Mediated Uncoupled Learning: Learning Functions without Direct Input-output Correspondences

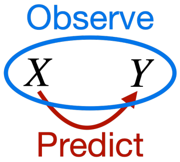
Ikko Yamane[1 2]   Junya Honda[3 2]
Florian Yger[1 2]   Masashi Sugiyama[2 4]

[1] LAMSADE, PSL Université Paris-Dauphine, Paris, FRANCE
[2] RIKEN AIP, Tokyo, JAPAN
[3] Kyoto University, Kyoto, JAPAN
[4] The University of Tokyo, Tokyo, JAPAN

The 38th International Conference on Machine Learning (ICML 2021)
July 22, 2021

Limitation of standard supervised learning

- Standard supervised learning
  - Goal: predict output r.v. Y from input r.v. X
  - Requires: (X, Y)-pairs

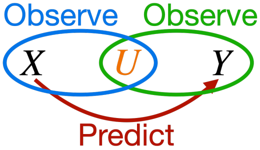


- However, they can be difficult to collect
  - Annotating images X with sentiment labels Y
  - Collecting translations between minor languages X and Y
  - Labeling low-quality images X (Y being the class labels)

Our Setup: Mediated Uncoupled Learning

- *Mediated uncoupled learning*
  - Goal: predict output r.v. Y from input r.v. X
  - Requires: (X, U)-pairs and (U', Y')-pairs with some r.v. U

    (U', Y') and (U, Y) are independent but follow the same distribution



- It can be easier to collect such data:
  - (image X with text caption U) & (text U', sentiment label Y')
  - (text in minor language X, translation in a major language U) & (translation in a major language U', text in minor language Y)
  - (low-quality image X, its high-quality version U) & (high-quality image U', class label Y')

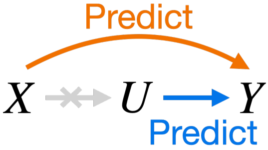
Naive Method

$$X \xrightarrow[\text{Predict}]{} U \xrightarrow[\text{Predict}]{} Y$$

1. Learn $X \longrightarrow U$ as $\widehat{u}(x)$ using (X, U)-pairs
2. Learn $U \longrightarrow Y$ as $\widehat{y}(u)$ using (U, Y)-pairs

- Prediction: $\widehat{y}(\widehat{u}(X))$
- However, it is statistically inconsistent
  - Generative models with numerical integration are necessary to correct the inconsistency
  - which needs delicate modeling and training
  - and prediction will be computationally expensive
- When U is complex such as images or texts, learning $X \longrightarrow U$ is not easy

Proposed Method



- Avoid predicting U which causes the issues
- *Two-step Regressed Regression* (*2Step-RR*)

1. Learn $U \longrightarrow Y$ as $\widehat{y}(u)$ using (U, Y)-pairs
2. Learn $X \longrightarrow \widehat{y}(U)$ as $\widehat{y}_{\mathrm{RR}}(u)$ using (X, U)-pairs

- Statistical consistency and an error bound when $\mathbf{E}[Y \mid U, X] = \mathbf{E}[Y \mid U]$

Experiments on Low-quality Image Classification

- We use downsampled images for X and the original ones for U
- U-Net (Ronneberger et al. 2015) for $X \longrightarrow U$
- ResNets (He et al. 2016) for $U \longrightarrow Y$ and $X \longrightarrow Y$
- *Joint-RR*: variant that jointly performs the two steps of 2Step-RR

Y: “automobile

X

U

Dataset	Naive	2Step-RR	Joint-RR
MNIST	88.06% (1.13)	**96.19%** (0.26)	**96.34%** (0.28)
Fashion-MNIST	73.12% (0.37)	85.53% (0.16)	**86.93%** (0.14)
CIFAR-10	48.35% (0.28)	67.60% (0.13)	**69.10%** (0.11)
CIFAR-100	19.97% (0.12)	27.43% (0.08)	**28.05%** (0.08)

Summary

- Learning with separate observations of X and Y with a mediating variable U
- The proposed approach that avoids predicting U outperforms the naive method
- Joint training can further improve the performance
- Latest version: `https://arxiv.org/pdf/2107.08135.pdf`
- Our code can be found at `https://github.com/i-yamane/mediated_uncoupled_learning`