Learning from Positive & Unlabeled Data with Arbitrary Positive Shift

Zayd Hammoudeh (zayd@cs.uoregon.edu)
Daniel Lowd (lowd@cs.uoregon.edu)

NEURAL INFORMATION PROCESSING SYSTEMS

Motivation

Two Joint Data Distributions: Source (training) & Target (test)

Positive-Unlabeled (PU) Learning: Trains a binary classifier (g) using only positive-labeled and unlabeled data

- *Common Simplifying Assumption*: Positive-labeled set is representative of the target positive class

Biased-Positive, Unlabeled (bPU) Learning: Positive-labeled set is *biased* w.r.t. the target positive class

- Positive bias is commonly formulated as a selection bias (e.g., PUSB [1]) or covariate shift (e.g., PUc [5]) problem

Our Proposed Problem Setting

Arbitrary-Positive, Unlabeled (aPU) Learning: Positive-labeled set is biased *arbitrarily* w.r.t. the target positive class

- More general and harder than bPU learning
- **Our Key Insight**: aPU learning is possible provided two unlabeled sets as in [5] when *all negative examples are generated from a single distribution*
- **Real-World aPU Learning Applications**: Land-cover classification, epidemiology, and adversarial domains

Simplifying PU Empirical Risk Estimation

Unbiased PU (uPU) Risk Estimator [3]: For positive prior $\pi := p(y = 1)$

$$\widehat{R}_{\text{uPU}}(g) := \pi \widehat{R}_{\text{p}}^{+}(g) + \widehat{R}_{\text{u}}^{-}(g) - \pi \widehat{R}_{\text{p}}^{-}(g)$$

Non-Negative PU (nnPU) Estimator [4]: Addresses uPU's propensity to implausibly overfit. Biased but *consistent*. Needs custom ERM framework

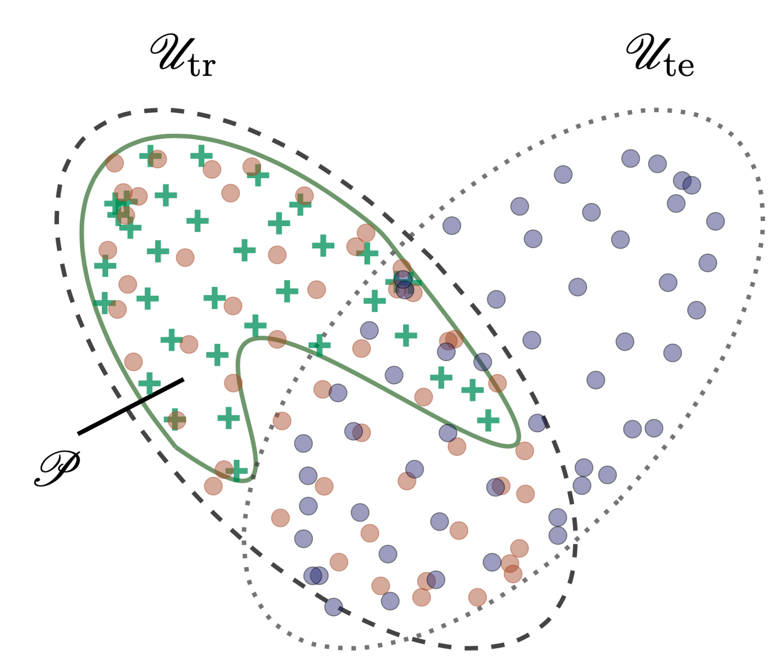
$$\widehat{R}_{\text{nnPU}}(g) := \pi \widehat{R}_{\text{p}}^{+}(g) + \max\{0, \widehat{R}_{\text{u}}^{-}(g) - \pi \widehat{R}_{\text{p}}^{-}(g)\}$$

Our Absolute-value PU (abs-PU) Estimator: Statistically consistent. Yields models as good or better than nnPU with much simpler optimization.

$$\widehat{R}_{\text{abs-PU}}(g) := \pi \widehat{R}_{\text{p}}^{+}(g) + \left|\widehat{R}_{\text{u}}^{-}(g) - \pi \widehat{R}_{\text{p}}^{-}(g)\right|$$

UNIVERSITY OF OREGON

What is an aPU Learning Dataset?



Three Independently Sampled Datasets

- $\mathscr{P}$: Positive-labeled set (biased) sample of training pos. class-conditional
- $\mathscr{U}_{\text{tr}}$: Training unlabeled set i.i.d. sample of *training* marginal distribution
- $\mathscr{U}_{\text{te}}$: Test unlabeled set i.i.d. sample of *test* marginal distribution

Solution #1: Unlabeled-Unlabeled Learning

Main Idea: Train final classifier g in *two-steps* by first extracting *surrogate negative set* $\mathscr{N}$ from unlabeled training set $\mathscr{U}_{\text{tr}}$

Step #1: Train PU probabilistic classifier $\hat{\sigma}(x) \approx \Pr_{\text{tr}}[y = 1 \,|\, x]$ using datasets $\mathscr{P}$ and $\mathscr{U}_{\text{tr}}$

- *Surrogate negative set* $\mathscr{N}$ is a statistically consistent estimate of negative-class risk. $\mathscr{N}$ soft weights unlabeled training set $\mathscr{U}_{\text{tr}}$'s loss $\ell : \mathbb{R}^d \to \mathbb{R}$ via $\hat{\sigma}$:

$$\widetilde{R}_{\text{n}}^{\hat{y}}(g) := \frac{1}{|\mathscr{U}_{\text{tr}}|} \sum_{x_i \in \mathscr{U}_{\text{tr}}} \frac{\hat{\sigma}(x_i)\ell(\hat{y}g(x_i))}{1 - \pi_{\text{tr}}}$$

