Domain Adaptation summary

E. Kondyli, Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr
Version 2.5.4

Artificial Intelligence &
Information Analysis Lab

Domain Adaptation

- Domain Adaptation (DA) has as an objective to eradicate the dissimilarity between the way the labeled data is distributed across the source domain and the unlabeled or partially labeled in the target domain.

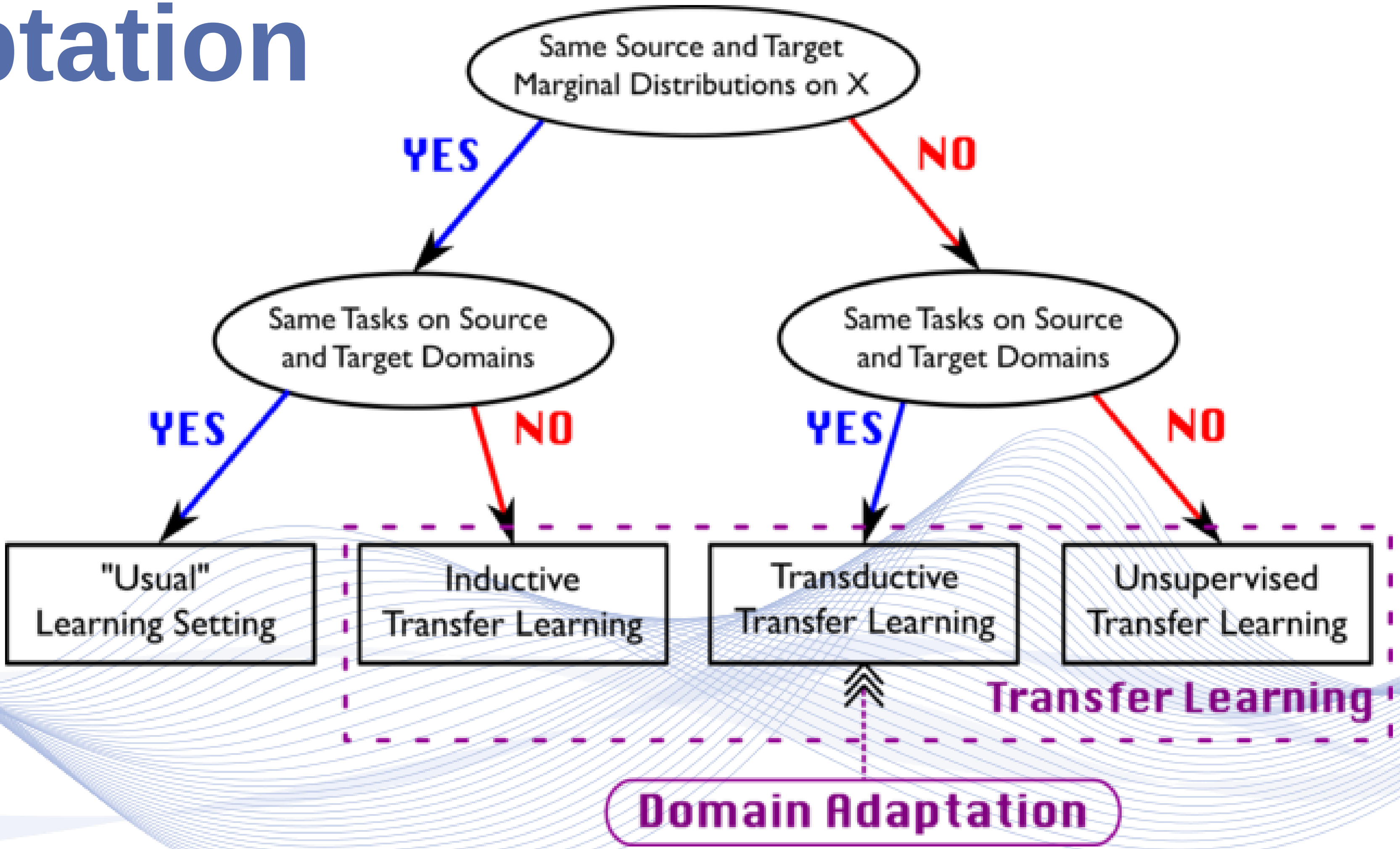
- The classifier is trained on the source domain.

