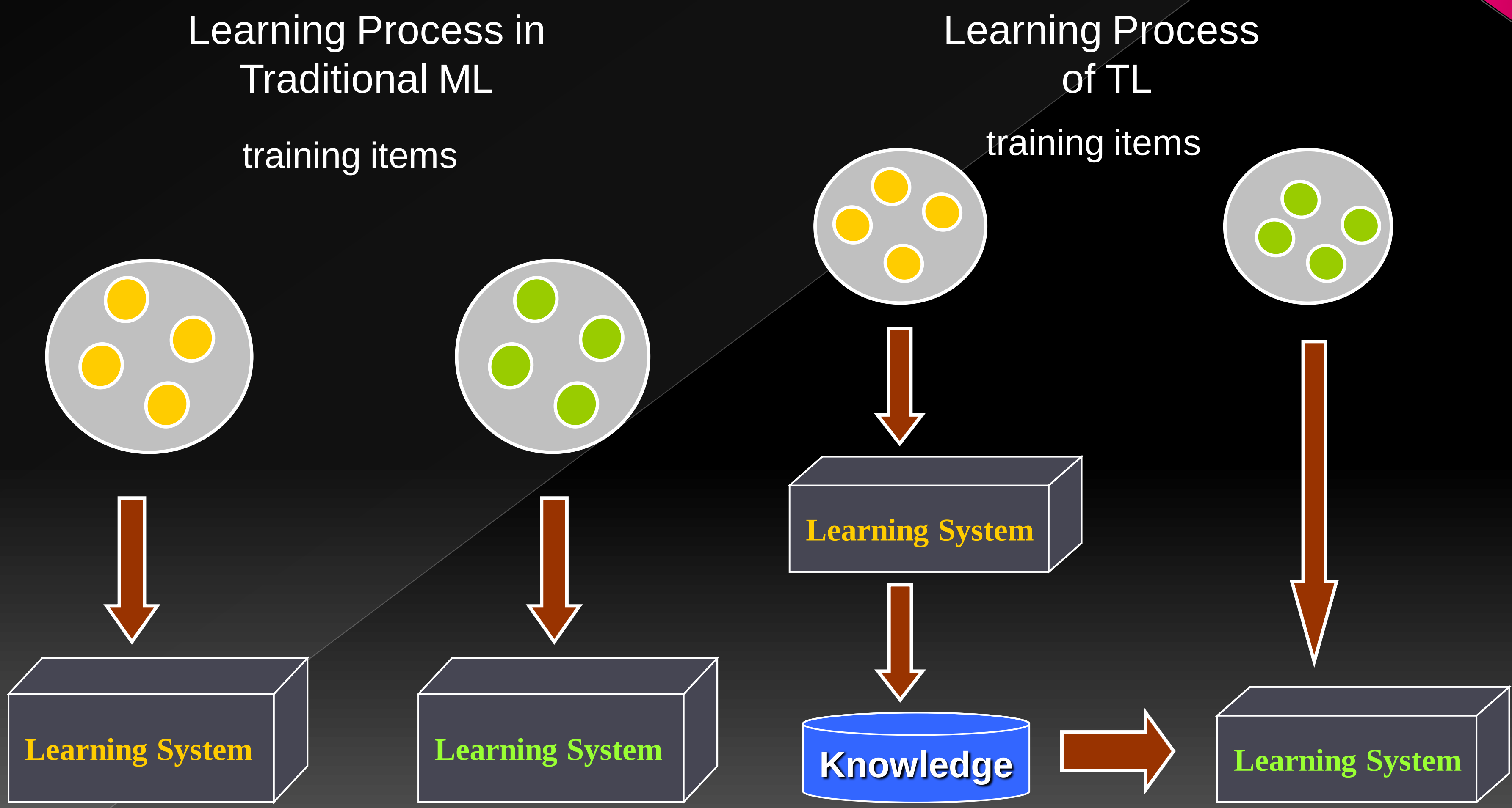
Transfer Learning for Action Recognition

Nazli FarajiDavar
CVSSP, FEPS, University of Surrey
Supervisors: Dr. T. DeCampos, Prof. J. Kittler

Outlines:

- Brief Literature review on transfer learning
- Theory of implemented algorithm
- Discussion
- Future work

Traditional ML VS. TL



Notation:

Domain:

It consists of two components: A feature space $\mathcal{X}$, a marginal distribution $P(x)$, where $x=\{x1, x2,... , xn\} \in \mathcal{X}$.

In general when two domains are different then they may have different feature spaces or different marginal distributions.

