1 Title:
Weighted Maximum Mean Discrepancy for Unsupervised Domain Adaptation

2 Summary:
This paper proposes to extend MMD based approaches to the case where prior probabilities are different across domains. A CEM framework is proposed to alternatively generate pseudo-labels for target, re-weight MMD based on prior probabilities and optimize the network. Experimental results on Office-31, Digit Classification and ImageCLEF indicate an improved performance over non-reweighted MMD on different network architectures.

3 Strengths:
i) Handles both label shift and covariate shift unlike previous approaches. In many practical cases, such as in Figure 1. where house numbers naturally have more '1's and '2's as compared to MNIST where no such preference exists, this method is helpful.ii) Strengths of DAN including multi-kernel, multi-layer and O(n) approximation for MMD carry forward.iii) Classification step is computationally inexpensive because it just needs one forward pass to classify and assign weights. So, the overhead over DAN per optimization step is minimal

4 Weaknesses:
i) Does not handle imbalanced data, i.e. when support of target classes is different from support of source classes. ii) Theoretical properties such as convergence of CEM, variance of the estimators etc. are not discussed iii) Weaknesses of DAN such as computational overhead of choosing the optimal kernel carry forward.iv) Possible misreporting of DAN accuracy in Table 4.

5 Analysis of Experiments:
i) Table 1.,2.,3.,4. show improvement over DAN using different architectures such as GoogLeNet, LeNet and AlexNet. So, the method is not very much dependent on architecture. However, dependency might still exist as the percentage improvement isn't the same. ii) Table 3. suggests that digit classification/adaptation over MNIST, SVHN is a difficult task and maximum improvement is shown in that task (53.5% to 57.2%). So, weighting class priors during MMD is very helpful in this case. iii) Figure 4. clearly brings out the advantage of this method by showing a decay in performance of DAN by increasing class bias while performance of WDAN is more stable. iv) T-SNE visualizations show improved discriminativeness over DAN adding more credibility to the approach.

6 Possible Extensions:
i) Varying the support of priors across source and target: One idea is to generate an 'unknown class' label and use the 'confidence' in predictions as proxy for unknown class during classification maximization. This handles the case when target has unseen classes.
ii) Covariate assumption can be further relaxed by allowing the conditional distributions to vary as well. This leads to domain dependent networks, i.e., network weights will not be shared at some point. However, we can utilize the pseudo labels as proposed by this method, or strengthen these pseudo labels by ensembling, to train separate networks while preserving domain invariance using MMD.