Transfer Learning

Pranav Goel
14075039

CSE B.Tech. Pt. IV, IIT (BHU) Varanasi

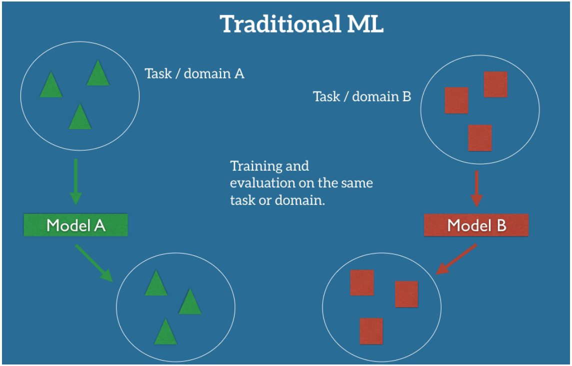
February 2018

Underlying Motivation

- We have become great at mapping inputs to outputs in the classic supervised setting.
- But we greatly lack the ability to generalize to conditions that are different from the ones encountered during training - and the real world *is* messy!
- We need the ability to transfer knowledge to new conditions - **Transfer Learning**.

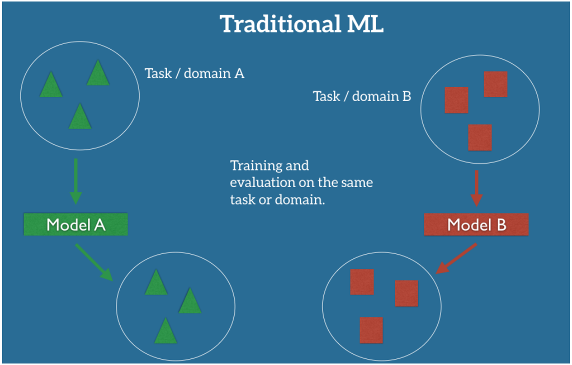
Underlying Motivation

- Classic Supervised Learning - generally assumes that the training and test set examples are from same task and domain.



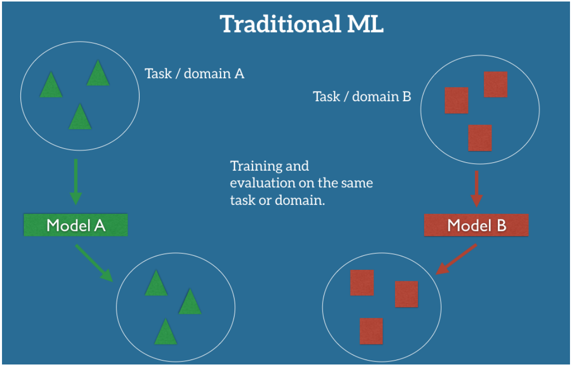
Underlying Motivation

- Task: objective of our model (like image recognition).
- Domain: where the data comes from (like images of Indian footpaths).



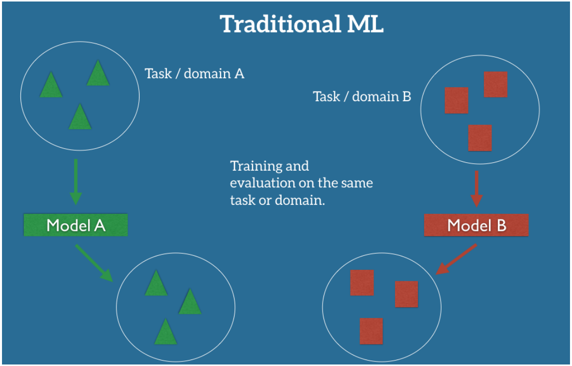
Underlying Motivation

- Traditional Supervised ML breaks down when we do NOT have *sufficient labeled data* for the task or domain we care about.



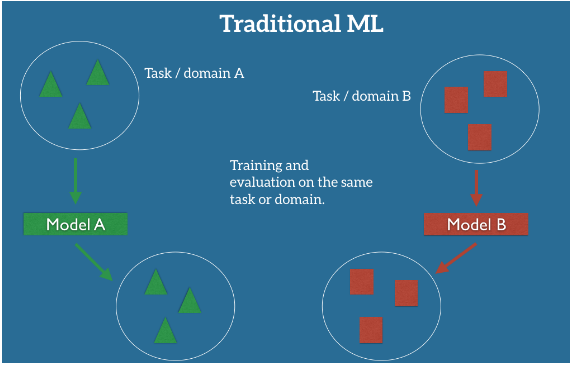
Underlying Motivation

- Traditional Supervised ML breaks down when we do NOT have **sufficient** labeled **data** for the task or **domain** we care about.
- Performance Deterioration!



