A Discrepancy-Based Theory of Adaptation

Joint work with:
Pranjal Awasthi (Google Research), Corinna Cortes (Google Research), Andres Muñoz Medina (Google), Yishay Mansour (Google Research & Tel-Aviv), Afshin Rosamizadeh (Google Research)

MEHRYAR MOHRI MOHRI@
GOOGLE RESEARCH & COURANT INSTITUTE

Motivation: Domain Adaptation

- Distribution mismatch: in many real-world problems, source and target domains differ.
- Challenges: collecting labeled data for target domains is costly, generalization problem.
- Special instances: sample bias correction, covariate-shift problems, fine-tuning for LLMs.
- Real-world applications: healthcare, autonomous driving, speech recognition, best-effort fairness.
- Can we design a theoretical framework to guide adaptation methods?

This Talk

- Discrepancy.
- Reweighting algorithms.
- Experimental results.

Multiple-Source Adaptation

- Multiple-source adaptation problem: no labeled data.
  - Theoretical analysis (Mansour, MM, and Rostamizadeh, 2008, 2009).
  - Theory and algorithms (Hoffman, MM, and Zhang, 2021), (Cortes, MM, Suresh, Zhang, 2021).
- Learning with multiple source distribution: labeled data.
  - Theoretical analysis and algorithms, application to federated learning (MM, Sivek, and Suresh, 2019).
  - Boosting with multiple sources (Cortes, MM, Storcheus, Suresh, 2021).
  - Limited target data (Mansour, MM, Ro, Suresh, Wu, 2021).

Adaptation Scenario

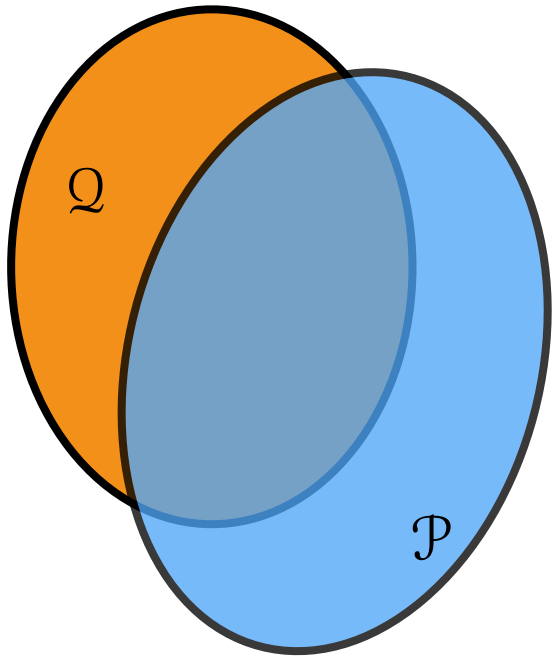
- Insput space $\mathcal{X}$, output space $\mathcal{Y}$.
- Loss function $\ell \colon \mathcal{Y} \times \mathcal{Y} \to [0, 1]$.
- Hypothesis set $\mathcal{H}$ of functions mapping from $\mathcal{X}$ to $\mathcal{Y}$.
- Learner receives:
  - Labeled sample from source domain, distribution $\mathcal{Q}$.
  - Labeled points from target domain, distribution $\mathcal{P}$: supervised adaptation (fair amount), weakly supervised (only some), unsupervised (none).
  - Typically large unlabeled sample from $\mathcal{P}$.

Adaptation Problem

- Learning problem:
  - Use labeled samples from $\mathcal{Q}$ and $\mathcal{P}$ (different scenarios) as well as typically large unlabeled sample from $\mathcal{P}$ to find hypothesis $h \in \mathcal{H}$ with small target expected loss

$$\mathcal{L}(\mathcal{P}, h) = \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{P}}[\ell(h(x), y)].$$

Challenging Problem



- Which divergence between distributions should we use?

Divergence

- Some key desiderata:
  - Tailored to adaptation problem.
  - Captures structure: loss function, hypothesis set.
  - Can be estimated from finite samples.
  - Can be leveraged algorithmically.

