; \boldsymbol{X}_L, \boldsymbol{Y}_L)$

Feasible Set
Intersection of all constraints

Feasible set is *soft labelings*
Use *expected loss* over *label distributions*

The Adversarial Multiclass Learning Algorithm

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu

Minimax optimization of the empirical risk $\hat{\text{Risk}}(h_{\boldsymbol{\theta}}; \boldsymbol{X}_U, \boldsymbol{Y})$:

- **Minimize**: model parameters $\boldsymbol{\theta}$ in model space $\boldsymbol{\Theta}$
- **Maximize**: feasible labeling $\boldsymbol{Y}$ in the feasible set

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu

The Adversarial Multiclass Learning Algorithm

Minimax optimization of the empirical risk $\hat{\text{Risk}}(h_{\boldsymbol{\theta}}; \boldsymbol{X}_U, \boldsymbol{Y})$:

- **Minimize**: model parameters $\boldsymbol{\theta}$ in model space $\boldsymbol{\Theta}$
- **Maximize**: feasible labeling $\boldsymbol{Y}$ in the feasible set
- **Feasibility**: $\hat{\text{Risk}}(\phi_i; \boldsymbol{X}_U, \boldsymbol{Y}) \approx \hat{\text{Risk}}(\phi_i; \boldsymbol{X}_L, \boldsymbol{Y}_L)$ for each weak labeler ϕ_i
  - Linear constraints based on probabilistic tail bounds
  - With high probability: true $\boldsymbol{Y}^*$ of $\boldsymbol{X}_U$ is feasible

The Adversarial Multiclass Learning Algorithm

ICML 2021
Adversarial Multiclass Learning under Weak Supervision with Performance Guarantees

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu

Minimax optimization of the empirical risk $\hat{\text{Risk}}(h_{\boldsymbol{\theta}}; \boldsymbol{X}_U, \boldsymbol{Y})$:

- **Minimize**: model parameters $\boldsymbol{\theta}$ in model space $\boldsymbol{\Theta}$
- **Maximize**: feasible labeling $\boldsymbol{Y}$ in the feasible set
- **Feasibility**: $\hat{\text{Risk}}(\phi_i; \boldsymbol{X}_U, \boldsymbol{Y}) \approx \hat{\text{Risk}}(\phi_i; \boldsymbol{X}_L, \boldsymbol{Y}_L)$ for each weak labeler ϕ_i
  - Linear constraints based on probabilistic tail bounds
  - With high probability: true $\boldsymbol{Y}^*$ of $\boldsymbol{X}_U$ is feasible

$$\overbrace{\max_{\text{feasible } \boldsymbol{Y}} \hat{\text{Risk}}(h_{\boldsymbol{\theta}}; \boldsymbol{X}_U, \boldsymbol{Y})}^{\textbf{Linear Programming}}$$

The Adversarial Multiclass Learning Algorithm

ICML 2021
Adversarial Multiclass Learning under Weak Supervision with Performance Guarantees

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu

Minimax optimization of the empirical risk $\hat{\text{Risk}}(h_{\boldsymbol{\theta}}; \boldsymbol{X}_U, \boldsymbol{Y})$:

- **Minimize**: model parameters $\boldsymbol{\theta}$ in model space $\boldsymbol{\Theta}$
- **Maximize**: feasible labeling $\boldsymbol{Y}$ in the feasible set
- **Feasibility**: $\hat{\text{Risk}}(\phi_i; \boldsymbol{X}_U, \boldsymbol{Y}) \approx \hat{\text{Risk}}(\phi_i; \boldsymbol{X}_L, \boldsymbol{Y}_L)$ for each weak labeler ϕ_i
  - Linear constraints based on probabilistic tail bounds
  - With high probability: true $\boldsymbol{Y}^*$ of $\boldsymbol{X}_U$ is feasible

$$\underbrace{\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}}}_{\text{Subgradient Steps}} \overbrace{\max_{\text{feasible } \boldsymbol{Y}} \hat{\text{Risk}}(h_{\boldsymbol{\theta}}; \boldsymbol{X}_U, \boldsymbol{Y})}^{\text{Linear Programming}}$$