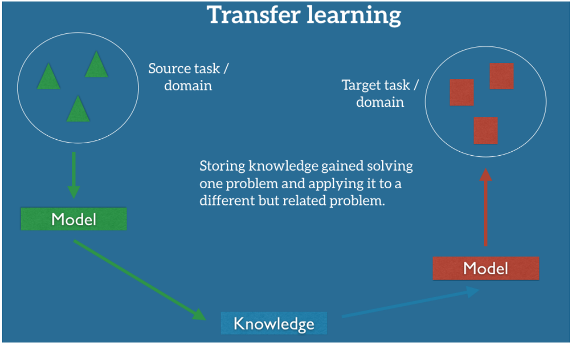
Underlying Motivation

- Traditional Supervised ML breaks down when we do NOT have sufficient **labeled data** for the **task** or domain we care about.
- Cannot reuse existing model!



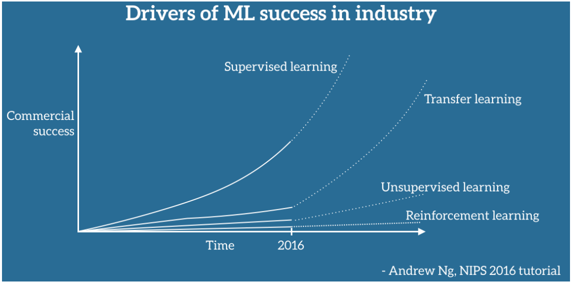
Underlying Motivation

- Transfer Learning allows us to leverage already existing labeled data of some related task or domain.



Why Transfer Learning NOW?

- "Transfer Learning will be the next driver of ML success" - Andrew Ng, NIPS 2016.



Transfer Learning *Defined*

- Consider binary document classification as a running example.

Transfer Learning *Defined*

- Consider binary document classification as a running example.
- A domain D consists of a feature space χ and a marginal probability distribution $P(X)$ over the feature space, where $X = x_1, .., x_n \in \chi$.

Transfer Learning *Defined*

- Consider binary document classification as a running example.
- A domain D consists of a feature space χ and a marginal probability distribution $P(X)$ over the feature space, where $X = x_1, .., x_n \in \chi$.
- For document classification, $\chi =$ space of all document representations, x_i is the i-th term of feature vector corresponding to some document and X is the sample of documents used for training.

Transfer Learning *Defined*

- Consider binary document classification as a running example.
- A domain D consists of a feature space χ and a marginal probability distribution $P(X)$ over the feature space, where $X = x_1, .., x_n \in \chi$.
- For document classification, $\chi =$ space of all document representations, x_i is the i-th term of feature vector corresponding to some document and X is the sample of documents used for training.
- So, Domain $D = \{\chi, P(X)\}$.

Transfer Learning *Defined*

- Task $T = \{\gamma, P(Y|X)\}$, where γ is the label space and $P(Y|X)$ is the conditional probability distribution typically learned from the training data.

Transfer Learning *Defined*

- Task $T = \{\gamma, P(Y|X)\}$, where γ is the label space and $P(Y|X)$ is the conditional probability distribution typically learned from the training data.
- For transfer learning, we are given the source domain D_S, source task T_S, target domain D_T and the target task T_T.

Transfer Learning *Defined*

- Task $T = \{\gamma, P(Y|X)\}$, where γ is the label space and $P(Y|X)$ is the conditional probability distribution typically learned from the training data.
- For transfer learning, we are given the source domain D_S, source task T_S, target domain D_T and the target task T_T.
- Now, the objective of transfer learning is to enable us to learn the target conditional probability distribution $P(Y_T|X_T)$ in D_T with the information gained from D_S and T_S where $D_S \neq D_T$ or $T_S \neq T_T$.