Discrepancy

Discrepancy

- Labeled discrepancy:

$$\mathrm{dis}(\mathcal{P}, \mathcal{Q}) = \sup_{h \in \mathcal{H}} \left\{ \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{P}} [\ell(h(x), y)] - \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{Q}} [\ell(h(x), y)] \right\}.$$

- Unlabeled discrepancy:

$$\overline{\mathrm{dis}}(\mathcal{P}, \mathcal{Q}) = \sup_{h, h' \in \mathcal{H}} \left\{ \mathop{\mathbb{E}}_{x\sim\mathcal{P}_X} [\ell(h(x), h'(x))] - \mathop{\mathbb{E}}_{x\sim\mathcal{Q}_X} [\ell(h(x), h'(x))] \right\}.$$

  - also finer local labeled or unlabeled discrepancy (Cortes et al., 2019).
  - unlabeled discrepancy coincides with d_A-*distance* of (Kifer et al., 2004), for zero-one loss.

Discrepancy - Properties

- Takes into account hypothesis set and loss function.
- Can be accurately estimated from finite samples for a hypothesis set with favorable complexity:

$$\left|\mathrm{dis}(\mathcal{P}, \mathcal{Q}) - \mathrm{dis}(\widehat{\mathcal{P}}, \widehat{\mathcal{Q}})\right| = O\left(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}}\right).$$

- Triangle inequality, distance under some assumptions.
- Upper bounds in terms of ℓ_1-distance, relative entropy, Wassertein distance.
- Coincides with d_A-distance of (Kifer et al., 2004), for zero-one loss.

Discrepancy Estimation

- Notation:
  - $\widehat{\mathcal{P}}$ empirical distribution for sample drawn from $\mathcal{P}^n$.
  - $\widehat{\mathcal{Q}}$ empirical distribution for sample drawn from $\mathcal{Q}^n$.
- Theorem: With high probability, the following holds:

$$\left|\mathrm{dis}(\mathcal{P},\mathcal{Q}) - \mathrm{dis}(\widehat{\mathcal{P}},\widehat{\mathcal{Q}})\right| \leq 2\mathfrak{R}_n(\ell \circ \mathcal{H}) + 2\mathfrak{R}_m(\ell \circ \mathcal{H}) + O\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\right).$$

- Proof:

$$\left|\mathrm{dis}(\mathcal{P},\mathcal{Q}) - \mathrm{dis}(\widehat{\mathcal{P}},\widehat{\mathcal{Q}})\right| \leq \sup_{h\in\mathcal{H}}\left|\left[\mathcal{L}(\mathcal{P},h) - \mathcal{L}(\widehat{\mathcal{P}},h)\right] - \left[\mathcal{L}(\mathcal{Q},h) - \mathcal{L}(\widehat{\mathcal{Q}},h)\right]\right|.$$

Discrepancy - Upper Bounds

- Upper bounded by ℓ_1-distance and relative entropy:

$$\begin{aligned}\mathrm{dis}(\mathcal{P},\mathcal{Q}) &= \sup_{h\in\mathcal{H}}\left\{\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{P}}[\ell(h(x),y)] - \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{Q}}[\ell(h(x),y)]\right\}\\ &\leq \sup_{h\in\mathcal{H}}\iint_{\mathcal{X}\times\mathcal{Y}}|p(x,y)-q(x,y)||\ell(h(x),y)|\,dxdy \leq \ell_1(\mathcal{P},\mathcal{Q}).\end{aligned}$$

- Upper bounded via importance weights $w(x,y)=\dfrac{p(x,y)}{q(x,y)}$:

$$\begin{aligned}\mathrm{dis}(\mathcal{P},\mathcal{Q}) &= \sup_{h\in\mathcal{H}}\iint_{\mathcal{X}\times\mathcal{Y}}[w(x,y)-1]\,q(x,y)\,\ell(h(x),y)dxdy\\ &= \sup_{h\in\mathcal{H}}\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{Q}}[\Delta w(x,y)\,\ell(h(x),y)]\\ &\leq \sqrt{\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{Q}}[\Delta^2 w(x,y)]\sup_{h\in\mathcal{H}}\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{Q}}[\ell^2(h(x),y)]}.\end{aligned}$$

