Lecture 9:
Transfer Learning & Ensemble Learning

Yanbin Liu

Auckland University of Technology

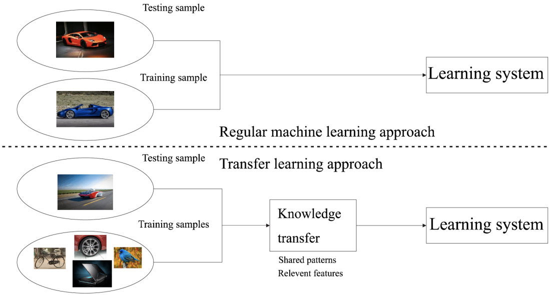
May 9, 2024

Table of Contents

Transfer Learning

Transfer Learning

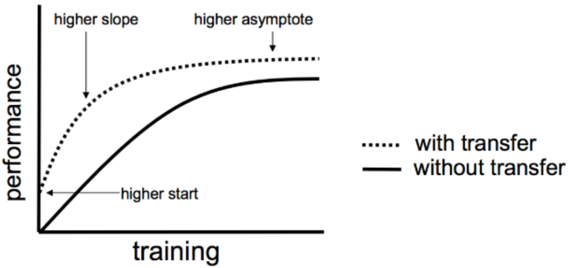
Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task.



Shao, L., et al. (2015) Transfer learning for visual categorization: A survey. IEEE Transactions on Neural Networks and Learning Systems, 26(5).

Transfer Learning

Transfer learning is a machine learning method where a model developed for a task is reused as the starting point for a model on a second task.

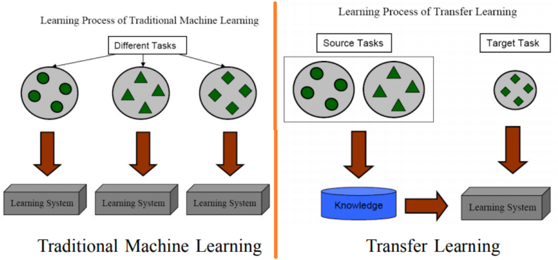


Web: https://machinelearningmastery.com/transfer-learning-for-deep-learning/

Transfer Learning

Transfer Learning

Transfer learning allows the domains $\mathcal{D}$, tasks $\mathcal{T}$, and distributions used in training and test to be different.

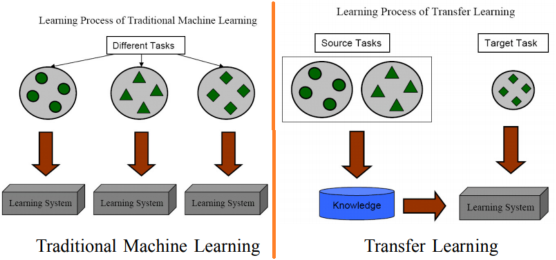


Pan, S. and Yang, Q. (2010) A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10):1345-1359.

Transfer Learning

Transfer Learning

Transfer learning aims to extract the knowledge from one or more source tasks $\mathcal{T}_S$ and applies the knowledge to a target task $\mathcal{T}_T$.



Pan, S. and Yang, Q. (2010) A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10):1345-1359

Transfer Learning *v.s.* Machine Learning

- What to transfer?
- How to transfer?
- When to transfer?

- Inductive transfer learning ($\mathcal{T}_T \neq \mathcal{T}_S$ and $\mathcal{D}_T = \mathcal{D}_S$)
- Transductive transfer learning ($\mathcal{T}_T = \mathcal{T}_S$ and $\mathcal{D}_T \neq \mathcal{D}_S$)
- Unsupervised transfer learning ($\mathcal{T}_T \neq \mathcal{T}_S$ and $\mathcal{D}_T \neq \mathcal{D}_S$)

Relationship between Traditional Machine Learning and Various Transfer Learning Settings

Learning Settings		Source and Target Domains	Source and Target Tasks
Traditional Machine Learning		the same	the same
Transfer Learning	*Inductive Transfer Learning*	the same	different but related
	Unsupervised Transfer Learning	different but related	different but related
	Transductive Transfer Learning	different but related	the same

Pan, S. and Yang, Q. (2010) A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10):1345-1359

Categories of Transfer Learning

- What to transfer?
- How to transfer?
- When to transfer?