Transfer Learning *Defined*

- Task $T = \{\gamma, P(Y|X)\}$, where γ is the label space and $P(Y|X)$ is the conditional probability distribution typically learned from the training data.
- For transfer learning, we are given the source domain D_S, source task T_S, target domain D_T and the target task T_T.
- Now, the objective of transfer learning is to enable us to learn the target conditional probability distribution $P(Y_T|X_T)$ in D_T with the information gained from D_S and T_S where $D_S \neq D_T$ or $T_S \neq T_T$.
- In most cases, we have a limited number of labeled target examples (exponentially smaller than the number of labeled source examples).

Transfer Learning *Scenarios*

- Given source and target domains D_S and D_T where $D = \{\chi, P(X)\}$ and source and target tasks T_S and T_T where $T = \{\gamma, P(Y|X)\}$, the source and target conditions can vary in **four** ways -

Transfer Learning *Scenarios*

- Given source and target domains D_S and D_T where $D = \{\chi, P(X)\}$ and source and target tasks T_S and T_T where $T = \{\gamma, P(Y|X)\}$, the source and target conditions can vary in **four** ways -
  1. $\chi_S \neq \chi_T$ (cross-lingual adaption in NLP).

Transfer Learning *Scenarios*

- Given source and target domains D_S and D_T where $D = \{\chi, P(X)\}$ and source and target tasks T_S and T_T where $T = \{\gamma, P(Y|X)\}$, the source and target conditions can vary in **four** ways -
  1. $\chi_S \neq \chi_T$ (cross-lingual adaption in NLP).
  2. $P(X_S) \neq P(X_T)$ (domain adaptation).

Transfer Learning *Scenarios*

- Given source and target domains D_S and D_T where $D = \{\chi, P(X)\}$ and source and target tasks T_S and T_T where $T = \{\gamma, P(Y|X)\}$, the source and target conditions can vary in **four** ways -
  1. $\chi_S \neq \chi_T$ (cross-lingual adaption in NLP).
  2. $P(X_S) \neq P(X_T)$ (domain adaptation).
  3. $\gamma_S \neq \gamma_T$ (practically occurs with scenario 4).

Transfer Learning *Scenarios*

- Given source and target domains D_S and D_T where $D = \{\chi, P(X)\}$ and source and target tasks T_S and T_T where $T = \{\gamma, P(Y|X)\}$, the source and target conditions can vary in **four** ways -
  1. $\chi_S \neq \chi_T$ (cross-lingual adaption in NLP).
  2. $P(X_S) \neq P(X_T)$ (domain adaptation).
  3. $\gamma_S \neq \gamma_T$ (practically occurs with scenario 4).
  4. $P(Y_S|X_S) \neq P(Y_T|X_T)$ (imbalance with respect to the classes, usually handled with oversampling or undersampling).

Transfer Learning Methods

- Deep Learning has brought about many set of approaches for transfer learning.

Transfer Learning Methods

Using *pre-trained* CNN features

Transfer Learning Methods

Using *pre-trained* CNN features

- Image Detection on ImageNet using CNN - features/representations learned by the layers.

Transfer Learning Methods

Using *pre-trained* CNN features

- Image Detection on ImageNet using CNN - features/representations learned by the layers.
- CNN trained on ImageNet seems to capture details about the way animals and objects are structured and composed, which is relevant when dealing with images in general.

Transfer Learning Methods

Using *pre-trained* CNN features

- Image Detection on ImageNet using CNN - features/representations learned by the layers.
- CNN trained on ImageNet seems to capture details about the way animals and objects are structured and composed, which is relevant when dealing with images in general.
- Astounding success in vision

Transfer Learning Methods

Using *pre-trained* CNN features

- Image Detection on ImageNet using CNN - features/representations learned by the layers.
- CNN trained on ImageNet seems to capture details about the way animals and objects are structured and composed, which is relevant when dealing with images in general.
- Astounding success in vision
- Beyond Vision?

Transfer Learning Methods

Using *pre-trained* CNN features

- Image Detection on ImageNet using CNN - features/representations learned by the layers.
- CNN trained on ImageNet seems to capture details about the way animals and objects are structured and composed, which is relevant when dealing with images in general.
- Astounding success in vision
- Beyond Vision?
- Mostly used in scenario 3 - adapting to new tasks.

Transfer Learning Methods

Making Representations *more similar*

Transfer Learning Methods

Making Representations *more similar*

- Representations for the two domains should be as similar as possible.