Discrepancy - Upper Bounds

- Upper-bound in terms of Wasserstein distance

$$\mathcal{W}(\mathcal{P}, \mathcal{Q}) = \sup_{\|f\|_{\text{Lip}} \leq 1} \left\{ \mathop{\mathbb{E}}_{\mathbf{x} \sim \mathcal{P}_X}[f(x)] - \mathop{\mathbb{E}}_{\mathbf{x} \sim \mathcal{Q}_X}[f(x)] \right\}.$$

- For ℓ μ_ℓ-Lipschitz, $\ell(h(x), h'(x)) \leq \mu_\ell |h(x) - h'(x)|$, and for a hypothesis set $\mathcal{H}$ of $\mu_{\mathcal{H}}$-Lipschitz functions,

$$|h(x') - h'(x')| - |h(x) - h'(x)| \leq 2\mu_{\mathcal{H}} |x' - x|,$$

$$\begin{aligned} \overline{\text{dis}}(\mathcal{P}, \mathcal{Q}) &= \sup_{h, h' \in \mathcal{H}} \left\{ \mathop{\mathbb{E}}_{x \in \mathcal{P}_X}[\ell(h(x), h'(x))] - \mathop{\mathbb{E}}_{x \in \mathcal{Q}_X}[\ell(h(x), h'(x))] \right\} \\ &\leq 2\mu_\ell \, \mu_{\mathcal{H}} \mathcal{W}(\mathcal{P}, \mathcal{Q}). \end{aligned}$$

Discrepancy-Based Guarantee

- Notation:

$$\mathcal{L}(\mathcal{P}, h) = \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{P}}[\ell(h(x), y)] \quad \mathcal{L}(\mathcal{P}, h, h') = \mathop{\mathbb{E}}_{x\sim\mathcal{P}_X}[\ell(h(x), h'(x))].$$

- Theorem: Assume that ℓ verifies the triangle inequality. Then, the following inequality holds for all $h \in \mathcal{H}$:

$$\mathcal{L}(\mathcal{P}, h) \leq \inf_{\substack{(h_\mathcal{Q}, h_\mathcal{P}) \in \mathcal{H}_{\text{all}} \times \mathcal{H} \\ \vee (h_\mathcal{Q}, h_\mathcal{P}) \in \mathcal{H} \times \mathcal{H}_{\text{all}}}} \Big\{ \mathcal{L}(\mathcal{Q}, h, h_\mathcal{Q}) + \overline{\text{dis}}(\mathcal{P}, \mathcal{Q}) + \mathcal{L}(\mathcal{P}, h_\mathcal{P}) + \min\big\{\mathcal{L}(\mathcal{Q}, h_\mathcal{Q}, h_\mathcal{P}), \mathcal{L}(\mathcal{P}, h_\mathcal{Q}, h_\mathcal{P})\big\}\Big\}.$$

Discrepancy-Based Guarantee

- Properties:
  - always tighter than bound of (Ben-David et al., 2010):

$$\mathcal{L}(\mathcal{P}, h) \leq \mathcal{L}(\mathcal{Q}, h) + \overline{\mathrm{dis}}(\mathcal{P}, \mathcal{Q}) + \min_{h' \in \mathcal{H}} \{\mathcal{L}(\mathcal{Q}, h') + \mathcal{L}(\mathcal{P}, h')\}.$$

  - for same best-in class hypotheses $h^*_{\mathcal{Q}} = h^*_{\mathcal{P}} = h^*$, bound becomes:

$$\mathcal{L}(\mathcal{P}, h) \leq \mathcal{L}(\mathcal{Q}, h, h^*) + \overline{\mathrm{dis}}(\mathcal{P}, \mathcal{Q}) + \mathcal{L}(\mathcal{P}, h^*).$$

  - for consistent case, bound becomes:

$$\mathcal{L}(\mathcal{P}, h) \leq \mathcal{L}(\mathcal{Q}, h, f_{\mathcal{P}}) + \overline{\mathrm{dis}}(\mathcal{P}, \mathcal{Q}).$$

