

迁移学习





主要内容

一

迁移学习简介

二

迁移学习的分类方法

三

代表性研究工作

四

问题与展望

五

迁移学习资源

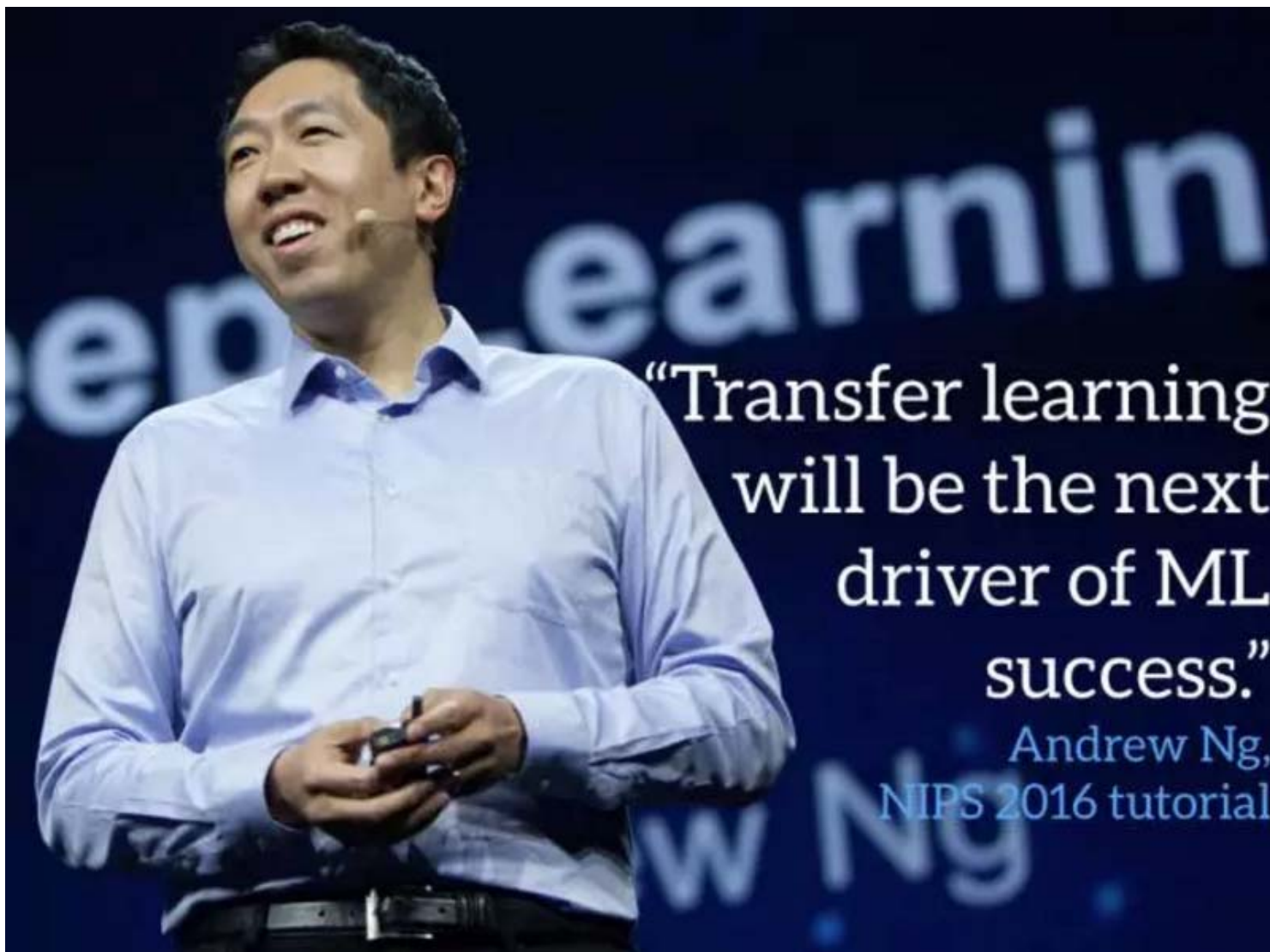


迁移学习简介

- 定义与概念
- 迁移学习 vs 传统机器学习
- 与其他概念的对比



引子



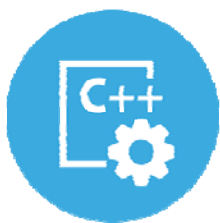
迁移学习将会是引领下一次机器学习热潮的驱动力。——吴恩达,NIPS 2016



迁移学习简介

- 什么是迁移学习？

- **心理学角度**：人们利用之前的经验和知识进行推理和学习的能力。
- **机器学习角度**：一个系统将别的相关领域中的知识应用到本应用中的学习模式。[DARPA]
- 举例：C++→Java；骑自行车→骑摩托车
- 关键词：举一反三



- 迁移学习要解决的问题：

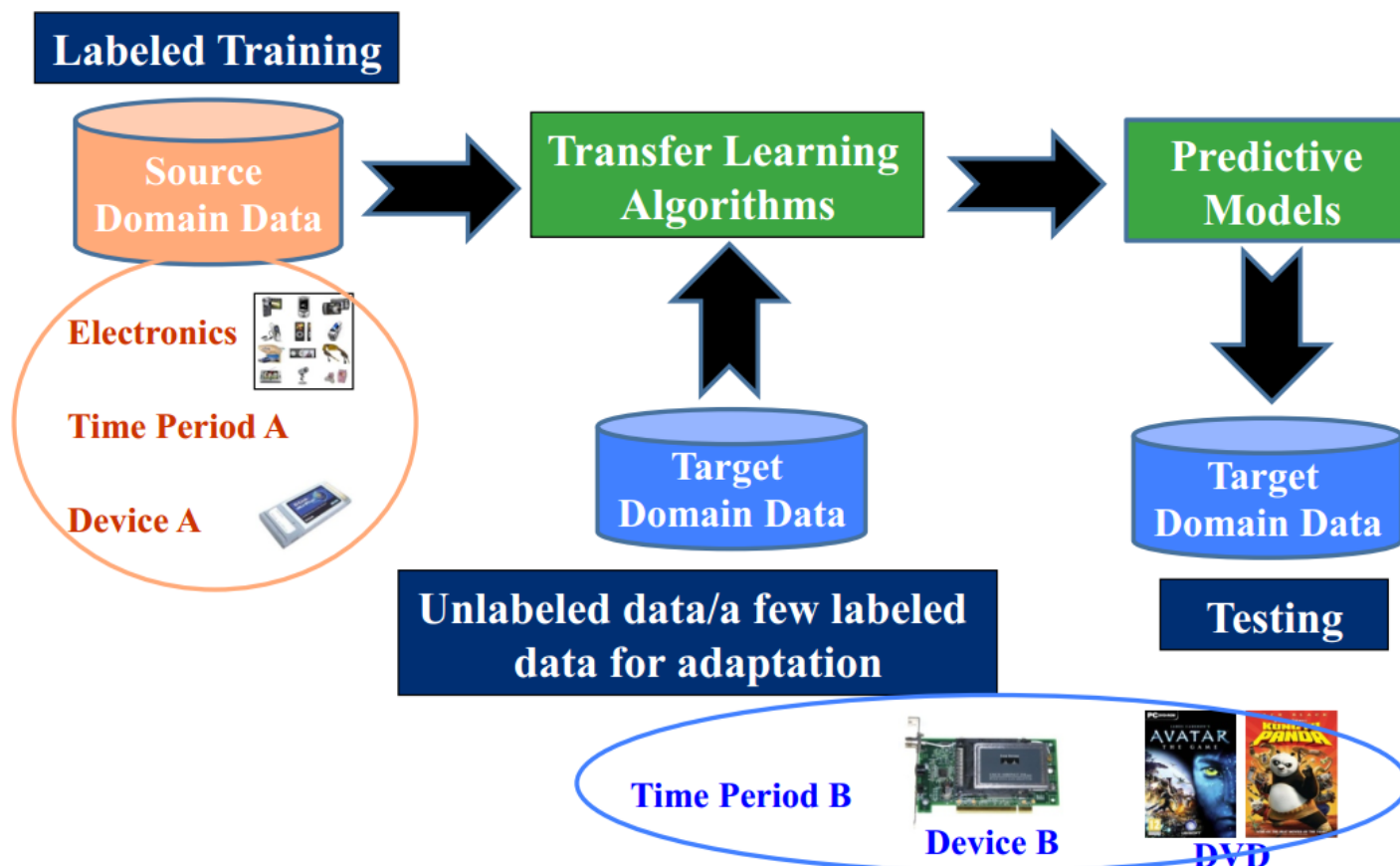
- 给定一个研究领域和任务，如何利用相似领域进行知识的迁移，从而达成目标？



迁移学习简介

- 为什么要进行迁移学习？
 - 数据的标签很难获取
 - 从头建立模型是复杂和耗时的

对已有知识的重用是必要的

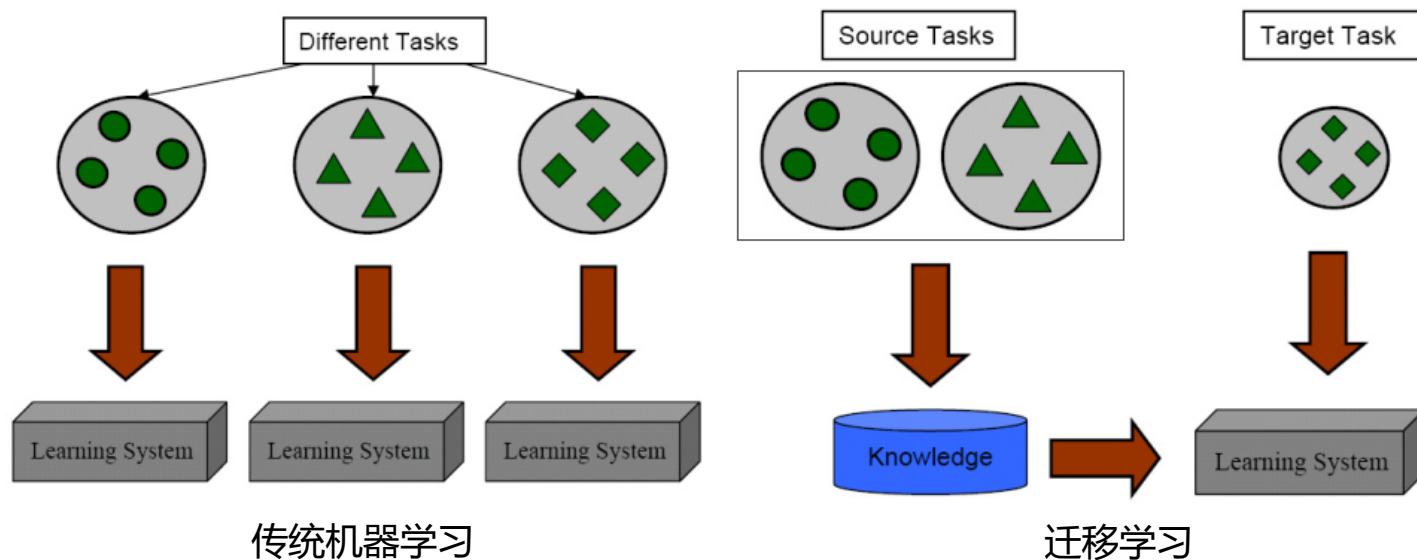




迁移学习简介

- 迁移学习 vs 传统机器学习

	传统机器学习	迁移学习
数据分布	训练和测试数据同分布	训练和测试数据不需要同分布
数据标签	足够的数据标注	不需要足够的数据标注
建模	每个任务分别建模	可以重用之前的模型





迁移学习简介

- 与其他概念的对比
 - Life-long learning (终身学习) :连续不断地在一个域上学习
 - Multi-task learning (多任务学习) : 两个任务同时完成
 - Domain adaptation (域适配) : 迁移学习的子类
 - Incremental learning (增量学习) : 一个域上的不断学习
 - Self-taught learning (自我学习) : 从自身数据中学习
 - Covariance shift (协方差漂移) : 迁移学习的子类
 -
- 迁移学习与其他已有概念相比，着重强调学习任务之间的**相关性**，并利用这种相关性完成知识之间的迁移。



