# Domain Adaptation with Optimal Transport for Extended Variable Space

**Toshimitsu Aritake[1], Hideitsu Hino[1, 2]**

[1]The Institute of Statistical Mathematics, [2]RIKEN AIP

## Abstract

### Domain Adaptation

In supervised learning, the accuracy of predictions degrades when the distributions of the training and test data are different. Domain adaptation aims at learning a model from training data that well performs on the test data.

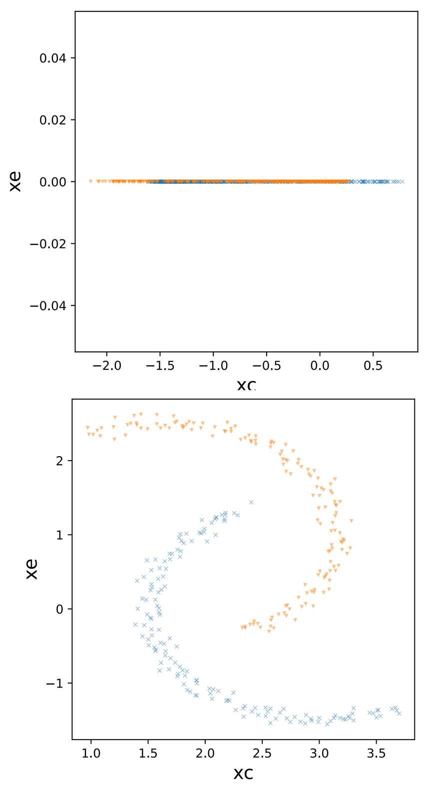
### Problem Setup

We consider the case where new features are observed for test data; hence, the distribution of training and test data have different dimensionalities. We also assume the labels are given only for the training data (unsupervised domain adaptation).

### Methods & Results

We consider bi-directional optimal transport (OT) to transfer label information from the training data to the test data. Our proposed method enables the use of extra features of test data for better prediction.

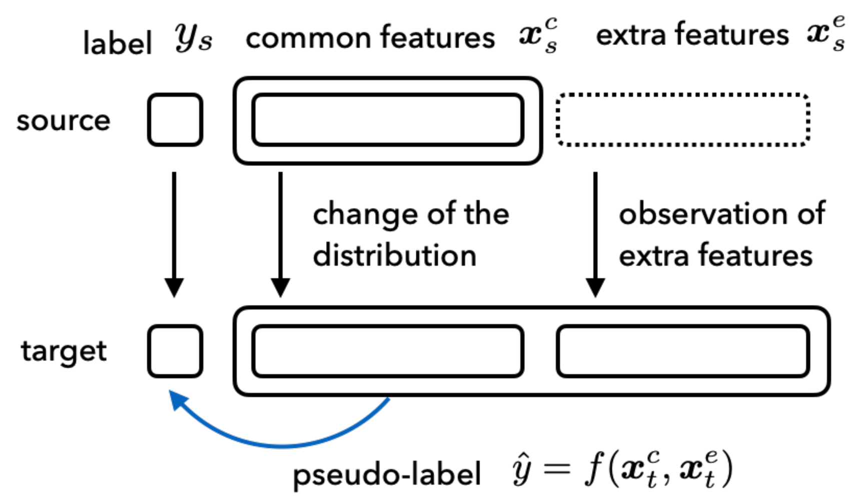
Experimental results show that the proposed method successfully estimate the good decision boundary using extra features.




## Method

### Problem

- Common variables are observed in both the source and target domains
  - Distribution of common variables may differ
- Extra variables are observed in the target domain

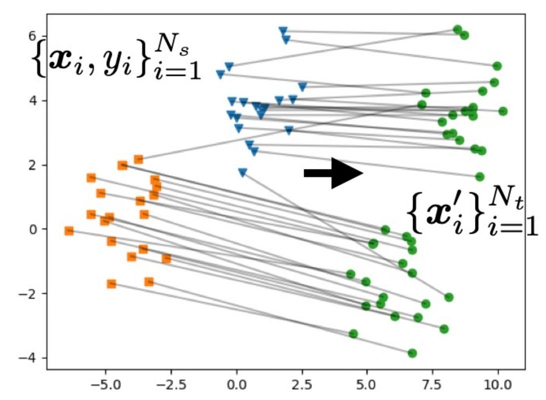



### Joint Distribution Optimal Transport (JDOT)

$$\hat{\pi}^* = \arg\min_{f \in \mathcal{F}, \hat{\pi} \in \hat{\Pi}(\mathcal{D}_S, \mathcal{D}_T)} \sum_{i,j} \hat{\pi}_{ij} \mathcal{E}_{\alpha,ij} \quad (1)$$

