# Transfer Learning based Activity Recognition via Domain Adaptation

Jindong Wang, Liping Jia

wangjindong,jialiping@ict.ac.cn

June 27, 2016

This project is based on the following papers:

- Pan S J, Kwok J T, Yang Q. Transfer Learning via Dimensionality Reduction[C] //AAAI. 2008, 8: 677-682.[PKY08]
- Pan S J, Tsang I W, Kwok J T, et al. Domain adaptation via transfer component analysis[C] //IJCAI 2009: 1187-1192.[PTKY09]

Activity recognition aims to seek the **profound high-level knowledge** about human activity through **low-level signals**, like:

- Motion sensor (accelerometer, gyroscope $\cdots$)
- Ambient sensor (microphone, light, camera $\cdots$)
- Context sensor (Wi-Fi, Bluetooth $\cdots$)
- Medical equipment (EMG $\cdots$)

For example:

- SmartGPA [WHH$^+$15], StudentLife [WCC$^+$14]
- ContextSense [CCW$^+$13], DoppleSleep [RAR$^+$15]
- Sound detect [RAZ$^+$14], safety test [JBR$^+$15]

## Traditional ML Assumptions

- Training and testing samples must be in the **same** feature distributions.
- Training samples must be **enough**.

## TL conditions

- Source and target domains do **not** need to be in the same distributions.
- **Less** training samples, even **none**.
- Example: getting labeled samples is time-consuming and expensive.

## Common Definition

- Wikipedia: research problem in machine learning that focuses on storing knowledge gained while solving one problem and applying it to a **different** but **related** problem [wik].

## Proceedings

- Data mining: ACM SIGKDD, IEEE ICDM, PKDD
- Machine learning: ICML, AAAI, IJCAI,NIPS, ECML
- Applications: ACM SIGIR, CVPR, ACL, IEEE TKDE

# Background
## TL notations

- Introduction
- Background
  - Activity Recognition
  - Transfer Learning
- Method
  - Algorithm
  - Domain Adaptation
  - TCA
  - Model Training
- Experiment
- Conclusion
- Resources

### Basic notations

- Domain: $\boldsymbol{D} = (\mathbf{X}, P(X))$,$\mathbf{X}$: feature space, $P(X)$: marginal distribution where $\mathbf{X} = \{X_1, X_2, \cdots, X_n\}$
- Task: $\boldsymbol{T} = (Y, f(\cdot))$,$Y$: label space, $f(\cdot)$: objective predictive function.

### Transfer learning

- Source domain: $\boldsymbol{D}_S = \{\mathbf{X}_S, P(X_S)\}$
- Source task: $\boldsymbol{T}_S = \{Y_S, f_S(\cdot)\}$
- Target domain: $\boldsymbol{D}_T = \{\mathbf{X}_T, P(X_T)\}$
- Target task: $\boldsymbol{T}_T = \{Y_T, f_T(\cdot)\}$
- Goal: $\min \epsilon(f_T(\mathbf{X}_T), Y_T)$

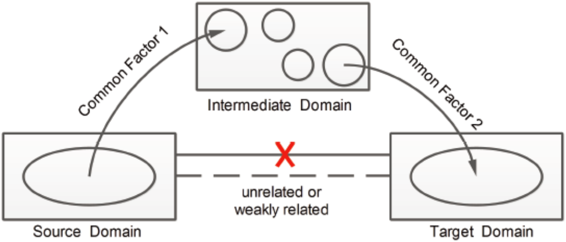
## Problem Definition

Given:

- Labeled $\boldsymbol{D}_{src} = \{\mathbf{X}_{src}, P(X_{src})\}$ with $Y = \{y_{src}\}$
- Unlabeled $\boldsymbol{D}_{tar} = \{\mathbf{X}_{tar}\}$

Task:

- Predict labels for $\mathbf{X}_{tar}$



The algorithm involves two steps:

1. Domain adaptation: brings two domains on the same feature space
   - Two domains must be close enough
   - No loss of structural information
2. Train a model on the new feature set

## Maximum Mean Discrepancy Embedding

$$\text{dist}(\mathbf{X}'_{src}, \mathbf{X}'_{tar}) = \|\frac{1}{n_1}\sum_{i=1}^{n_1}\phi(x'_{src_i}) - \frac{1}{n_2}\sum_{i=1}^{n_2}\phi(x'_{tar_i})\|_{\mathcal{H}}$$

