Class-conditioned Domain Generalization via Wasserstein Distributional Robust Optimization

Tsinghua University

Jingge Wang, Yang Li, Liyan Xie, Yao Xie

Abstract

- Given multiple source domains, **domain generalization** aims at learning a universal model adaptable to any unseen target domain.
- Existing works do not work well in certain domain generalization cases due to variable domain shifts.
- We propose a novel class-conditional domain generalization algorithm based on **Wasserstein robust hypothesis testing** and show promising results on simulated and real data.

Overview

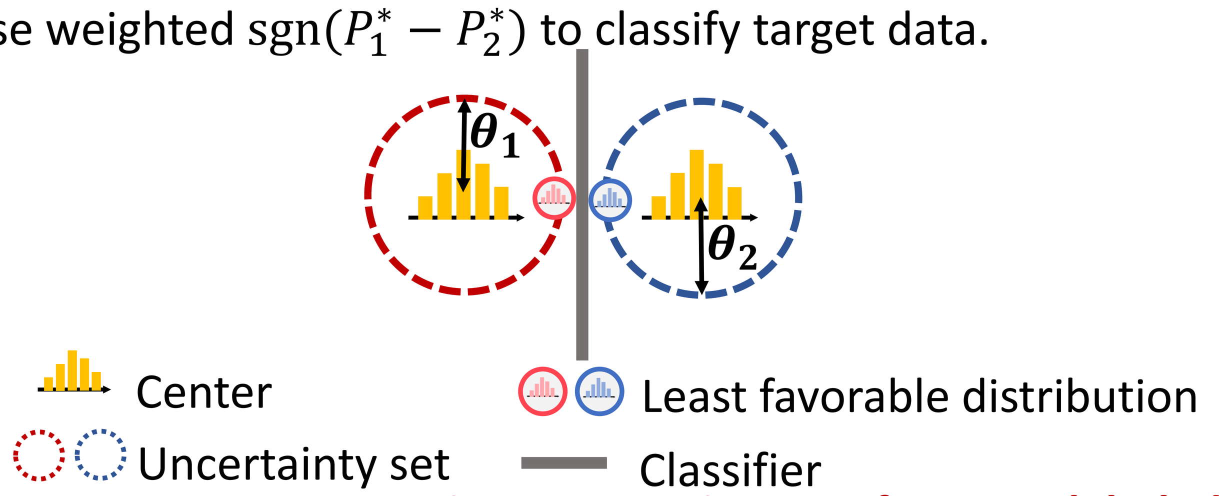
Domain generalization

- Given M source domains with labeled data and class-conditional distributions $S_m(X|Y=y), m = 1, \dots, M$.
- Generalize to any unseen target domain T without any labeled training data.

Wasserstein distributional robust optimization [1]:

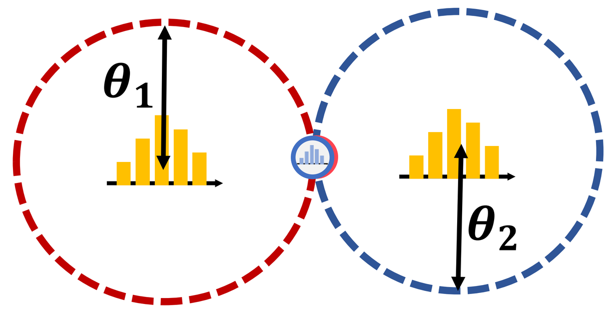
$$\min_{\phi} \max_{P_1 \in \mathcal{P}_1, P_2 \in \mathcal{P}_2} \Phi(\phi; P_1, P_2)$$

- Use Wasserstein distance to construct appropriate uncertainty sets around some center distribution Q_y for class-conditional distributions, i.e.,
  $\mathcal{P}_y = \{P(X|Y=y): \mathcal{W}_2(P, Q_y) \le \theta_y\},\ y = 1, 2.$
- Optimal detector ϕ^* minimizes classifier risk under least favorable distributions (LFDs) P_1^*, P_2^*.
- Use weighted $\mathrm{sgn}(P_1^* - P_2^*)$ to classify target data.



How to construct appropriate uncertainty set for an unlabeled target task?

Trade-off needed between robustness and discriminability:



Too large radius: Undistinguishable LFDs!

