Pervasive Computing Research Center,
Institute of Computing Technology, Chinese Academy of Sciences

Introduction to Transfer Learning

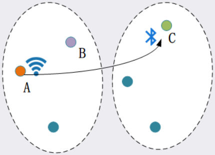
Jindong Wang
wangjindong@ict.ac.cn

Jun. 3, 2016

Content

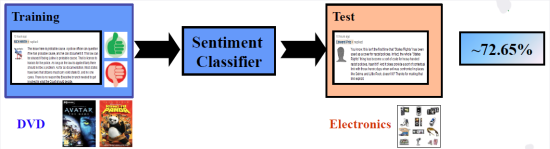
Activity Recognition: An Example

A: labeled Wi-fi $\rightarrow$ C:unlabeled Bluetooth, but how?



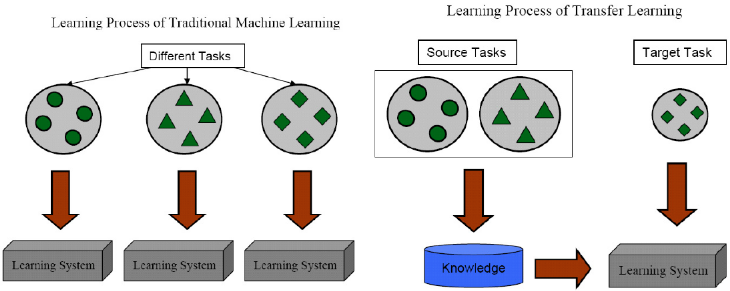
Sentiment Classification: Another Example

Only sentiments on DVD, how to obtain those on Electronics?



Basics

- Building every model from scratch is time-consuming and expensive.
- But there are many existing knowledge. Can we reuse them?



(a) Traditional Machine Learning (b) Transfer Learning

The Origin of TL

- *Thorndike* and *Woodworth* in 1901:
  how individuals **transfer** in one context to another context that share **similar** characteristics [TW01].

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, NIPS, ECML, AAAI, IJCAI
- Applications: ACM SIGIR, WWW, ACL, IEEE TKDE

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.

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: $\min \epsilon(f_T(\mathbf{X}_T), Y_T)$
- Conditions: $\boldsymbol{D}_T \neq \boldsymbol{D}_S$ or $\boldsymbol{T}_T \neq \boldsymbol{T}_S$ with $(\boldsymbol{D}_T, \boldsymbol{D}_S, Y_T, Y_S)$ may be unknown, respectively

By Data Distribution

- Inductive TL
- Transductive TL
- Unsupervised TL

By Methodology

- Instance based TL
- Feature based TL
- Parameter/model based TL
- Relational TL

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

Figure: Transfer learning settings[PY10]

Inductive transfer learning

Given $\boldsymbol{T}_S \neq \boldsymbol{T}_T$ under conditions:

- A lot of labeled $\boldsymbol{D}_S$ or
- No labeled $\boldsymbol{D}_S$

Transductive transfer learning

Given $\boldsymbol{T}_S = \boldsymbol{T}_T$ under conditions:

- $\mathbf{X}_S \neq \mathbf{X}_T$ or
- $\mathbf{X}_S = \mathbf{X}_T$ and $P(X_S) \neq P(X_T)$

Unsupervised transfer learning

Given $\boldsymbol{T}_S \neq \boldsymbol{T}_T$ under conditions:

- No labeled $\boldsymbol{D}_S$ and $\boldsymbol{D}_T$

Instance based transfer learning

Reuse source domain: instance re-weighting and importance sampling

Feature based transfer learning

Learn good feature representation of target domain

Parameter based transfer learning

Transfer models between source and target domains

Relational transfer learning

Relationships are same in source and target domains

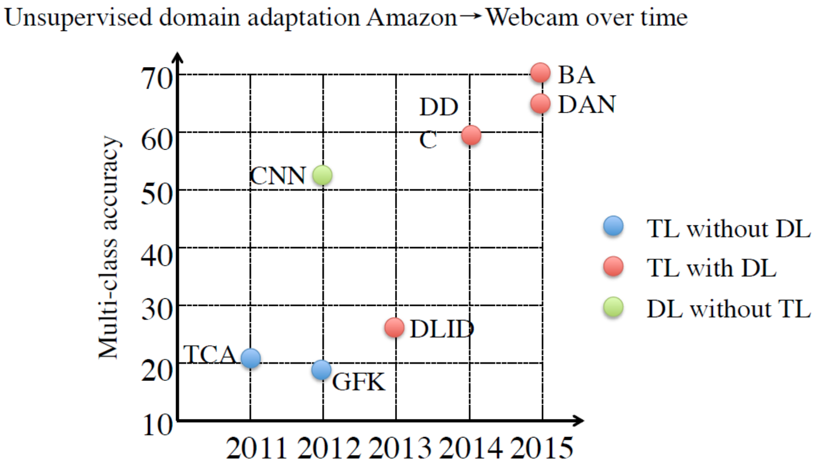
Deep Learning: Nonlinear Representations

- **Hierarchical** network.
- **disentangle** different explanatory factors of variation behind data samples.