## Using Kernel Matrix

$$\text{dist}(\mathbf{X}'_{src}, \mathbf{X}'_{tar}) = \text{trace}(KL)$$

$$K = \begin{bmatrix} K_{S,S} & K_{S,T} \\ K_{T,S} & K_{T,T} \end{bmatrix}, L = \begin{cases} \frac{1}{n_1^2} & x_i, x_j \in \mathbf{X}_{src} \\ \frac{1}{n_2^2} & x_i, x_j \in \mathbf{X}_{tar} \\ -\frac{1}{n_1 n_2} & \text{otherwise} \end{cases}$$

## Problem Induction

$$\begin{aligned} \min \quad & \text{trace}(KL) - \lambda \text{trace}(K) \\ \text{s.t.} \quad & K_{ii} + K_{jj} - 2K_{ij} + 2\epsilon = d_{ij}^2 \\ & K\mathbf{1} = -\epsilon\mathbf{1} \end{aligned}$$

- This is an **semidefinite program**, can be solved using standard SDP solvers.
- We used CVXPY: `http://cvxpy.com/`

SDP is probably the most exciting development in mathematical programming in the last ten years.

## Standard Form of SDP

$$\begin{aligned} \min \quad & C \bullet X \\ \text{s.t.} \quad & A_i \bullet X = b_i, i = 1, \cdots, m, \\ & X \succeq 0 \end{aligned}$$

where

$$C \bullet X := \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} X_{ij} = \text{trace}(CX)$$

$LP \in QP \in QCQP \in SDP \in CP$.

## Problem Induction

$$\begin{aligned} \min \quad & \text{trace}(KL) - \lambda \text{trace}(K) \\ \text{s.t.} \quad & K_{ii} + K_{jj} - 2K_{ij} + 2\epsilon = d_{ij}^2 \\ & K\mathbf{1} = -\epsilon \mathbf{1} \end{aligned}$$

## SDP Induction (Our Work)

$$\begin{aligned} \min \quad & \text{trace}((L - \lambda I)K) \\ \text{s.t.} \quad & A^{(m)} \bullet K = D_{ij} \end{aligned}$$

where

$$A^{(m)} = \begin{cases} A_{ii}^{(m)} = A_{jj}^{(m)} = 1 \\ A_{ij}^{(m)} = A_{ji}^{(m)} = -1 \end{cases}$$

Solving such an SDP problem is very expensive.
In fact, it's $O(n_1 + n_2)^{6.5}$.

### Transfer Component Analysis

$$\min_{W} \quad \mathrm{tr}(W^T KLKW) + \mu \mathrm{tr}(W^T W)$$
$$\text{s.t.} \quad W^T KHKW = I_m$$

where $H = I_{n_1+n_2} - (\frac{1}{n_1+n_2})\mathbf{1}\mathbf{1}^T$, and $W = K^{-1/2}\widetilde{W}$ where $\widetilde{W} \in \mathbb{R}^{(n_1+n_2)\times m}$ transforms the empirical kernel map features to an $m$-dimensional space.

TCA takes only $O(m(n_1 + n_2))$ time when $m$-dimensional eigenvectors are to be extracted.

After obtaining $K$:

- Apply PCA to $K$ to get new representations $\{x'_{src_i}\}$ and $\{x'_{tar_i}\}$
- Learn a classifier or regressor $f : x'_{src_i} \rightarrow y_{src_i}$
- Use $f$ to predict the labels of $\boldsymbol{D}_{tar}$, as $y_{tar_i} = f(x'_{tar_i})$
- Use harmonic functions to predict new data $\boldsymbol{D}^{new}_{tar}$

Here we choose $f$ to be a **random forest** classifier.

### Dataset

We use UCI ADL(activity of daily living) [AB10] dataset to perform evaluation using MATLAB.

| | | | |
|---|---|---|---|
| 8 persons | 19 activites | 3 sensors | 5 body parts |
| 45 columns | 3 axis | 25 Hz | 5 min/act |

### Experiments

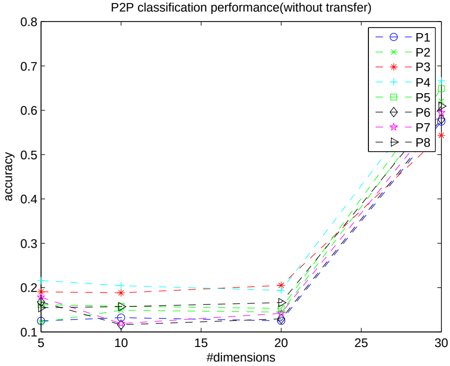
We performed 3 kinds of experiments:

- Basic classification without transfer
- P2P: Same feature space, different person
- S2S: Varied feature space, same person

Before feature extraction, we integrate the 3 axis as 1 for every sensor using $a = \sqrt{x^2 + y^2 + z^2}$.