Optimization Guarantees

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu

Theorem

Assume that the loss function is convex and L-Lipschitz Continous w.r.t. $\boldsymbol{\theta}$. If we run $T = \Omega(L^2/\varepsilon^2)$ iterations of the subgradient method using step size $\alpha = \varepsilon/L^2$, then:

$$\underbrace{\max_{\text{feasible } \boldsymbol{Y}} \hat{\text{Risk}}(h_{\tilde{\boldsymbol{\theta}}}, \boldsymbol{X}_U, \boldsymbol{Y})}_{\text{Subgradient solution}} \leq \underbrace{\min_{\boldsymbol{\theta}} \max_{\text{feasible } \boldsymbol{Y}} \hat{\text{Risk}}(h_{\boldsymbol{\theta}}, \boldsymbol{X}_U, \boldsymbol{Y})}_{\text{Minimax optimal solution}} + \varepsilon \ .$$

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu

Optimization Guarantees

Theorem

Assume that the loss function is convex and L-Lipschitz Continous w.r.t. $\boldsymbol{\theta}$. If we run $T = \Omega(L^2/\varepsilon^2)$ iterations of the subgradient method using step size $\alpha = \varepsilon/L^2$, then:

$$\underbrace{\max_{\text{feasible } \boldsymbol{Y}} \hat{\text{Risk}}(h_{\tilde{\boldsymbol{\theta}}}, \boldsymbol{X}_U, \boldsymbol{Y})}_{\text{Subgradient solution}} \leq \underbrace{\min_{\boldsymbol{\theta}} \max_{\text{feasible } \boldsymbol{Y}} \hat{\text{Risk}}(h_{\boldsymbol{\theta}}, \boldsymbol{X}_U, \boldsymbol{Y})}_{\text{Minimax optimal solution}} + \varepsilon \ .$$

Example Applications:

- *softmax* with *cross-entropy loss*
- *convex combination* of weak labelers with *Brier loss*

Generalization Bounds

ICML 2021
Adversarial Multiclass Learning under Weak Supervision with Performance Guarantees

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu

$$\hat{\boldsymbol{\theta}} = \operatorname*{arg\,min}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \max_{\text{feasible } \boldsymbol{Y}} \hat{\text{Risk}}(h_{\boldsymbol{\theta}}; \boldsymbol{X}_U, \boldsymbol{Y}) \quad \textbf{vs} \quad \boldsymbol{\theta}^* = \operatorname*{arg\,min}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \text{Risk}(h_{\boldsymbol{\theta}})$$

Assume:

- loss function codomain is $[0, B]$
- true labeling of the unlabeled data is feasible

$$R(h_{\hat{\boldsymbol{\theta}}}) \leq R(h_{\boldsymbol{\theta}^*})$$

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu

Generalization Bounds

$$\hat{\boldsymbol{\theta}} = \operatorname*{arg\,min}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \max_{\text{feasible } \boldsymbol{Y}} \hat{\text{Risk}}(h_{\boldsymbol{\theta}}; \boldsymbol{X}_U, \boldsymbol{Y}) \quad \textbf{vs} \quad \boldsymbol{\theta}^* = \operatorname*{arg\,min}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \text{Risk}(h_{\boldsymbol{\theta}})$$

Assume:

- loss function codomain is $[0, B]$
- true labeling of the unlabeled data is feasible

$$R(h_{\hat{\boldsymbol{\theta}}}) \leq R(h_{\boldsymbol{\theta}^*}) + B \cdot D_f$$

The **average diameter** of the feasible set **geometrically** quantifies the information provided by the weak labelers:

$$D_f = \frac{1}{m_U} \sup_{\text{feasible } \boldsymbol{Y}', \boldsymbol{Y}''} \sum_{j=1}^{m_U} ||\boldsymbol{y}'_j - \boldsymbol{y}''_j||_1$$

Generalization Bounds

ICML 2021
Adversarial Multiclass Learning under Weak Supervision with Performance Guarantees