Transfer Learning Methods

Making Representations *more similar*

- Representations for the two domains should be as similar as possible.
- Capture commonalities between domains, do NOT focus on domain specific characteristics

Transfer Learning Methods

Making Representations *more similar*

- Representations for the two domains should be as similar as possible.
- Capture commonalities between domains, do NOT focus on domain specific characteristics
- Actively encourage our autoencoder/neural model to learn similar representations -

Transfer Learning Methods

Making Representations *more similar*

- Representations for the two domains should be as similar as possible.
- Capture commonalities between domains, do NOT focus on domain specific characteristics
- Actively encourage our autoencoder/neural model to learn similar representations -
    1. A pre-processing step to get new representation to then use for training.

Transfer Learning Methods

Making Representations *more similar*

- Representations for the two domains should be as similar as possible.
- Capture commonalities between domains, do NOT focus on domain specific characteristics
- Actively encourage our autoencoder/neural model to learn similar representations -
  1. A pre-processing step to get new representation to then use for training.
  2. Modify the learning objective or loss function.

Transfer Learning Methods

Confusing Domains

Transfer Learning Methods

Confusing Domains

- Add another objective to an existing model that encourages it to *confuse* the two domains.

Transfer Learning Methods

Confusing Domains

- Add another objective to an existing model that encourages it to *confuse* the two domains.
- Original task objective has an associated loss function (based on predicted and actual labels).

Transfer Learning Methods

Confusing Domains

- Add another objective to an existing model that encourages it to *confuse* the two domains.
- Original task objective has an associated loss function (based on predicted and actual labels).
- Domain Confusion Loss - classification loss when the model tries to *predict the domain* of the input example.

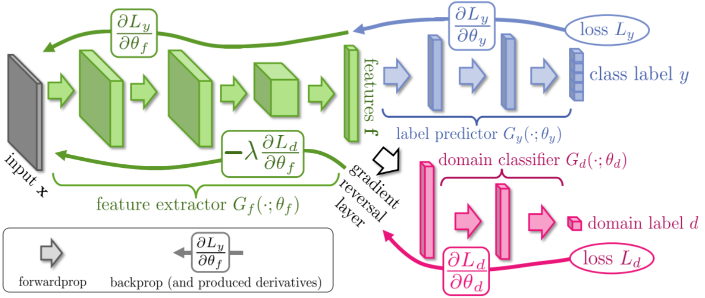
Transfer Learning Methods

Confusing Domains

- Add another objective to an existing model that encourages it to *confuse* the two domains.
- Original task objective has an associated loss function (based on predicted and actual labels).
- Domain Confusion Loss - classification loss when the model tries to *predict the domain* of the input example.
- Difference from regular loss? Gradients that flow from the loss to the rest of the network are reversed - gradient reversal layer - causes model to try and *maximize* error.

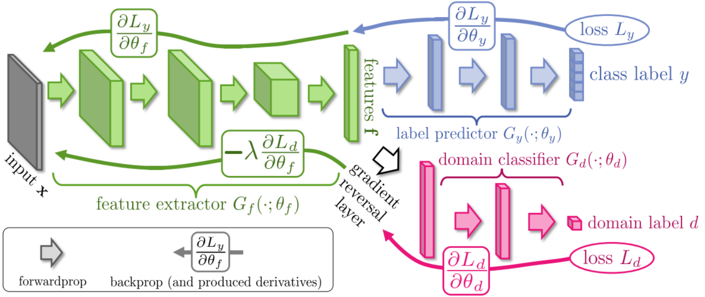
Transfer Learning Methods

Confusing Domains



Transfer Learning Methods

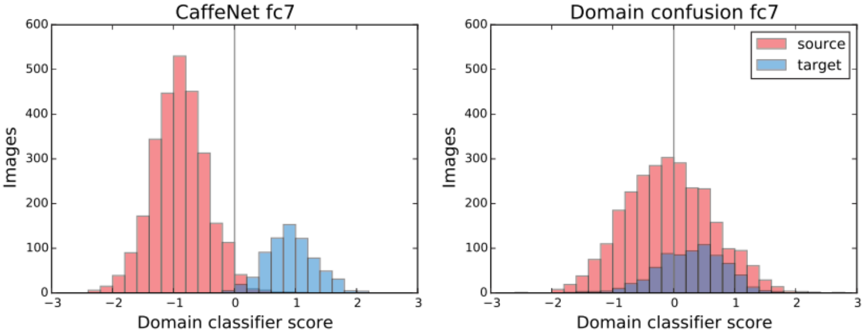
Confusing Domains



Model learns representations that allow it to **minimize** its **original objective**, while **not allowing it to differentiate between the two domains**, which is beneficial for knowledge transfer!

