Transfer Learning

Part I: Overview

Sinno Jialin Pan
Institute for Infocomm Research (I2R), Singapore

Agency for Science, Technology and Research
SINGAPORE

1991-2011
Creating Growth
Enhancing Lives

I²R
A*STAR

Transfer of Learning

A psychological point of view

- The study of dependency of human conduct, learning or performance on prior experience.

- [Thorndike and Woodworth, 1901] explored how individuals would transfer in one context to another context that share similar characteristics.

- C++ → Java
- Maths/Physics → Computer Science/Economics

Transfer Learning

In the machine learning community

- The ability of a system to recognize and apply knowledge and skills learned in previous tasks to novel tasks or new domains, which share some commonality.

- Given a target task, how to identify the commonality between the task and previous (source) tasks, and transfer knowledge from the previous tasks to the target one?

A*STAR I²R

Fields of Transfer Learning

- Transfer learning for reinforcement learning.

[Taylor and Stone, Transfer Learning for Reinforcement Learning Domains: A Survey, JMLR 2009]

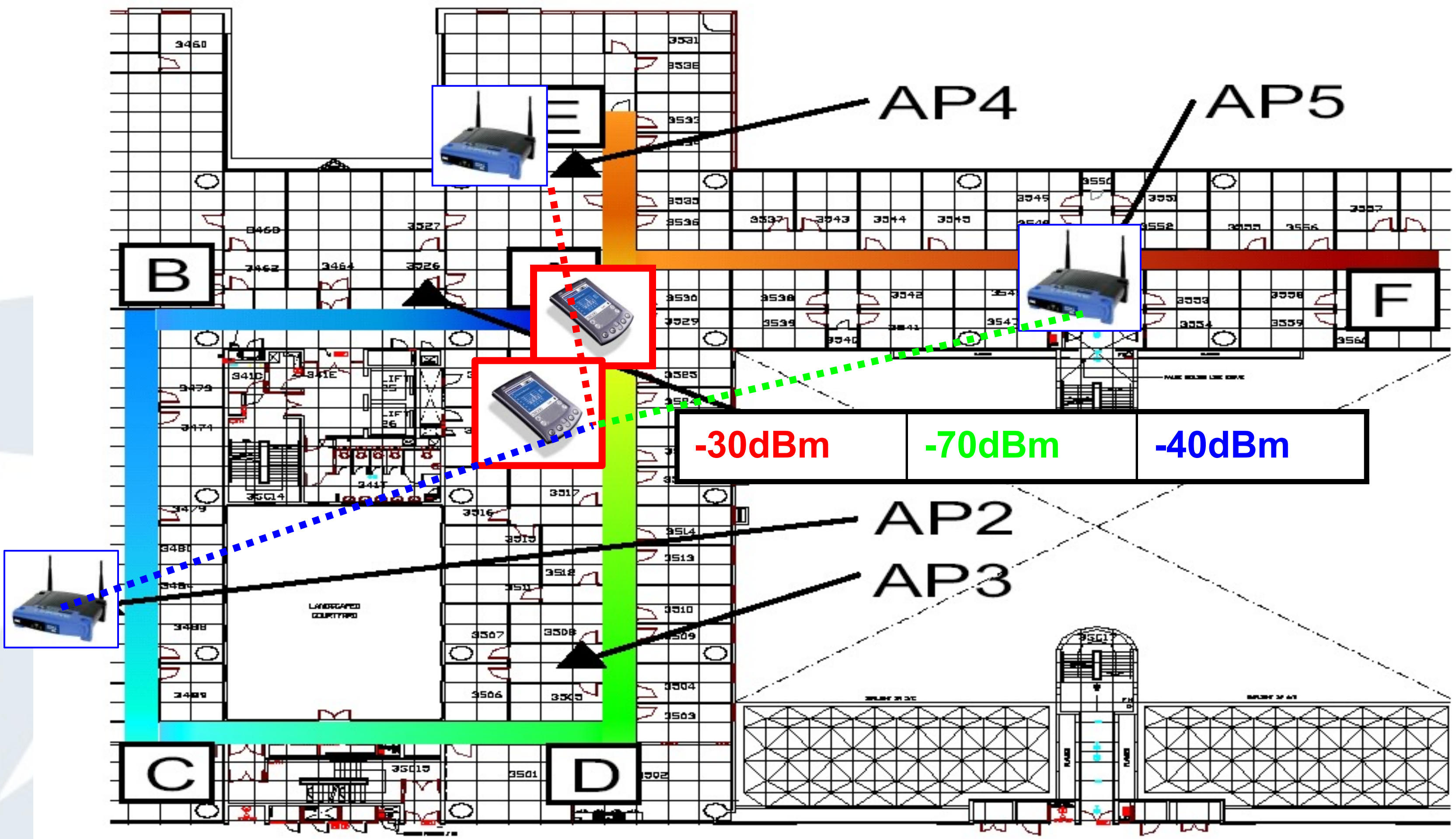
- Transfer learning for classification and regression problems.