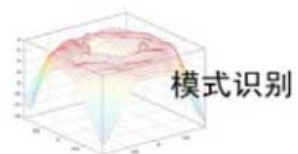
迁移学习形式化概念

- 迁移学习常用概念
 - **Domain (域)**：由数据特征和特征分布组成，是学习的主体
 - Source domain (源域)：已有知识的域
 - Target domain (目标域)：要进行学习的域
 - **Task (任务)**：由目标函数和学习结果组成，是学习的结果
- 迁移学习的形式化定义
 - 条件：给定一个源域 $\mathcal{D}_S$ 和源域上的学习任务 $\mathcal{T}_S$ ，目标域 $\mathcal{D}_T$ 和目标域上的学习任务 $\mathcal{T}_T$
 - 目标：利用 $\mathcal{D}_S$ 和 $\mathcal{T}_S$ 学习在目标域上的预测函数 $f(\cdot)$ 。
 - 限制条件： $\mathcal{D}_S \neq \mathcal{D}_T$ 或 $\mathcal{T}_S \neq \mathcal{T}_T$



迁移学习

- 应用领域



机器学习
迁移学习





迁移学习的分类方法

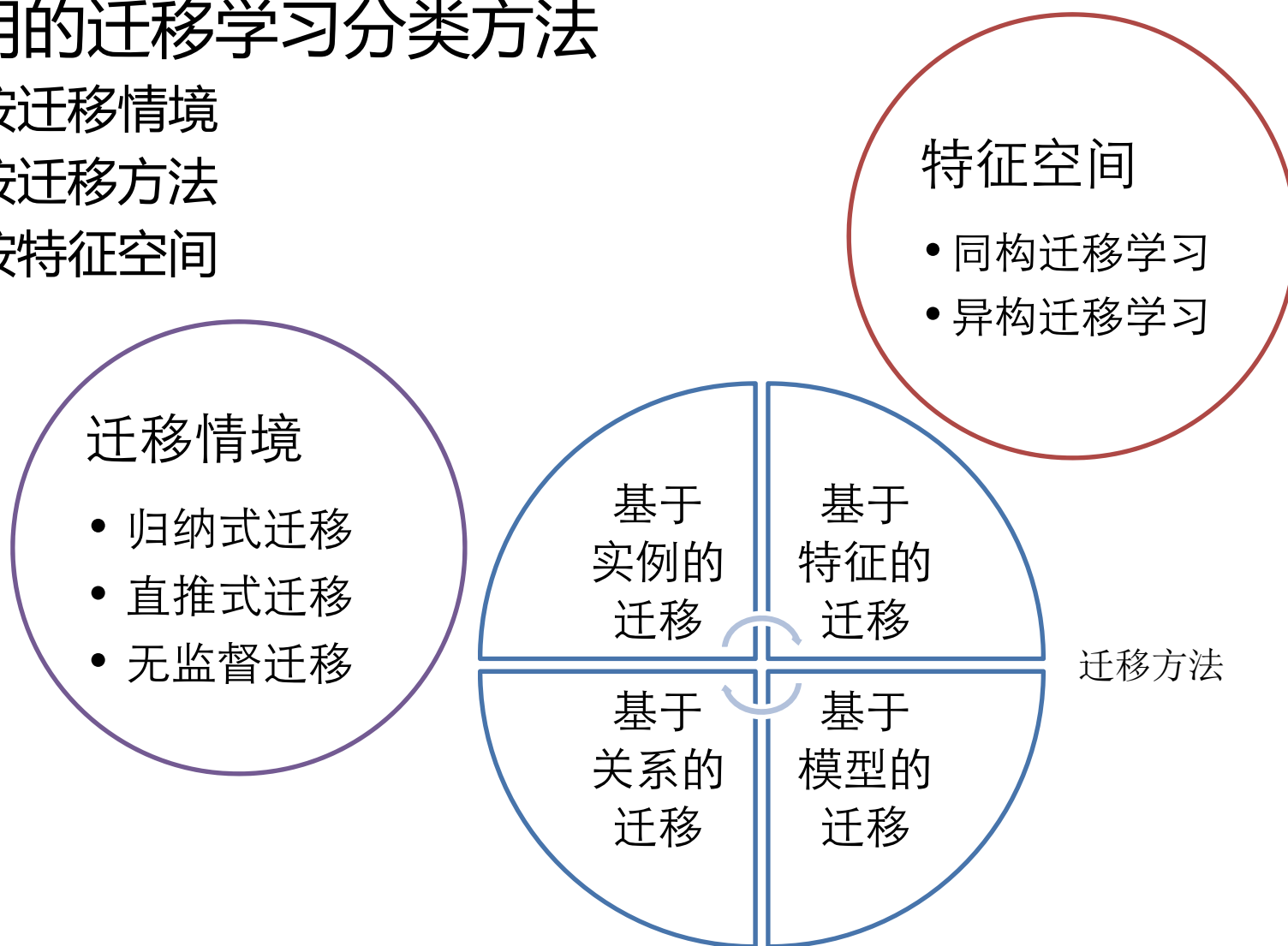
- 按迁移情境
- 按特征空间
- 按迁移方法



迁移学习的分类方法

- 常用的迁移学习分类方法

- 按迁移情境
- 按迁移方法
- 按特征空间





迁移学习的分类方法

- 按迁移情境分类：

归纳式迁移学习 (inductive transfer learning)

- 源域和目标域的学习任务不同

直推式迁移学习 (transductive transfer learning)

- 源域和目标域不同，学习任务相同

无监督迁移学习 (unsupervised transfer learning)

- 源域和目标域均没有标签

学习情境		源域和目标域	源域和目标域任务
传统机器学习		相同	相同
迁移学习	归纳式迁移/ 无监督迁移	相同	不同但相关
		不同但相关	不同但相关
	直推式迁移	不同但相关	相同



迁移学习的分类方法

- 按特征空间进行分类：

同构迁移学习

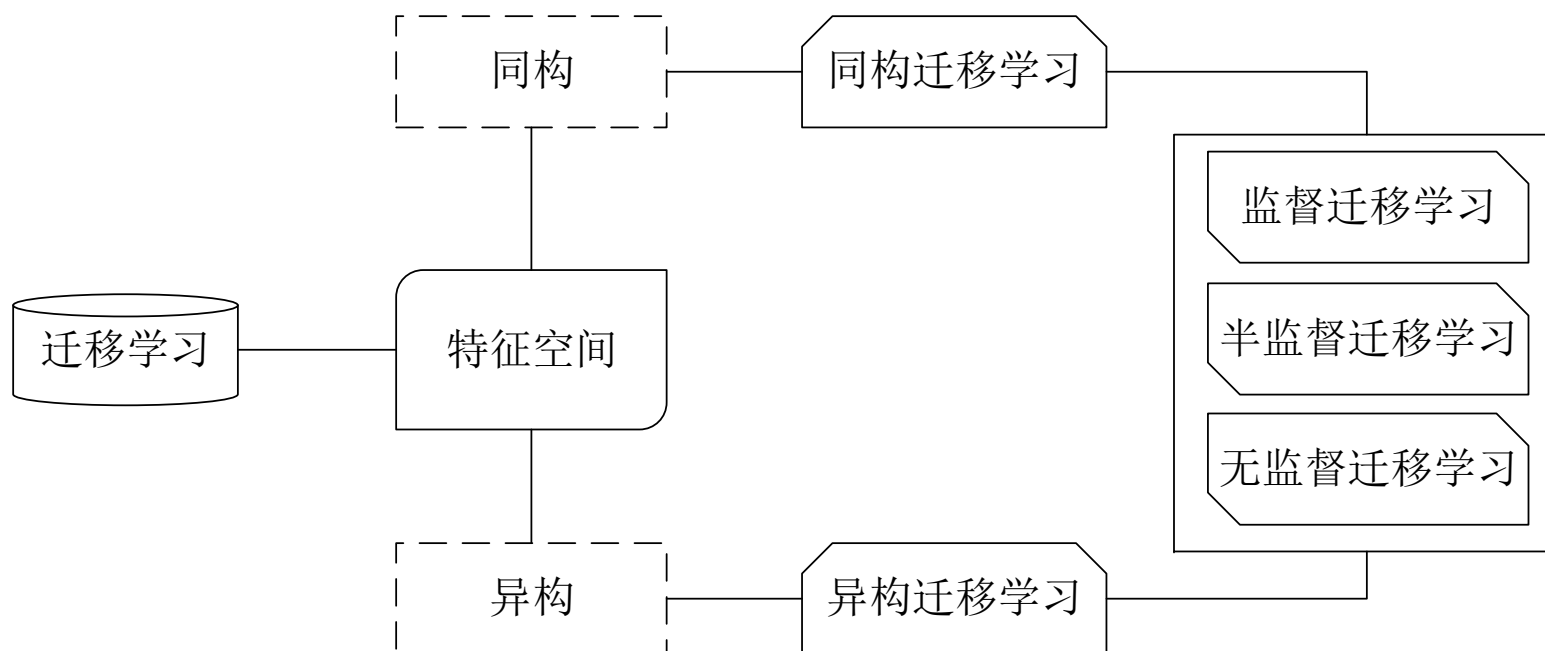
异构迁移学习

■ Homogeneous TL

■ Heterogeneous TL

□ 特征维度相同分布不同

□ 特征维度不同





迁移学习的分类方法

- 按迁移方法分类：

基于实例的迁移 (instance based TL)

- 通过权重重用源域和目标域的样例进行迁移

基于特征的迁移 (feature based TL)

- 将源域和目标域的特征变换到相同空间

基于模型的迁移 (parameter based TL)

- 利用源域和目标域的参数共享模型

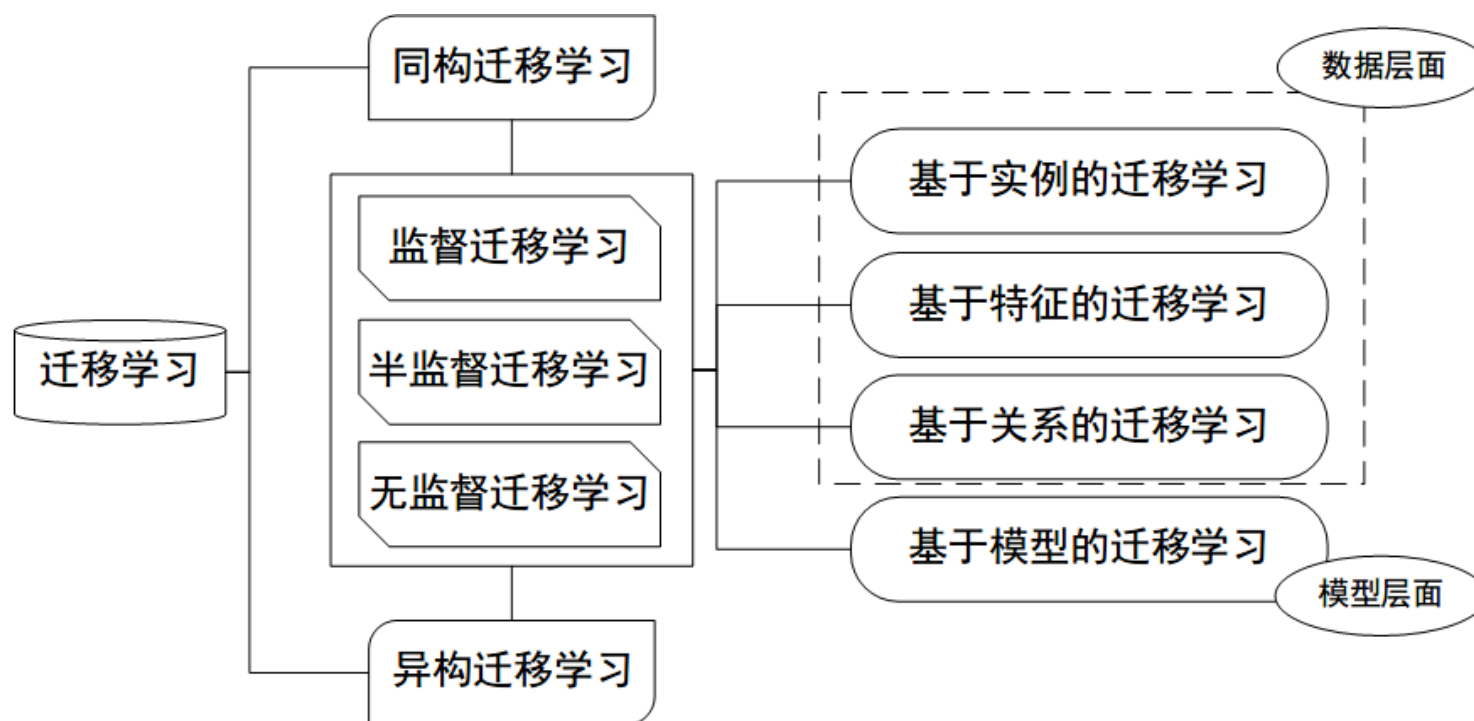
基于关系的迁移 (relation based TL)