Class-conditioned domain generalization

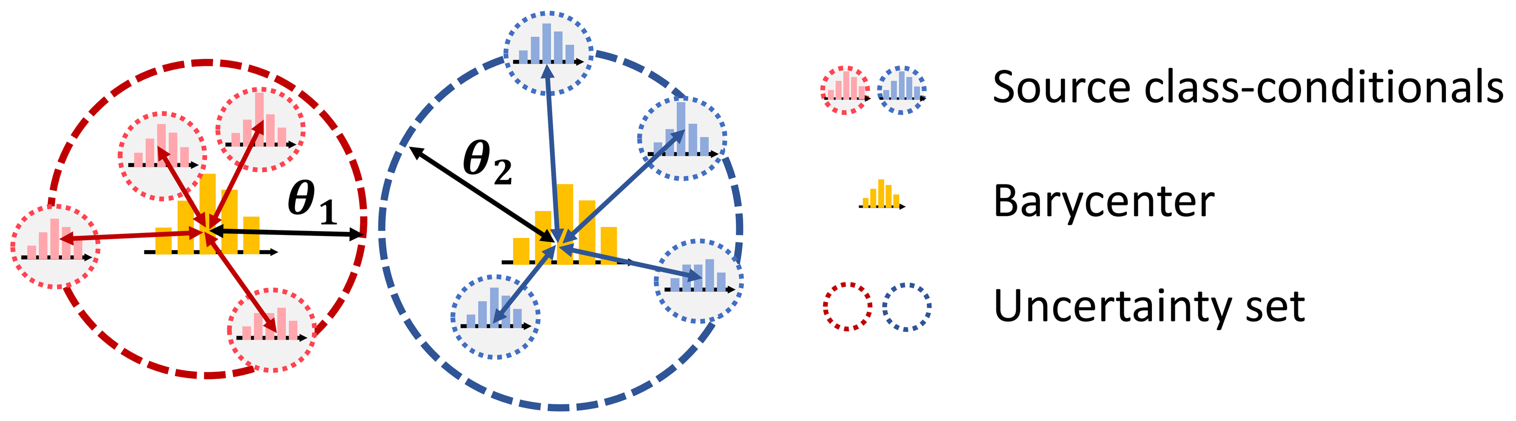
Construct uncertainty sets

Reassign label for source domains

- Estimate clusters for unlabeled target domain using GMM.
- Project source clusters to target clusters using optimal transport.

Estimation of uncertainty sets

- Find Wasserstein barycenter[2] distribution for each class as uncertainty set center.
- Take the maximum distance as uncertain set radius.



Trade-off between robustness and discriminability

In practice, the radius tends to be too large and the optimal class-conditionals may become indistinguishable.

- Add constraint to ensure the LFDs are significantly different characterized by Wasserstein distance:

$$\mathcal{W}_2(P_1, P_2) \ge \gamma$$

Results

Datasets

Synthesis data

- Gaussian-like 2-dimensional class-conditional data with different sample size for 4 source domains and one target domain.

Battery binary dataset

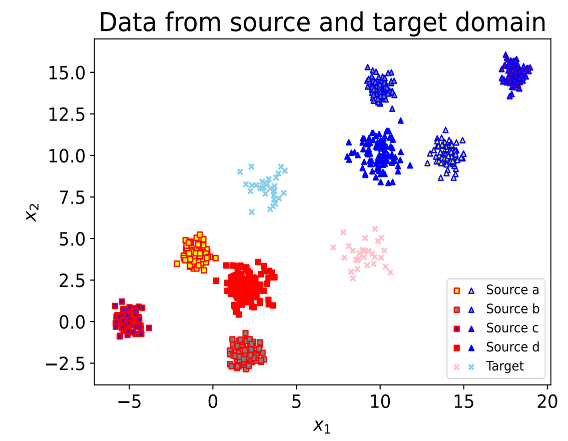
- 5 domains from different testing environments.
- 17-dimensional electric charge-discharge tests data of power batteries.

Results

Baselines

- Single domain:
  - Target only: use k-NN on target training data.
  - Source only: use data from one of the source domains.
- Domain generalization:
  - All sources: simply mix all source training data.
  - All sources + target: mix source and target training data.

Results on synthesis data



	Method	Accuracy
Single domain	Source = a	0.035
	Source = b	0.712
	Source = c	0.212
	Source = d	0.180
Domain generalization	All source (unsupervised)	0.175
	All source + target (semi-supervised)	1
	Proposed method (unsupervised)	**0.878**

- Remark: Our method outperforms method of mixing available source data.

Results on battery binary classification task

	Method	Target = 5	Target = 4	Target = 3	Target = 2	Target = 1	average
Single domain	Source = 1	0.069	0.127	0.061	0.665	0.996	0.384
	Source = 2	0.090	0.230	0.130	0.982	0.956	0.478
	Source = 3	0.945	0.896	0.963	0.408	0.008	0.644
	Source = 4	0.916	0.881	0.939	0.363	0.020	0.624
	Source = 5	0.940	0.889	0.950	0.292	0.011	0.616
Domain generalization	All source (unsupervised)	0.655	0.668	0.561	0.602	0.026	0.502
	All source + target (semi-supervised)	0.718	0.720	0.670	0.848	0.818	**0.755**
	Proposed method (unsupervised)	0.920	0.880	0.913	0.649	0.926	**0.858**

- Remark:
  - Our method outperforms directly mixing available data in both unsupervised and semi-supervised way.
  - The best single domain result outperforms domain generalization, but it requires prior knowledge of target domain.

Conclusion and future work

- Proposed a robust domain generalization method that alleviates adverse impact of domain shift in total absence of target data.
- Realized heuristic tradeoff between robustness and discriminability.
- Future works: improve the significance test and extend it to multi-class classification.

Contact:
wangjg19@mails.tsinghua.edu.cn, yangli@sz.tsinghua.edu.cn
lxie49@gatech.edu, yao.xie@isye.gatech.edu

References:
[1] Rui Gao, Liyan Xie, Yao Xie, and Huan Xu. Robust hypothesis testing using Wasserstein uncertainty sets. In NeurIPS, pp. 7913–7923, 2018.
[2] Courty, N., Flamary, R., Tuia, D., & Rakotomamonjy, A. (2017). Optimal Transport for Domain Adaptation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(9), 1853–1865.