Discrepancy-Based Guarantee

- Proof: By definition of the triangle inequality and the discrepancy,

$$\begin{aligned}
\mathcal{L}(\mathcal{P}, h) &\leq \inf_{h_{\mathcal{P}} \in H} \Big\{ \mathcal{L}(\mathcal{P}, h, h_{\mathcal{P}}) + \mathcal{L}(\mathcal{P}, h_{\mathcal{P}}) \Big\} && \text{(triangle ineq.)} \\
&\leq \inf_{h_{\mathcal{P}} \in H} \Big\{ \mathcal{L}(\mathcal{Q}, h, h_{\mathcal{P}}) + \overline{\mathrm{dis}}(\mathcal{P}, \mathcal{Q}) + \mathcal{L}(\mathcal{P}, h_{\mathcal{P}}) \Big\} && \text{(def. of discrepancy)} \\
&\leq \inf_{h_{\mathcal{Q}} \in \mathcal{H}_{\mathrm{all}}, h_{\mathcal{P}} \in H} \Big\{ \mathcal{L}(\mathcal{Q}, h, h_{\mathcal{Q}}) + \mathcal{L}(\mathcal{Q}, h_{\mathcal{Q}}, h_{\mathcal{P}}) + \overline{\mathrm{dis}}(\mathcal{P}, \mathcal{Q}) + \mathcal{L}(\mathcal{P}, h_{\mathcal{P}}) \Big\}. \\
& && \text{(triangle ineq.)}
\end{aligned}$$

  - Combining similar inequalities yields:

$$\mathcal{L}(\mathcal{P}, h) \leq \inf_{\substack{(h_{\mathcal{Q}}, h_{\mathcal{P}}) \in \mathcal{H}_{\mathrm{all}} \times \mathcal{H} \\ \vee (h_{\mathcal{Q}}, h_{\mathcal{P}}) \in \mathcal{H} \times \mathcal{H}_{\mathrm{all}}}} \Big\{ \mathcal{L}(\mathcal{Q}, h, h_{\mathcal{Q}}) + \overline{\mathrm{dis}}(\mathcal{P}, \mathcal{Q}) + \mathcal{L}(\mathcal{P}, h_{\mathcal{P}}) + \min\big\{ \mathcal{L}(\mathcal{Q}, h_{\mathcal{Q}}, h_{\mathcal{P}}), \mathcal{L}(\mathcal{P}, h_{\mathcal{Q}}, h_{\mathcal{P}}) \big\} \Big\}.$$

Reweighting Algorithms

Reweighting Algorithms

- Ideas:
  - Sample weights to reduce empirical discrepancy.
  - Weights can affect weighted empirical loss.
  - Select weights and predictor jointly.
- General class of adaptation algorithm:
  - KMM (Huang et al., 2006).
  - KLIEP (Sugiyama et al., 2007).
  - Importance weighting (analysis by (Cortes et al., 2010)).
  - Discrepancy minimization (Cortes & MM, 2014).
  - Generalized disc. minimization (Cortes et al., 2019).

Learning Setup

- General supervised adaptation scenario:
  - Labeled sample $S = ((x_1, y_m), \ldots, (x_m, y_m)) \sim \mathcal{Q}^m$.
  - Labeled sample $S' = ((x_{m+1}, y_{m+1}), \ldots, (x_{m+n}, y_{m+n})) \sim \mathcal{P}^n$.
  - Non-negative weight vector $\mathsf{q} \in [0,1]^{m+n}$.
  - Total weight of first m samples: $\overline{\mathsf{q}} = \sum_{i=1}^{m} \mathsf{q}_i$.
- Problem:
  - Find weights $\mathsf{q} \in [0,1]^{m+n}$ and $h \in \mathcal{H}$ to achieve small target domain expected loss $\mathcal{L}(\mathcal{P}, h)$.

Weighted Rademacher Comp.

