Transfer Learning and Domain Adaptation

Sargur N. Srihari
srihari@cedar.buffalo.edu

Topics in Representation Learning

1. Greedy Layer-Wise Unsupervised Pretraining
2. Transfer Learning and Domain Adaptation
3. Semi-supervised Disentangling of Causal Factors
4. Distributed Representation
5. Exponential Gains from depth
6. Providing Clues to Discover Underlying Causes

What is transfer learning?

- It is the situation where what has been learned in one setting is exploited to improve generalization in another
- Ex: Pretraining
  - Transfer unsupervised task to supervised task
- Ex: Visual classification
  - There is more data in distribution P_1 (cats and dogs)
  - and very few in distribution P_2 (ants and wasps)
  - Visual categories share low-level notions of edges and visual shapes, geometric changes, lighting

Architecture for Transfer Learning

- Output variable y has the same semantics for all classes

  $\boldsymbol{x}$ has different meaning dimension for each task

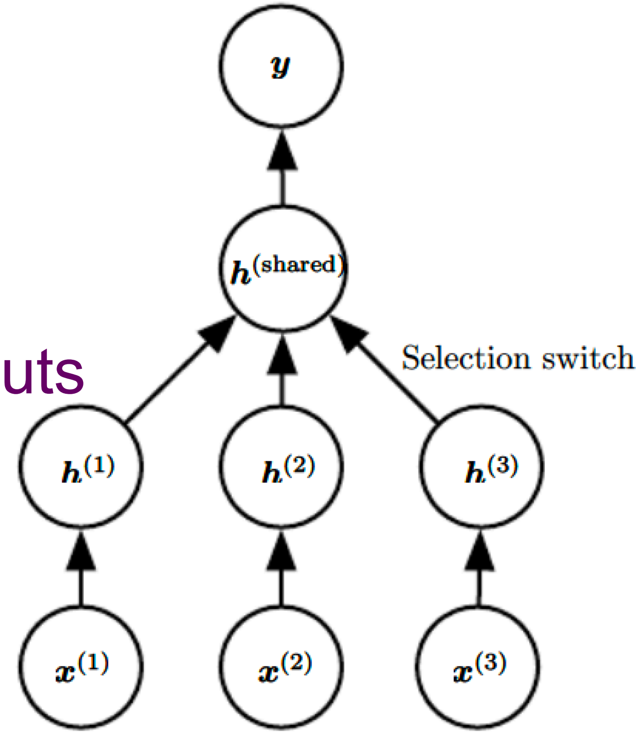
  - Three tasks $\boldsymbol{x}^{(1)}$, $\boldsymbol{x}^{(2)}$ and $\boldsymbol{x}^{(3)}$ are inputs

- Lower levels upto selection switch are task-specific

  – Upper levels are shared

  - Semantics of output are shared, not semantics of input

    as in speech recognition where vocalizations are based on different speakers



Success of Transfer Learning

- Unsupervised deep learning for transfer learning has found success in ML competitions
  - Each participant is given data from distribution P_1 illustrating some set of categories
  - Participants learn a feature space
    - Mapping raw input to a representation space
  - This transformation is applied to samples from P_2
  - A linear classifier is trained from very few samples
- As deeper representations used (learned purely unsupervised from P_1) performance improves
  - For deeper representations fewer samples needed

Domain Adaptation

- Related to transfer learning between settings
- Task remains the same between each setting, but the input distribution is slightly different
- Ex: Sentiment Analysis
  - Task: determine if comment is positive/negative
    - Sentiment predictor is trained on customer reviews of media content such as books, videos and music
    - Later used to analyze comments about consumer electronics such as televisions and smartphones
    - Vocabulary and style may vary from one domain to other
    - Simple unsupervised pretraining (with denoising autoencoders) found useful with domain adaptation

Concept Drift

- A form of transfer learning where there are gradual changes in data over time
- Both concept drift and transfer learning can be regarded as different forms of multi-task learning
  - Typically refers to supervised learning
  - Also applicable to unsupervised and reinforcement learning

One-shot learning

- Only one labeled example of the transfer task
  - Possible because the representation learns to cleanly separate underlying classes during Stage 1
  - During transfer learning, only one labeled example is needed to infer the label of many possible test examples that cluster around the same point in representation space
- Works to the extent that factors of variation corresponding to these invariances have been cleanly separated from the other factors in the learned representation space

Zero-shot learning

- No labeled examples
- Ex: A learner reads a large collection of text and then solves object recognition problems
  - Having read that a cat has four legs and pointed ears, learner guesses that an image is a cat without having seen a cat before

Zero-data learning explained

- Possible because additional data exploited
- Zero-data learning scenario includes three random variables
  1. Traditional inputs $\boldsymbol{x}$
     - Unlabeled text data containing sentences such as "cats have four legs", "cats have pointy ears')
  2. Traditional outputs y (y=1 indicating yes, y=0 for no)
  3. Description of task T (represents questions to be answered)
     - Is there a cat in this image?
- Model trained to determine conditional $p(y|\boldsymbol{x}, T)$

Type of Representation of T

- Zero-shot learning requires T to be represented in a way that allows some sort of generalization
  - T cannot be just a one-hot code indicating an object category
  - Instead a distributed representation of object categories by using a learned word embedding for the word associated with each category

Similar phenomenon in Machine Translation

- We have words in one language
  - Word relationships learned from a unilingual corpus
  - We have translated sentences that relate words in one language with words in the other
- No labeled word translations available
  - i.e., word A in lang. $\boldsymbol{X}$ to word B in lang. $\boldsymbol{Y}$
- Can guess a translation for word A because
  - We have learned distributed representations for words in $\boldsymbol{X}$ and for words in $\boldsymbol{Y}$ then created a link relating the two spaces via training examples of matched pairs of sentences
    - Works best when two representations and relations are learned jointly

Transfer learning enables zero-shot

- Labeled or unlabeled examples of $\boldsymbol{x}$ allow:
  - Learning a representation function f_x and similarly with examples of $\boldsymbol{y}$ to learn f_y
    - Each application of f_x and f_y appears as an upward arrow
    - Distances in h_x and h_y space provide a similarity metric
    - Image $\boldsymbol{x}_{\text{test}}$ is associated with word $\boldsymbol{y}_{\text{test}}$ even if no image of that word was ever presented