- 利用源域中的逻辑网络关系进行迁移



迁移学习的分类方法

- 分类方法总结

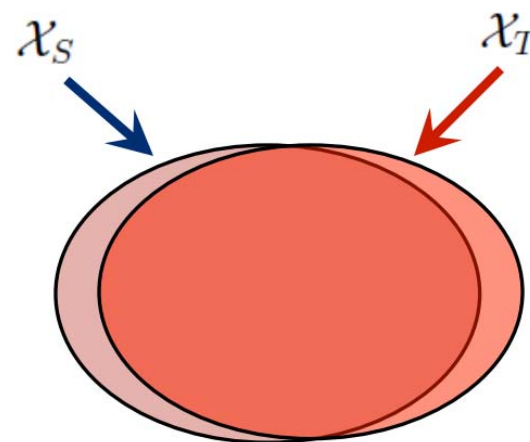




迁移学习：迁移方法

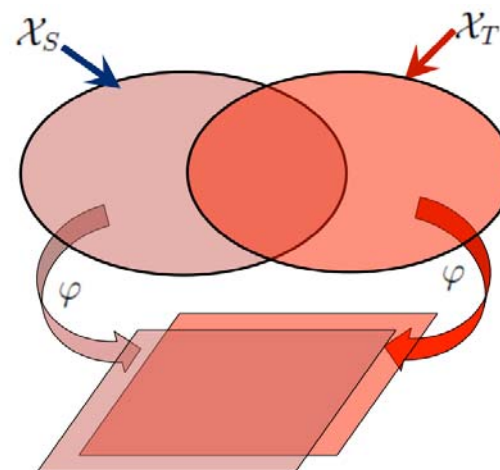
- 基于实例的迁移学习方法

- 假设：源域中的一些数据和目标域会共享很多共同的特征
- 方法：对源域进行instance reweighting，筛选出与目标域数据相似度高的数据，然后进行训练学习
- 代表工作：
 - TrAdaBoost [Dai, ICML-07]
 - Kernel Mean Matching (KMM) [Smola, ICML-08]
 - Density ratio estimation [Sugiyama, NIPS-07]
- 优点：
 - 方法较简单，实现容易
- 缺点：
 - 权重选择与相似度度量依赖经验
 - 源域和目标域的数据分布往往不同



- 基于特征的迁移学习方法

- 假设：源域和目标域仅仅有一些交叉特征
- 方法：通过特征变换，将两个域的数据变换到同一特征空间，然后进行传统的机器学习
- 代表工作：
 - Transfer component analysis (TCA) [Pan, TKDE-11]
 - Spectral Feature Alignment (SFA) [Pan, WWW-10]
 - Geodesic flow kernel (GFK) [Duan, CVPR-12]
 - Transfer kernel learning (TKL) [Long, TKDE-15]
- 优点：
 - 大多数方法采用
 - 特征选择与变换可以取得好效果
- 缺点：
 - 往往是一个优化问题，难求解
 - 容易发生过适配





迁移学习：迁移方法

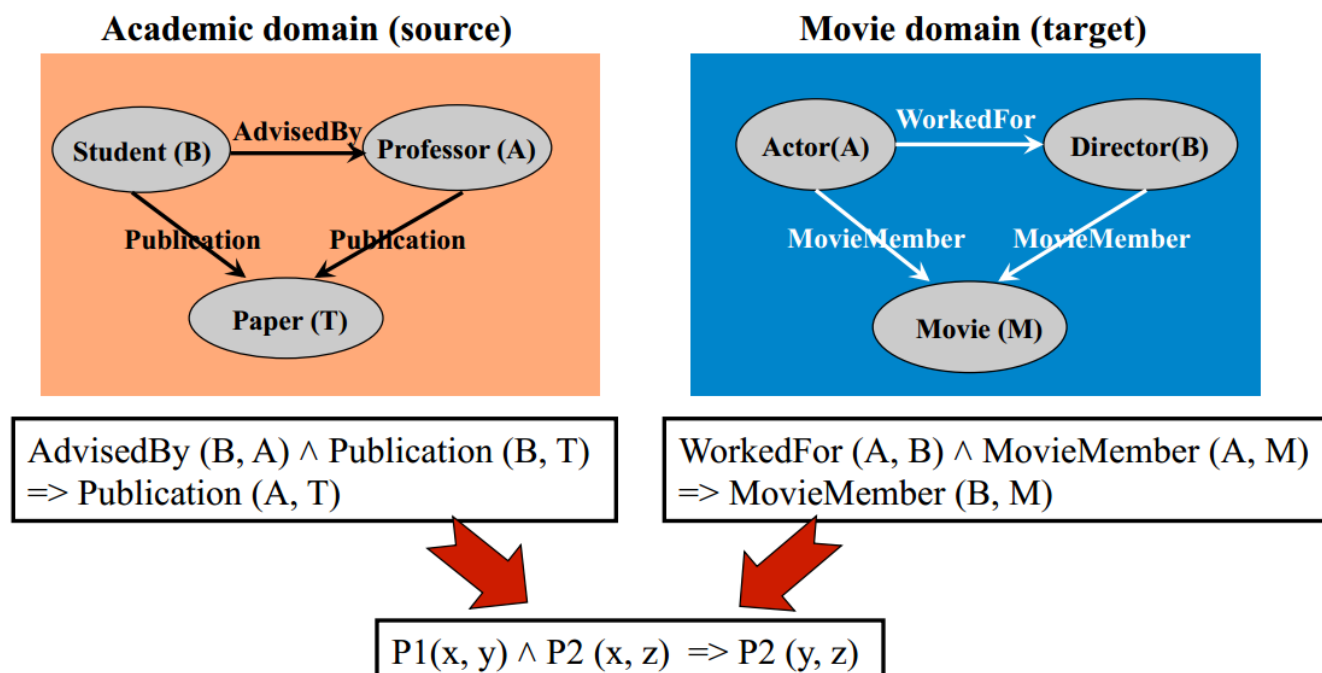
- 基于模型的迁移学习方法
 - 假设：源域和目标域可以共享一些模型参数
 - 方法：由源域学习到的模型运用到目标域上，再根据目标域学习新的模型
 - 代表工作：
 - TransEMDT [Zhao, IJCAI-11]
 - TRCNN [Oquab, CVPR-14]
 - TaskTrAdaBoost [Yao, CVPR-10]
 - 优点：
 - 模型间存在相似性，可以被利用
 - 缺点：
 - 模型参数不易收敛



迁移学习：迁移方法

- 基于关系的迁移学习方法

- 假设：如果两个域是相似的，那么它们会共享某种相似关系
- 方法：利用源域学习逻辑关系网络，再应用于目标域上
- 代表工作：
 - Predicate mapping and revising [Mihalkova, AAAI-07],
 - Second-order Markov Logic [Davis, ICML-09]





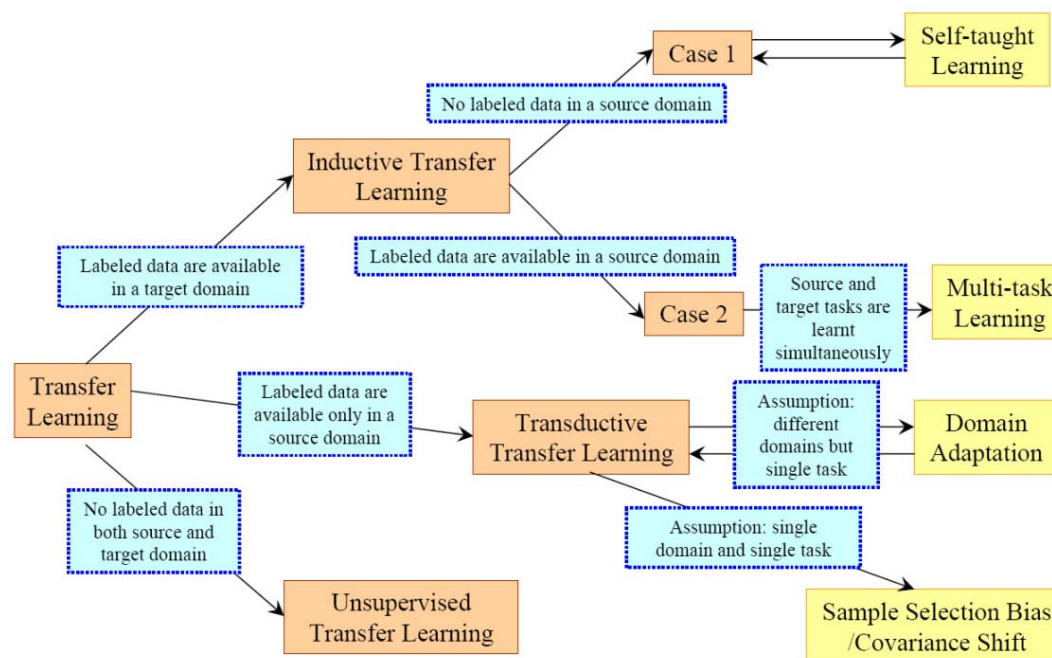
代表性研究成果

- 域适配问题
- 多源迁移学习
- 深度迁移学习



迁移学习：研究现状

- 迁移学习的热门研究领域
 - 域适配问题 (domain adaptation)
 - 多源迁移学习 (multi-source TL)
 - 深度迁移学习 (deep TL)
 - 异构迁移学习 (heterogeneous TL)





迁移学习：域适配问题

- 域适配问题：
 - domain adaptation; cross-domain learning
 - 问题定义：有标签的源域和无标签的目标域共享相同的特征和类别，但是特征分布不同，如何利用源域标定目标域
- $\mathcal{D}_S \neq \mathcal{D}_T: P_S(X) \neq P_T(X)$