- For $\mathsf{q} \in [0,1]^{[m+n]}$,

$$\mathfrak{R}_{\mathsf{q}}(\ell \circ \mathcal{H}) = \mathop{\mathbb{E}}_{S,S',\boldsymbol{\sigma}}\left[\sup_{h \in \mathcal{H}} \sum_{i=1}^{m+n} \sigma_i \mathsf{q}_i \ell(h(x_i), y_i)\right].$$

  - By Talagrand's contraction lemma,

$$\mathfrak{R}_{\mathsf{q}}(\ell \circ \mathcal{H}) \leq \|\mathsf{q}\|_\infty (m+n)\, \mathfrak{R}_{m+n}(\ell \circ \mathcal{H}).$$

Reweighting Learning Bound

- Theorem: fix weights $\mathsf{q} \in [0,1]^{[m+n]}$. Then, with probability at least $1-\delta$ over the draw of a sample $S \sim \mathcal{Q}^m$ from the source domain and $S' \sim \mathcal{P}^n$, for any $h \in \mathcal{H}$,

$$\mathcal{L}(\mathcal{P}, h) \leq \sum_{i=1}^{m+n} \mathsf{q}_i \ell(h(x_i), y_i) + \mathrm{dis}\Big(\big[(1 - \|\mathsf{q}\|_1) + \overline{\mathsf{q}}\big]\mathcal{P}, \overline{\mathsf{q}}\mathcal{Q}\Big)$$

$$+ 2\mathfrak{R}_{\mathsf{q}}(\ell \circ \mathcal{H}) + \|\mathsf{q}\|_2 \sqrt{\frac{\log \frac{1}{\delta}}{2}}.$$

q distribution

$$\overline{\mathsf{q}}\mathrm{dis}(\mathcal{P}, \mathcal{Q})$$

Reweighting Lower Bound

- Theorem: fix distribution $\mathsf{q} \in \Delta_{m+n}$. Then, for any with $\epsilon > 0$, there exists $h \in \mathcal{H}$ such that for any $\delta > 0$, with probability at least $1 - \delta$ over the draw of a sample $S \sim \mathcal{Q}^m$ from the source domain and $S' \sim \mathcal{P}^n$,

$$\mathcal{L}(\mathcal{P}, h) \geq \sum_{i=1}^{m+n} \mathsf{q}_i \ell(h(x_i), y_i) + \overline{\mathsf{q}} \mathrm{dis}(\mathcal{P}, \mathcal{Q}) + \Omega\left(\frac{1}{\sqrt{m+n}}\right).$$

  - for $\|\mathsf{q}\|_2, \mathfrak{R}_{\mathsf{q}}(\ell \circ \mathcal{H}) \in O\left(\frac{1}{\sqrt{m+n}}\right)$.

Reweighting Uniform Bound

- Theorem: For any $\delta > 0$, with probability at least $1 - \delta$ over the draw of a sample $S \sim \mathcal{Q}^m$ and sample $S' \sim \mathcal{P}^n$, the following holds for all $h \in \mathcal{H}$ and $\mathsf{q} \in \mathsf{B}_1(\mathsf{p}_0, 1)$,

$$\mathcal{L}(\mathcal{P}, h) \leq \sum_{i=1}^{m+n} \mathsf{q}_i \ell(h(x_i), y_i) + \overline{\mathsf{q}} \mathrm{dis}(\mathcal{P}, \mathcal{Q}) + \mathrm{dis}(\mathsf{p}^0, \mathsf{q}) + 2\mathfrak{R}_{\mathsf{q}}(\ell \circ \mathcal{H})$$

$$+ 8\|\mathsf{q} - \mathsf{p}^0\|_1 + \left[\|\mathsf{q}\|_2 + 2\|\mathsf{q} - \mathsf{p}^0\|_1\right] \left[\sqrt{\log \log_2 \frac{2}{1 - \|\mathsf{q} - \mathsf{p}^0\|_1}} + \sqrt{\frac{\log \frac{2}{\delta}}{2}}\right].$$

  - p_0: reference weights.