Transfer Learning: Alleviation

- Doesn’t need a large **amount** of data.

Applying transfer learning with deep learning **outperforms** directly applying Deep Learning.



Negative Transfer

Negative transfer happens when source domain data and task contribute to **reduced** performance of learning in the target domain.

Negative Transfer Conditions

- Domains are too dissimilar [RMKD05]
- Conditional Kolmogorov complexity is not related [BH03]
- Tasks are not well-related [BH03]

Transitive transfer learning [TSZY15]

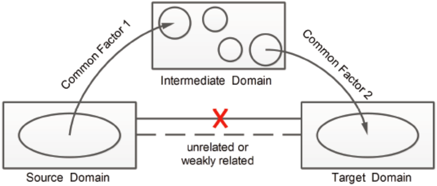


Figure: TTL tries to bridge the source and target domain using auxiliary sources

Can we transfer from existing activities domain to a different but related domain?[ZHY09]

Problem Formulation

- Labeled source activities: $\mathbf{A}_{src} = \{a_1, a_2, \cdots, a_m\}$
- Unlabeled target activities: $\mathbf{A}_{tar} = \{a_{m+1}, a_{m+2}, \cdots, a_n\}, \mathbf{A}_{src} \cap \mathbf{A}_{tar} = \emptyset$.

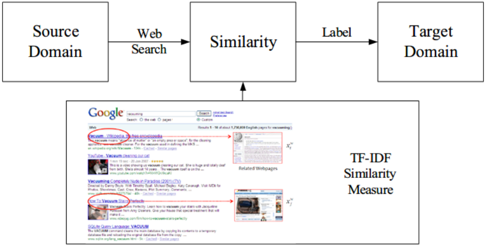


Figure: Cross domain activity recognition

Similarity Measure

What is a good similarity measure?

Maximum Mean Discrepancy [BGR+06]

- $X = \{x_1, x_2, \cdots, x_n\}, Y = \{y_1, y_2, \cdots, y_m\}$, i.i.d
- $\|\cdot\|_{\mathcal{H}}$: Reproducing Hilbert Kernel Space
- ϕ: kernel function, like Gaussian

$$MMD^2(X,Y) = \|\frac{1}{n}\sum_{i=1}^{n}\phi(x_i) - \frac{1}{m}\sum_{i=1}^{m}\phi(y_i)\|_{\mathcal{H}}^2$$

Other Measures

- Cosine similarity: $sim(X,Y) = \frac{X \cdot Y}{|X||Y|}$
- Kullback-Leibler (KL) divergence: $D_{KL}(P\|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$
- Jensen-Shannon divergence (JSD): let $M = \frac{1}{2}(P+Q)$, then $JSD(P\|Q) = \frac{1}{2}D_{KL}(P\|M) + \frac{1}{2}D_{KL}(Q\|M)$

Text Mining

- Unified clustering and shared knowledge transfer [DXYY07a]
- Transfered Bayes [DXYY07b]

Image Processing

- Text to image clustering [DCX+08]
- Heterogeneous transfer learning [ZCL+11b]

Collaborative Filtering

- Sub feature space transfer [PXY12]
- Latent feature sharing [CLY10]

Indoor Localization

- Transfer similar floors [WZZY10]
- A transfer learning framework [ZY14]

Activity Recognition

- Cross-people mobile phone based AR [ZCL+11a]
- Community similarity network [LXL+11]

Resources

- Open source program: `http://www.cse.ust.hk/TL/`
- *Qiang Yang*: `http://www.cs.ust.hk/~qyang/`
- *Sinno Jialin Pan*: `http://www.ntu.edu.sg/home/sinnopan/`
- *Wenyuan Dai*: `http://www.4paradigm.com/homepage.html`

Survey

- A survey on Transfer Learning [PY10].
- A survey of Transfer Learning [WKW16].
- Transfer learning for activity recognition: A survey [CFK13].
- Fuzzy Transfer Learning: Methodology and application [SC15].