Task:

Given a specific domain and label space y, for each xi on domain, task is to predict its corresponding label yi where yi $\in$ Y and in mathematical manner we need $P(y|x)$, where $y=\{y1, y2,... , yn\} \in$ Y

Notation

For simplicity, we only consider at most two domains and two tasks.

Source domain: $\mathcal{P}(X_S)$, where $X_S = \{x_{S_1}, x_{S_2}, ..., x_{S_{n_S}}\} \in \mathcal{X}_S$

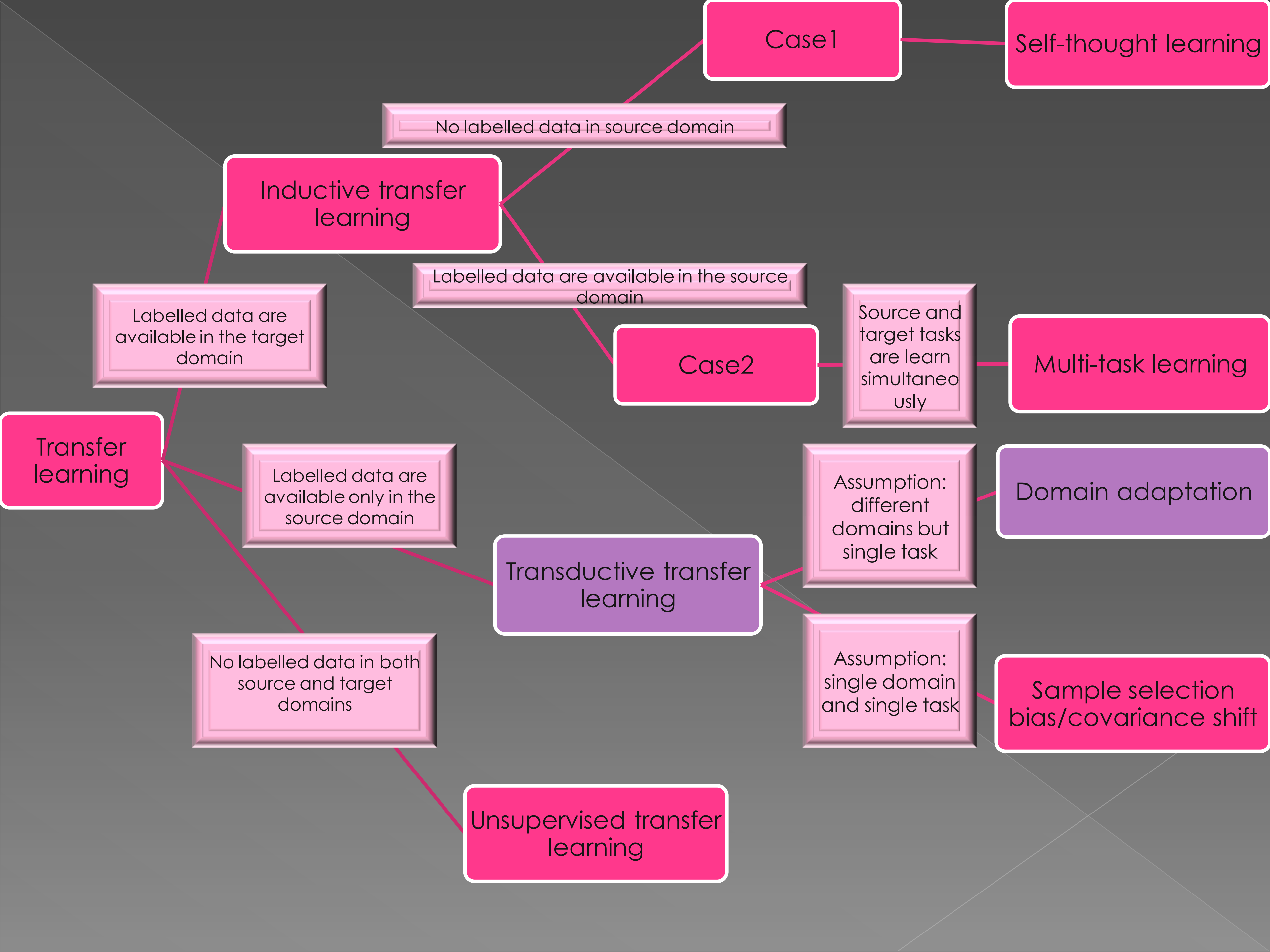
Task in the source domain:

$$\mathcal{P}(Y_S|X_S), \text{ where } Y_S = \{y_{S_1}, y_{S_2}, ..., y_{S_{n_S}}\} \text{ and } y_{S_i} \in \mathcal{Y}_S$$

