Unsupervised Domain Adaptation by Backpropagation (ICML 2014)

Main Idea - Learn event/domain independent embeddings during training by adversarial training.

Keywords - Domain Independent, Unsupervised Representations

Summary / Key Points of the work

- Propose a new approach to domain adaptation in deep architectures that can be trained on large amounts of labelled data from the source domain and large amounts of unlabelled data from the target domain (no labelled data is necesary)
- The learnt features are
    - Discriminative for the main learning tasks on the source domain
    - Invariant with respect to the shift between domains
- Focus on learning features that combine
    - Discriminativeness
    - Domain Invariance

This is achieved by jointly optimizing the underlying features as well as 2 discriminative classifiers operating on these features

- Label Predictor that predicts class labels and is used both during training and test time
- Domain Classifier that discriminates between the source and the target domains during training

- The Parameters of the classifiers are optimized in order to minimize their error on the training set
    - The parameters of the underlying deep features mapping are optimized in order to **minimize the loss of the label classifier** and **maximize the loss of the domain classifer.**
        - This encourages **domain invariant** feaures.
    - Training is done using standard **backpropagation algorithms** based on **SGD**.
- Introduction of a **GRL (Gradient Reversal Layer)** - This leaves the input unchanged during forward propagation, and reverses the gradient by multiplying it by a negative scalar during backpropagation.
- Approach Combines the following
    - **Feature Learning + Domain Adaptation + Classifier Learning** - all into a single / unified architecture.
    - Using a single learning algorithm (backpropagation)

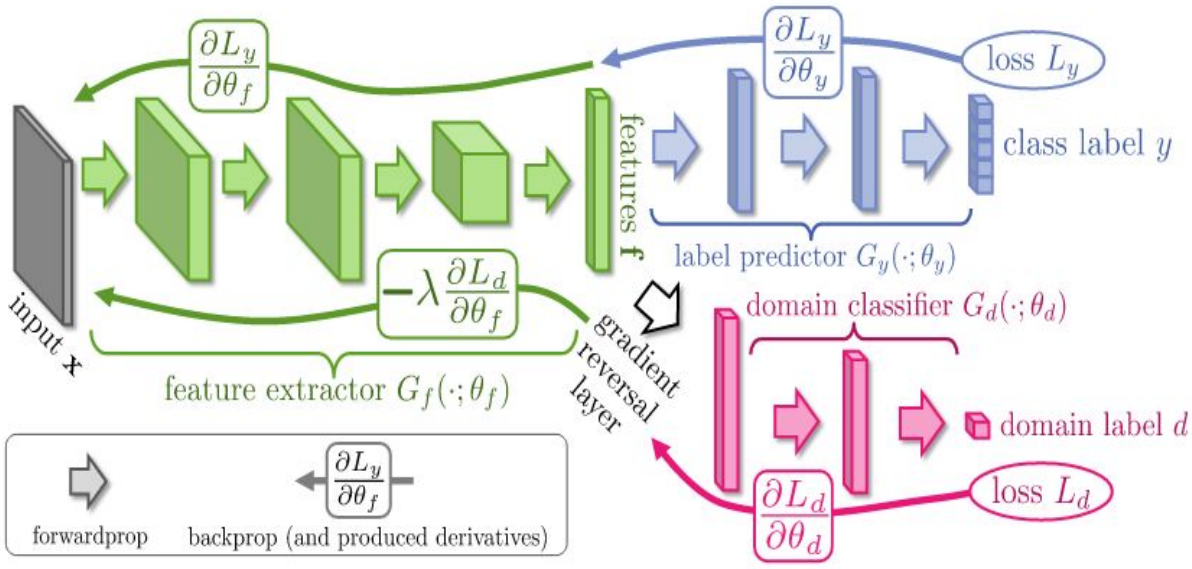
- Unsupervised - does not require labeled target domain data (+ can easily incorporate such a data when it is available)

----------Model Description Begins Here----------

- Input Samples x ϵ X, Output Labels y ϵ Y ; Y = {1,2,..., L}
- 2 distributions S(x,y) and T(x,y) on X□Y -> Source distributions + Target distributions
  - These 2 distributions are similar but different (in other words, S is “shifted” from T by some domain shift)

GOAL -> Predict labels “y” given the input “x” for target distributions.

----------Architecture----------



References:

1. Unsupervised Domain Adaptation by Backpropagation - Yaroslav Ganin, Victor Lempitsky; Proceedings of the 32nd International Conference on Machine Learning, PMLR 37:1180-1189, 2015.