- The classifier is applied on the target domain.

Domain Adaptation

- The types of Domain Adaptation differ because of the information that is considered for the target task. The DA types are the following.
  - Unsupervised Domain Adaptation (UDA). Zero labels are available.
  - Semi-supervised Domain Adaptation (SSDA). Few labeled data from target domain is used in the training of the classifier.
  - The Supervised Domain Adaptation (SDA). All the target data has to be labeled.

Positioning of Domain Adaptation



Domain Adaptation

- **Domain Shift**
- Unsupervised Domain Adaptation
  - Domain-specific Whitening Transform
  - Min-entropy Consensus loss
  - Maximum Classification Discrepancy
  - Sliced Wasserstein Discrepancy
- Deep Learning methods for Unsupervised Domain Adaptation
  - DLID: Deep learning for DA by Interpolating between Domains
- Semi-supervised Adaptation
  - Entropy minimization, minmax entropy
- Supervised Domain Adaptation
- Bayesian Domain Adaptation
- Margin Disparity Discrepancy

Domain Shift

- Generalization is an important issue of Machine Learning (ML).

- It concerns the way we can ensure that the trained models perform as well as possible on new and unseen data.

- If a predictor trained on a dataset is tested on new domains and performs poorly, Domain Shift can be noticed.

Domain Shift

- The dissimilarity betwixt the source and target datas' marginal feature distributions is caused by domain shift.

- Loss functions are created and used to deal with this issue.

- Recently domain shift is dealt with by promptly inserting in a deep network domain, alignment layers which make use of BN.

Domain Shift

- In domain 1 there is no background.
- In domain 2 the background is complex.
- If the classifier is trained on the first domain the performance will be degraded when tested on the second and vice versa.

[XIN2016]

Domain Adaptation

- Domain Shift
- **Unsupervised Domain Adaptation**
  - **Domain-specific Whitening Transform**
  - Min-entropy Consensus loss
  - Maximum Classification Discrepancy
  - Sliced Wasserstein Discrepancy
- Deep Learning methods for Unsupervised Domain Adaptation
  - DLID: Deep learning for DA by Interpolating between Domains
- Semi-supervised Adaptation
  - Entropy minimization, minmax entropy
- Supervised Domain Adaptation
- Bayesian Domain Adaptation
- Margin Disparity Discrepancy

Unsupervised Domain Adaptation

- $\mathrm{D_s} = \{(I_j^s, y_j^s)\}_{j=1}^{n_s}$ is the labeled source dataset, where I_j^s is an image and $y_j^s \in y = \{1,2,\dots,C\}$ its associated label.

- $D_T = \{I_i^t\}_{i=1}^{n_t}$ is the target dataset without labels.

- UDA aims to find a predictor for the target domain by using samples from both D_S and D_T.

Domain-specific Whitening Transform

The Whitening is performed by

$$\hat{x}_i = \mathbf{W}_\mathrm{B}(x_i - \boldsymbol{\mu}_B), \qquad (3)$$

where:

- $\boldsymbol{\mu}_\mathrm{B}$ vector is the mean of the components in B.
- Matrix $\mathbf{W}_\mathrm{B}$ is such that: $\mathbf{W}_B^T\mathbf{W}_B = \boldsymbol{\Sigma}_\mathrm{B}^{-1}$, $\boldsymbol{\Sigma}_B$ being the covariance matrix computed using *B*.
- $\Omega = (\boldsymbol{\mu}_B, \boldsymbol{\Sigma}_B)$ are the Batch-dependent first and second-order statistics.

Domain Adaptation

- Domain Shift
- **Unsupervised Domain Adaptation**
  - Domain-specific Whitening Transform
  - **Min-entropy Consensus loss**
  - Maximum Classification Discrepancy
  - Sliced Wasserstein Discrepancy
- Deep Learning methods for Unsupervised Domain Adaptation
  - DLID: Deep learning for DA by Interpolating between Domains
- Semi-supervised Adaptation
  - Entropy minimization, minmax entropy
- Supervised Domain Adaptation
- Bayesian Domain Adaptation
- Margin Disparity Discrepancy

Min-Entropy Consensus Loss

- The predictor greatly divides the target data through the entropy loss minimization.
- Consistent predictions are forced, for the target samples from indistinguishable and unseen categories, as the consistency loss is minimized.
- The Min-Entropy Consensus loss (MEC), introduced in [ROY2019], combines both of the above (entropy and consistency loss) to one function, within the framework of UDA.