Empirical Disc. Estimation

- Optimization problem:

$$\widehat{d} = \max_{h \in \mathcal{H}} \left\{ \frac{1}{n} \sum_{i=m+1}^{m+n} \ell(h(x_i), y_i) - \frac{1}{m} \sum_{i=1}^{m} \ell(h(x_i), y_i) \right\}.$$

  - for a convex loss, can be cast as DC-programming problem and solved via DCA (Tao and An, 1988).
  - for squared loss, global optimum convergence guarantee.

$$h_{t+1} \in \operatorname*{argmin}_{h \in \mathcal{H}} \left\{ \frac{1}{m} \sum_{i=1}^{m} \ell(h(x_i), y_i) - \frac{1}{n} \sum_{i=m+1}^{m+n} \nabla \ell(h_t(x_i), y_i) \cdot (h - h_t) \right\}.$$

Algorithm

- Optimization problem: SBEST algorithm.

$$\min_{h\in\mathcal{H},\mathsf{q}\in[0,1]^{m+n}} \sum_{i=1}^{m+n} \mathsf{q}_i \ell(h(x_i), y_i) + \overline{\mathsf{q}}\mathrm{dis}(\widehat{\mathcal{P}}, \widehat{\mathcal{Q}}) + \lambda_\infty \|\mathsf{q}\|_\infty \|h\|^2 + \lambda_1 \|\mathsf{q} - \mathsf{p}^0\|_1 + \lambda_2 \|\mathsf{q}\|_2^2.$$

  - Alternate minimization solution.
  - For squared loss with linear predictors, convex optimization problem.
  - Empirical discrepancy estimation via DC-programming.
  - Extension to weakly or unsupervised adaptation.

Labeled Disc. Upper Bounds

- Theorem: for squared loss, for any $h_0 \in \mathcal{H}$,

$$\mathrm{dis}(\widehat{\mathcal{P}}, \widehat{\mathcal{Q}}) \leq \overline{\mathrm{dis}}_{\mathcal{H} \times \{h_0\}}(\widehat{\mathcal{P}}, \widehat{\mathcal{Q}}) + 2\delta_{\mathcal{H}, h_0}(\widehat{\mathcal{P}}, \widehat{\mathcal{Q}}).$$

  - where:

$$\delta_{\mathcal{H}, h_0}(\widehat{\mathcal{P}}, \widehat{\mathcal{Q}}) = \sup_{h \in \mathcal{H}} \left| \mathop{\mathbb{E}}_{(x,y) \sim \widehat{\mathcal{P}}} [h(x)(y - h_0(x))] - \mathop{\mathbb{E}}_{(x,y) \sim \widehat{\mathcal{Q}}} [h(x)(y - h_0(x))] \right|.$$

  - favorable when h_0 can be chosen so that $|y - h_0(x)|$ is relatively small for both samples.
  - note that $\delta_{\mathcal{H}, h_0}(\widehat{\mathcal{P}}, \widehat{\mathcal{Q}}) = 0$ for $\widehat{\mathcal{P}} = \widehat{\mathcal{Q}}$.

Experimental Results

Classification Tasks

Dataset	Train source $\mathcal{Q}$	Train target $\mathcal{P}$	KMM	gapBoost	SBEST
Adult	82.72 ± 0.10	81.61 ± 0.42	81.24 ± 0.01	*83.1 ± 0.02*	*83.30 ± 0.28*
German	68.24 ± 0.21	69.87 ± 0.27	65.7 ± 0.01	69.8 ± 0.03	$\mathbf{71.26 \pm 0.11}$
Accent	27.20 ± 0.26	81.64 ± 0.22	53.1 ± 0.03	81.2 ± 0.04	$\mathbf{84.15 \pm 0.30}$
comp vs sci	83.2 ± 0.004	89.4 ± 0.03	83.1 ± 0.004	92.08 ± 0.01	$\mathbf{94.4 \pm 0.01}$
rec vs sci	79.2 ± 0.007	91.3 ± 0.02	79.7 ± 0.004	92.2 ± 0.01	$\mathbf{92.4 \pm 0.004}$
comp vs talk	71.4 ± 0.002	89.9 ± 0.02	71 ± 0.006	90.6 ± 0.01	$\mathbf{91 \pm 0.02}$
comp vs rec	65.4 ± 0.007	85.2 ± 0.01	67.7 ± 0.007	85.9 ± 0.01	$\mathbf{88 \pm 0.01}$
rec vs talk	81.3 ± 0.004	88 ± 0.02	81.2 ± 0.005	89.2 ± 0.01	$\mathbf{92.3 \pm 0.03}$
sci vs talk	88.2 ± 0.005	93.3 ± 0.008	88.5 ± 0.003	*94.6 ± 0.01*	*94.6 ± 0.02*

- Details of experimental results in (Awasthi, Cortes, and MM, 2024).