Focus!

[Pan and Yang, A Survey on Transfer Learning, IEEE TKDE 2009]

Motivating Example I:

Indoor WiFi localization



Indoor WiFi Localization (cont.)

Average Error Distance

Training

S=(-37dbm, .., -77dbm), L=(1, 3)
S=(-41dbm, .., -83dbm), L=(1, 4)
…
S=(-49dbm, .., -34dbm), L=(9, 10)
S=(-61dbm, .., -28dbm), L=(15,22)

Time Period A

Localization model

Test

S=(-37dbm, .., -77dbm)
S=(-41dbm, .., -83dbm)
…
S=(-49dbm, .., -34dbm)
S=(-61dbm, .., -28dbm)

Time Period A

~1.5 meters

Drop!

Training

S=(-37dbm, .., -77dbm), L=(1, 3)
S=(-41dbm, .., -83dbm), L=(1, 4)
…
S=(-49dbm, .., -34dbm), L=(9, 10)
S=(-61dbm, .., -28dbm), L=(15,22)

Time Period B

Localization model

Test

S=(-37dbm, .., -77dbm)
S=(-41dbm, .., -83dbm)
…
S=(-49dbm, .., -34dbm)
S=(-61dbm, .., -28dbm)

Time Period A

~6 meters

A*STAR I²R

Indoor WiFi Localization (cont.)

Average Error Distance

Training

S=(-37dbm, .., -77dbm), L=(1, 3)
S=(-41dbm, .., -83dbm), L=(1, 4)
…
S=(-49dbm, .., -34dbm), L=(9, 10)
S=(-61dbm, .., -28dbm), L=(15,22)

Device A

Localization model

Test

S=(-37dbm, .., -77dbm)
S=(-41dbm, .., -83dbm)
…
S=(-49dbm, .., -34dbm)
S=(-61dbm, .., -28dbm)

Device A

~ 1.5 meters

Drop!

Training

S=(-33dbm, .., -82dbm), L=(1, 3)
…
S=(-57dbm, .., -63dbm), L=(10, 23)

Device B

Localization model

Test

S=(-37dbm, .., -77dbm)
S=(-41dbm, .., -83dbm)
…
S=(-49dbm, .., -34dbm)
S=(-61dbm, .., -28dbm)