Training



Testing



Caltech-256



DSLR



Amazon



Webcam





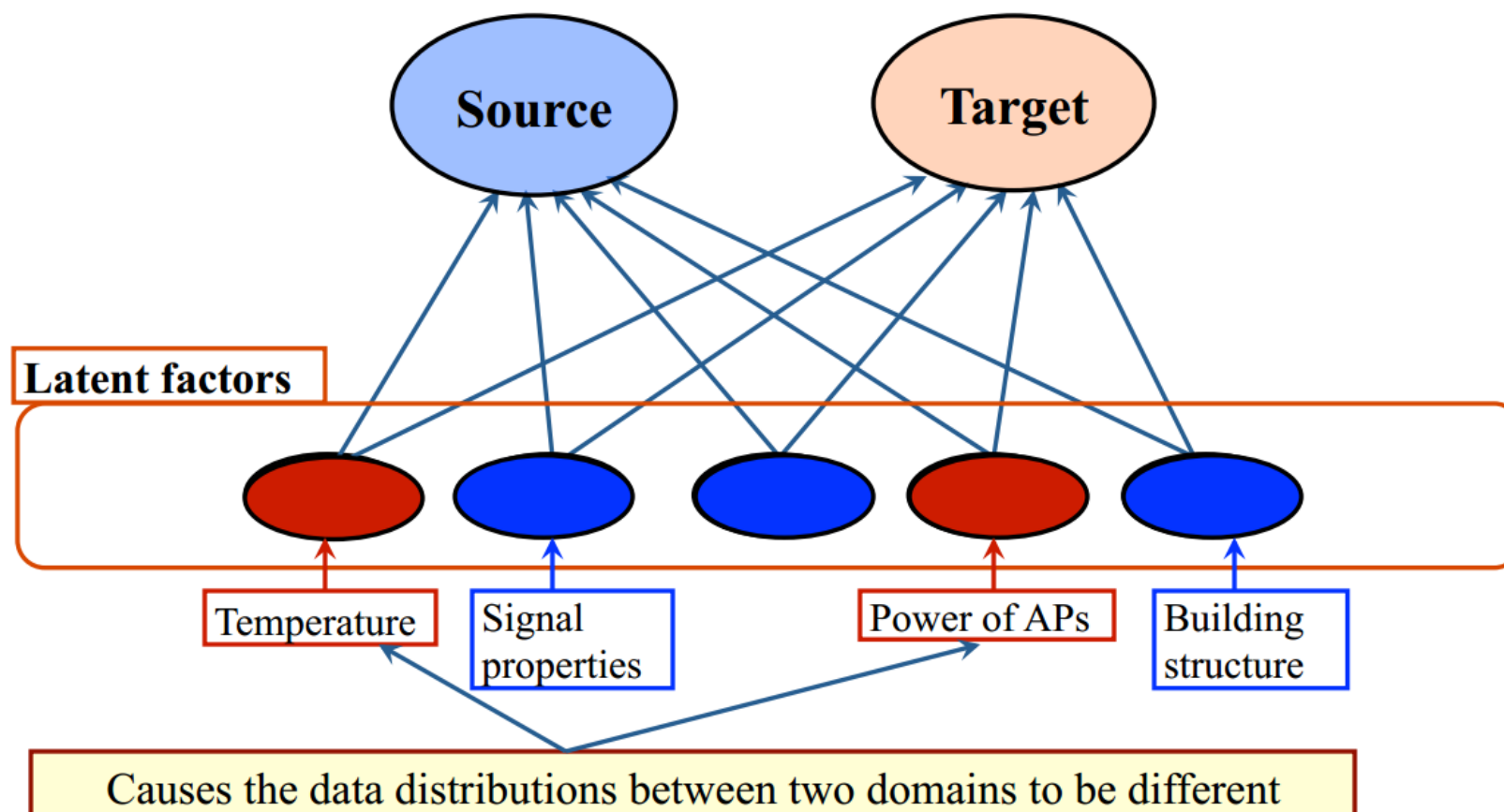
迁移学习：域适配问题

- 域适配问题：
 - 基于特征的迁移方法：
 - Transfer component analysis [Pan, TKDE-11]
 - Geodesic flow kernel [Duan, CVPR-12]
 - Transfer kernel learning [Long, TKDE-15]
 - TransEMDT [Zhao, IJCAI-11]
 - 基于实例的迁移方法：
 - Kernel mean matching [Huang, NIPS-06]
 - Covariate Shift Adaptation [Sugiyama, JMLR-07]
 - 基于模型的迁移方法：
 - Adaptive SVM (ASVM) [Yang et al, ACM Multimedia-07]
 - Multiple Convex Combination (MCC) [Schweikert, NIPS-09]
 - Domain Adaptation Machine (DAM) [Duan, TNNLS-12]



迁移学习：域适配问题

- 迁移成分分析 (TCA, transfer component analysis) [Pan, TKDE-11]
 - 将源域和目标域变换到相同空间，最小化它们的距离



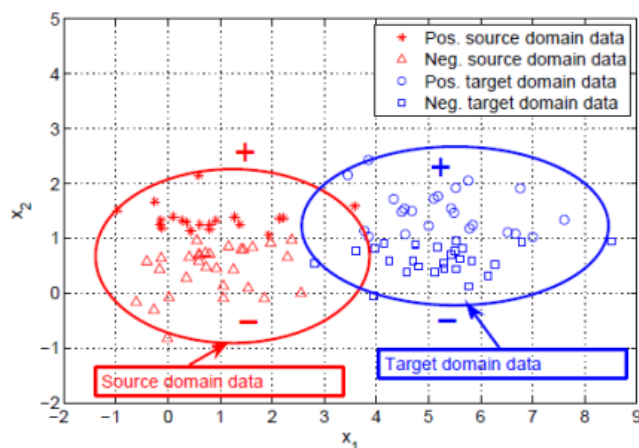


迁移学习：域适配问题

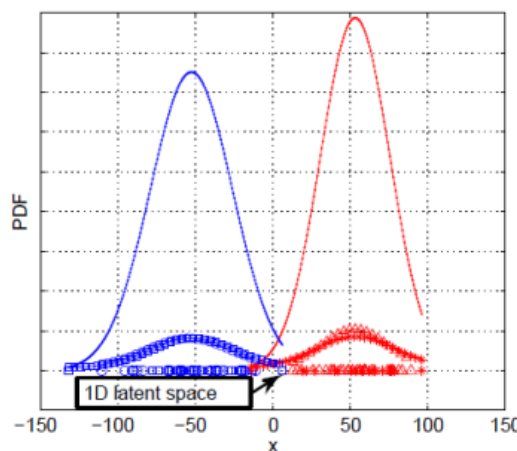
- 迁移成分分析：

- 优化目标：
$$\min_{\varphi} \text{Dist}(\varphi(\mathbf{X}_S), \varphi(\mathbf{X}_T)) + \lambda \Omega(\varphi)$$
$$\text{s.t. constraints on } \varphi(\mathbf{X}_S) \text{ and } \varphi(\mathbf{X}_T)$$
- Maximum mean discrepancy (MMD)

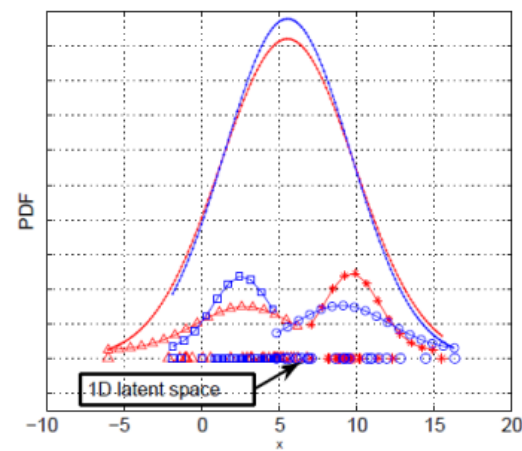
$$\text{Dist}(P(\mathbf{X}_S), P(\mathbf{X}_T)) = \left\| \frac{1}{n_S} \sum_{i=1}^{n_S} \Phi(x_{S_i}) - \frac{1}{n_T} \sum_{j=1}^{n_T} \Phi(x_{T_j}) \right\|_{\mathcal{H}}$$



Original feature space



PCA

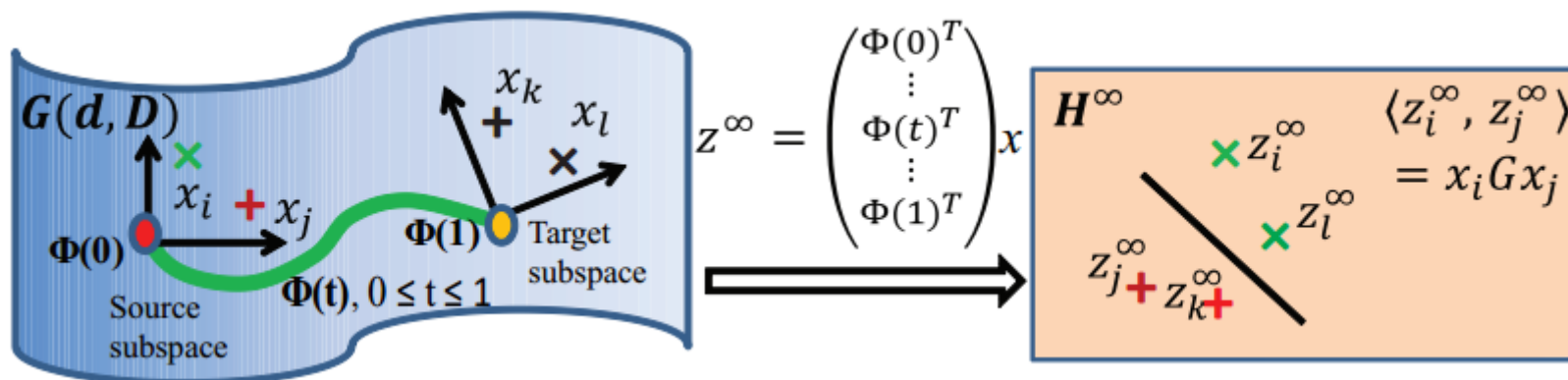


TCA



迁移学习：域适配问题

- GFK (geodesic flow kernel) [Duan, CVPR-12]
 - 利用流形学习，将数据映射到高维空间中，然后测量其距离，使得源域和目标域差异最大
 - 优化目标： $\Phi(t) = P_S U_1 \Gamma(t) - R_S U_2 \Sigma(t)$
$$P_S^T P_T = U_1 \Gamma V^T, \quad R_S^T P_T = -U_2 \Sigma V^T$$
 - 流形正则项： $\mathcal{R}(\mathcal{S}, \mathcal{T}) = \frac{1}{d^*} \sum_i^{d^*} \theta_i [KL(\mathcal{S}_i \| \mathcal{T}_i) + KL(\mathcal{T}_i \| \mathcal{S}_i)]$





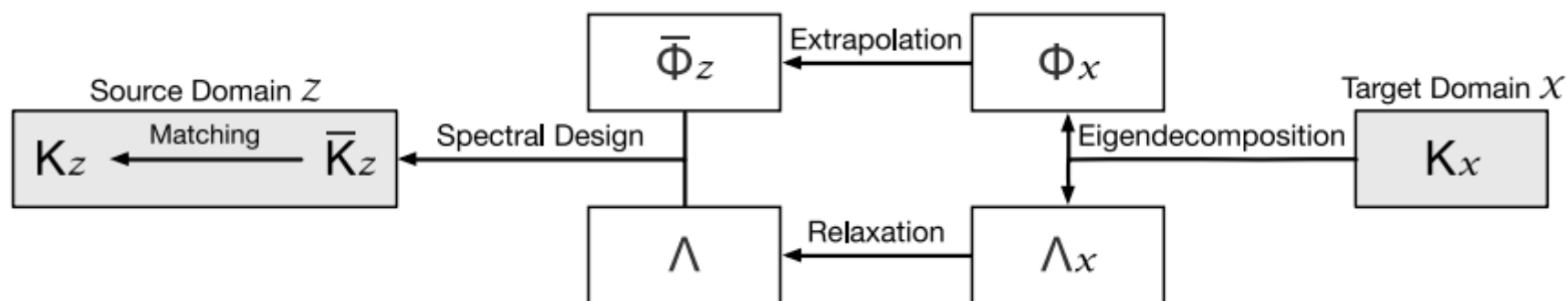
迁移学习：域适配问题

- Transfer Kernel Learning (TKL) [Long, TKDE-15]
 - 在再生核希尔伯特空间中学习一个领域不变核矩阵，从而实现源域和目标域的适配

