Cross-Domain Cross-Task Transferability Estimation

Yang Tan, TBSI Lab2C

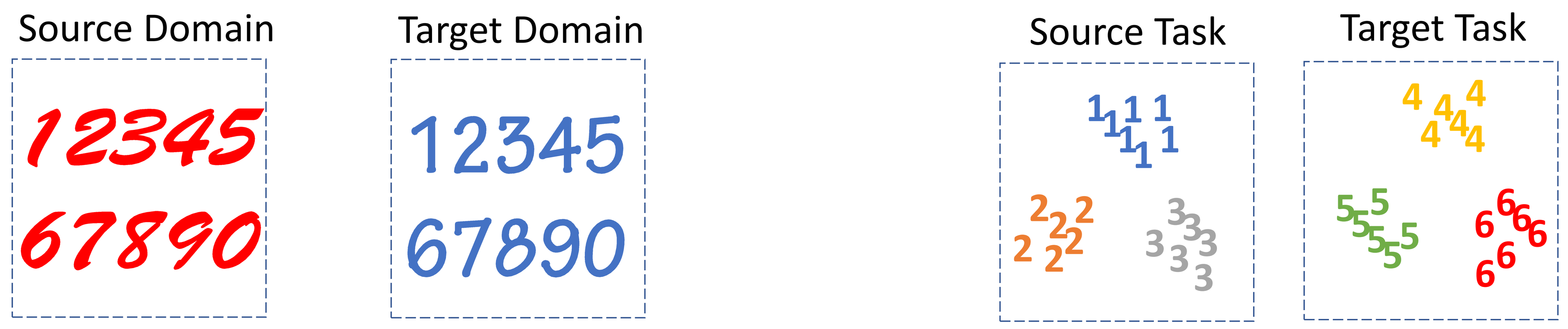
Advisor: Prof. Yang Li & Prof. Shao-Lun Huang

TBSI-WOLT The 2nd TBSI Workshop on Learning Theory

TBSI 清华-伯克利深圳学院 Tsinghua-Berkeley Shenzhen Institute

Abstract

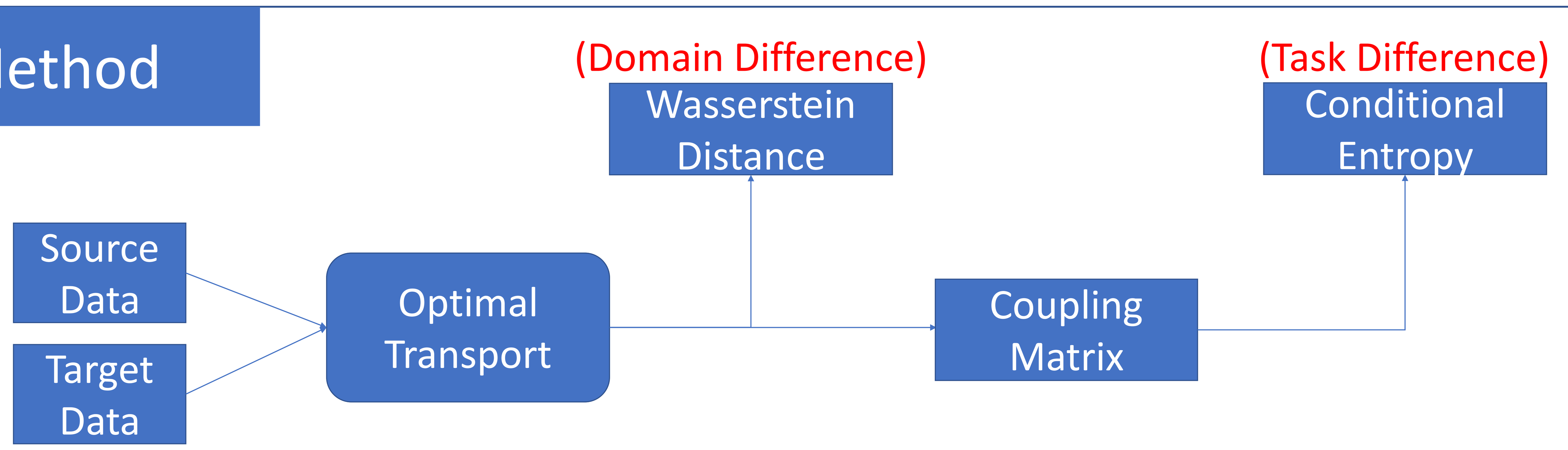
Transferability can reveal how easy it is to transfer knowledge learned from one classification task to another by observing the differences of data. Usually, there exists **domain difference** and **task difference** between source and target tasks. Previous works did not decompose the total difference[1] or just ignored the domain difference[2,3] leading to inaccurate estimation and limited application scenarios. To address this problem, we propose a pipeline adopting Wasserstein distance to evaluate domain difference and Conditional Entropy to represent task difference. Experiments using hand-written digit recognition and image classification datasets have demonstrated the effectiveness of our proposed method on simple tasks.