Min-Entropy Consensus Loss

- When the MEC loss is used the network is given three batches.
- These are the B^s and two different target batches, B_1^t and B_2^t.
- The target batches enclose duplicate pairs of images that are identical with the exception of the adopted image perturbation.

Domain Adaptation

- Domain Shift
- **Unsupervised Domain Adaptation**
  - Domain-specific Whitening Transform
  - Min-entropy Consensus loss
  - **Maximum Classification Discrepancy**
  - Sliced Wasserstein Discrepancy
- Deep Learning methods for Unsupervised Domain Adaptation
  - DLID: Deep learning for DA by Interpolating between Domains
- Semi-supervised Adaptation
  - Entropy minimization, minmax entropy
- Supervised Domain Adaptation
- Bayesian Domain Adaptation
- Margin Disparity Discrepancy

Maximum Classification Discrepancy

Numerous adversarial learning methods train domain classifier networks to identify the features as a source or a target and train a feature generator network to mimic the discriminator. These methods cause the following problems:

- The domain classifier attempts to distinguish the features as a source or a target and so, it does not take into consideration the task-specific decision margins in between classes.

Maximum Classification Discrepancy

- The above techniques aspire to make the feature distributions between different domains thoroughly match, which is difficult due to the domain's characteristics.

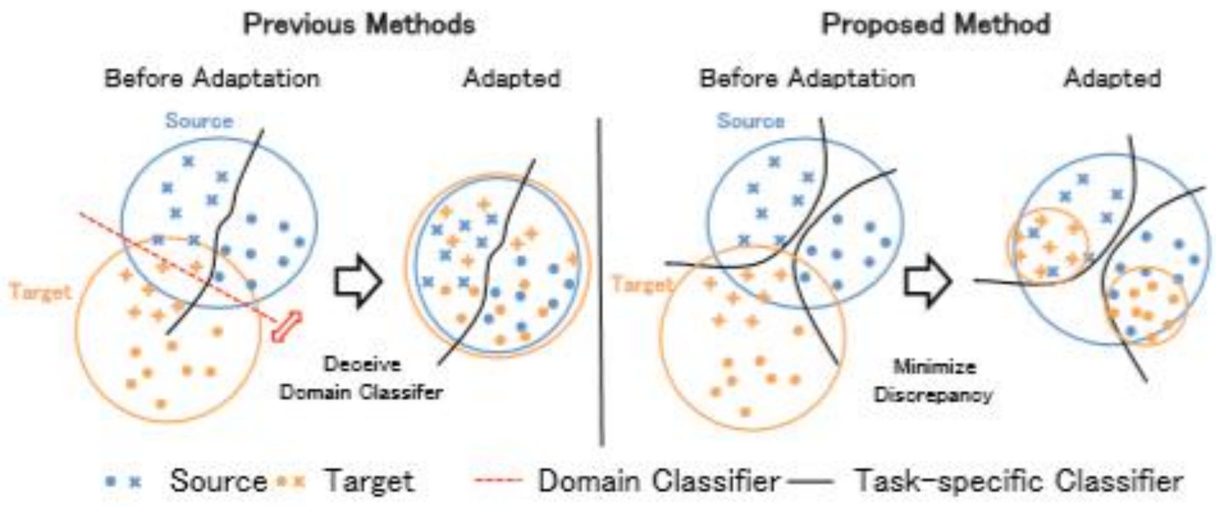
- To solve these problems a new approach is introduced in [SAITO2018], that aligns distributions of source and target by making use of the task-specific decision boundaries.

Maximum Classification Discrepancy

- This technique maximizes the discrepancy between two classifiers' outputs to find target samples that are distant from the support of the source.