- Inductive transfer learning ($\mathcal{T}_T \neq \mathcal{T}_S$ and $\mathcal{D}_T = \mathcal{D}_S$)
- Transductive transfer learning ($\mathcal{T}_T = \mathcal{T}_S$ and $\mathcal{D}_T \neq \mathcal{D}_S$)
- Unsupervised transfer learning ($\mathcal{T}_T \neq \mathcal{T}_S$ and $\mathcal{D}_T \neq \mathcal{D}_S$)

Different Settings of Transfer Learning

Transfer Learning Settings	Related Areas	Source Domain Labels	Target Domain Labels	Tasks
Inductive Transfer Learning	Multi-task Learning	Available	Available	Regression, Classification
	Self-taught Learning	Unavailable	Available	Regression, Classification
Transductive Transfer Learning	Domain Adaptation, Sample Selection Bias, Co-variate Shift	Available	Unavailable	Regression, Classification
Unsupervised Transfer Learning		Unavailable	Unavailable	Clustering, Dimensionality Reduction

Pan, S. and Yang, Q. (2010) A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10):1345-1359

Approaches of Transfer Learning

- What to transfer?
- How to transfer?
- When to transfer?

- Inductive transfer learnin ($\mathcal{T}_T \neq \mathcal{T}_S$ and $\mathcal{D}_T = \mathcal{D}_S$)
- Transductive transfer learning ($\mathcal{T}_T = \mathcal{T}_S$) and ($\mathcal{D}_T \neq \mathcal{D}_S$)
- Unsupervised transfer learning ($\mathcal{T}_T \neq \mathcal{T}_S$ and $\mathcal{D}_T \neq \mathcal{D}_S$)

Transfer Learning Approaches	Brief Description
Instance-transfer	To re-weight some labelled data in the source domain for use in the target domain
Feature-representation-transfer	Find a good feature representation that reduces difference between the source and the target domains and the error of classification and regression models
Parameter-transfer	Discover shared parameters or priors between the source domain and the target domain models
Relational-knowledge-transfer	Build mapping of relational knowledge between the source domain and the target domain. Both domains are relational domains.

Approaches of Transfer Learning

- What to transfer?
- How to transfer?
- When to transfer?

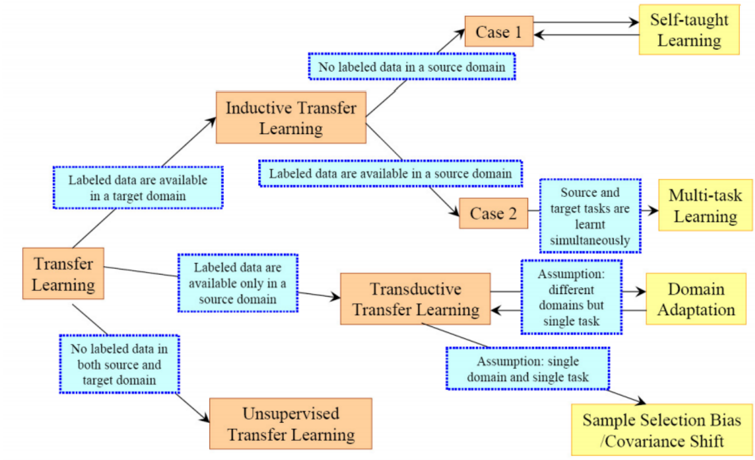
- Inductive transfer learnin ($\mathcal{T}_T \neq \mathcal{T}_S$ and $\mathcal{D}_T = \mathcal{D}_S$)
- Transductive transfer learning ($\mathcal{T}_T = \mathcal{T}_S$) and ($\mathcal{D}_T \neq \mathcal{D}_S$)
- Unsupervised transfer learning ($\mathcal{T}_T \neq \mathcal{T}_S$ and $\mathcal{D}_T \neq \mathcal{D}_S$)

Different Approaches Used in Different Settings

	Inductive Transfer Learning	Transductive Transfer Learning	Unsupervised Transfer Learning
Instance-transfer	√	√	
Feature-representation-transfer	√	√	√
Parameter-transfer	√		
Relational-knowledge-transfer	√		

Pan, S. and Yang, Q. (2010) A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10):1345-1359

