<table>
<tr><th colspan="2" rowspan="2">Transfer Learning Settings</th><th rowspan="2">Description</th><th colspan="2">Source & Target Domains</th><th colspan="2">Source &Target Tasks</th><th rowspan="2">Source Domain Labels</th><th rowspan="2">Target Domain Labels</th><th rowspan="2">Ex.</th><th rowspan="2">Algorithms</th><th rowspan="2">Related Papers</th></tr>
<tr><th>$\mathcal{X}$</th><th>P(X)</th><th>$\mathcal{Y}$</th><th>P(Y|X) /f(.)</th></tr>
<tr><td rowspan="5">Label-setting-based</td><td rowspan="2">Inductive Transfer Learning</td><td rowspan="2">Adapt existing supervised training model on the new labeled dataset</td><td colspan="2" rowspan="2">=/≠, related</td><td colspan="2" rowspan="2">≠, related</td><td>Unlabeled</td><td>Labeled</td><td>Self-taught Learning</td><td>▪ Use unlabeled data to construct a new representation with sparse coding algorithm.
▪ Apply this new representation to labeled data.
▪ Use existing classification methods in this new space.</td><td>Self-taught Learning: http://ai.stanford.edu/~hllee/icml07-selftaughtlearning.pdf</td></tr>
<tr><td>Labeled</td><td>Labeled</td><td>Question Answering (Multitask Learning)</td><td>▪ Build a Multitask Q/A Network trained jointly on decaNLP.
- Takes in a question and context, encodes both with a BiLSTM, uses dual coattention to condition representations for both sequences, compresses all info with another two BiLSTMs, and applies self-attention
- Uses a final two BiLSTMs to get representations of the question & context.
- Creates Multi-pointer-generator decoder.</td><td>The Natural Language Decathlon Multitask Learning as Question Answering: https://arxiv.org/pdf/1806.08730.pdf</td></tr>
<tr><td rowspan="2">Transductive Transfer Learning</td><td rowspan="2">Adapt existing supervised training model on the new unlabeled dataset</td><td colspan="2" rowspan="2">≠, related</td><td colspan="2" rowspan="2">=</td><td rowspan="2">Labeled</td><td>Unlabeled/ Limited</td><td>Supervised Domain Adaptation</td><td>▪ Create pairs of source and target instances to handle target data.
▪ Extend adversarial learning to align the semantic information of classes.
- Create a discriminator to distinguish between samples of the source and target distributions.
- Create an inference function to map a target sample to a feature space.</td><td>Few-shot Adversarial Domain Adaptation: https://arxiv.org/pdf/1711.02536.pdf</td></tr>
<tr><td>Limited</td><td>Binary Classification (Sample Selection Bias)</td><td>▪ Build a minimax estimator, Robust Bias-aware classifier, measured by the conditional logloss (for Sample Selection Bias)
▪ Datasets: Generate biased subsets of used datasets as source and unbiased subsets as target samples.</td><td>Robust Classification Under Sample Selection Bias: https://papers.nips.cc/paper/5458-robust-classification-under-sample-selection-bias.pdf</td></tr>
<tr><td>Unsupervised Transfer Learning</td><td>Adapt existing unsupervised training model on the new unlabeled dataset</td><td colspan="2">=/≠, related</td><td colspan="2">≠, related</td><td>Unlabeled</td><td>Unlabeled</td><td>Self-taught Clustering</td><td>▪ A clustering problem of self-taught learning
▪ Learn the feature representation of data in the source domain as auxiliary data using Sparse Coding technique.
▪ Build two objective functions by extending the information theoretic co-clustering and share feature clustering.</td><td>Self-taught clustering: https://dl.acm.org/doi/pdf/10.1145/1390156.1390182</td></tr>
</table>

Space-setting-based	**Homogeneous Transfer Learning**	Adapting supervised training model on new target dataset that the feature and label spaces between the domains are identical.	=	≠	=	≠	Labeled	Limited	Fatigue Detection	▪ CP-MDA: correcting the conditional distribution differences. - Build classifiers for all source domains, and obtain a weight value for each classifier. - Create a learning task (to find pseudo labels for target data) by summing weighted source classifiers. - Build target learner by using labeled and pseudo labeled target data.	Multi-source Domain Adaptation and its Application to Early Detection of Fatigue: https://www.cs.ucdavis.edu/~davidson/Publications/TKDD.pdf
								Unlabeled		▪ 2SW-MDA: addresses marginal and conditional distribution differences. - Compute weights for source domains based on marginal distribution differences. - Modify the weights as a function of conditional distribution differences performed in CP-MDA. - A target classifier is learned based on reweighted source instances and any labeled (if available) target data.	
	Heterogeneous Transfer Learning	Adapting supervised training model on new target dataset that the feature and label spaces between the domains can be alternatively same or different.	≠		=		Labeled	Limited		The objective to minimize the risk function. ▪ Build a translator to connect two different feature spaces using co-occurrence data. ▪ Use the *language model* and combine feature translation with nearest neighbor learning. ▪ Model is built using a Markov chain c → y → x.	Translated learning: Transfer learning across different feature spaces: https://papers.nips.cc/paper/3492-translated-learning-transfer-learning-across-different-feature-spaces.pdf
			=		≠		Labeled	Labeled		▪ Initialize the weights from a pre-trained model ▪ Use the target data to fine-tune the parameters for the target task. - Fine-tune all the layers of DNN - Freeze several layers of the DNN and only fine-tune the layers that can reduce overfitting.	How transferable are features in deep neural networks?: https://papers.nips.cc/paper/5347-how-transferable-are-features-in-deep-neural-networks.pdf
			≠		≠		Labeled	Limited		▪ Find a common latent input spaces using *spectral mapping*. ▪ Apply a clustering-based sample selection method to select new related instances. ▪ Use Bayesian-based method to find the relationship and resolve the differences in the output space.	Transfer Learning on Heterogeneous Feature Spaces via Spectral Transformation: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.298.7133&rep=rep1&type=pdf
	Teacher-Student Network								Knowledge Distillation	▪ Distillation between teacher and student network. - Train the teacher model to produce class probabilities by using output layer that convert the logit. - Establish the correspondence between the intermediate outputs of the student and teacher network by matching the logit. - Forward pass through the teacher network - Backpropagate through the student network.	Distilling the Knowledge in a Neural Network: https://arxiv.org/pdf/1503.02531v1.pdf

Domain Adaptation refers to predicting the labels of samples drawn from a target domain, given labeled samples drawn from a source domain and unlabeled samples drawn from the target domain itself.

Note: We sometimes can consider the domain adaptation as special case of transfer learning, where differences between feature spaces and label spaces are allowed.

Note:

(a). In most medical classification problem that the training requires the big amount of data. Even though there exists the dataset in medication, it is still not enough to obtain the satisfied performance. Hence, Transfer Learning is the best idea in this problem.

For example, Kidney detection that we have a pre-trained model on ImageNet dataset. What we care about is training a classifier in target domain that lacks data for training (GE Healthcare LOGIQ E9 scanner). We can address this problem as "Inductive Transfer Learning" because the domains and the task differ.

References

[1]. Wouter M. Kouw, Marco Loog, "A review of domain adaptation without target labels"

Freeze the weights of low-level layers of the network, while updating the weights of other layers until it saturates.