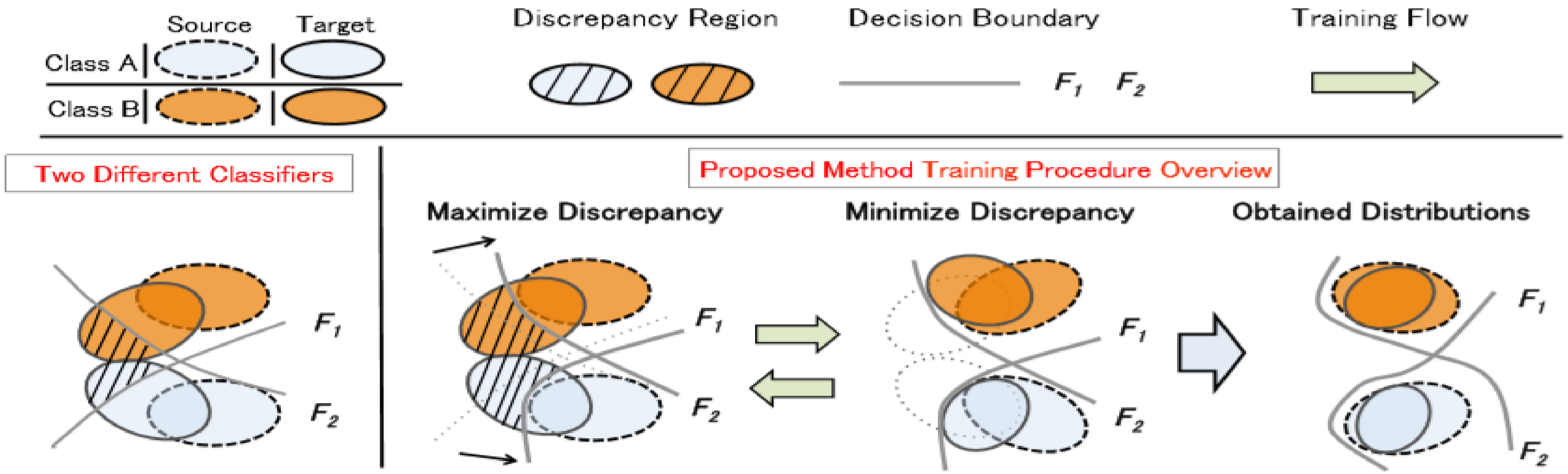
- To eliminate discrepancy, the feature generator is taught to generate the target features in close proximity to the source.

Maximum Classification Discrepancy



- Other methods (left) aspire to match different distributions by mimicking the domain classifier, without considering the decision boundary.
- This method (right) aims to identify target samples' distributions outside the support of the source, with the use of classifiers that are task-specific.

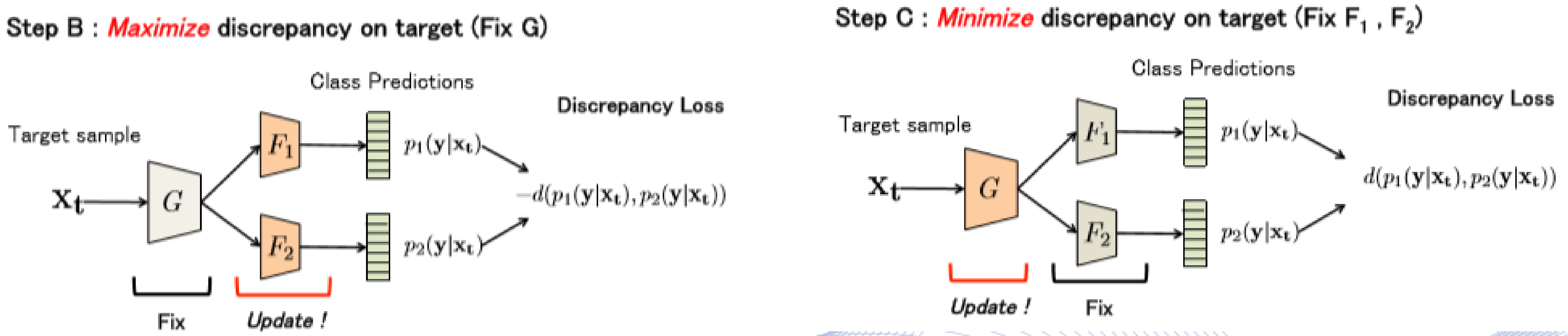
Maximum Classification Discrepancy



The target samples outside the support of the source can be assessed by the two classifiers (F_1 is the first classifier, F_2 is the second classifier).