– 优化目标：
$$\min_{\Lambda} \|\bar{\mathbf{K}}_Z - \mathbf{K}_Z\|_F^2 = \|\bar{\Phi}_Z \Lambda \bar{\Phi}_Z^T - \mathbf{K}_Z\|_F^2$$

$$\lambda_i \geq \zeta \lambda_{i+1}, i = 1, \dots, n-1$$

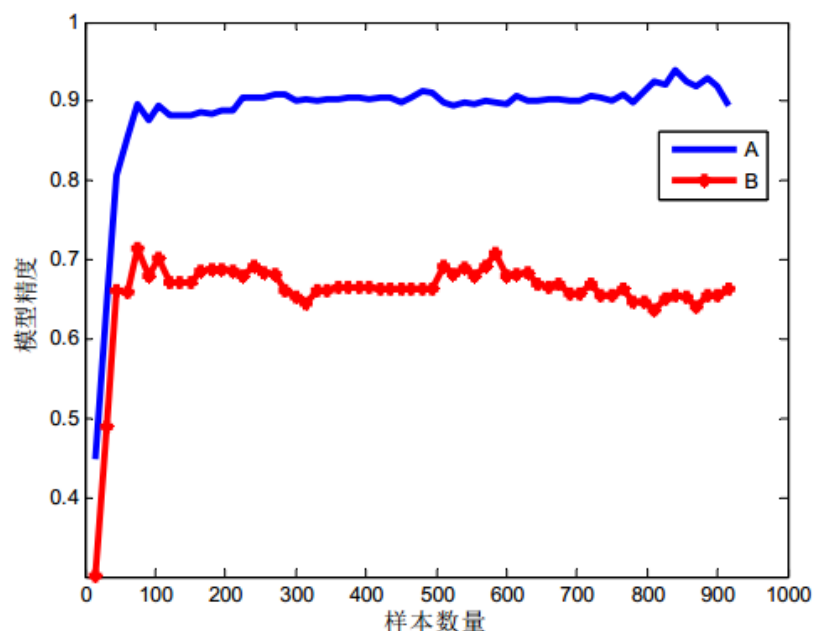
$$\lambda_i \geq 0, i = 1, \dots, n,$$



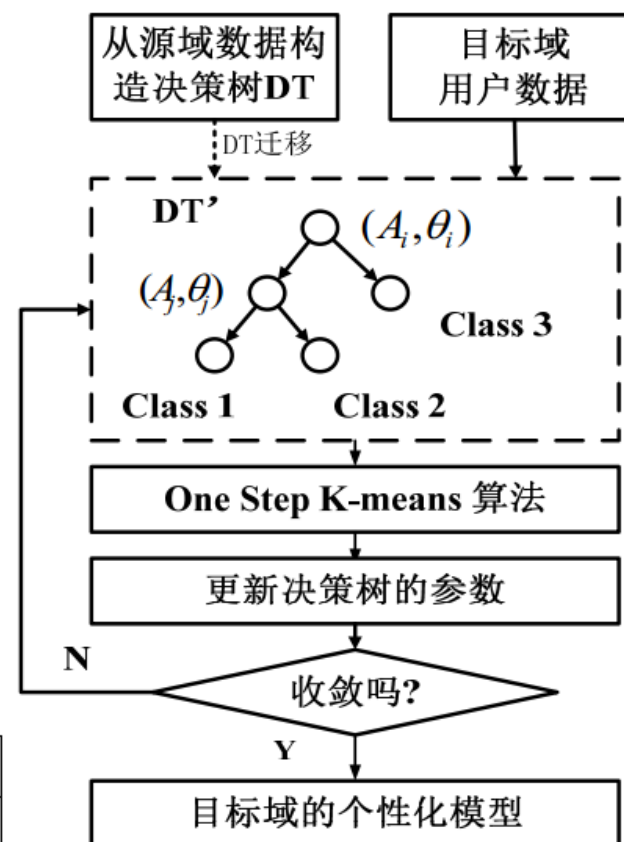


迁移学习：域适配问题

- 嵌入决策树算法 (TransEMDT) [Zhao, IJCAI-11]
 - 首先通过聚类得到初始的目标域决策树模型，然后迭代更新决策树的参数直到收敛为止



用户对	应用算法前	应用算法后
132	63.30%±4.35%	83.26%±6.27%



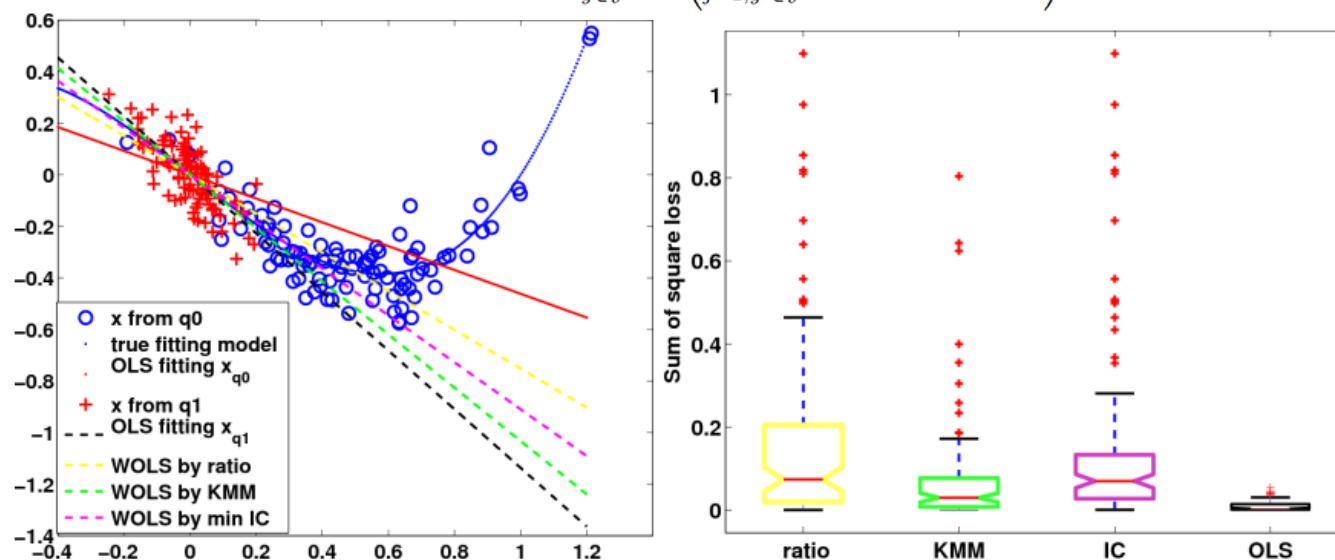


迁移学习：域适配问题

- Kernel mean matching [Huang, NIPS-06]
 - 在再生希尔伯特空间中计算源域和目标域的协方差分布差异，然后用二次规划求解样本权重

- 优化目标：
$$\underset{\alpha}{\text{minimize}} \sum_{i=1}^{n_{tr}} \beta_i g(\alpha | x_i^{tr}) - \sum_{i,j=1; y \in \mathcal{Y}}^{n_{tr}} \alpha_{iy} \beta_j k(x_i^{tr}, y, x_j^{tr}, y_j^{tr})$$
$$+ \sum_{i,j=1; y, y' \in \mathcal{Y}}^{n_{tr}} \alpha_{iy} \alpha_{jy'} k(x_i^{tr}, y, x_j^{tr}, y')$$

$$\text{where } g(\alpha | x_i^{tr}) := \log \sum_{y \in \mathcal{Y}} \exp \left(\sum_{j=1; y' \in \mathcal{Y}}^{n_{tr}} \alpha_{jy'} k(x_i^{tr}, y, x_j^{tr}, y') \right)$$



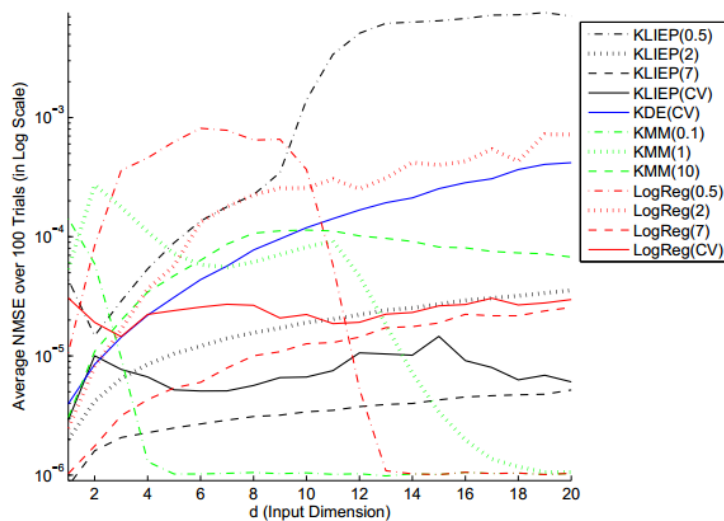


迁移学习：域适配问题

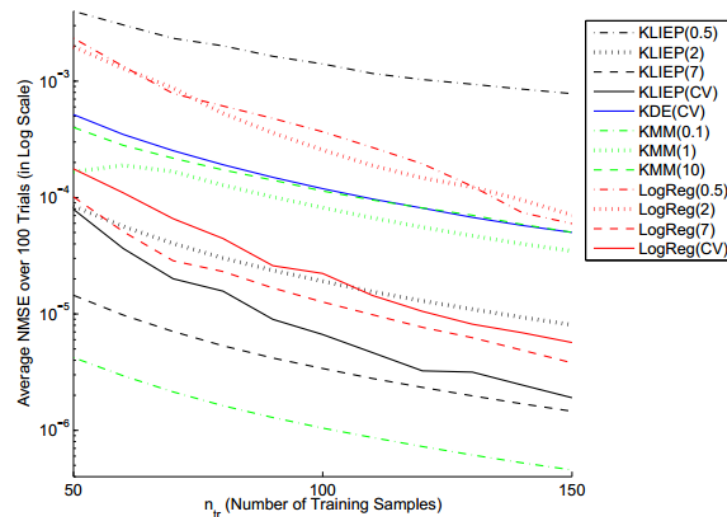
- Covariate Shift Adaptation [Sugiyama, JMLR-07]

- 采用自然估计法估计源域和目标域的密度比例，然后进行实例权重的分配，最后迁移

- 优化目标：
$$\underset{\{\alpha_\ell\}_{\ell=1}^b}{\text{maximize}} \left[\sum_{j=1}^{n_{te}} \log \left(\sum_{\ell=1}^b \alpha_\ell \varphi_\ell(\mathbf{x}_j^{te}) \right) \right]$$
$$\text{subject to } \sum_{i=1}^{n_{tr}} \sum_{\ell=1}^b \alpha_\ell \varphi_\ell(\mathbf{x}_i^{tr}) = n_{tr} \text{ and } \alpha_1, \alpha_2, \dots, \alpha_b \geq 0.$$



(a) When input dimension is changed



(b) When training sample size is changed



迁移学习：域适配问题

- Adaptive SVM (ASVM) [Yang et al, ACM Multimedia-07]
 - 使用SVM模型，在适配和原始模型之间学习“数据函数”，达到模型迁移效果
 - 优化目标：
$$\min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i$$
$$\text{s.t. } \xi_i \geq 0, \quad y_i \sum_{k=1}^M t_k f_k^a(\mathbf{x}_i) + y_i \mathbf{w}^T \phi(\mathbf{x}_i) \geq 1 - \xi_i$$
- Multiple Convex Combination (MCC) [Schweikert, NIPS-09]
 - 对一些域适配的方法做集成学习
 - 优化目标：
$$F(\mathbf{x}) = \alpha f_T(\mathbf{x}) + (1 - \alpha) \frac{1}{|\mathcal{S}|} \sum_{S \in \mathcal{S}} f_S(\mathbf{x})$$



迁移学习：域适配问题

- 总结

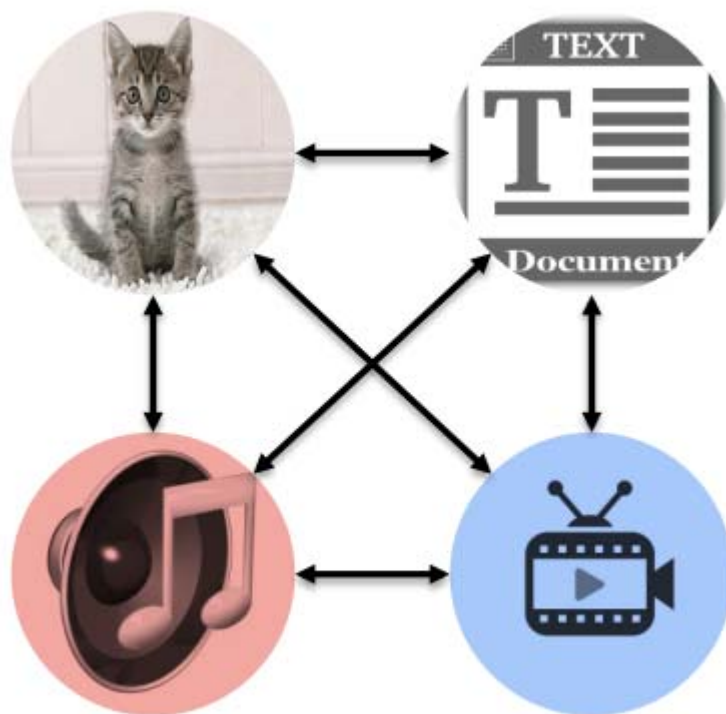
- 通常假设源域和目标域的数据有着相同的条件分布，或者在高维空间里，有着相同的条件分布
- 这个假设是有一定局限性的，无法衡量源域和目标域之间相似性，可能发生负迁移