Future Work

- Reliable similarity measure
- Transfer within different algorithms
- More accurate theoretical support

Figure: The future of machine learning [Yan16]

References I

[BGR+06] Karsten M Borgwardt, Arthur Gretton, Malte J Rasch, Hans-Peter Kriegel, Bernhard Schölkopf, and Alex J Smola.
Integrating structured biological data by kernel maximum mean discrepancy.
Bioinformatics, 22(14):e49–e57, 2006.

[BH03] Bart Bakker and Tom Heskes.
Task clustering and gating for bayesian multitask learning.
The Journal of Machine Learning Research, 4:83–99, 2003.

[CFK13] 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.

[CLY10] Bin Cao, Nathan N Liu, and Qiang Yang.
Transfer learning for collective link prediction in multiple heterogenous domains.
In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 159–166, 2010.

[DCX+08] Wenyuan Dai, Yuqiang Chen, Gui-Rong Xue, Qiang Yang, and Yong Yu.
Translated learning: Transfer learning across different feature spaces.
In *Advances in neural information processing systems*, pages 353–360, 2008.

[DXYY07a] Wenyuan Dai, Gui-Rong Xue, Qiang Yang, and Yong Yu.
Co-clustering based classification for out-of-domain documents.
In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 210–219. ACM, 2007.

[DXYY07b] Wenyuan Dai, Gui-Rong Xue, Qiang Yang, and Yong Yu.
Transferring naive bayes classifiers for text classification.
In *Proceedings of the national conference on artificial intelligence*, volume 22, page 540. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2007.

[LXL+11] Nicholas D Lane, Ye Xu, Hong Lu, Shaohan Hu, Tanzeem Choudhury, Andrew T Campbell, and Feng Zhao.
Enabling large-scale human activity inference on smartphones using community similarity networks (csn).
In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 355–364. ACM, 2011.

[PXY12] Weike Pan, Evan Wei Xiang, and Qiang Yang.
Transfer learning in collaborative filtering with uncertain ratings.
In *AAAI*, 2012.

[PY10] Sinno Jialin Pan and Qiang Yang.
A survey on transfer learning.
Knowledge and Data Engineering, IEEE Transactions on, 22(10):1345–1359, 2010.

[RMKD05] Michael T Rosenstein, Zvika Marx, Leslie Pack Kaelbling, and Thomas G Dietterich.
To transfer or not to transfer.
In *NIPS 2005 Workshop on Transfer Learning*, volume 898, 2005.

References II

[SC15] Jethro Shell and Simon Coupland.
Fuzzy transfer learning: methodology and application.
Information Sciences, 293:59–79, 2015.

[TSZY15] 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.

[TW01] Edward L Thorndike and RS Woodworth.
The influence of improvement in one mental function upon the efficiency of other functions. ii. the estimation of magnitudes.
Psychological Review, 8(4):384, 1901.

[wik] https://en.wikipedia.org/wiki/Inductive_transfer.

[WKW16] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang.
A survey of transfer learning.
Journal of Big Data, 3(1):1–40, 2016.

[WZZY10] Hua-Yan Wang, Vincent W Zheng, Junhui Zhao, and Qiang Yang.
Indoor localization in multi-floor environments with reduced effort.
In *Pervasive Computing and Communications (PerCom), 2010 IEEE International Conference on*, pages 244–252. IEEE, 2010.

[Yan16] Qiang Yang.
Transfer learning report.
2016.

[ZCL+11a] Zhongtang Zhao, Yiqiang Chen, Junfa Liu, Zhiqi Shen, and Mingjie Liu.
Cross-people mobile-phone based activity recognition.
In *IJCAI*, volume 11, pages 2545–250. Citeseer, 2011.

[ZCL+11b] Yin Zhu, Yuqiang Chen, Zhongqi Lu, Sinno Jialin Pan, Gui-Rong Xue, Yong Yu, and Qiang Yang.
Heterogeneous transfer learning for image classification.
In *AAAI*, 2011.

[ZHY09] Vincent Wenchen Zheng, Derek Hao Hu, and Qiang Yang.
Cross-domain activity recognition.
In *Proceedings of the 11th international conference on Ubiquitous computing*, pages 61–70. ACM, 2009.

[ZY14] Hankz Hankui Zhuo and Qiang Yang.
Action-model acquisition for planning via transfer learning.
Artificial intelligence, 212:80–103, 2014.

Q & A

Thank you for your listening