$$\hat{\Pi}(\mathcal{D}_S, \mathcal{D}_T) \equiv \left\{ \pi \in \mathbb{R}_+^{N_s \times N_t} \,\middle|\, \sum_{i=1}^{N_s} \pi_{ij} = \frac{1}{N_t}, \sum_{j=1}^{N_t} \pi_{ij} = \frac{1}{N_s} \right\}$$

- Minimize the sum of transportation cost
  - Use distance between variable and label discrepancy as transportation cost (Joint Distribution Optimal Transport; [1]) $\mathcal{E}_{\alpha,ij} = \alpha \|x_{si}^c - x_{tj}^c\|_2^2 + \mathcal{L}(y_{si}, \hat{y}_{tj} = f(x_{tj}^c))$
  - Use the output of the model $f$ as a pseudo-label of the target label
- Train the model using the transported source labels




### Proposed Method

1. Transfer the target extra features to the source domain
$$\mathcal{E}_{\alpha,ij} = \alpha \|x_{si}^c - x_{tj}^c\|_2^2 + \mathcal{L}(y_{si}, f(x_{tj}^c, x_{tj}^e)) \quad (2)$$
2. Transfer the source labels to the target domain
$$\mathcal{E}_{\alpha,ij} = \alpha(\|x_{si}^c - x_{tj}^c\|_2^2 + \|\hat{x}_{si}^e - x_{tj}^e\|_2^2) + \mathcal{L}(y_{si}, f(x_{tj}^c, x_{tj}^e))$$

- **Assumption** $\mathcal{P}_S(x^e | x^c, y) = \mathcal{P}_T(x^e | \mathcal{T}(x^c, y))$
  - The distribution of extra features given common features is identical for each class before and after OT
  - There always exists the destination of OT whose transportation cost of extra variables is zero
- **The above bi-directional OT is reduced to the unidirectional OT from the source to the target domain**

### Algorithm

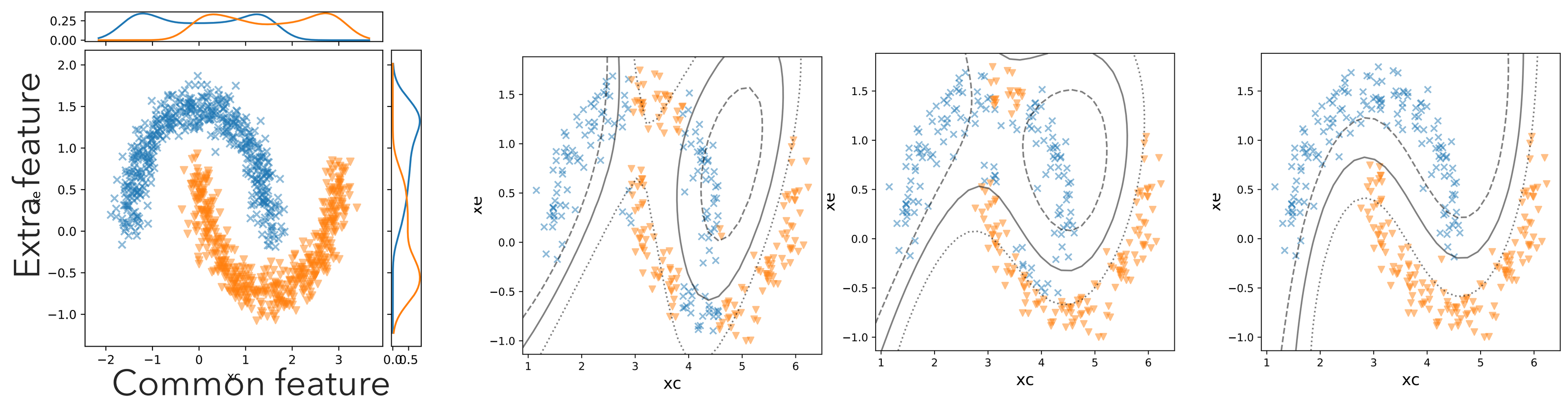
Repeat the following two steps until the stopping criterion is met

