1 Summary

Methods such as MMD look for a transformation T that find a shared feature representation of the input data and then aligning the distributions, ie, $P_s(T(X)) -> P_t(T(X))$. JDOT however works by finding a an optimal coupling distribution T that minimizes the cost of directly aligning the **joint** distribution $P_s(X, Y)$ with $P_t(X, f(X))$, where f is the classification function on target domain which is simultaneously learned since target has no labels.

2 Strengths

- Classic Domain discrepancy measures like KL divergence and MMD require that the source and target distributions $P(X)$ share a common support. Wasserstein metric alleviates this shared support requirement.
- The papers outlines a very general approach to Domain Adaptation which can be used with a large class of optimization functions like kernel SVMs and neural networks.
- The method takes into account both marginal as well as conditional distribution shifts across domains.

3 Weaknesses

- The optimization is computationally expensive with calculation of $N_s N_t$ number of terms. Only for some specific Loss functions can this optimization be reduced to $O(N_t)$ computations (like least square and hinge losses as they have mentioned).
- For simplification, the paper assumes that the joint cost measure is separable in features and label spaces, ie,

$$D(x_1, y_1; x_2, y_2) = \alpha d(x_1, x_2) + L(y_1, y_2)$$

  It does not solve this problem for any generic joint cost function.

4 Critique of Experiments

- The authors first test their model on Caltech-Office dataset for classification. They compare their model with 4 other standard Domain adaptation methods. All of the methods use SVM with linear kernel for the task. It is observed that JDOT consistently outperforms naive baseline but does not consistently beat other SOTA methods. However, the reported p-values show that JDOT is statistically better than other methods.
- **JDOT** is also tested on Amazon review classification dataset and results are compared against the SOTA **DANN** neural network. JDOT beats DANN by 2.4% and this shows the versatility of JDOT which can be adopted to any type of classifier.
- The authors also test their approach for regression problem on Wifi localization dataset and compare their results to 7 other SOTA methods. For domain transfer across periods, JDOT performs at par with other methods. However, for transfer across devices task, JDOT beats other methods by upto 10% which shows it can handle large shifts in distribution much better than other methods.

5 Follow Ups/Extensions

- Since JDOT seems like a versatile adaptation method, incorporating this schema into a CNN seems like a good idea for more powerful adaptation performances. Joint distributions of the CNN activations from deeper layer can be used for aligning source and target domains.
- The paper assumes that joint cost measure is separable in feature and label spaces. We can extend this method for generic cost measure by minimizing the cost function using gradient descent in neural net.