[SAITO2018]

Maximum Classification Discrepancy



Steps B and C of the adversarial training.
The generator G and classifiers F_1, F_2 compose the network.
[SAITO2018]

Domain Adaptation

- Domain Shift
- **Unsupervised Domain Adaptation**
  - Domain-specific Whitening Transform
  - Min-entropy Consensus loss
  - Maximum Classification Discrepancy
  - **Sliced Wasserstein Discrepancy**
- Deep Learning methods for Unsupervised Domain Adaptation
  - DLID: Deep learning for DA by Interpolating between Domains
- Semi-supervised Adaptation
  - Entropy minimization, minmax entropy
- Supervised Domain Adaptation
- Bayesian Domain Adaptation
- Margin Disparity Discrepancy

Sliced Wasserstein Discrepancy

In [LEE2019] two easily distinguishable concepts of UDA are combined:

- the feature distribution alignment between domains by employing the task-specific decision limit and
- the Wasserstein metric.

Sliced Wasserstein Discrepancy

- The sliced Wasserstein discrepancy (SWD) aims to express the natural idea of divergence between the outputs of task-specific classifiers.
- It delivers information with geometrical importance to determine target samples that are away from the support of the source.
- It also authorizes a systematic distribution alignment in a completely trainable manner.

Sliced Wasserstein Discrepancy

- The SWD technique aspires to make the damage of moving the distributions close to the boundaries between the task-specific classifiers become the least possible.

- This is done by making use of the Wasserstein metric, which delivers an idea of dissimilarity of the probability distributions that has greater meaningfulness.

Domain Adaptation

- Domain Shift
- Unsupervised Domain Adaptation
    - Domain-specific Whitening Transform
    - Min-entropy Consensus loss
    - Maximum Classification Discrepancy
    - Sliced Wasserstein Discrepancy
- **Deep Learning methods for Unsupervised Domain Adaptation**
    - DLID: Deep learning for DA by Interpolating between Domains
- Semi-supervised Adaptation
    - Entropy minimization, minmax entropy
- Supervised Domain Adaptation
- Bayesian Domain Adaptation
- Margin Disparity Discrepancy

Deep Learning methods for Unsupervised Domain Adaptation

We observe the differences between each technique in the way the reduction of the discrepancy of the source and target feature distributions is done. Some of the different categories that exist are presented bellow.

- Methods presenting the domain distributions with regard to their first and second order statistics.

- Methods that learn domain-invariant deep representations.

Deep Learning methods for Unsupervised Domain Adaptation

- Methods that train deep networks by adopting the entropy-loss.

- Methods based on Generative Adversarial Networks (GANs). The overall idea of these methods is to promptly transform images from the target domain to the source domain.

Domain Adaptation

- Domain Shift
- Unsupervised Domain Adaptation
  - Domain-specific Whitening Transform
  - Min-entropy Consensus loss
  - Maximum Classification Discrepancy
  - Sliced Wasserstein Discrepancy
- **Deep Learning methods for Unsupervised Domain Adaptation**
  - **DLID: Deep learning for DA by Interpolating between Domains**
- Semi-supervised Adaptation
  - Entropy minimization, minmax entropy
- Supervised Domain Adaptation
- Bayesian Domain Adaptation
- Margin Disparity Discrepancy

DLID: Deep Learning for DA by Interpolating between Domains

This deep learning method of DA, proposed in [CHOPRA2013], aims to learn a predictively functional depiction of the data by including information from the distribution shift linking the test and training data.

- By operating in the deep learning model, we learn hierarchical non-linear depiction of the source and target inputs.
- We clearly define and use an “interpolating path” between the source and target domains. This path represents information about structures in-between the source and target domains.

DLID: Deep Learning for DA by Interpolating between Domains

There are multiple benefits to this approach.