1. Solve the joint distribution OT problem Eq. (1), and transport the source label information to the target domain
   - At the initial iteration, the distance between features is used as transportation cost $\mathcal{E}_{\alpha,ij} = \alpha \|x_{si}^c - x_{tj}^c\|_2^2$
   - Label discrepancy is added to the cost function after the initial iteration, $\mathcal{E}_{\alpha,ij} = \alpha \|x_{si}^c - x_{tj}^c\|_2^2 + \mathcal{L}(y_{si}, f^{(n)}(x_{tj}^c, x_{tj}^e))$
2. Train the model $f$ using the transported source label
   - Barycentric mapping is used to determine target labels to train the model $(X^t, (\mathrm{diag}(\hat{\pi}^\top \mathbf{1}))^{-1} \hat{\pi}^\top \hat{Y}^s)$
   - $\hat{Y}_s$ : a matrix whose rows are one-hot encoded source labels

## Experiments

### Artificial Data

- Our proposed method successfully learned a good prediction model in the target domain that uses extra features
  - Incorrectly transported labels are modified iteratively




### Experiment with Realistic Dataset

- Gas Sensor Array Drift Dataset[2] is used
  - Binary classification of the types of gasses
  - Original data is timeseries of sensor values and transient features are extracted from original data
  - Each domain has different degree of deterioration, which results in domain shift
- Comparison with baseline (no domain adaptation) and prior works [1, 3, 4]

| domains | Baseline | JDOT no extra | CCA | DSFT | Proposed | JDOT ideal |
|---|---|---|---|---|---|---|
| 1→ 2 | **83.33** | 77.71 | 66.87 | 42.97 | 78.31 | 83.73 |
| 1→ 3 | 52.28 | 93.45 | 43.27 | 57.78 | **96.02** | 94.15 |
| 1→ 4 | 64.49 | 60.75 | 58.88 | **94.39** | 87.85 | 85.98 |
| 2→ 1 | 52.13 | 79.26 | 56.91 | 62.77 | **84.04** | 85.64 |
| 2→ 3 | 56.84 | 89.36 | 25.03 | 84.56 | **89.47** | 90.99 |
| 2→ 4 | 63.55 | 69.16 | 51.40 | 41.12 | **71.96** | 74.77 |
| 3→ 1 | 51.06 | 92.02 | 92.55 | 66.49 | **94.15** | 95.74 |
| 3→ 2 | 68.67 | 81.92 | 67.67 | 86.94 | **88.76** | 88.55 |
| 3→ 4 | 94.39 | 77.57 | **96.26** | 40.19 | 81.31 | 80.37 |
| 4→ 1 | 50.00 | 52.66 | 48.93 | 51.60 | **53.72** | 80.85 |
| 4→ 2 | 42.97 | 71.08 | 32.93 | 32.93 | **74.30** | 73.89 |
| 4→ 3 | **92.98** | 82.81 | 57.31 | 42.69 | 82.57 | 82.57 |

## Conclusion

- Domain adaptation method to use the extra variables given for the target domain is proposed
- Bi-directional OT for domain adaptation is reduced to the unidirectional OT under the assumption $\mathcal{P}_S(x^e | x^c, y) = \mathcal{P}_T(x^e | \mathcal{T}(x^c, y))$
- Upper-bound on the target error is provided for the proposed algorithm

## References

[1] Courty, et al., "Joint distribution optimal transportation for domain adaptation", Adv. Neural Inf. Process. Syst. (2017)
[2] Vergara, et al., "Chemical gas sensor drift compensation using classifier ensembles", Sensors and Actuators B: Chemical (2012)
[3] Yeh, et al., "Heterogeneous domain adaptation and classification by exploiting the correlation subspace", IEEE Trans. Image Process. ( 2014)
[4] Wei, et al., "A General Domain Specific Feature Transfer Framework for Hybrid Domain Adaptation", IEEE Trans Knowl Data Eng., (2019)