Divyanshu Mund
A53321339

Contrastive Adaptation Network for
Unsupervised Domain Adaptation

2020/02/14

1 Summary

The paper tackles the issue of Domain adaptation by defining a new Domain Discrepancy metric which takes class information into account while aligning different domains. Authors propose Contrastive Adaptation Network and end-to-end training framework which involves alternatively optimizing psuedo target labels and feature representations of the neural net. The model achieves SOTA performance on Office-31 dataset and performs comparable to second ranked model on VisDA-2017 benchmark.

2 Strengths

- Proposes novel Contrastive Domain Discrepancy metric which minimizes the intra-class domain discrepancy as well as maximize the inter-class domain discrepancy.
- Ablation studies section does a great job in explaining the impact of various methodologies that the paper uses. The importance of Alternate Optimization, Class Aware sampling and Inter class domain discrepancy term etc are clearly shown.

3 Weaknesses

- The method aligns the conditional data distributions across domains. The model will still be susceptible to class weight bias unless it is explicitly taken into account.
- There are many hyperparameters in the model and it doesn't seem like it is very robust to hyperparameter tuning. For example $b = 0.75$, works for Office, while $b = 2.25$ works for VisDA-2017. Also, the hyperparameters (D_0, N_0) should be varied to study their effect on model performance.

4 Critique of Experiments

- Authors use a pre-trained ResNet-50 for training their CAN model and compare the performance with standard DA methods like DAN, JAN and MADA. It is found that CAN beats all methods by over 5% margin. t-SNE plots reveal that CAN demonstrate higher intra-class compactness and much larger inter-class margin compared to JAN which shows that CDD is a better metric for producing more distinctive features for target domain.
- Pre-trained ResNet-101 is taken for training CAN model on VisDA-2017 dataset. The models achieves test accuracy performance comparable to second ranked model.
- Training the network without alternatively optimizing the psuedo target labels using K-means and CDD loss degrades the model performance. However, the model still outperforms class-agnostic methods which reveals that CAN is noise robust to some extent.
- It is also shown that inter-class domain discrepancy term in CDD loss and Class aware sampling enhances model performance.
- Plots for ground truth CDD vs epochs during training are plotted for JAN and CAN models. The plots reveal that CDD metric for CAN steadily decreases with epochs while JAN cannot minimize the contrastive domain discrepancy effectively. This, however, is very surprising since CAN explicitly reduces CDD loss.

5 Follow Ups/Extensions

- Like in JAN, we can try and align the joint distribution distributions across domains by using the tensor product from the layers and using the CDD metric.
- Data augmentation and Self ensembling methods can be used to further improve generalization performance of the given model.