迁移学习：多源迁移学习

- 多源迁移学习

- 问题定义：多个源域和目标域，如何进行有效的域筛选，从而进行迁移？





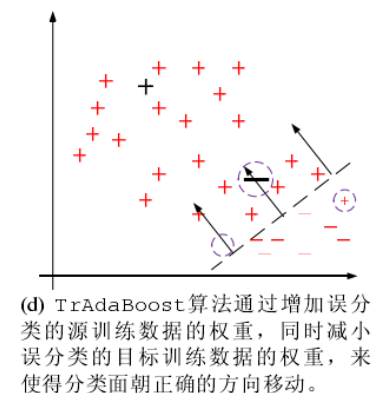
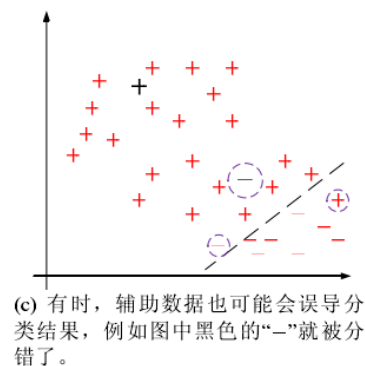
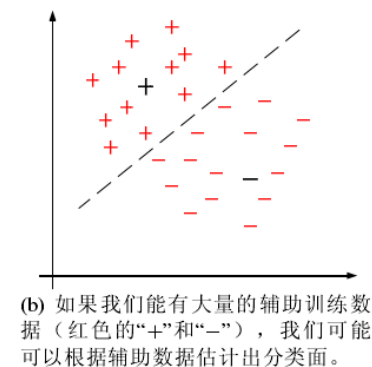
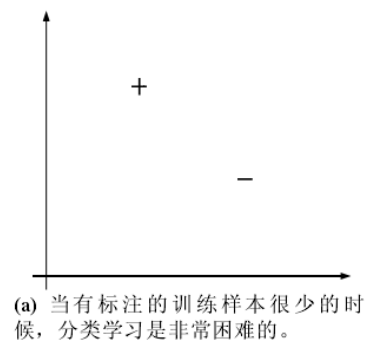
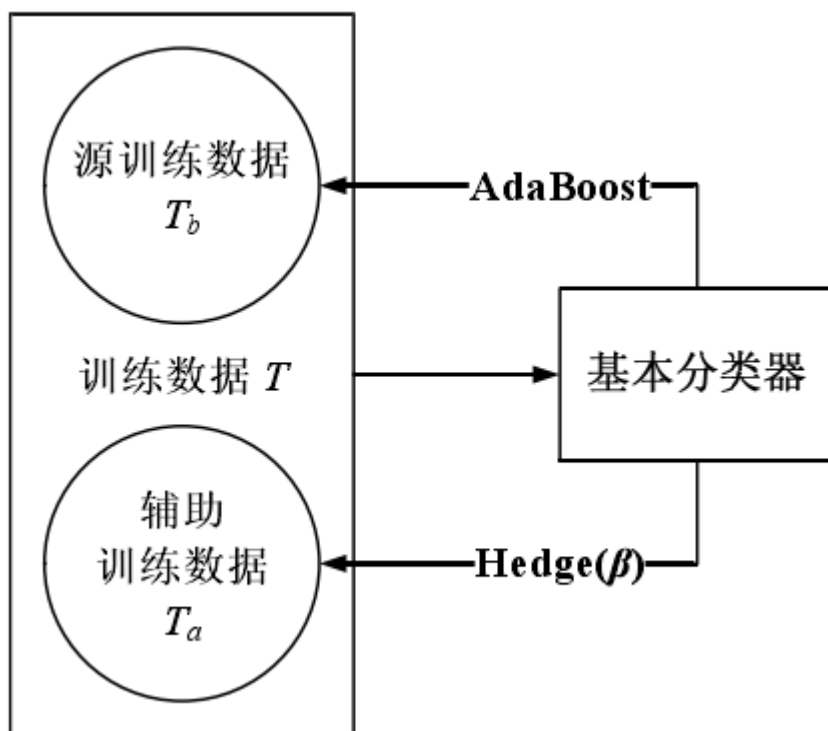
迁移学习：多源迁移学习

- 多源迁移学习
 - TrAdaBoost [Dai, ICML-07]
 - MsTL-MvAdaboost [Xu, ICONIP-12]
 - Consensus regularization [Luo, CIKM-08]
 - Transitive transfer learning [Tan, KDD-15]
 - Distant domain TL [Tan, AAAI-17]



迁移学习：多源迁移学习

- TrAdaBoost [Dai, ICML-07]
 - 利用Boost的技术过滤掉多个源域中与目标域不相似的样本，然后进行实例迁移学习





迁移学习：多源迁移学习

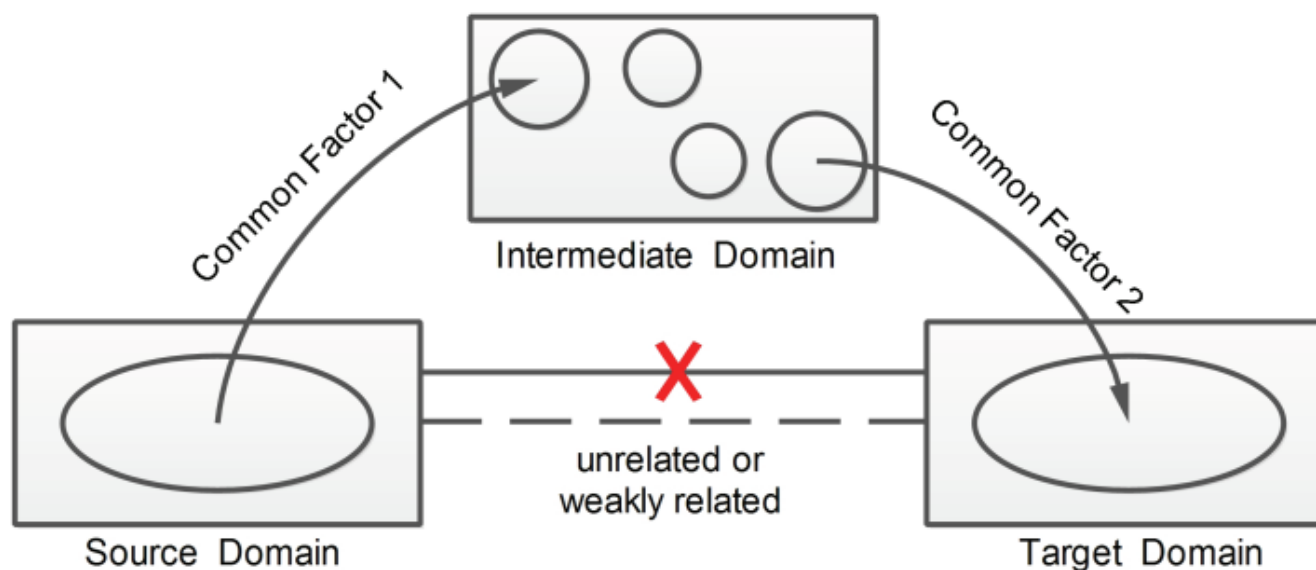
- MsTL-MvAdaboost [Xu, ICONIP-12]
 - 不仅考虑源域和目标域的样本相似度情况，同时，以多视图学习的目标来进行统一的迁移
- Consensus regularization [Luo, CIKM-08]
 - 同时在源域和伪标注的目标域上训练分类器，利用一致性约束进行知识的迁移

$$\begin{aligned}\max P(h^l | \mathcal{D}_s^l) &= \max \frac{P(\mathcal{D}_s^l | h^l) P(h^l)}{P(\mathcal{D}_s^l)} = \max P(\mathcal{D}_s^l | h^l) P(h^l) \\ &= \max P(h^l) \cdot \prod_{i=1}^{n^l} P(y_i^l | \mathbf{x}_i; h^l) = \max(\log P(h^l) + \sum_{i=1}^{n^l} \log P(y_i^l | \mathbf{x}_i; h^l)).\end{aligned}$$



迁移学习：多源迁移学习

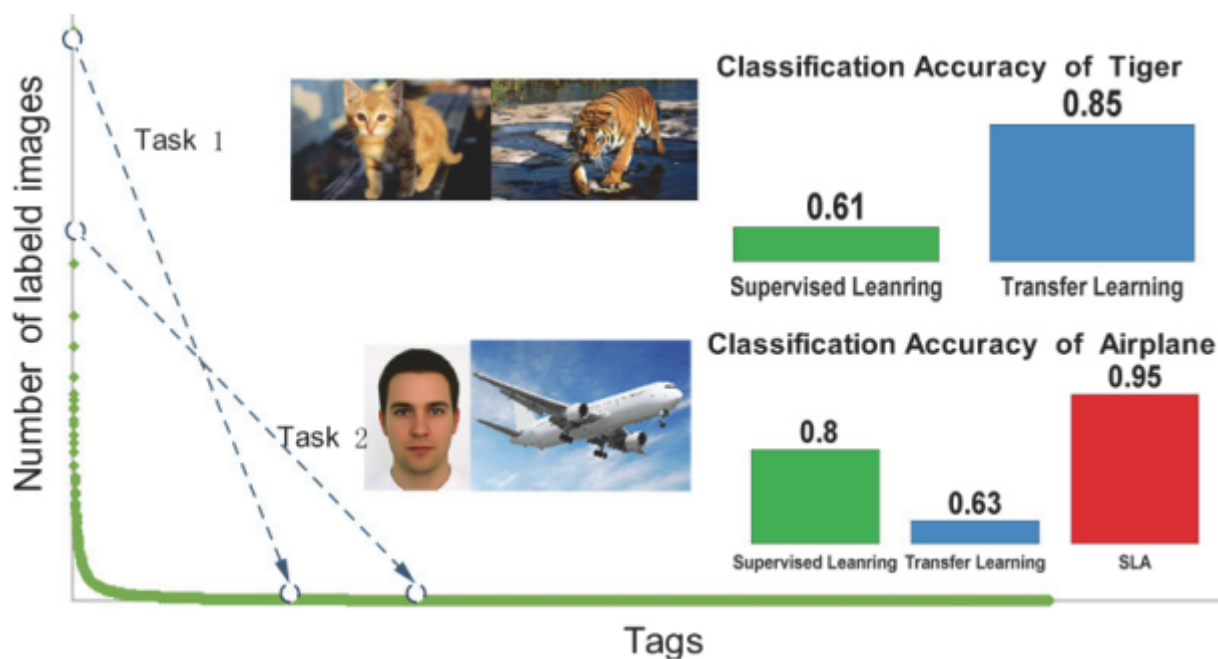
- Transitive transfer learning [Tan, KDD-15]
 - 在两个相似度不高的域中，利用从第三方中学习到的相似度关系，完成知识的传递迁移





迁移学习：多源迁移学习

- Distant domain TL [Tan, AAAI-17]
 - 在相似度极低的两个域进行迁移时，用autoencoder自动从多个中间辅助域中选择知识，完成迁移





迁移学习：多源迁移学习

- 总结：
 - 多源迁移学习可以有效利用存在的多个可用域，综合起来进行迁移，达到较好的效果
 - 如何衡量多个域之间的相关性还是一个问题
 - 对多个域的利用方法也存在一定挑战性



迁移学习：深度迁移学习

- 深度迁移学习
 - 利用深度神经网络的结构进行迁移学习
 - 神经网络学习非线性的特征表示
 - 层次性
 - 使得数据具有不可解释性
 - 表明在数据中具有某些不可变的成分，可以用来迁移
- 代表方法
 - Joint CNN [Tzeng, ICCV-15]
 - SHL-MDNN [Huang, ICASSP-13]
 - Deep Adaptation Network (DAN) [Long, ICML-15]
 - Joint Adaptation Networks [Long, CVPR-13]