Device A

~10 meters

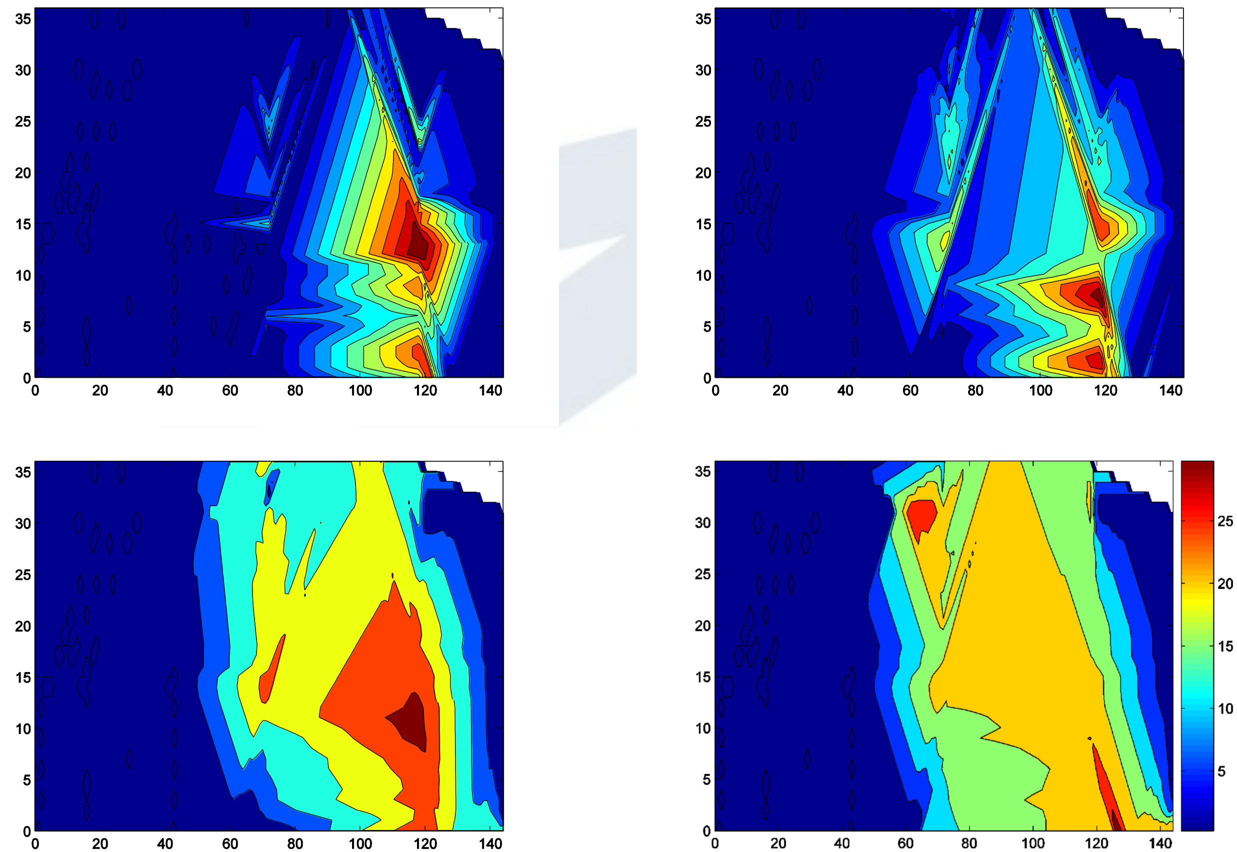
Difference between Tasks/Domains

Time Period A

Time Period B

Device A

Device B



Motivating Example II:

Sentiment Classification

10 hours ago

Edward Priz replied:

You know, this isn't the first time that "States Rights" has been used as a cover for racist policies. In fact, the whole "States Rights" thing has become a sort of code for heavy-handed racist policies, hasn't it? And it does provide a sort of contextual

10 hours ago

RICH HIRTH replied:

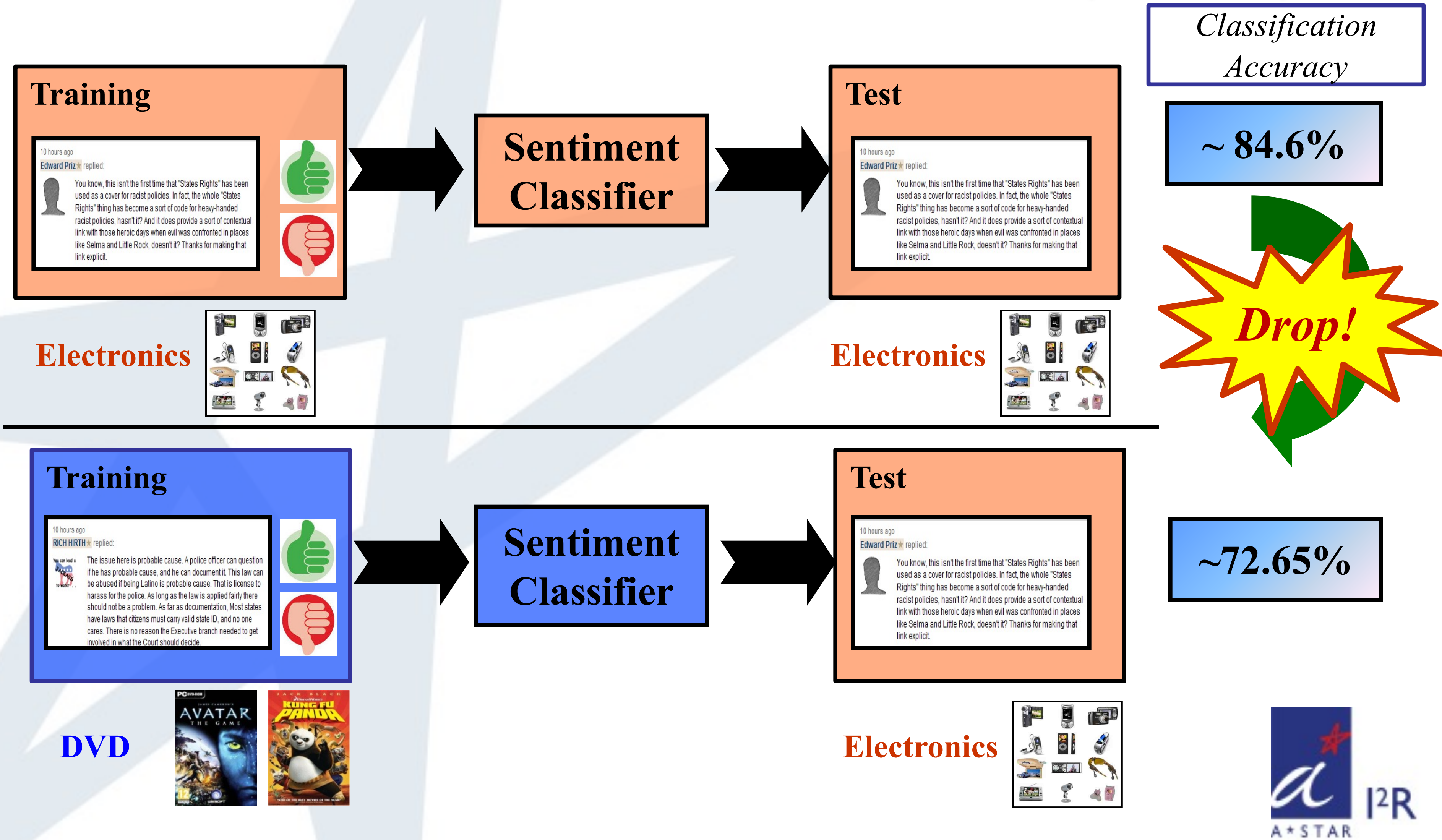
The issue here is probable cause. A police officer can question if he has probable cause, and he can document it. This law can be abused if being Latino is probable cause. That is license to harass for the police. As long as the law is applied fairly there

2 hours ago

Julia Gomez replied:

The Arizona law is so clearly unconstitutional that I do not think it will ever reach the point of being enforced. The article did not say so, but the Republican governor is afraid of a GOP primary electorate that is even more reactionary than usual. That is why she signed the bill, not because she thinks it is legally defensible.

Sentiment Classification (cont.)