- We have the ability to train complex non-linear representations of the input, while clearly modeling the change between the source and target domains.

- This approach can effortlessly handle additional training data that become available later in the future, by simply fine-tuning the model with new data.

DLID: Deep Learning for DA by Interpolating between Domains

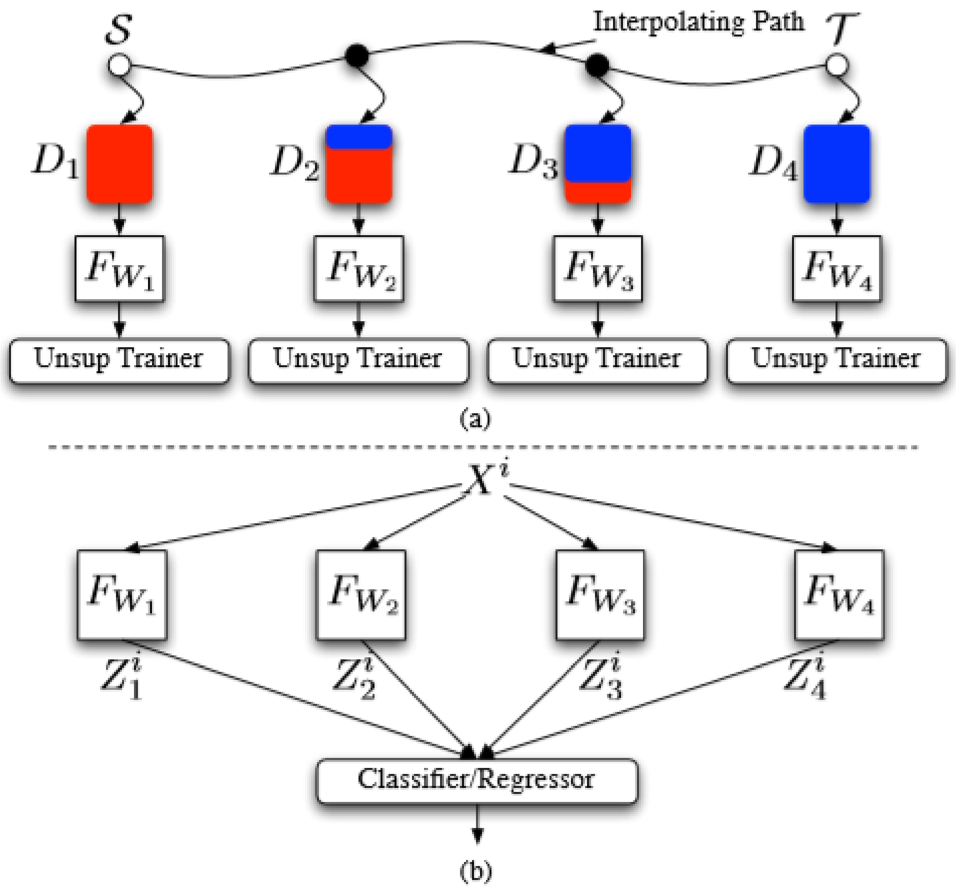
- Rather than being taught a representation which is independent of the last task, our model can learn depictions with information from the last classification/regression task.

- This is made possible by fine-tuning the previously trained intermediate feature extractors utilizing reactions from the final task.

DLID: Deep Learning for DA by Interpolating between Domains

An overview of the DLID model.

[CHOPRA2013]



Domain Adaptation

- Domain Shift
- Unsupervised Domain Adaptation
  - Domain-specific Whitening Transform
  - Min-entropy Consensus loss
  - Maximum Classification Discrepancy
  - Sliced Wasserstein Discrepancy
- Deep Learning methods for Unsupervised Domain Adaptation
  - DLID: Deep learning for DA by Interpolating between Domains
- **Semi-supervised Adaptation**
  - **Entropy minimization, minmax entropy**
- Supervised Domain Adaptation
- Bayesian Domain Adaptation
- Margin Disparity Discrepancy

Semi-Supervised Domain Adaptation

- Semi-supervised domain adaptation (SSDA) is a crucial setting.
- It has not been entirely researched, specifically with reference to deep learning based techniques.
- In SSDA a few target labels are available.