Cross Domain

Cross Task

Method



- Two datasets

$$D_A = (x_A^i, y_A^i)_{i=1}^m \sim P_A(x,y) \quad D_B = (x_B^j, y_B^j)_{j=1}^n \sim P_B(x,y)$$

- Domain difference

$$OT(D_A, D_B) \triangleq \min_{\pi} \int_{\mathcal{X}\times\mathcal{X}} c(x_A, x_B) d\pi(x_A, x_B) + \epsilon H(\pi)$$

$$c(x_A^i, x_B^j) = \|x_A^i - x_B^j\|_2^2$$

$$W(D_A, D_B) = \sum_{i,j=1}^{m,n} \pi^*(x_A^i, x_B^j) c(x_A^i, x_B^j)$$ (Domain Difference)

- Task difference

$$\pi^* = \begin{bmatrix} P(x_A^1, x_B^1) & P(x_A^1, x_B^2) & \dots & P(x_A^1, x_B^n) \\ P(x_A^2, x_B^1) & P(x_A^2, x_B^2) & \dots & P(x_A^2, x_B^n) \\ \vdots & \vdots & \ddots & \vdots \\ P(x_A^n, x_B^1) & P(x_A^n, x_B^2) & \dots & P(x_A^n, x_B^n) \end{bmatrix}$$ Coupling Matrix

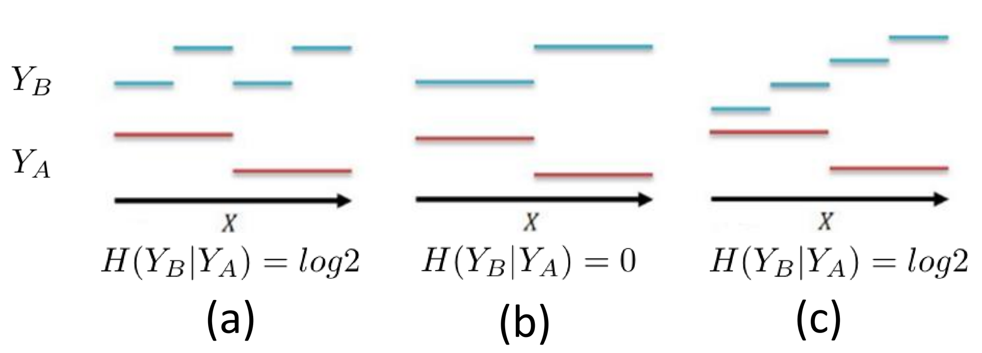
$$\{((x_A^k, y_A^k), (x_B^l, y_B^l))\}_{k,l}^n$$ Building **n** pairs of correspondences

$$\hat{P}(y_A, y_B) = \frac{1}{n}|\{(k,l) : y_A^k = y_A \ and \ y_B^l = y_B\}|$$

$$\hat{P}(y_A) = \sum_{y_B \in \mathcal{Y}_B} \hat{P}(y_A, y_B)$$ Empirical Distribution

$$H(Y_B|Y_A) = H(Y_B, Y_A) - H(Y_A)$$ (Task Difference)