Difference between Tasks/Domains

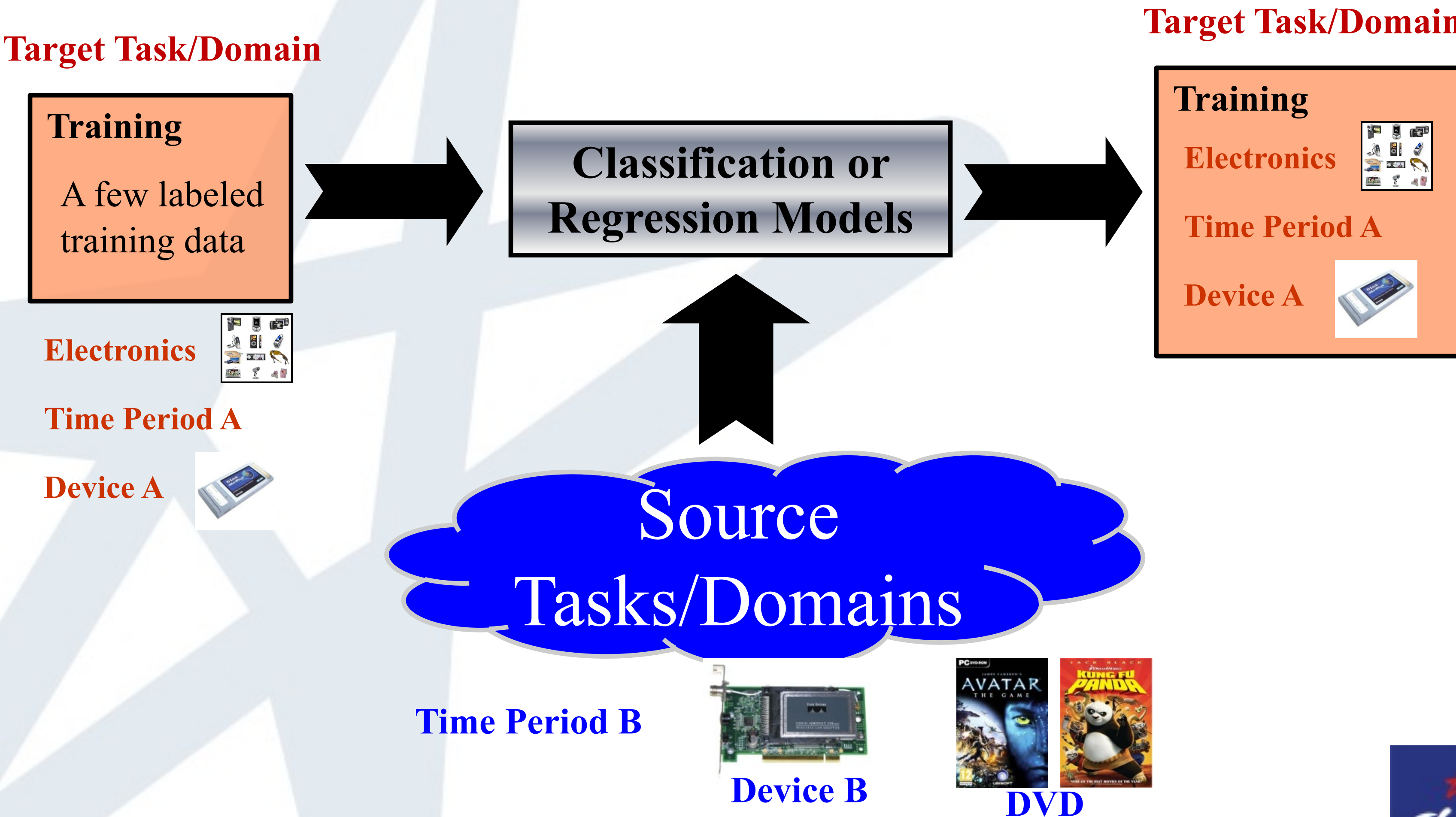
Electronics	Video Games
(1) **Compact**; easy to operate; very good picture quality; looks **sharp**!	(2) A very good game! It is action packed and full of excitement. I am very much **hooked** on this game.
(3) I purchased this unit from Circuit City and I was very excited about the quality of the picture. It is really nice and **sharp**.	(4) Very **realistic** shooting action and good plots. We played this and were **hooked**.
(5) It is also quite **blurry** in very dark settings. I will never buy HP again.	(6) The game is so **boring**. I am extremely unhappy and will probably never buy UbiSoft again.

A Major Assumption

Training and future (test) data come from a same task and a same domain.