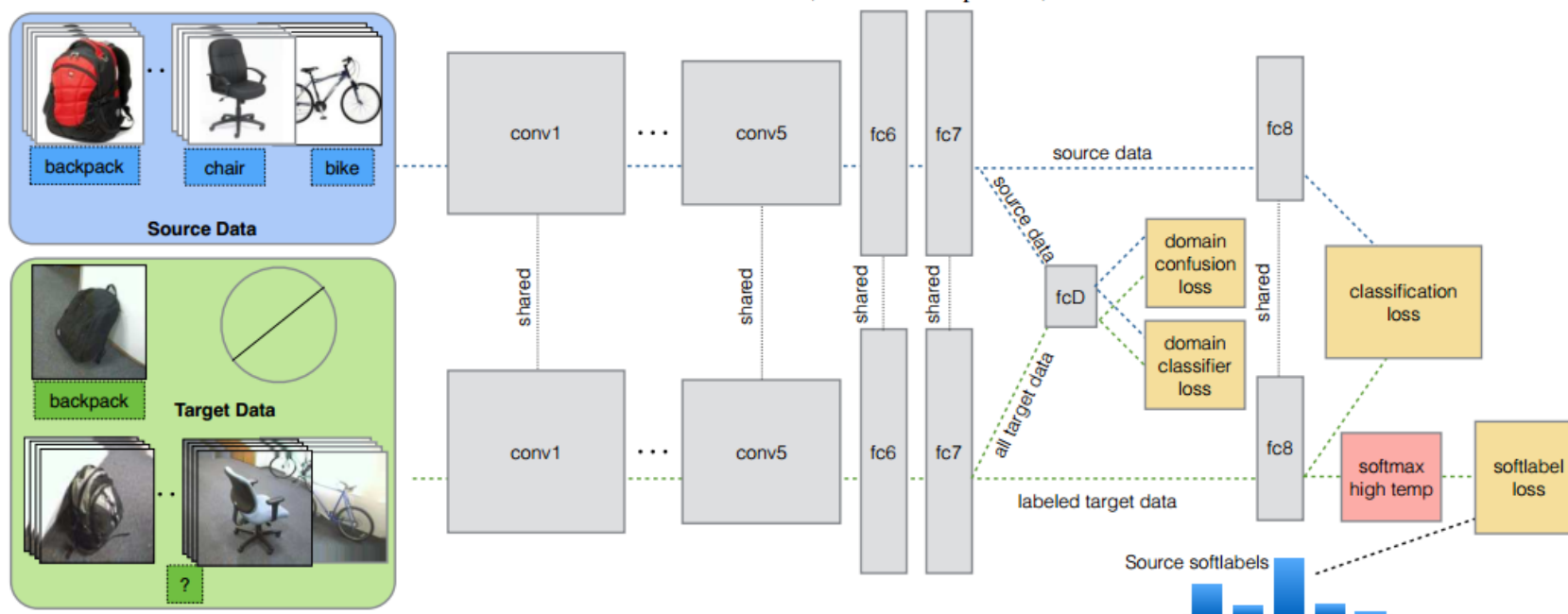
- Joint CNN [Tzeng, ICCV-15]
 - 针对有稀疏标记的目标域数据，用CNN同时优化域之间的距离和迁移学习任务的损失

$$\mathcal{L}(x_S, y_S, x_T, y_T, \theta_D; \theta_{\text{repr}}, \theta_C) =$$

$$\mathcal{L}_C(x_S, y_S, x_T, y_T; \theta_{\text{repr}}, \theta_C)$$

$$+ \lambda \mathcal{L}_{\text{conf}}(x_S, x_T, \theta_D; \theta_{\text{repr}})$$

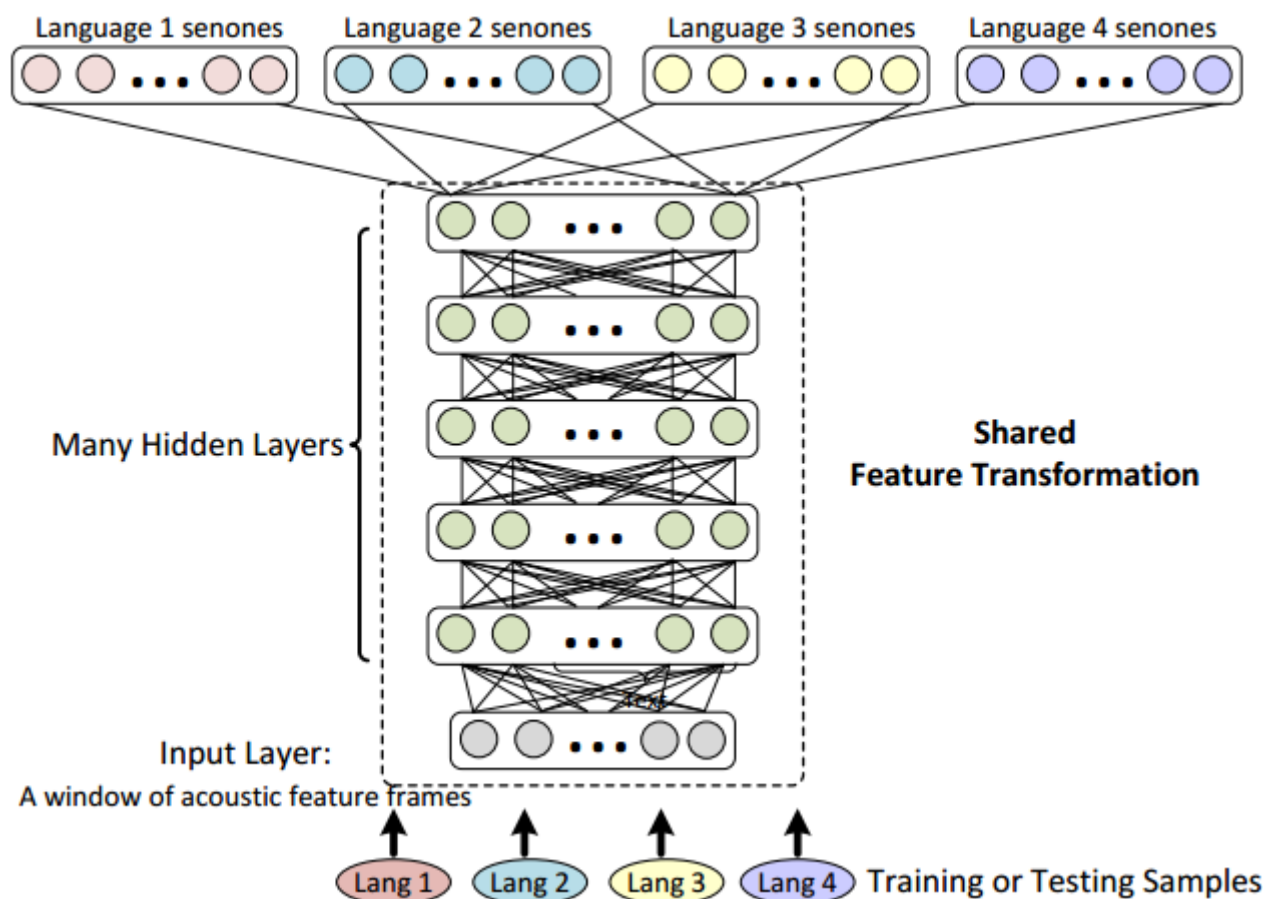
$$+ \nu \mathcal{L}_{\text{soft}}(x_T, y_T; \theta_{\text{repr}}, \theta_C).$$





迁移学习：深度迁移学习

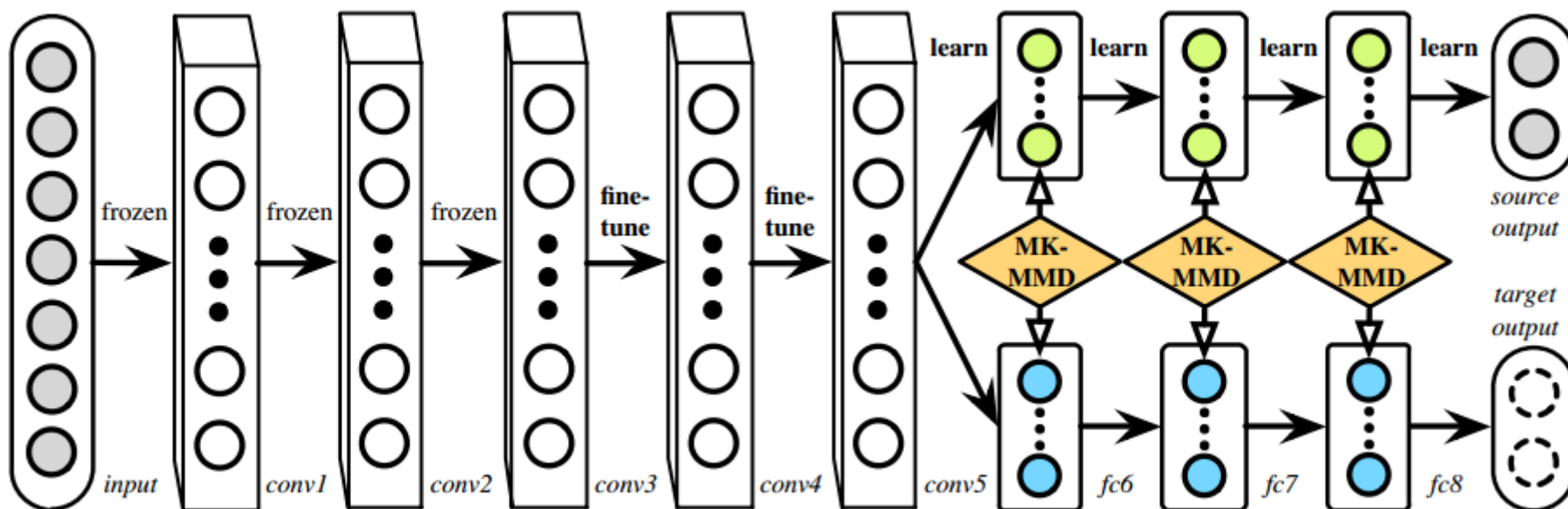
- SHL-MDNN [Huang, ICASSP-13]
 - 在不同的学习网络之间共享隐藏层，通过不同的softmax层控制学习任务的不同



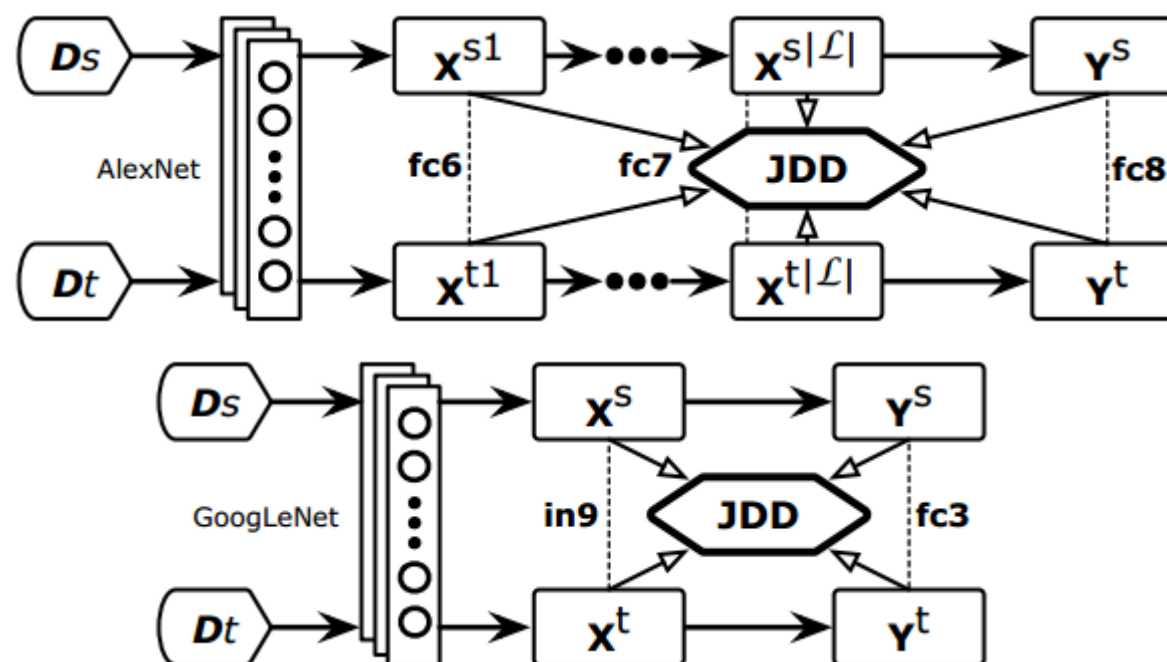


迁移学习：深度迁移学习

- Deep Adaptation Network (DAN) [Long, ICML-15]
 - 将CNN中与学习任务相关的隐藏层映射到再生核希尔伯特空间中，通过多核优化的方法最小化不同域之间的距离