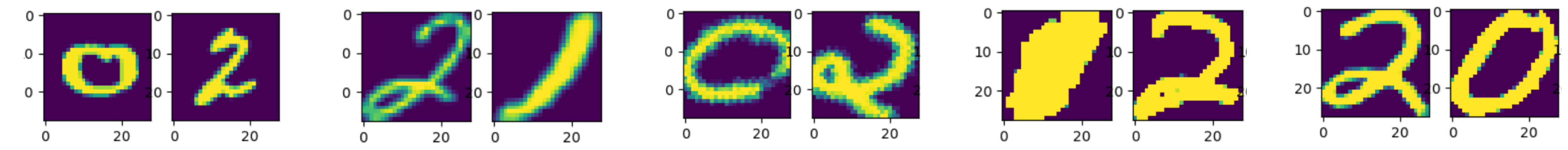
$$= -\sum_{y_B \in \mathcal{Y}_B} \sum_{y_A \in \mathcal{Y}_A} \hat{P}(y_A, y_B) log \frac{\hat{P}(y_A, y_B)}{\hat{P}(y_A)}$$



Task Difference: (a) = (c) > (b)

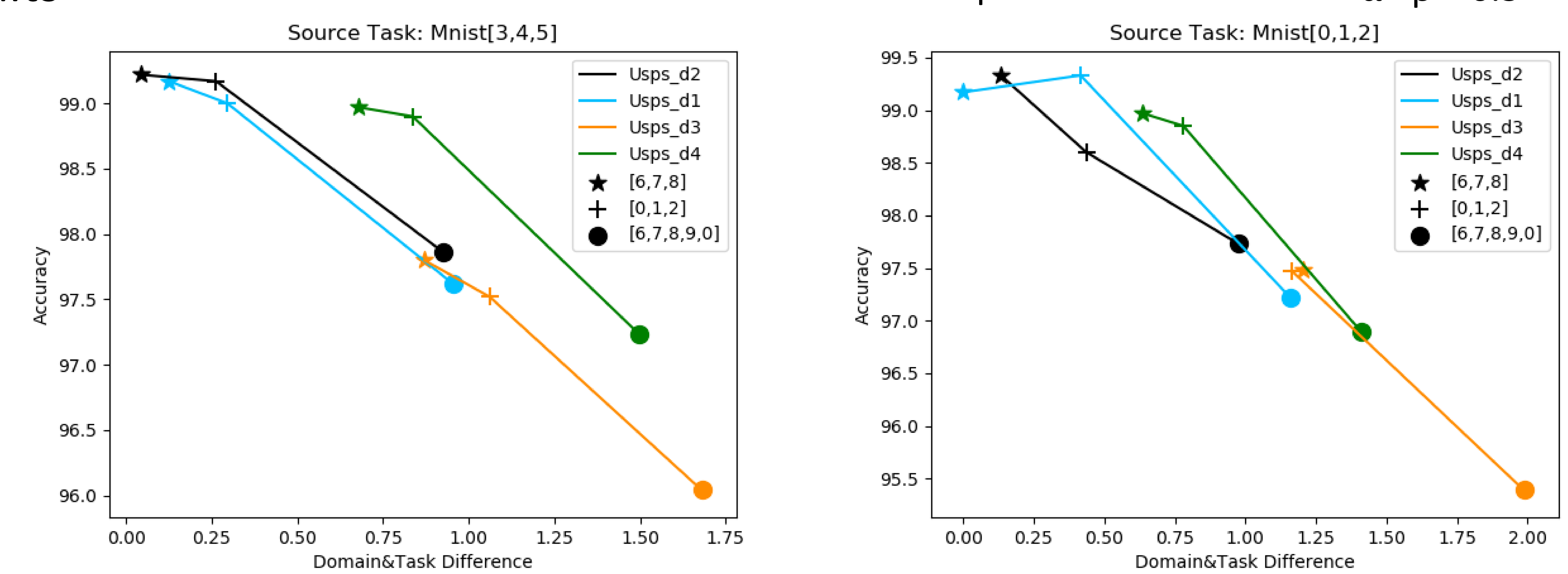
Experiment I

- Source tasks
  - Domain: **Mnist**
  - Categories: **[0,1,2], [3,4,5]**
- Target tasks
  - Domain: **Usps_d1, Usps_d2, Usps_d3, Usps_d4**
  - Categories: **[0,1,2], [6,7,8], [6,7,8,9,0]**