- *Represented in same feature and label spaces.*
- *Follow a same distribution.*

The Goal of Transfer Learning



Notations

Domain:

- Feature space $\mathcal{X}$;
- $P(x)$, where $x \in \mathcal{X}$.

Two domains are different $\Rightarrow$

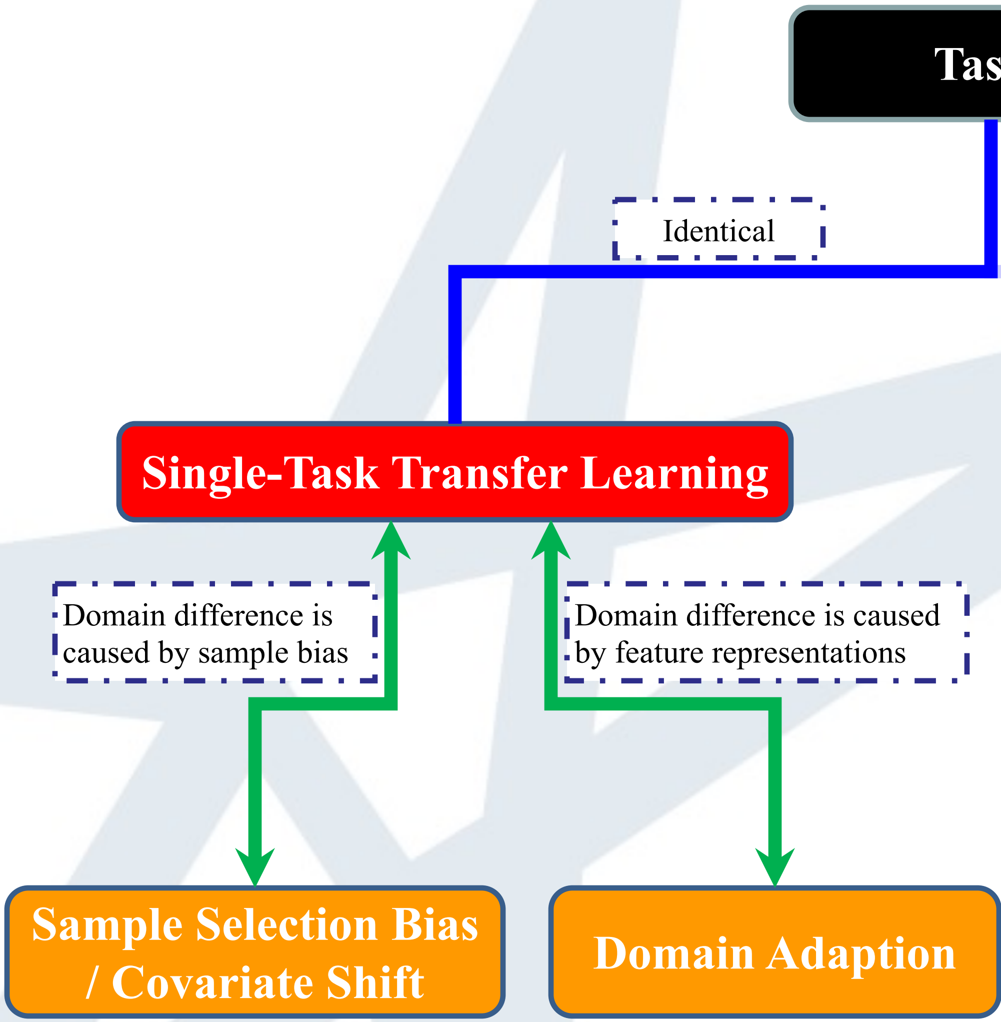
$\mathcal{X}_S \neq \mathcal{X}_T$, or $P_S(x) \neq P_T(x)$.

Task:

- Given $\mathcal{X}$ and label space $\mathcal{Y}$;
- To learn $f : x \rightarrow y$, or estimate $P(y|x)$, where $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.

Two tasks are different $\Rightarrow$

$\mathcal{Y}_S \neq \mathcal{Y}_T$, or $f_S \neq f_T$ $(P_S(y|x) \neq P_T(y|x))$.

