A Theoretical Analysis on Independence-driven Importance Weighting for Covariate-shift Generalization

Renzhe Xu	Xingxuan Zhang	Zheyan Shen	Tong Zhang	Peng Cui
THU	THU	THU	HKUST	THU

General idea

- Recently, independence-driven importance weighting algorithms have shown empirical effectiveness to deal with covariate-shift generalization.
- The theoretical explanations for such effectiveness are still missing.
- In this paper, we propose to explain these algorithms as processes of feature selection.
  - The set of selected features is minimal and optimal to deal with covariate-shift generalization for common loss functions.
  - This variable set is named as the minimal stable variable set.
  - The minimal stable variable set is closely related to the Markov boundary.

Background --- covariate-shift generalization

- Covariate shift
  - Suppose the test distribution P^{te} differs from the training distribution P^{tr} in covariate shift only, *i.e.*,

$$P^{te}(X,Y) = P^{te}(X)P^{tr}(Y|X)$$

- Covariate-shift generalization problem
  - To guarantee the performance on the unknown test distribution P^{te}.
  - P^{te} differs from P^{tr} in covariate shift.

Background --- Independence-driven IW algorithms

- The algorithms usually consist of two steps
  - Importance weighting
    - Learn weight $w(X)$ to make X statistically independent of each other.
  - Weighted least squares

$$\beta_w = \arg\min_{\beta} \mathbb{E}_{P^{tr}}[w(X)(\beta^T X - Y)^2]$$

- To prove the effectiveness of independence-driven IW algorithms, answer:
  - To deal with covariate-shift generalization, what set of variables is optimal?
  - Can independence-driven IW algorithms identify the variable set?

Optimal variable set for covariate-shift generalization

- Observe that $\mathbb{E}_{P^{te}}[Y|X]$ is the optimal solution under several common loss functions in the test distribution P^{te}.

- Properties of the optimal variable set --- minimal stable variable set
  - $\mathbb{E}_{P^{tr}}[Y|X] = \mathbb{E}_{P^{tr}}[Y|S]$
    - A subset of variables $S \subseteq X$ that can fit the target $\mathbb{E}_{P^{te}}[Y|X]$ if and only if it satisfies $\mathbb{E}_{P^{tr}}[Y|X] = \mathbb{E}_{P^{tr}}[Y|S]$.
  - the minimal set of variables that satisfies $\mathbb{E}_{P^{tr}}[Y|X] = \mathbb{E}_{P^{tr}}[Y|S]$
    - To relieve the negative impact of other variables in the test distribution

Independence-driven IW algorithms: feature selection

- Recall the process of Independence-driven IW algorithms
  - Importance weighting
    - Learn weight $w(X)$ to make X are statistically independent of each other.
  - Weighted least squares

$$\beta_w = \arg\min_{\beta} \mathbb{E}_{P^{tr}}[w(X)(\beta^T X - Y)^2]$$

- Use β_w to select features
  - Select feature X_i if the corresponding coefficient $\beta_w(X_i)$ is not 0.

Independence-driven IW algorithms: feature selection

- Under ideal conditions (perfectly learned sample weights, and infinite samples),
  - if a variable X_i is not in the minimal stable variable set, then independence-driven IW algorithms could filter it out with any weighting function that satisfies the independence condition, and
  - if a variable X_i is in the minimal stable variable set, then there exist weighting functions with which independence-driven IW algorithms could identify X_i.

- We further provide non-asymptotic properties when the ideal conditions are not satisfied.

Experimental results

- The minimal stable variable set is optimal for the covariate-shift generalization.
- Independence-driven IW algorithms (DWR [1] and SRDO [2]) could identify the minimal stable variable set better than other baselines.