- **5s**: We use sliding window (5s) to perform feature extraction on time and frequency domains [fea].
- **405 features**: For every sensor, we extracted 27 features, that's 405 features in total.
- **30 features**: We applied PCA to perform dimensionality reduction: 405 $\rightarrow$ 30 [AB10].

This experiment involves no transfer.
Train a model on person $i$, and test on the others.

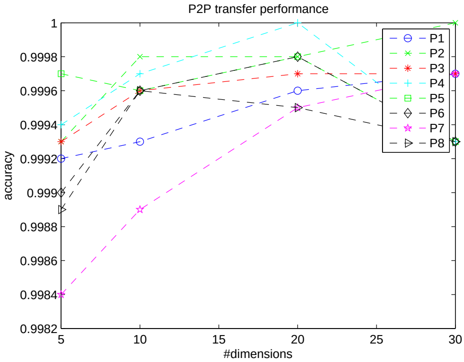



- Apply classification directly leads to poor performance.
- The performance decreases with dimensions.

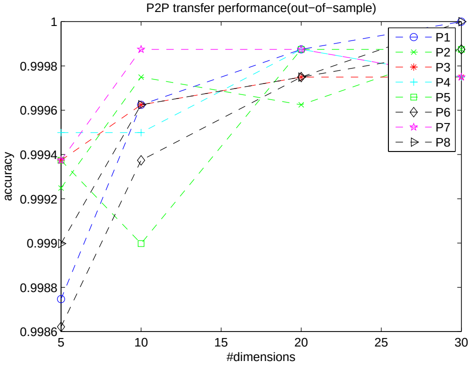
We split the target domain into 2 parts:

- Unlabeled part: to perform TCA and to test
- Out-of-sample part: only to test

Result tested on the unlabeled data:



Result tested on the out-of-sample data:

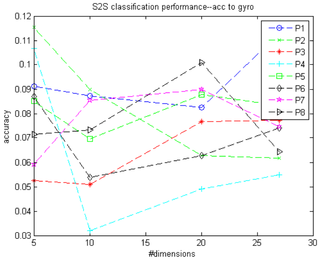



- Transfer works. No subject to dimensions.
- For new data (out-of-sample), transfer still works.

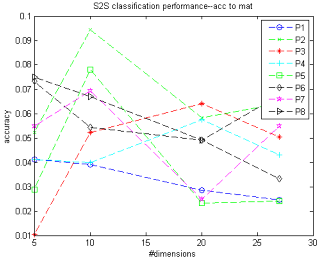
For torso part, transfer from sensor $i$ to sensor $j$.

Same as P2P, we split the unlabeled data into 2 sets.

Result tested on the unlabeled data:



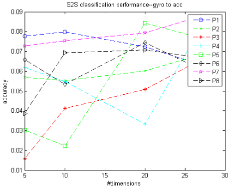

(a) Unlabeled:a-g



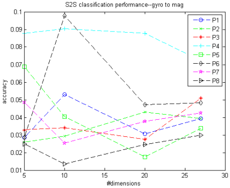

(b) Unlabeled:a-m

# Experiment
## Results of sensor to sensor transfer

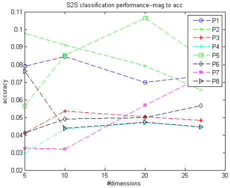




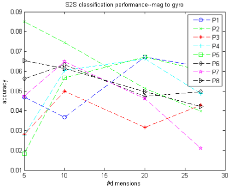

(c) Unlabeled:g-a




(d) Unlabeled:g-m

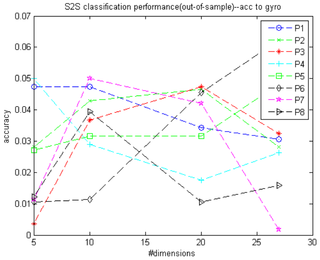



(e) Unlabeled:m-a

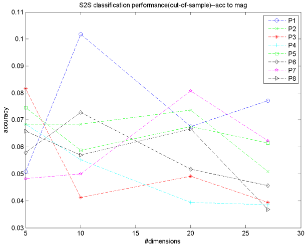



(f) Unlabeled:m-g

Testing results on out-of-sample data:



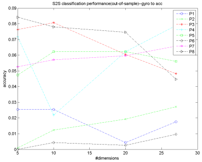

(g) OOS:a-g



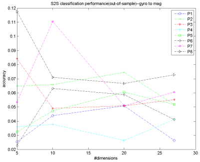

(h) OOS:a-m

# Experiment

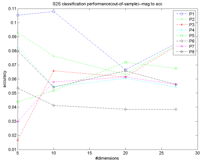
## Results of sensor to sensor transfer



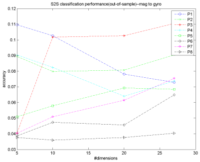

(i) OOS:g-a




(j) OOS:g-m




(k) OOS:m-a




(l) OOS:m-g

- Basic classification experiment:
  - Apply classification directly leads to poor performance.
  - The performance decreases with dimensions.
- P2P transfer (w/o) experiment:
  - Transfer works. No subject to dimensions.
  - For new data (out-of-sample), transfer still works.
- S2S transfer experiment:
  - Poor performance for S2S.
  - Still need some sensor specified info.

