Aggregating From Multiple Target-Shifted Sources

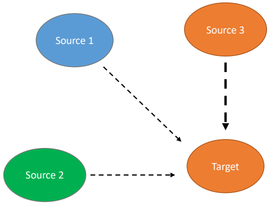
Changjian Shui, Zijian Li, Jiaqi Li, Christian Gagné, Charles X.Ling, Boyu Wang

ICML 2021

Université Laval
Institut intelligence et données

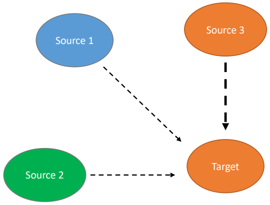
Multiple-Source Domain Adaptation

- Learning a target domain with limited or even no label information through multiple *related* sources.



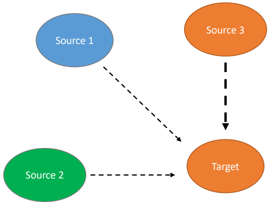
Multiple-Source Domain Adaptation

- Learning a target domain with limited or even no label information through multiple *related* sources.
- Widely applied in image segmentation, crowd sourcing and personal medicine.



Multiple-Source Domain Adaptation

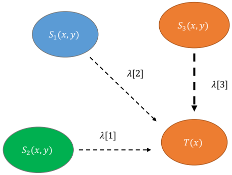
- Learning a target domain with limited or even no label information through multiple *related* sources.
- Widely applied in image segmentation, crowd sourcing and personal medicine.
- Key Question:
  How to select relevant sources to avoid negative transfer ?



Selection through domain similarity

Conventional theories in multi-source domain adaptation:

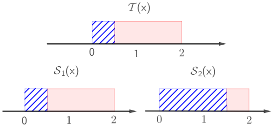
$$R_{\mathcal{T}}(h) \leq \frac{1}{T}\sum_{t=1}^{T} R_{\mathcal{S}_t}(h) + \sum_{t=1}^{T} \boldsymbol{\lambda}[t]\mathrm{dist}(\mathcal{S}_{\mathrm{t}}(\mathrm{x}), \mathcal{T}(\mathrm{x}))$$



- $\boldsymbol{\lambda}$ is a simplex, measuring source-target relations.
- If marginal distribution distance $\mathrm{dist}(\mathcal{S}_{\mathrm{t}}(\mathrm{x}), \mathcal{T}(\mathrm{x}))$ is small, assigning higher $\boldsymbol{\lambda}[t]$.

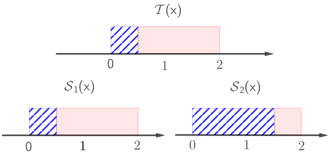
Limitation of adopting $\mathrm{dist}(\mathcal{S}_t(x), \mathcal{T}(x))$

- $\mathrm{dist}(\mathcal{S}_1(x), \mathcal{T}(x)) = \mathrm{dist}(\mathcal{S}_2(x), \mathcal{T}(x)))$
- $\boldsymbol{\lambda}[1] = \boldsymbol{\lambda}[2]$
- $\mathcal{S}_2$ is a unreliable source: label propotion between sources-target is different.



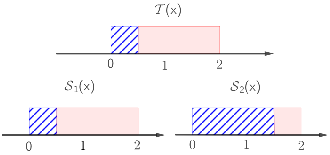
Research Goal: Leveraging from different label (y)-shifted sources.

One Solution



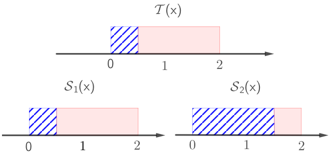
- $\mathrm{dist}(\mathcal{S}_1(\mathrm{x}|\mathrm{y}), \mathcal{T}(\mathrm{x}|\mathrm{y})) < \mathrm{dist}(\mathcal{S}_2(\mathrm{x}|\mathrm{y}), \mathcal{T}(\mathrm{x}|\mathrm{y})))$

One Solution



- $\mathrm{dist}(\mathcal{S}_1(\mathrm{x}|\mathrm{y}), \mathcal{T}(\mathrm{x}|\mathrm{y})) < \mathrm{dist}(\mathcal{S}_2(\mathrm{x}|\mathrm{y}), \mathcal{T}(\mathrm{x}|\mathrm{y})))$
- Assigning higher $\boldsymbol{\lambda}[1]$ for $\mathcal{T}$.

One Solution



- $\mathrm{dist}(\mathcal{S}_1(\mathrm{x}|\mathrm{y}), \mathcal{T}(\mathrm{x}|\mathrm{y})) < \mathrm{dist}(\mathcal{S}_2(\mathrm{x}|\mathrm{y}), \mathcal{T}(\mathrm{x}|\mathrm{y})))$
- Assigning higher $\boldsymbol{\lambda}[1]$ for $\mathcal{T}$.
- Adopting the similarity of *conditional* distribution is more reliable.

Our Contributions

- Analyze multi-sources domain adaptation with $\mathcal{S}_t(y) \neq \mathcal{T}(y)$, $\mathcal{S}_t(x|y) \neq \mathcal{T}(x|y)$

Our Contributions

- Analyze multi-sources domain adaptation with $\mathcal{S}_t(y) \neq \mathcal{T}(y)$, $\mathcal{S}_t(x|y) \neq \mathcal{T}(x|y)$
- A theoretically grounded approach with compelling empirical results, compared with modern baselines.

Our Contributions

- Analyze multi-sources domain adaptation with $\mathcal{S}_t(y) \neq \mathcal{T}(y)$, $\mathcal{S}_t(x|y) \neq \mathcal{T}(x|y)$
- A theoretically grounded approach with compelling empirical results, compared with modern baselines.
- A unified method for handling different scenarios, where previous works generally treated as separate problems.

Unified Approach

Table 1: Three multi-source domain adaptation (DA) scenarios

	Target label	$\mathrm{supp}(\mathcal{S}_\mathrm{t}(\mathrm{y})) = \mathrm{supp}(\mathcal{T}(\mathrm{y}))$	Additional Assumption
DA Limited label	✓	✓	✗
Unsupervised DA	✗	✓	✓
Partial Unsupervised DA	✗	✗	✓

- Require additional assumptions in unsupervised scenarios when label and conditional distribution shift.
- Partial Unsupervised DA $\mathrm{supp}(\mathcal{T}(\mathrm{y})) \subseteq \mathrm{supp}(\mathcal{S}_\mathrm{t}(\mathrm{y}))$

Thank You