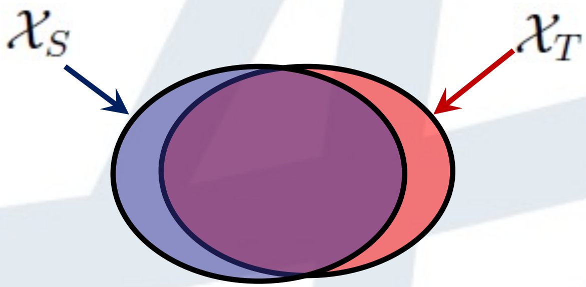


Assumption

- $\mathcal{Y}_S = \mathcal{Y}_T$,
- $P_S(y|x) = P_T(y|x)$.
- But, $\mathcal{X}_S \neq \mathcal{X}_T$ or $P_S(x) \neq P_T(x)$.

Single-Task Transfer Learning

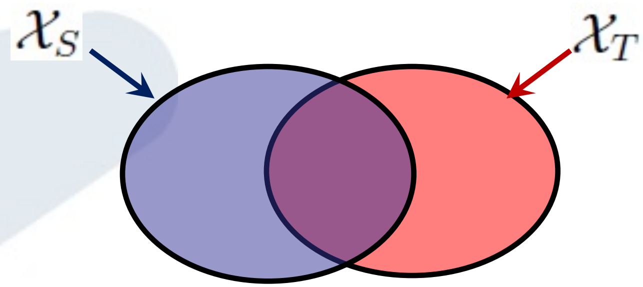
Case 1



Sample Selection Bias / Covariate Shift

Instance-based Transfer Learning Approaches

Case 2



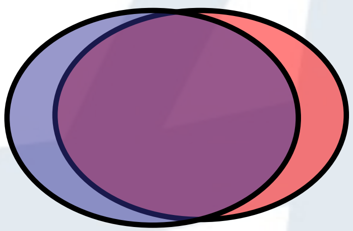
Domain Adaption in NLP

Feature-based Transfer Learning Approaches

A*STAR I²R

Single-Task Transfer Learning

Case 1



Sample Selection Bias / Covariate Shift

Instance-based Transfer Learning Approaches

Problem Setting

Given $\mathrm{D}_S = \{x_{S_i}, y_{S_i}\}_{i=1}^{n_S}$, $\mathrm{D}_T = \{x_{T_i}\}_{i=1}^{n_T}$,

Learn f_T, s.t. $\sum_i \epsilon(f_T(x_{T_i}), y_{T_i})$ is small,

where y_{T_i} is unknown.

Assumption

- $\mathcal{Y}_S = \mathcal{Y}_T$, and $P(Y_S|X_S) = P(Y_T|X_T)$,
- $\mathcal{X}_S \approx \mathcal{X}_T$,
- $P(X_S) \neq P(X_T)$.

Single-Task Transfer Learning

Instance-based Approaches

Recall, given a target task,

$$\begin{aligned}
\theta^* &= \arg\min \mathbb{E}_{(x,y)\sim P_T}[l(x,y,\theta)] \\
&= \arg\min \mathbb{E}_{(x,y)\sim P_T}\left[\frac{P_S(x,y)}{P_S(x,y)} l(x,y,\theta)\right] \\
&= \arg\min \int_y \int_x P_T(x,y)\left(\frac{P_S(x,y)}{P_S(x,y)} l(x,y,\theta)\right) dxdy \\
&= \arg\min \int_y \int_x P_S(x,y)\left(\frac{P_T(x,y)}{P_S(x,y)} l(x,y,\theta)\right) dxdy \\
&= \arg\min \mathbb{E}_{(x,y)\sim P_S}\left[\frac{P_T(x,y)}{P_S(x,y)} l(x,y,\theta)\right]
\end{aligned}$$

Single-Task Transfer Learning

Instance-based Approaches (cont.)