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu

$$\hat{\boldsymbol{\theta}} = \operatorname*{arg\,min}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \max_{\text{feasible } \boldsymbol{Y}} \hat{\text{Risk}}(h_{\boldsymbol{\theta}}; \boldsymbol{X}_U, \boldsymbol{Y}) \quad \textbf{vs} \quad \boldsymbol{\theta}^* = \operatorname*{arg\,min}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \text{Risk}(h_{\boldsymbol{\theta}})$$

Assume:

- loss function codomain is $[0, B]$
- true labeling of the unlabeled data is feasible

$$R(h_{\hat{\boldsymbol{\theta}}}) \leq R(h_{\boldsymbol{\theta}^*}) + B \cdot D_f + \sup_{\text{feasible } \boldsymbol{Y}} 4\hat{\mathfrak{R}}_{m_U}(\mathcal{L}; \boldsymbol{X}_U, \boldsymbol{Y})$$

The **average diameter** of the feasible set **geometrically** quantifies the information provided by the weak labelers:

$$D_f = \frac{1}{m_U} \sup_{\text{feasible } \boldsymbol{Y}', \boldsymbol{Y}''} \sum_{j=1}^{m_U} ||\boldsymbol{y}'_j - \boldsymbol{y}''_j||_1$$

Generalization Bounds

ICML 2021
Adversarial Multiclass Learning under Weak Supervision with Performance Guarantees

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \max_{\text{feasible } \boldsymbol{Y}} \hat{\text{Risk}}(h_{\boldsymbol{\theta}}; \boldsymbol{X}_U, \boldsymbol{Y}) \quad \textbf{vs} \quad \boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \text{Risk}(h_{\boldsymbol{\theta}})$$

Assume:

- loss function codomain is $[0, B]$
- true labeling of the unlabeled data is feasible

$$R(h_{\hat{\boldsymbol{\theta}}}) \leq R(h_{\boldsymbol{\theta}^*}) + B \cdot D_f + \sup_{\text{feasible } \boldsymbol{Y}} 4\hat{\mathfrak{R}}_{m_U}(\mathcal{L}; \boldsymbol{X}_U, \boldsymbol{Y}) + O\left(B\sqrt{\frac{\ln \frac{1}{\delta}}{m_U}}\right)$$

The **average diameter** of the feasible set **geometrically** quantifies the information provided by the weak labelers:

$$D_f = \frac{1}{m_U} \sup_{\text{feasible } \boldsymbol{Y}', \boldsymbol{Y}''} \sum_{j=1}^{m_U} ||\boldsymbol{y}'_j - \boldsymbol{y}''_j||_1$$

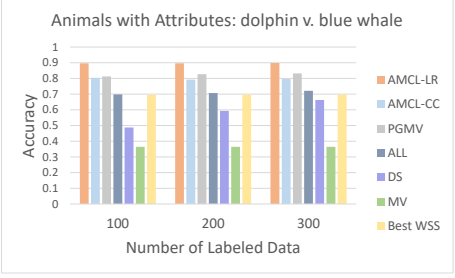
Experiments

ICML 2021
Adversarial Multiclass Learning under Weak Supervision with Performance Guarantees

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu

Two datasets:

- Animals with Attributes (binary classification)
- Domain Net (multiclass classification)



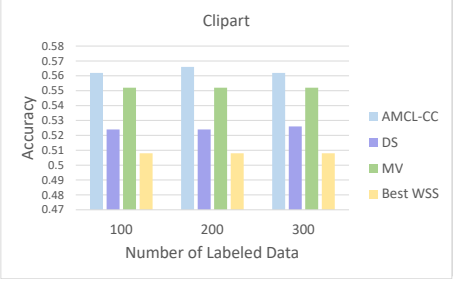


Figure: Our method often achieves better results than the baselines and the state-of-the-art

Check out our paper for more results and details

Conclusion

A. Mazzetto*
C. Cousins*
D. Sam
S. Bach
E. Upfal
cyrus_cousins@brown.edu

Thanks for your attention!