Overview of Transfer Learning

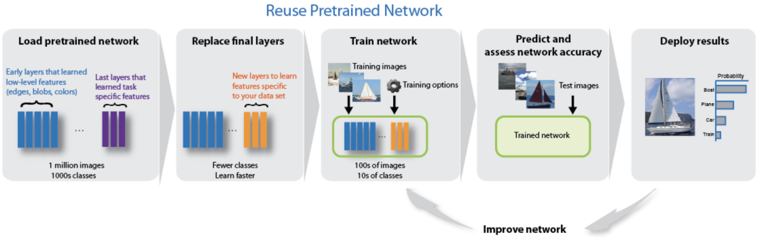


Pan, S. and Yang, Q. (2010) A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10):1345-1359

Transfer Learning

MATLAB Transfer Learning

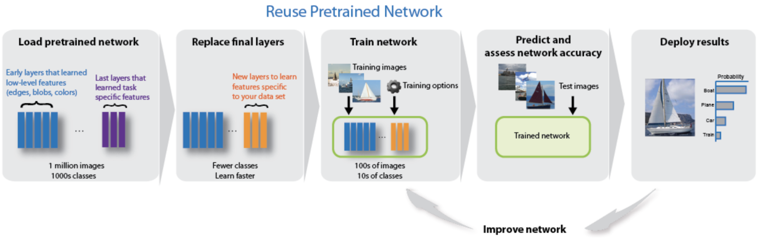
- Transfer the learned features of a pretrained network to a new problem.
- Transfer learning is faster and easier than training a new network.
- Reduce training time and data size.
- Perform deep learning without needing to learn how to create a whole new network.



Transfer Learning

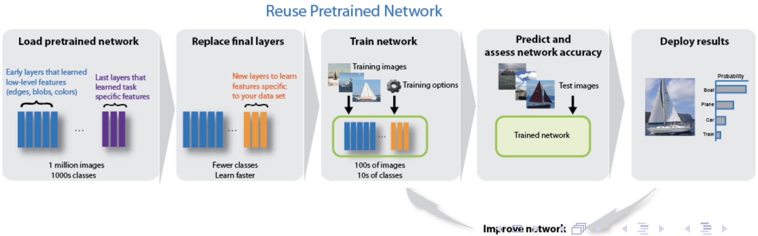
MATLAB Transfer Learning

- Choose a pretrained network and import it into the application.
- Replace the final layer with a new layer adapted to the new dataset:
  - Specify the new number of classes in training dataset.
  - Set learning rates to learn faster in the new layers than in the transferred layers.
- Export the network using the command line.



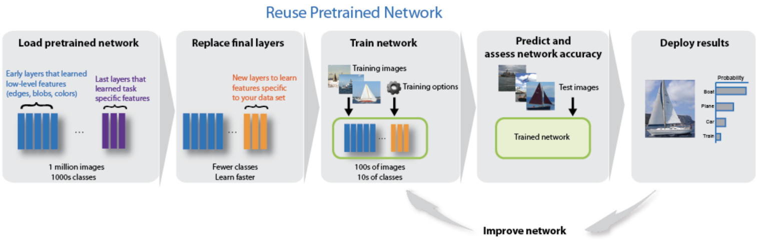
MATLAB Transfer Learning

- AlexNet has eight layers with learnable weights: Five convolutional layers and three fully connected layers. It has been trainied over 1M images and classifies images into 1,000 object categories (ILSVRC 2012)
- VGG-16 has 16 layers with learnable weights: 13 convolutional layers and three fully connected layers.
- VGG-19 has 19 layers with learnable weights: 16 convolutional layers and three fully connected layers.
- GoogLeNet is 22 layers depth, won the ILSVRC in 2014.



MATLAB Transfer Learning

- Transfer learning is used in deep learning applications.
- A pretrained network is used as a starting point to learn a new task.
- Fine-tuning a network with transfer learning is usually much faster and easier than training a network with randomly initialized weights from scratch.
- Learned features are transferred to a new task using a smaller number of training images.



Web:

Transfer Learning

Ensemble Learning

Ensemble Learning

- Each learning algorithm dictates a model that comes with a set of assumptions.
- By suitably generating and combining multiple base learners, the accuracy can be improved.
- The usual approach is to choose the one that performs the best on a separate validation set.