If $P_S(x, y) = P_T(x, y)$

$$\theta^* = \arg\min \mathbb{E}_{(x_T, y_T) \sim P_T}[l(x_T, y_T, \theta)]$$

$$\Updownarrow$$

$$\theta^* = \arg\min \mathbb{E}_{(x_S, y_S) \sim P_S}[l(x_S, y_S, \theta)]$$

$$\Downarrow$$

$$\theta^* = \arg\min \sum_{i=1}^{n_S} l(x_{S_i}, y_{S_i}, \theta) + \lambda \Omega(\theta)$$

A*STAR I²R

Single-Task Transfer Learning

Instance-based Approaches (cont.)

Assumption: $\{P_S(x) \neq P_T(x),\ P_S(y|x) = P_T(y|x)\} \Rightarrow P_S(x,y) \neq P_T(x,y)$

$$
\begin{aligned}
\theta^* &= \arg\min \mathbb{E}_{(x,y)\sim P_S}\left[\frac{P_T(x,y)}{P_S(x,y)} l(x,y,\theta)\right] \\
&= \arg\min \mathbb{E}_{(x,y)\sim P_S}\left[\frac{P_T(x)P_T(y|x)}{P_S(x)P_S(y|x)} l(x,y,\theta)\right] \\
&= \arg\min \mathbb{E}_{(x,y)\sim P_S}\left[\frac{P_T(x)}{P_S(x)} l(x,y,\theta)\right]
\end{aligned}
$$

$$
\text{Denote } \beta_i = \frac{P_T(x_{S_i})}{P_S(x_{S_i})},
$$

$$
\theta^* = \arg\min \sum_{i=1}^{n_S} \beta_i l(x_{S_i}, y_{S_i}, \theta) + \lambda\Omega(\theta)
$$

Single-Task Transfer Learning

Instance-based Approaches (cont.)

How to estimate $\beta_i = \frac{P_T(x_{S_i})}{P_S(x_{S_i})}$?

A simple solution is to first estimate $P_T(x), \; P_S(x),$ respectively,

and calculate $\frac{P_T(x_{S_i})}{P_S(x_{S_i})}$.

Sample Selection Bias / Covariate Shift
[Quionero-Candela, *etal,* Data Shift in Machine Learning, MIT Press 2009]

Reference

- [Thorndike and Woodworth, The Influence of Improvement in one mental function upon the efficiency of the other functions, 1901]
- [Taylor and Stone, Transfer Learning for Reinforcement Learning Domains: A Survey, JMLR 2009]
- [Pan and Yang, A Survey on Transfer Learning, IEEE TKDE 2009]
- [Quionero-Candela, *etal,* Data Shift in Machine Learning, MIT Press 2009]
- [Biltzer *etal*.. Domain Adaptation with Structural Correspondence Learning, *EMNLP* 2006]
- [Pan *etal*., Cross-Domain Sentiment Classification via Spectral Feature Alignment, WWW 2010]
- [Pan *etal*., Transfer Learning via Dimensionality Reduction, AAAI 2008]

Reference (cont.)

- [Pan *etal.*, Domain Adaptation via Transfer Component Analysis, IJCAI 2009]
- [Evgeniou and Pontil, Regularized Multi-Task Learning, KDD 2004]
- [Zhang and Yeung, A Convex Formulation for Learning Task Relationships in Multi-Task Learning, UAI 2010]
- [Saha *etal*, Learning Multiple Tasks using Manifold Regularization, NIPS 2010]
- [Argyriou *etal.*, Multi-Task Feature Learning, NIPS 2007]
- [Ando and Zhang, A Framework for Learning Predictive Structures from Multiple Tasks and Unlabeled Data, JMLR 2005]
- [Ji *etal*, Extracting Shared Subspace for Multi-label Classification, KDD 2008]

Reference (cont.)

- [Raina *etal.*, Self-taught Learning: Transfer Learning from Unlabeled Data, ICML 2007]
- [Dai *etal.,* Boosting for Transfer Learning, ICML 2007]
- [Glorot *etal.*, Domain Adaptation for Large-Scale Sentiment Classification: A Deep Learning Approach, ICML 2011]

Thank You