Target domain: $\mathcal{P}(X_T)$, where $X_T = \{x_{T_1}, x_{T_2}, ..., x_{T_{n_T}}\} \in \mathcal{X}_T$

Task in the target domain:

$$\mathcal{P}(Y_T|X_T), \text{ where } Y_T = \{y_{T_1}, y_{T_2}, ..., y_{T_{n_T}}\} \text{ and } y_{T_i} \in \mathcal{Y}_T$$


Transfer learning
Labelled data are available in the target domain
Inductive transfer learning
No labelled data in source domain
Case1
Self-thought learning
Labelled data are available in the source domain
Case2
Source and target tasks are learn simultaneously
Multi-task learning
Labelled data are available only in the source domain
Transductive transfer learning
Assumption: different domains but single task
Domain adaptation
Assumption: single domain and single task
Sample selection bias/covariance shift
No labelled data in both source and target domains
Unsupervised transfer learning

Domain Adaptation

Motivation: If two domains are related to each other, then there may exist
some "pivot" features across both domain.

Main Idea: To identify correspondences among features from different domains by modelling their correlations with pivot features.

Feature enhancement methods: Clean the feature observed at test time so that they better match the train model

Model adaptation methods: Adapting the train model to better fit the new target/test domain

Arnold et. al.s Model Adaptation Algorithm

Input: *Dsrc={(xi, yi)}*, *Dtar* and the parameters of a classification model (KLDA model in our case)

Output: P(Yt|Xt)

Algorithm:

1.An initial estimation for target labels based on the model learnt for original source domain

2.Computing a transformation matrix for source data by rescaling either source features or samples based on a ratio of feature-label joint expectations in target to the expectation in source domain

3.Transforming the source domain and retraining on the new source space features

4.Use the learned classifier to predict the labels for the target data

The effect is to rescale each feature, putting more weights on features that occur frequently in the target but rarely in source, and down-weighting features that are common in the source but seldom seen in the target.

Mathematical details:

The transformation matrix derived as:

$$\forall i = 1 : N_{train}^{source}, G(f_j(x_i)) = f_j \frac{E_{\Lambda_{source}}^{target}[f_j, y_i]}{E^{source}[f_j, y_i]}$$

Where the *fj* is the *jth* feature of sample *xi* and *Λ-source* is the classification model learned on the original source data

A smoothing factor *Θ* has been used to to control the degree to which we use the target conditional estimates to alter the source conditionals

$$G'(f_j(x_i)) = \theta f_j(x_i) + (1 - \theta) G(f_j(x_i))$$

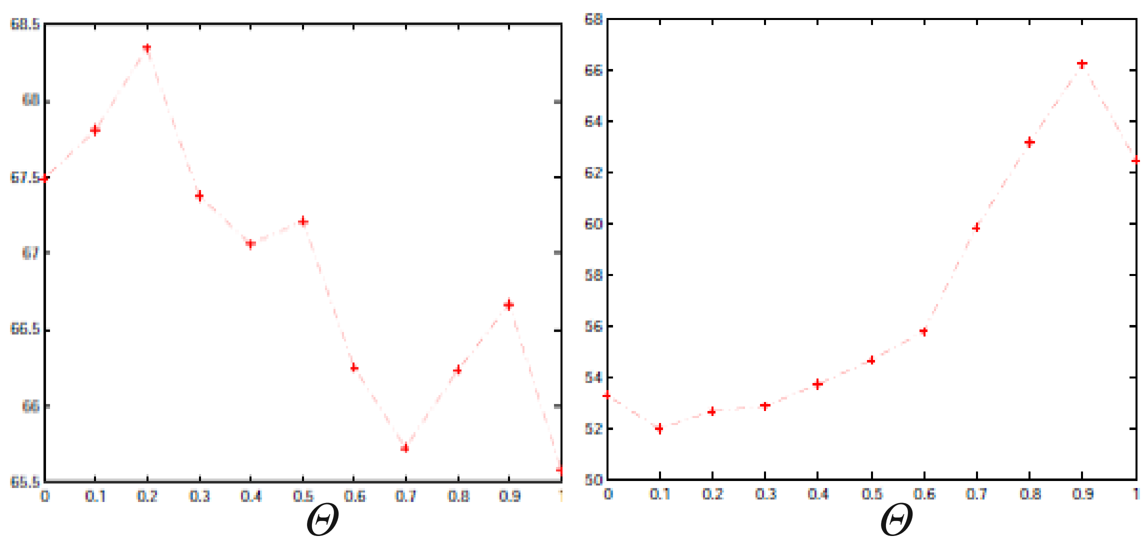
Experiential data sets:

Source domain data: Badminton game

Target domain data: Aus2003, Aus2003men, Japan 2009

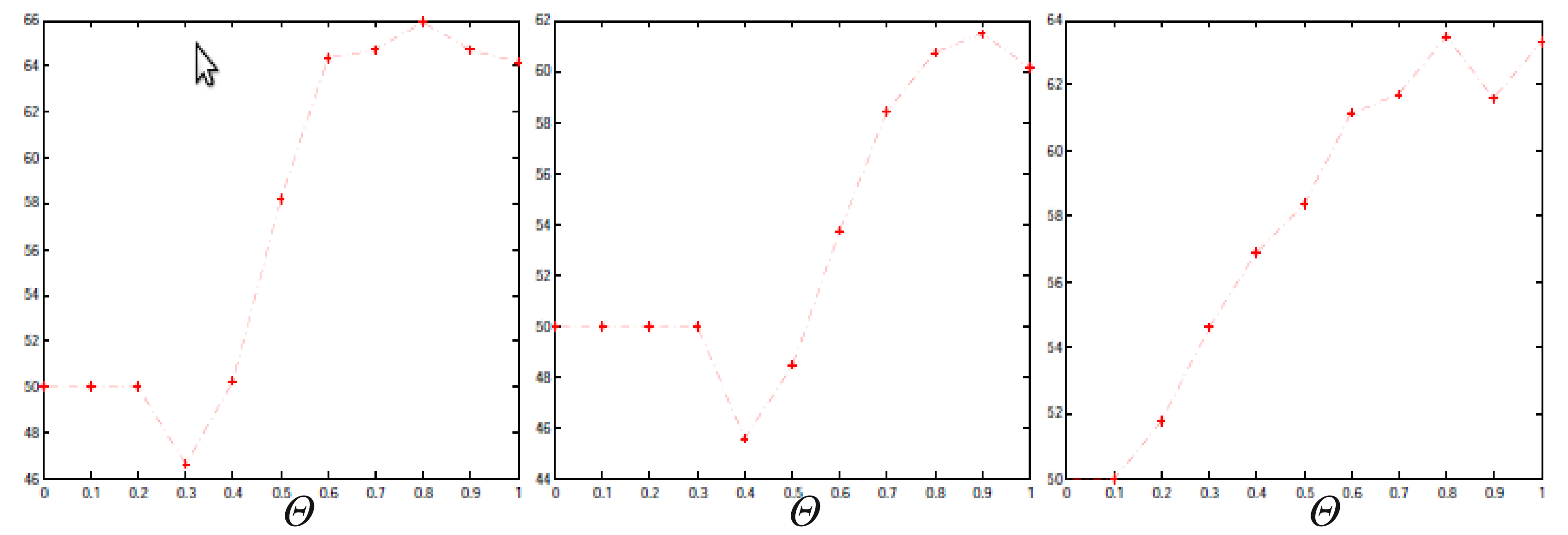
#	Source Data (Badminton)	Target (Aus2003)	Target (Aus2003 men)	Target (Japan2009)
Hit samples	704	946	1881	859
Non-Hit samples	461	284	590	283
Total number of samples	1165	1230	2471	1142

Feature re-weighting results:



(a) test:Aus2003 (b) test:Aus2003men (c) test:Japan2009

Kernel/sample re-weighting Results:



(a) test:Aus2003 (b) test:Aus2003men (c) test:Japan2009

Transfer learning results for kernel/sample re-weighting (*Θ-MA*), as it can be seen a particular *Θ* value (*Θ*=0.8) can be selected as a promising smoothing factor

Discussion:

Pros: A straight forward and fast method

Cons: The transformation model is too simple

Further experiments and tests:

-done experiments with lots of different setups of source/target matches

-fixed normalisation issues (Josef's suggestion)

-tried better transformations (translation+scaling)

Future work:

Based on the close intersection between this work and speaker adaptation in speech recognition problems some MLLR methods based on GMM are considered

References:

1. *Sinno Jialin Pan et al., ''A survey on transfer learning", IEEE transactions on knowladge and data engineering, 22:1345-1359, 2010*
2. *Andrew Arnolds et al., ''Acomparative Study of methods for transductive transfer learning", In proceeding of the 7th IEEE international conference on data mining workshops, ICDMW '7,pages 77-82, Washington DC, 2007*
3. *Alexander Klaser et al., ''A Spatio-Temporal descriptor based on 3D-gradients", British Machine Vision Conference, page995-1004, Sep2008*
4. *Rodriguz work*

Thank you!