1 Title:
Maximum Classifier Discrepancy for Unsupervised Domain Adaptation

2 Summary:
This paper proposes to utilize classifier outputs to align distribution of features so that the task specific decision boundaries are respected during domain adaptation. Concretely, two task specific classifiers are trained to correctly classify the downstream task for source domain features, and in parallel, maximize discrepancy between their outputs for target domain features. Then, a feature extractor is trained to minimize this classifier discrepancy. Optimality of this procedure requires the feature extractor to produce features in a shared support so that the maximum discrepancy on target features is minimized. In turn, shared support ensures that the class boundaries are well respected during domain adaptation.

3 Strengths:
i) Theoretical justification is provided and $H\Delta H$ bound corresponds well with the mini-max objective for the discrepancy between the two task-specific classifiers ii) Domain adaptation respects the class boundaries by implicitly forcing a common support for source and target features, unlike previous approaches which do not consider class boundaries at all. iii) Complete distribution matching between features is not required and useful when such objective is impossible (when domain discrepancy is large)

4 Weaknesses:
i) An important hyper-parameter is the balancing of maximizing domain discrepancy and classifying source examples reasonably correctly. Maximizing domain discrepancy can negatively effect the source domain classification and vice-versa. This effect is not discussed. ii) When training the feature extractor to minimize discrepancy using target samples only, the source feature distribution invariably changes because of weight-sharing. So, the convergence of this approach is in question. iii) Complete distribution matching between features is not achieved even when feasible. Other approaches outperform in this case (Table 4.)

5 Analysis of Experiments:
i) The results on Toy dataset in Figure 4., clearly demonstrate the algorithm. Figure 4b. shows maximum classifier discrepancy, and 4c. shows the effect of minimizing it. ii) Table 1. demonstrates that proposed approach outperforms contemporary approaches, namely DSN, DANN etc. considerably. SVHN to MNIST task which has maximum discrepancy visually, shows marked improvement justifying the algorithm. Other approaches do not necessarily ensure features form the same support in contrast and hence perform poorly. iii) VisDA results in Table 2. and Semantic Segmentation results in Table 4. and Table 5. show that this approach doesn't outperform DANN in every task. Possibly, DANN is more suited for tasks that do not have a huge domain discrepancy, and vice-versa.

6 Possible Extensions:
Although, not theoretically directly justified, Table 4. empirically suggests that DANN objective can be combined with this objective in order to achieve better adaptation. Intuitively, the proposed approach ensures the supports are matched while DANN ensures that the distributions are matched.