Open-set Domain Adaptation via Joint Error based Multi-class Positive and Unlabeled Learning ECCV 2024

Dexuan Zhang[1], Thomas Westfechtel[1], Tatsuya Harada[1,2]

[1]The University of Tokyo, [2]RIKEN

September 10, 2024

Problem Setting

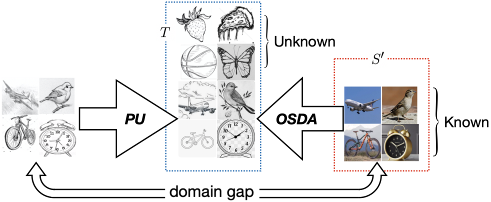


Figure: Given labeled data from source domain $S' : \mathcal{X} \times (\mathcal{Y}' = \{1, .., K-1\})$, and unlabeled data from target domain $T : \mathcal{X} \times (\mathcal{Y} = \{1, .., K\})$, the goal is to learn a target classifier $h : \mathcal{X} \rightarrow \mathcal{Y}$.

Remark

- OSDA can be considered as PU with covariate shift where $P_{S'}(x|y) \neq P_T(x|y)$ for $y \in \mathcal{Y}'$
- PU learning can be applied in OSDA if the gap is bridged

PU Learning induced Joint Error based OSDA (PUJE)

Theorem (Approximated Joint Error based Target Upper Bound)

Given $\mathcal{K} = \{k | k \in \mathbb{R}^K : \sum_{y \in \mathcal{Y}} k[y] = 1, k[y] \in [0,1]\}$, *let* $f_S, f_T : \mathcal{X} \to \mathcal{K}$ *be the true labeling functions for the source and target domains and* $\epsilon : \mathcal{K} \times \mathcal{K} \to \mathbb{R}$ *denote a distance metric and* $\epsilon_D(f, f') := \mathbb{E}_{x \sim D}\, \epsilon(f(x), f'(x))$ *measure the expected disagreement between the outputs of* $f, f' : \mathcal{X} \to \mathcal{K}$. *For* $\forall f_S^*, f_T^*, h \in \mathcal{H} : \mathcal{X} \to \mathcal{K}$, *the expected target error is bounded by*

$$\epsilon_T(h) \leq \epsilon_S(h) + \epsilon_T(f_S^*, f_T^*) + \epsilon_T(h, f_S^*) - |\epsilon_S(f_S^*, f_T^*) - \epsilon_S(h, f_T^*)| + \theta$$

$$\theta = 2\epsilon_T(f_S, f_S^*) + \epsilon_S(f_S, f_S^*) + 2\epsilon_S(f_T^*, f_T) + \epsilon_T(f_T^*, f_T) = \theta_{f_S} + \theta_{f_T}$$

Assumption (1)

Assume that there exists approximated labeling functions f_S^*, f_T^* *such that the empirical deviations* $\hat{\theta}_{f_S}, \hat{\theta}_{f_T}$ *measured on finite samples* $\hat{S}, \hat{T}$ *are close enough to zero.*

Multi-class PU Learning

Definition (Unknown Predictive Discrepancy)

Let $\upsilon : \mathcal{K} \times \mathcal{K} \to \mathbb{R}$ denote the Unknown Predictive Discrepancy as a distance metric and $\upsilon_D(f, f') := \mathbb{E}_{x\sim D}\, \upsilon(f(x), f'(x))$ measure the expected disagreement between the K-th outputs of $f, f' : \mathcal{X} \to \mathcal{K}$. Let $e^K : \mathcal{X} \to [0, ..., 1] \in \mathcal{K}$ denote a function that can predict any input as unknown. The deviation from e^K for a hypothesis $h \in \mathcal{H}$ is referred to as the shorthand $\upsilon_D(h) := \upsilon_D(h, e^K)$ that measures the probability that samples from D have not been classified as unknown.

Assumption (2)

Let $S^i = P_S(x|y = i), T^i = P_T(x|y = i)$ denote class conditional distributions, $S' = P_S(x|y \neq K), T' = P_T(x|y \neq K)$ indicate incomplete domains that do not contain unknown class S^K, T^K. Given a feature extractor $g : \mathcal{X} \to \mathcal{Z}$, assume that the feature space can be aligned: $P_{S^K}(z) = P_{T^K}(z), P_{S'}(z) = P_{T'}(z)$.

Multi-class PU Learning

Lemma (Estimated Source Error and Discrepancy)

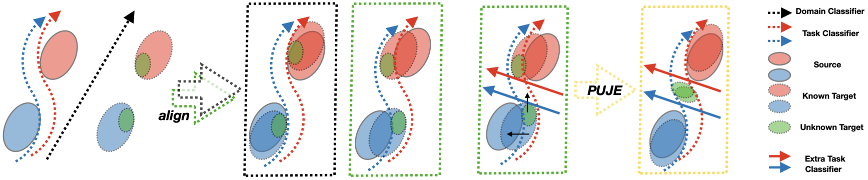
Let $\sum_{i=1}^{K} \pi_S^i = 1, \sum_{i=1}^{K} \pi_T^i = 1$ denote the label distribution of S and T respectively. Given feature extractor g, approximated labeling functions can be decomposed such that $f_S^* = f_S^\star \circ g, f_T^* = f^\star \circ g$. Given Assumption (2), the expected error on S can be estimated by the error on S' and Unknown Predictive Discrepancy on T with a mild condition that $\pi_S^K = \pi_T^K = 1 - \alpha$:

$$\epsilon_S(h \circ g) = \alpha[\epsilon_{S'}(h \circ g) - \upsilon_{S'}(h \circ g)] + \upsilon_T(h \circ g)$$

$$\epsilon_S(f_S^\star \circ g, f_T^\star \circ g) = \alpha[\epsilon_{S'}(f_S^\star \circ g, f_T^\star \circ g) - \upsilon_{S'}(f_S^\star \circ g, f_T^\star \circ g)] + \upsilon_T(f_S^\star \circ g, f_T^\star \circ g)$$

$$\epsilon_S(h \circ g, f_T^\star \circ g) = \alpha[\epsilon_{S'}(h \circ g, f_T^\star \circ g) - \upsilon_{S'}(h \circ g, f_T^\star \circ g)] + \upsilon_T(h \circ g, f_T^\star \circ g)$$

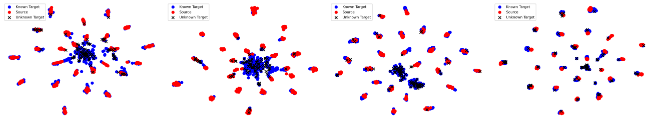
Mechanism



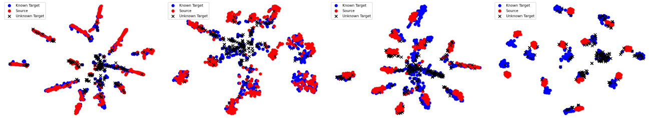
(a) existing methods (b) proposed method

Figure: Intuitive explanation of the difference between PUJE and existing methods. (a) existing methods do not explicitly minimize joint error and cannot group unknown class as a single cluster; (b) our proposal is an upper bound of joint error which can address large domain shift and group unknown class into a single cluster.

Results



(a) OSBP (b) PGL (c) ANNA (d) PUJE



(e) OSBP (f) PGL (g) ANNA (h) PUJE

Figure: T-SNE visualization of feature distributions in (a)-(d): Ar→Cl task (Office-Home dataset); (e)-(h): Syn2Real-O task.

Remark

- PUJE achieves a better alignment with a more discriminative class-wise decision boundary for unknown class, especially when the domain shift is large.