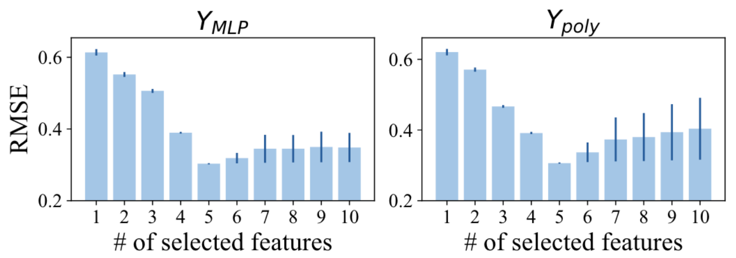
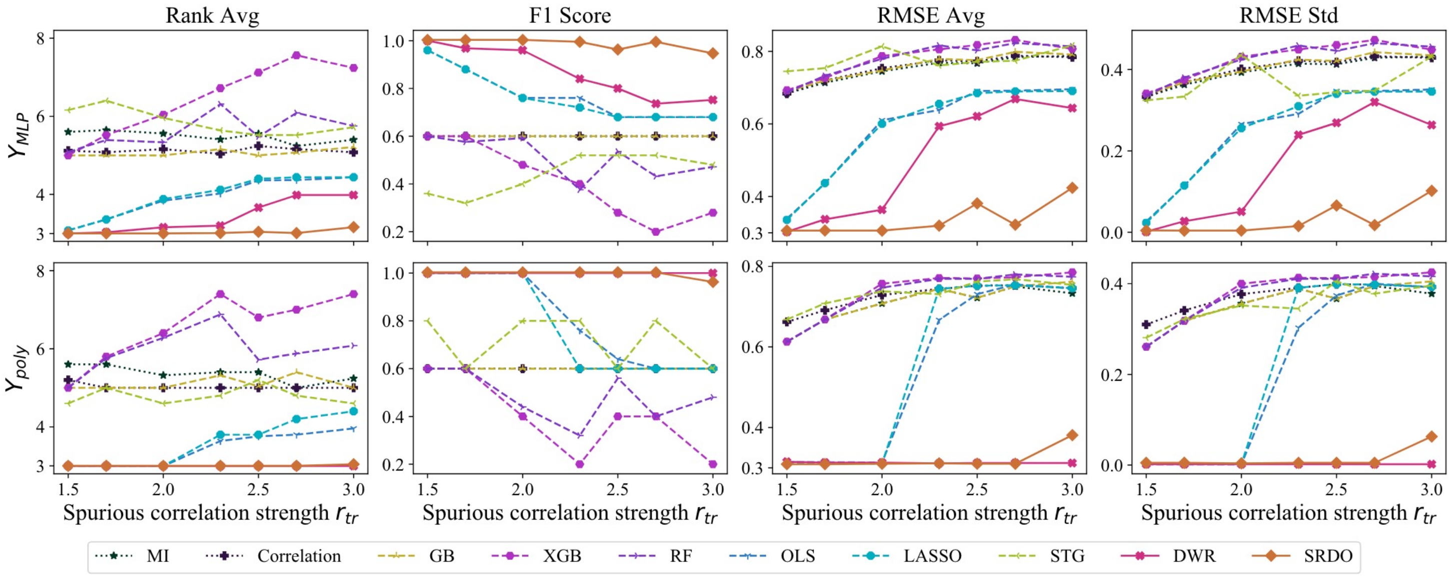


Figure 2. The covariate-shift generalization metrics (RMSE average and standard deviation) *w.r.t.* the number of selected features. Fix $r_{tr} = 2.5$ here and the feature ranking lists are provided by SRDO. The minimal stable variable set (5 features) achieves the optimal performance.



[1] Kuang, et al. Stable prediction with model misspecification and agnostic distribution shift. AAAI. 2020.
[2] Shen, et al. Stable learning via sample reweighting. AAAI. 2020.

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

ICML
International Conference On Machine Learning

Thanks!

Paper is available at https://arxiv.org/abs/2111.02355

Code is available at https://github.com/windxrz/independence-driven-IW