Semi-Supervised Domain Adaptation

- Some already existing concepts have attempted to deal with the SSDA setting.

- Such methods are generative, model-ensemble and adversarial, however they do not consider domain shift.

- To address the SSDA setting, the Minimax Entropy (MME) approach is proposed in [SAITO2019], which in an adversarial way improves an adaptive few-shot model.

Domain Adaptation

- Domain Shift
- Unsupervised Domain Adaptation
  - Domain-specific Whitening Transform
  - Min-entropy Consensus loss
  - Maximum Classification Discrepancy
  - Sliced Wasserstein Discrepancy
- Deep Learning methods for Unsupervised Domain Adaptation
  - DLID: Deep learning for DA by Interpolating between Domains
- Semi-supervised Adaptation
  - Entropy minimization, minmax entropy
- **Supervised Domain Adaptation**
- Bayesian Domain Adaptation
- Margin Disparity Discrepancy

Supervised Domain Adaptation

- UDA is usually preferred as the target data does not have to be labeled.

- Supervised Domain Adaptation (SDA) requires the target data to be labeled.

Supervised Domain Adaptation

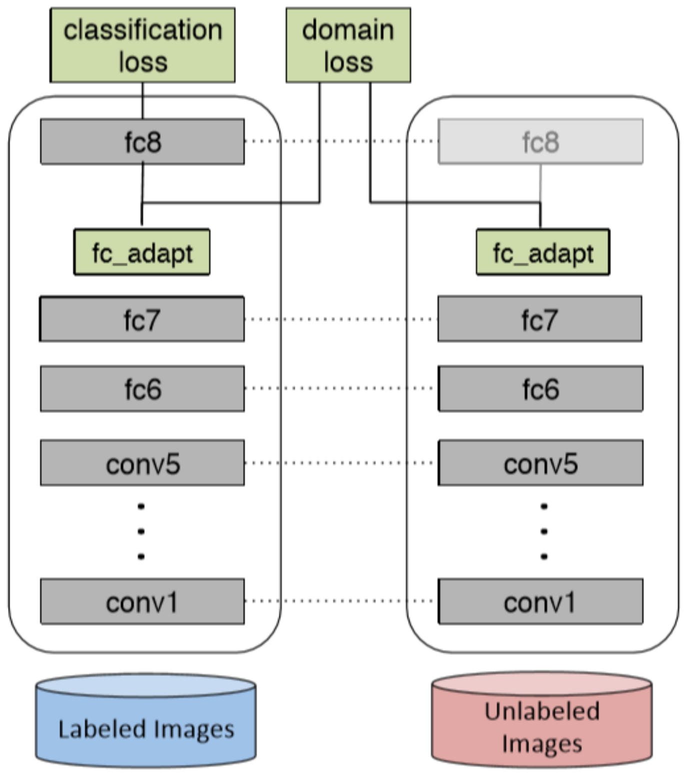
- This approach expects a small amount of labeled target samples per category during training.

- Even one sample can notably make the performance greater, and a few additional ones culminate it, showing a significant “speed” of adaptation.

Supervised Domain Adaptation

The CNN architecture which is outlined on the side uses:

- an adaptation layer and
- a domain confusion loss based on *Maximum Mean Discrepancy* (MMD) to immediately learn a depiction mutually trained to improve the classification and domain invariance.



Domain Adaptation

- Domain Shift
- Unsupervised Domain Adaptation
  - Domain-specific Whitening Transform
  - Min-entropy Consensus loss
  - Maximum Classification Discrepancy
  - Sliced Wasserstein Discrepancy
- Deep Learning methods for Unsupervised Domain Adaptation
  - DLID: Deep learning for DA by Interpolating between Domains
- Semi-supervised Adaptation
  - Entropy minimization, minmax entropy
- Supervised Domain Adaptation
- **Bayesian Domain Adaptation**
- Margin Disparity Discrepancy

Bayesian Domain Adaptation

VML

- There is a possibility for domain shift to still exist through label distribution shift at the classifier and so the model performances are reduced.