Fine-Tuning Tasks

Fine-tuning	Train on $\mathcal{P}$	gapBoost	SBEST
Last layer (CIFAR-10)	$88.61 \pm .43$	$87.1 \pm .01$	$\mathbf{89.62 \pm .32}$
Full model (CIFAR-10)	$90.18 \pm .31$	$90.8 \pm .02$	$\mathbf{92.30 \pm .24}$
Last layer (`Civil`)	$63.1 \pm .12$	$64.7 \pm .11$	$\mathbf{65.8 \pm .12}$
Full model (`Civil`)	$65.8 \pm .01$	$67.2 \pm .01$	$\mathbf{68.3 \pm .14}$

Regression Tasks

Dataset	KMM	DM	SBEST
Wind	1.2 ± 0.04	1.14 ± 0.03	$\mathbf{0.97 \pm 0.02}$
Airline	2.4 ± 0.09	1.72 ± 0.1	$\mathbf{0.952 \pm 0.03}$
Gas	0.41 ± 0.01	0.39 ± 0.01	$\mathbf{0.38 \pm 0.02}$
News	1.08 ± 0.01	1.1 ± 0.01	$\mathbf{0.99 \pm 0.01}$
Traffic	2.1 ± 0.1	2.08 ± 0.08	$\mathbf{0.99 \pm 0.002}$

Sentiment Analysis

$\mathcal{Q}$	$\mathcal{P}$	GDM	DM	KMM	Train on $\mathcal{Q}$
books	**dvd**	1.25 ± 0.01	1.26 ± 0.11	1.43 ± 0.08	2.34 ± 0.19
	elec	0.88 ± 0.01	0.89 ± 0.03	1.50 ± 0.05	2.13 ± 0.13
	ktchn	1.06 ± 0.03	1.08 ± 0.04	1.47 ± 0.01	1.55 ± 0.01
dvd	**books**	1.14 ± 0.02	1.17 ± 0.10	1.64 ± 0.14	2.18 ± 0.18
	elec	1.08 ± 0.01	1.10 ± 0.12	2.40 ± 0.05	3.26 ± 0.07
	ktchn	1.1 ± 0.03	1.12 ± 0.02	1.10 ± 0.02	2.34 ± 0.05
elec	books	0.98 ± 0.01	1.00 ± 0.01	1.33 ± 0.06	1.34 ± 0.04
	dvd	0.98 ± 0.02	1.00 ± 0.06	1.00 ± 0.06	1.04 ± 0.08
	ktchn	0.96 ± 0.01	0.98 ± 0.06	1.04 ± 0.01	1.14 ± 0.01
ktchn	**books**	1.00 ± 0.03	1.04 ± 0.07	1.27 ± 0.09	1.12 ± 0.08
	dvd	1.2 ± 0.002	1.33 ± 0.03	1.32 ± 0.03	1.42 ± 0.04
	elec	1.64 ± 0.02	1.67 ± 0.54	1.87 ± 0.56	1.89 ± 0.56

Conclusion

- Multiple-source adaptation problems.
- Discrepancy-based analysis of drifting (MM & Muñoz Medina, 2019; Awasthi, Cortes, and Mohri, 2022).
- Time series prediction and algorithms (MM & Kuznetsov, 2020).
- Differentially private adaptation from public to private domains or vice-versa (Bassily, Cortes, Mao, MM, 2024).
- Active learning (de Mathelin et al., 2022; Zhang et al., 2019, 2020).
- PAC-Bayesian analysis of adaptation (Germain et al., 2013).