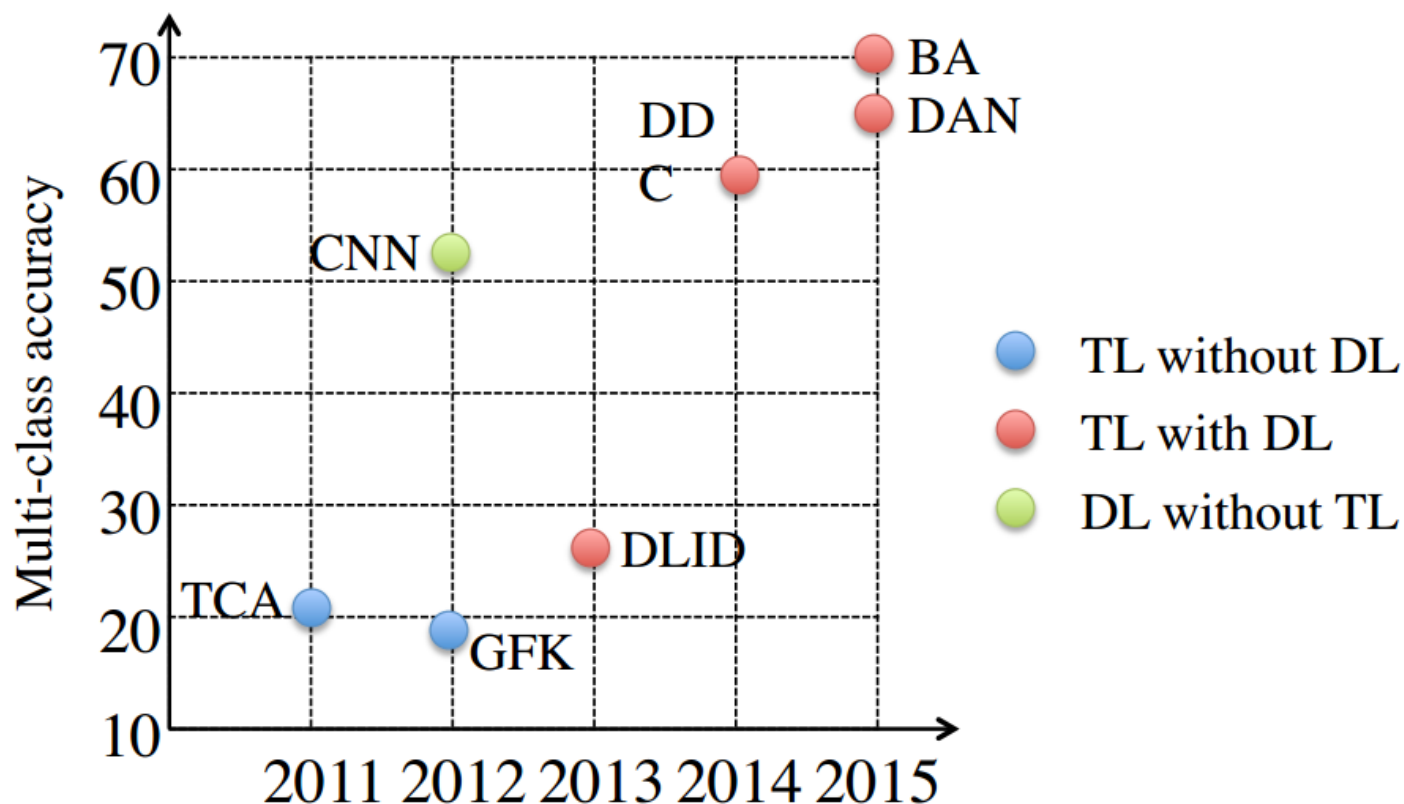
- Joint Adaptation Networks [Long, CVPR-15]
 - 提出一种新的联合分布距离度量关系，利用这种关系泛化深度模型的迁移学习能力，从而适配不同领域的数据分布。基于AlexNet和GoogLeNet重新优化了网络结构





迁移学习：深度迁移学习

- 总结：
 - 迁移学习大增强了模型的泛化能力
 - 深度学习可以深度表征域中的知识结构
 - 深度学习+迁移学习还可大有作为





问题与展望

- 迁移学习存在的问题：
 - 负迁移：无法判断域之间的相关性，导致负迁移
 - 缺乏理论支撑：尚未有统一的迁移学习理论
 - 相似度衡量：域之间的相似度通常依赖经验进行衡量，缺乏统一有效的相似度衡量方法
- 已有的基础
 - 负迁移：利用自编码器实现相关度较低的两个域之间的迁移（人脸→飞机）[Tan, AAAI-2017]
 - 理论支撑：利用物理学定律为迁移找到理论保证[Stewart, AAAI-17]
 - 相似度衡量：提出迁移度量学习，寻找行为之间相关性最高的域进行迁移[Al-Halah, ICPR-14]



迁移学习资源

- 综述
 - A survey on transfer learning [Pan and Yang, TKDE-10]
 - A survey of transfer learning [Weiss, Big data-15]
- 开源项目
 - <http://www.cse.ust.hk/TL/>
- 研究学者
 - Qiang Yang @ HKUST: <http://www.cs.ust.hk/~qyang/>
 - Sinno Jialin Pan @ NTU: <http://www.ntu.edu.sg/home/sinnopan/>
 - Fuzhen Zhuang @ ICT CAS: <http://www.intsci.ac.cn/users/zhuangfuzhen/>
 - Mingsheng Long @ THU: <http://ise.thss.tsinghua.edu.cn/~mlong/>
 - Lixin Duan @ Amazon: <http://www.lxduan.info/>
- 会议、期刊
 - 人工智能、机器学习：AAAI，ICML，ICJAI，NIPS，TNNLS，TIST，CVPR
 - 数据挖掘：TKDE，SIGKDD，ACL，WWW，SIGIR



- [DARPA] <http://www.darpa.mil/ipto/programs/tl/tl.asp>
- [Pan, TKDE-10] Pan S J, Yang Q. A survey on transfer learning[J]. IEEE Transactions on knowledge and data engineering, 2010, 22(10): 1345-1359.
- [Dai, ICML-07] Dai W, Yang Q, Xue G R, et al. Boosting for transfer learning[C]//Proceedings of the 24th international conference on Machine learning. ACM, 2007: 193-200.
- [Smola, ICML-08] Gretton A, Smola A, Huang J, et al. Covariate shift by kernel mean matching[J]. Dataset shift in machine learning, 2009, 3(4): 5.
- [Sugiyama, NIPS-07] Sugiyama M, Suzuki T, Kanamori T. Density ratio estimation in machine learning[M]. Cambridge University Press, 2012.
- [Pan, TKDE-11] Pan S J, Tsang I W, Kwok J T, et al. Domain adaptation via transfer component analysis[J]. IEEE Transactions on Neural Networks, 2011, 22(2): 199-210.
- [Pan, WWW-10] Pan S J, Ni X, Sun J T, et al. Cross-domain sentiment classification via spectral feature alignment[C]//Proceedings of the 19th international conference on World wide web. ACM, 2010: 751-760.
- [Duan, CVPR-12] Gong B, Shi Y, Sha F, et al. Geodesic flow kernel for unsupervised domain adaptation[C]//Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012: 2066-2073.
- [Long, TKDE-15] Long M, Wang J, Sun J, et al. Domain invariant transfer kernel learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2015, 27(6): 1519-1532.
- [Zhao, IJCAI-11] Zhao Z, Chen Y, Liu J, et al. Cross-people mobile-phone based activity recognition[C]//IJCAI. 2011, 2011(2011): 2545-2550.
- [Yao, CVPR-10] Yao Y, Doretto G. Boosting for transfer learning with multiple sources[C]//Computer vision and pattern recognition (CVPR), 2010 IEEE conference on. IEEE, 2010: 1855-1862.
- [Davis, ICML-09] Davis J, Domingos P. Deep transfer via second-order markov logic[C]//Proceedings of the 26th annual international conference on machine learning. ACM, 2009: 217-224.
- [Mihakova, AAAI-07] Mihalkova L, Huynh T, Mooney R J. Mapping and revising Markov logic networks for transfer learning[C]//AAAI. 2007, 7: 608-614.
- [Yang, ACM MM-07] Yang J, Yan R, Hauptmann A G. Cross-domain video concept detection using adaptive svms[C]//Proceedings of the 15th ACM international conference on Multimedia. ACM, 2007: 188-197.



- [Schweikert, NIPS-09] Schweikert G, Rätsch G, Widmer C, et al. An empirical analysis of domain adaptation algorithms for genomic sequence analysis[C]//Advances in Neural Information Processing Systems. 2009: 1433-1440.
- [Duan, TNNLS-12] Duan L, Tsang I W, Xu D, et al. Domain adaptation from multiple sources via auxiliary classifiers[C]//Proceedings of the 26th Annual International Conference on Machine Learning. ACM, 2009: 289-296.
- [Xu, ICONIP-12] Xu Z, Sun S. Multi-source transfer learning with multi-view adaboost[C]//International Conference on Neural Information Processing. Springer Berlin Heidelberg, 2012: 332-339.
- [Luo, CIKM-08] Luo P, Zhuang F, Xiong H, et al. Transfer learning from multiple source domains via consensus regularization[C]//Proceedings of the 17th ACM conference on Information and knowledge management. ACM, 2008: 103-112.
- [Tan, KDD-15] Tan B, Song Y, Zhong E, et al. Transitive transfer learning[C]//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015: 1155-1164.
- [Tan, AAI-17] Tan B, Qiang Y, et al. Distant domain transfer learning. AAAI 2017.
- [Tzeng, ICCV-15] Tzeng E, Hoffman J, Darrell T, et al. Simultaneous deep transfer across domains and tasks[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 4068-4076.
- [Huang, ICASSP-13] Huang J T, Li J, Yu D, et al. Cross-language knowledge transfer using multilingual deep neural network with shared hidden layers[C]//Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. IEEE, 2013: 7304-7308.
- [Long, ICML-15] Long M, Cao Y, Wang J, et al. Learning Transferable Features with Deep Adaptation Networks[C]//ICML. 2015: 97-105.
- [Long, CVPR-13] Long M, Wang J, Ding G, et al. Transfer feature learning with joint distribution adaptation[C]//Proceedings of the IEEE International Conference on Computer Vision. 2013: 2200-2207.
- [Stewart, AAAI-17] Stewart R, Ermon S. Label-free supervision of neural networks with physics and domain knowledge[C]//AAAI 2017.
- [Al-Halah, ICPR-14] Al-Halah Z, Rybok L, Stiefelbogen R. What to transfer? High-level semantics in transfer metric learning for action similarity[C]//Pattern Recognition (ICPR), 2014 22nd International Conference on. IEEE, 2014: 2775-2780.
- [Oquab, CVPR-14] Oquab, Bottou, Laptev, Sivic: Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks. CVPR 2014
- [Weiss, Bigdata-15] Weiss K, Khoshgoftaar T M, Wang D D. A survey of transfer learning[J]. Journal of Big Data, 2016, 3(1): 1-40.



谢谢！ 敬请批评指正！