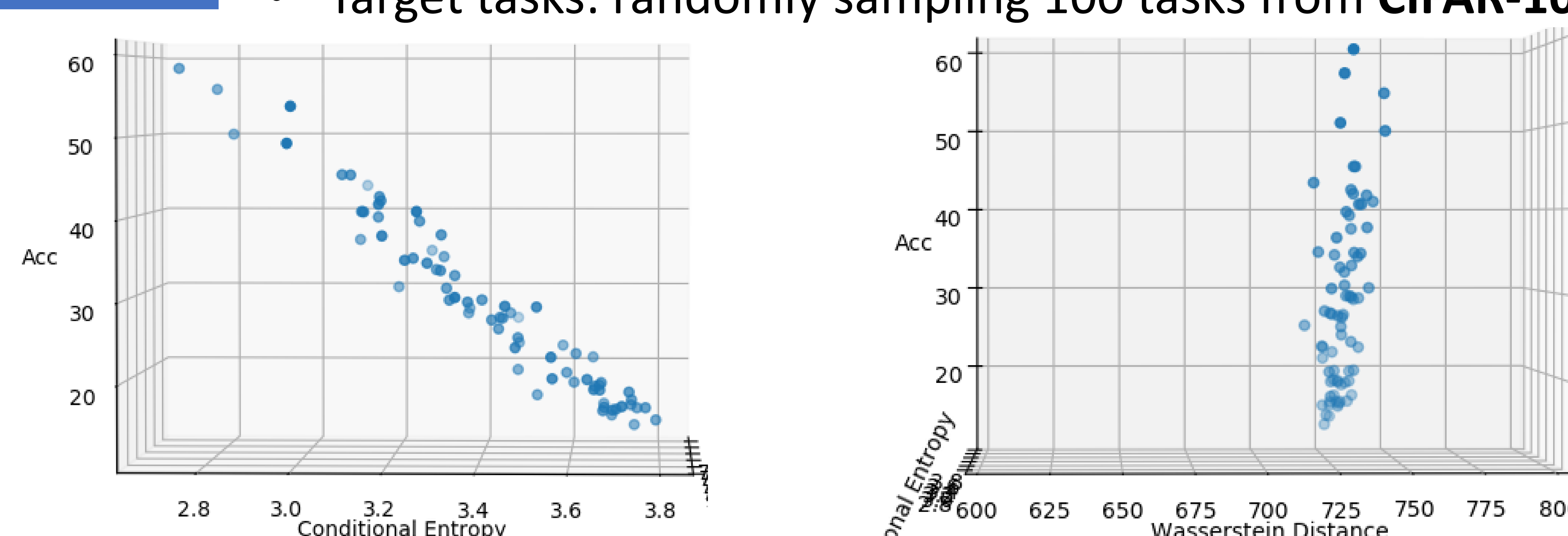
Visualizations of samples in Mnist, Usps_d1, Usps_d2, Usps_d3, Usps_d4 respectively.

- Results Domain&Task Difference = α · Domain difference + β · Task Difference α= β = 0.5



Experiment II

- Source task: 1000-classification on **ImageNet**
- Target tasks: randomly sampling 100 tasks from **CIFAR-100**



In this experiment, transferability is mainly revealed by task difference as expected since all target tasks are sampled from one dataset (domain). Meanwhile, the domain differences are stable in a small range.

Conclusion

- Basic experiments have shown the effectiveness of combining Wasserstein distance and conditional entropy for estimating transferability.
- Future works: strengthening the theoretical interpretations.

Reference

[1] Nguyen, Cuong V., et al. "LEEP: A New Measure to Evaluate Transferability of Learned Representations." *arXiv preprint arXiv:2002.12462* (2020).

[2] Bao, Yajie, et al. "An Information-Theoretic Approach to Transferability in Task Transfer Learning." *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019.

[3] Tran, Anh T., Cuong V. Nguyen, and Tal Hassner. "Transferability and hardness of supervised classification tasks." *Proceedings of the IEEE International Conference on Computer Vision*. 2019.