Transfer Learning Methods

Confusing Domains



Left - Model trained only with regular objective; learned rep. clearly separates the domains.

Right - Model objective augmented with domain confusion term.

Applications of Transfer Learning

Learning from Simulations

- For many ML applications that rely on hardware for interaction, gathering data and training a model in the real world is either expensive, time-consuming, or simply too dangerous.

Applications of Transfer Learning

Learning from Simulations

- For many ML applications that rely on hardware for interaction, gathering data and training a model in the real world is either expensive, time-consuming, or simply too dangerous.
- Simulation offers an alternative, less risky way to gather and use data.

Applications of Transfer Learning

Learning from Simulations

- For many ML applications that rely on hardware for interaction, gathering data and training a model in the real world is either expensive, time-consuming, or simply too dangerous.
- Simulation offers an alternative, less risky way to gather and use data.
- Learning from simulation, and applying the knowledge gained or making the transfer to real world - an example of scenario 2 (same feature space, different marginal probabilities).

Applications of Transfer Learning

Learning from Simulations - Benefits

- Makes data gathering easy (objects can be bounded and analyzed).

Applications of Transfer Learning

Learning from Simulations - Benefits

- Makes data gathering easy (objects can be bounded and analyzed).
- Enables fast training (learning can be parallelized across multiple instances).

Applications of Transfer Learning

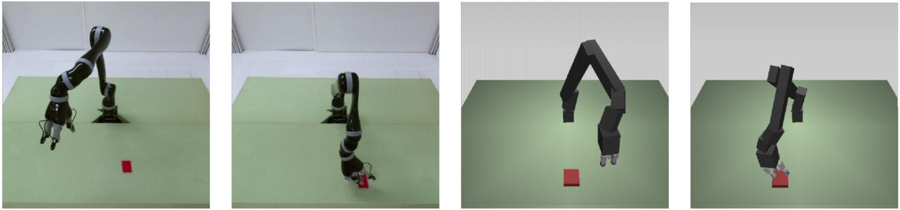
Learning from Simulations - Self-Driving Cars

Udacity's self-driving car simulator

Applications of Transfer Learning

Learning from Simulations - Robotics

Training models on a real robot is too slow and robots are expensive to train.



Robot and simulation images

Applications of Transfer Learning

Adapting to New Domains

- Data where labeled information is easily accessible and the data that we actually care about are often different.

Applications of Transfer Learning

Adapting to New Domains

- Data where labeled information is easily accessible and the data that we actually care about are often different.
- Even if training and test sets *look* the same, training data may contain some bias (which we cannot perceive) that will be exploited by model to overfit.

Applications of Transfer Learning

Adapting to New Domains

Different visual domains

Applications of Transfer Learning

Adapting to New Domains

Different text types/genres

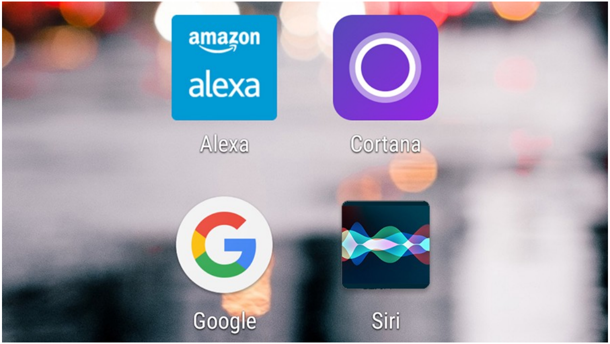
Applications of Transfer Learning

Next level challenge - domains pertaining to *individual* or *groups of users*

Applications of Transfer Learning

Next level challenge - domains pertaining to *individual* or *groups of users*

Consider **Automatic Speech Recognition (ASR)**



Thank You

Questions?