# Conclusion

TCA based activity recognition does achieve some good results, but:

## Pros

- Generate reliable results for different feature spaces.
- A new way of dimensionality reduction.

## Cons

- Lack theoretical support.
- Cannot generalize for new emerging data.
- Poor performance for low dimensional data.

### People

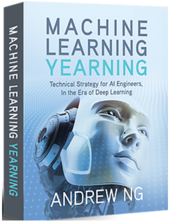
- *Qiang Yang*: IEEE/IAPR/AAAS fellow, AAAI councilor
- *Sinno Jialin Pan*: `http://ntu.edu.sg/home/sinnopan/`
- *Wenyuan Dai*: `http://www.4paradigm.com`

### Survey

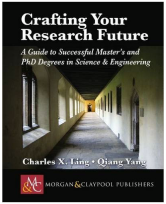
- A survey on Transfer Learning [PY10].
- Transfer learning for activity recognition: A survey [CFK13].
- Transitive Transfer Learning [TSZY15].
- Fuzzy Transfer Learning [SC15].

# Resources
## Book Sharing






Andrew Ng's new ML book: `Machine Learning Yearning.`




`Ling C X, Yang Q. Crafting Your Research Future: A Guide to Successful Master's and Ph. D. Degrees in Science & Engineering[J].`

# References I



Kerem Altun and Billur Barshan.
Human activity recognition using inertial/magnetic sensor units.
In *International Workshop on Human Behavior Understanding*, pages 38–51. Springer, 2010.

Zhenyu Chen, Yiqiang Chen, Shuangquan Wang, Junfa Liu, Xingyu Gao, and Andrew T Campbell.
Inferring social contextual behavior from bluetooth traces.
In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, pages 267–270. ACM, 2013.

Diane Cook, Kyle D Feuz, and Narayanan C Krishnan.
Transfer learning for activity recognition: A survey.
*Knowledge and information systems*, 36(3):537–556, 2013.

https://www.zhihu.com/question/41068341.

Shubham Jain, Carlo Borgiattino, Yanzhi Ren, Marco Gruteser, Yingying Chen, and Carla Fabiana Chiasserini.
Lookup: Enabling pedestrian safety services via shoe sensing.
In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, pages 257–271. ACM, 2015.

Sinno Jialin Pan, James T Kwok, and Qiang Yang.
Transfer learning via dimensionality reduction.
In *AAAI*, volume 8, pages 677–682, 2008.

Sinno Jialin Pan, Ivor W Tsang, James T Kwok, and Qiang Yang.
Domain adaptation via transfer component analysis.
In *Proceedings of the 21st international jont conference on Artifical intelligence*, pages 1187–1192. Morgan Kaufmann Publishers Inc., 2009.

Sinno Jialin Pan and Qiang Yang.
A survey on transfer learning.
*Knowledge and Data Engineering, IEEE Transactions on*, 22(10):1345–1359, 2010.

# References II



Tauhidur Rahman, Alexander T Adams, Ruth Vinisha Ravichandran, Mi Zhang, Shwetak N Patel, Julie A Kientz, and Tanzeem Choudhury.
Dopplesleep: A contactless unobtrusive sleep sensing system using short-range doppler radar.
In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 39–50. ACM, 2015.

Tauhidur Rahman, Alexander Travis Adams, Mi Zhang, Erin Cherry, Bobby Zhou, Huaishu Peng, and Tanzeem Choudhury.
Bodybeat: a mobile system for sensing non-speech body sounds.
In *MobiSys*, volume 14, pages 2–13, 2014.

Jethro Shell and Simon Coupland.
Fuzzy transfer learning: methodology and application.
*Information Sciences*, 293:59–79, 2015.

Ben Tan, Yangqiu Song, Erheng Zhong, and Qiang Yang.
Transitive transfer learning.
In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1155–1164. ACM, 2015.

Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell.
Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones.
In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 3–14. ACM, 2014.

Rui Wang, Gabriella Harari, Peilin Hao, Xia Zhou, and Andrew T Campbell.
Smartgpa: how smartphones can assess and predict academic performance of college students.
In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 295–306. ACM, 2015.

https://en.wikipedia.org/wiki/Inductive_transfer.

# Thank You