Step #2: Train final classifier using one of two novel risk estimators:

- **Weighted Unlabeled-Unlabeled (wUU)**: *Uses only unlabeled data*, i.e., unlabeled test set $\mathscr{U}_{\text{te}}$ and surrogate negative set $\mathscr{N}$ formed from $\mathscr{U}_{\text{tr}}$
- **Arbitrary-Positive, Negative, Unlabeled (aPNU)**: *Uses all available data*, i.e., arbitrary-positive $\mathscr{P}$, surrogate negative $\mathscr{N}$, & unlabeled test $\mathscr{U}_{\text{te}}$

Complete Two-Step Methods: PU2wUU† & PU2aPNU†

github.com/ZaydH/arbitrary_pu

Solution #2: Novel Recursive Risk Estimator

Main Idea: Train an aPU learner via a statistically consistent *joint method*

PURR†: Our *Positive-Unlabeled Recursive Risk* estimator

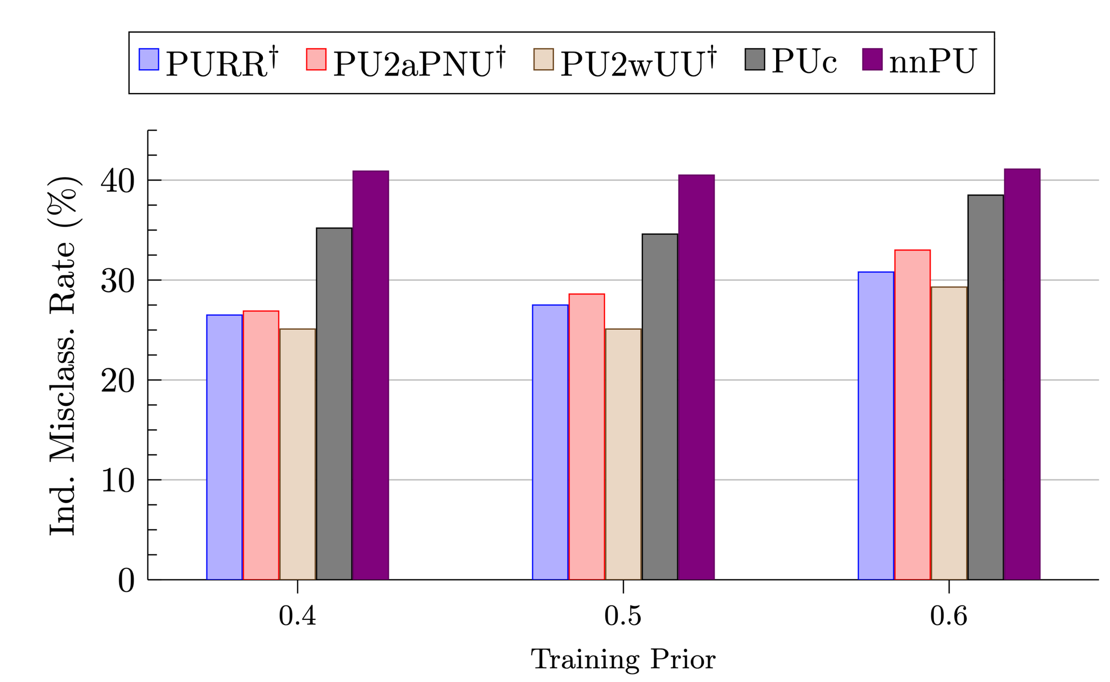
$$\widehat{R}_{\text{PURR}}(g) = \overbrace{\left| \widehat{R}_{\text{te-u}}^{+}(g) - (1 - \pi_{\text{te}}) \left| \overbrace{\frac{\widehat{R}_{\text{tr-u}}^{+}(g) - \pi_{\text{tr}} \widehat{R}_{\text{tr-p}}^{+}(g)}{1 - \pi_{\text{tr}}}}^{\widehat{R}_{\text{te-n}}^{+}(g)} \right| \right|}^{\pi_{\text{te}} \widehat{R}_{\text{te-p}}^{+}(g)}$$
$$+ (1 - \pi_{\text{te}}) \left| \underbrace{\frac{\widehat{R}_{\text{tr-u}}^{-}(g) - \pi_{\text{tr}} \widehat{R}_{\text{tr-p}}^{-}(g)}{1 - \pi_{\text{tr}}}}_{\widehat{R}_{\text{te-n}}^{-}(g)} \right|.$$

Intuition: Recursively nest du Plessis et al.'s [3] PU risk decomposition

Experimental Results

Real-World Datasets: Adversarial spam classification [2] with shift across two email datasets that are two years apart

- **Training Set**: TREC 2005 spam & ham emails
- **Test Set**: TREC 2007 spam & ham emails



Takeaway: All of our methods handle large positive shifts better than prior work, even in the realistic case of a shifting negative class

And a lot more! Additional baselines & many more datasets in the paper…

References

[1] Kato et al. "Learning from positive and unlabeled data with a selection bias." ICLR, 2019.
[2] Fusilier et al. "Using PU-learning to detect deceptive opinion spam" WASSA, 2013.
[3] du Plessis et al. "Analysis of learning for positive and unlabeled data" NeurIPS, 2014.
[4] Kiryo et al. "Positive-unlabeled learning with non-negative risk estimator." NeurIPS, 2017.
[5] Sakai & Shimizu. "Covariate shift adaptation on learning from positive and unlabeled data." AAAI, 2019.