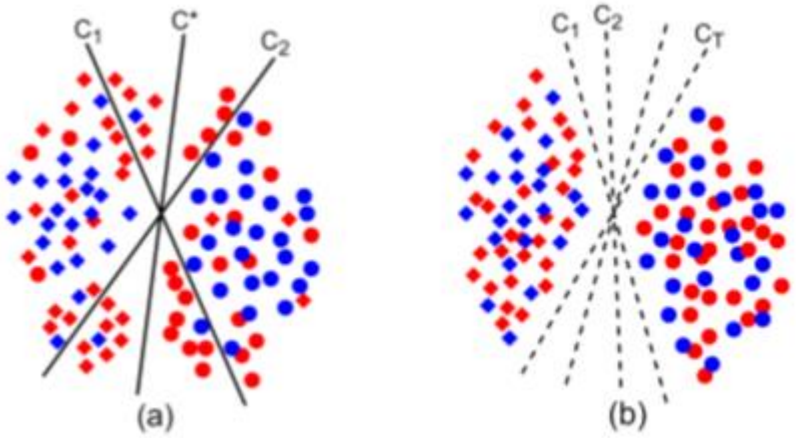
- To stop this from happening, we use an estimated joint distribution, introduced in [WEN2019], which can match the scheme by exploiting prediction uncertainty.

Bayesian Domain Adaptation

- A Bayesian Neural Network is used to do the quantification of the prediction uncertainty of a classifier.

- With the use of distribution matching on the features and the labels, the issue of label distribution mismatching in target and source data is successfully reduced, thus encouraging the classifier to create consistent predictions across domains.

Bayesian Domain Adaptation



Comparisons between conventional (a) and the Bayesian Domain-adaptation (b) methods.

- Blue: Source domain
- Red: Target domain
- Diamonds and circles: samples from two different categories

Domain Adaptation

- Domain Shift
- Unsupervised Domain Adaptation
  - Domain-specific Whitening Transform
  - Min-entropy Consensus loss
  - Maximum Classification Discrepancy
  - Sliced Wasserstein Discrepancy
- Deep Learning methods for Unsupervised Domain Adaptation
  - DLID: Deep learning for DA by Interpolating between Domains
- Semi-supervised Adaptation
  - Entropy minimization, minmax entropy
- Supervised Domain Adaptation
- Bayesian Domain Adaptation
- **Margin Disparity Discrepancy**

Margin Disparity Discrepancy

- The Margin Disparity Discrepancy is a measurement with meticulous generalization bounds, introduced in [ZHANG2019].

- It is adapted to the distribution with the asymmetric margin loss and the minimax development for effortless training.

- This attempts to bridge the gaps that exist amongst the theories and algorithms for DA and produce margin-aware generalization bounds based on Rademacher complexity.

Bibliography

[GOO2016] Goodfellow I, Bengio Y, Courville A, Bengio Y., *Deep learning* MIT press, 2016.
[HAY2009] S. Haykin, *Neural networks and learning machines*, Prentice Hall, 2009.
[BIS2006] C.M. Bishop, Pattern recognition and machine learning, Springer, 2006.
[GOO2016] I. Goodfellow, Y. Bengio, A. Courville, Deep learning, MIT press, 2016
[THEO2011] S. Theodoridis, K. Koutroumbas, Pattern Recognition, Elsevier, 2011.
[ZUR1992] J.M. Zurada, *Introduction to artificial neural systems*. Vol. 8. St. Paul: West publishing company, 1992.
[ROS1958] Rosenblatt, Frank. "The perceptron: a probabilistic model for information storage and organization in the brain." *Psychological review* 65.6 (1958): 386.
[YEG2009] Yegnanarayana, Bayya. *Artificial neural networks*. PHI Learning Pvt. Ltd., 2009.
[DAN2013] Daniel, Graupe. *Principles of artificial neural networks*. Vol. 7. World Scientific, 2013.
[HOP1988] Hopfield, John J. "Artificial neural networks." *IEEE Circuits and Devices Magazine* 4.5 (1988): 3-10.

Q & A

Thank you very much for your attention!

More material in
http://icarus.csd.auth.gr/cvml-web-lecture-series/

Contact: Prof. I. Pitas
pitas@csd.auth.gr