Questions

- How do we generate base learners that complement each other?
- How do we combine the outputs of base learners for the maximum accuracy?

Alpaydin, E. (2009) Introduction to Machine Learning, MIT Press.

Ensemble Learning

Generating Diverse Learners

A set of diverse learners that differ in their decisions so that they complement each other:

- *Different algorithms*: Combine multiple learners, free ourselves from taking a decision.
- *Different hyperparameters*: Train multiple base learners, average over this factor, reduce variance and error.
- *Different input representations*: Random subspace method.
- *Different training sets*: Randomly draw training sets from the given samples, train base learners, mixture of experts.

E. Alpaydin. (2009) Introduction to Machine Learning, MIT Press.

Ensemble Learning

Multi-Expert Combination

- Global approach: Learner fusion
- Local approach: Learner selection

Multi-Stage Combination

- Serial approach
- Cascading approach

Train Base Learners

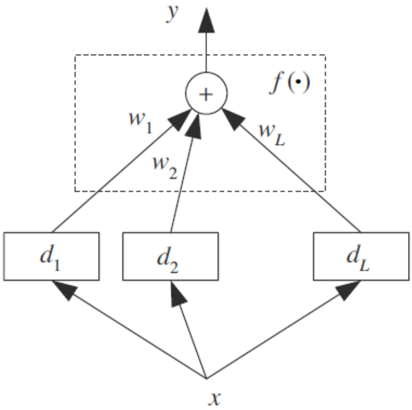
$$y = f(d_1, d_2, \cdots, d_L | \Phi);$$

$$c = \arg\max_{i=1,2\cdots,K} y_i$$

where $f(\cdot)$ is the combining function with Φ denoting its parameters, c is the returned class number.

Alpaydinm, E. (2009) Introduction to Machine Learning, MIT Press.

Base Learner



$$y = f(d_1, d_2, \cdots, d_L|\Phi)$$

$$c = \underset{i=1,2\cdots,K}{\arg\max}\; y_i$$

where $f(\cdot)$ is the combining function with Φ denoting its parameters, c is the returned class number.

Alpaydin, E. (2009) Introduction to Machine Learning, MIT Press.

Ensemble Learning

Combining Multiple Learners

Combining: $y_i = \sum_j w_j d_{ji}$, $\sum_j w_j = 1$, $w_j \geq 0$

- Ensembles and linear opinion pools.
- Classifier combination rules: $\sum, \max, \min, \prod$
- Simple voting: $w_i = w_j \in \{1, 0\}$

Classifier combination rules.

Rule	Fusion function $f(\cdot)$
Sum	$y_i = \frac{1}{L}\sum_{j=1}^{L} d_{ji}$
Weighted sum	$y_i = \sum_j w_j d_{ji}, w_j \geq 0, \sum_j w_j = 1$
Median	$y_i = \text{median}_j d_{ji}$
Minimum	$y_i = \min_j d_{ji}$
Maximum	$y_i = \max_j d_{ji}$
Product	$y_i = \prod_j d_{ji}$

Alpaydinm E. (2009) Introduction to Machine Learning, MIT Press.

Ensemble Learning

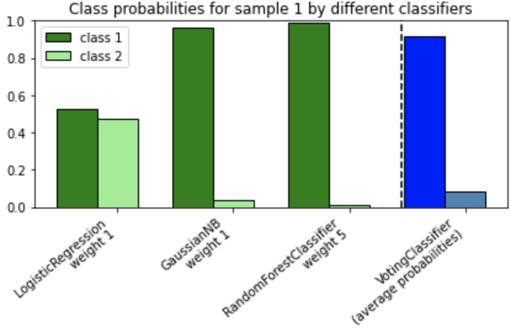
MATLAB Ensemble Learning

Ensemble learning can meld results from many weak learners into one high-quality ensemble predictor:

- Prepare the predictor data;
- Prepare the response data;
- Choose an applicable ensemble aggregation method;
- Set the number of ensemble members;
- Prepare the weak learners;
- Call an ensemble function.

Ensemble Learning: Example

```
Automatically created module for IPython interactive environment
```



Questions?

Ensemble Learning

Learning Objectives

- Design and analyse algorithms of deep neural networks.
- Demonstrate advanced understanding of the state-of-the-art in the practice of deep learning.