

极视角学术分享

迁移学习中的领域自适应方法

王晋东

中国科学院计算技术研究所

2017年12月14日

个人情况简介

- 中国科学院计算技术研究所 2014级直博生
- 主要研究迁移学习及其应用
- 在国际权威会议ICDM、UbiComp、PerCom等发表若干文章
- 知乎ID：王晋东不在家，乐于在知乎上分享相关知识
- 微博：@秦汉日记
- jindongwang@outlook.com
- 个人主页： <http://jd92.wang>
- **不是大牛，仅为分享**

目录

CONTENTS

- 1 迁移学习简介
- 2 领域自适应问题
- 3 领域自适应方法
- 4 最新研究成果
- 5 参考资料

1 迁移学习的背景

■ 智能大数据时代

- 数据量，以及数据类型不断增加
- 对机器学习模型的要求：快速构建和强泛化能力
- 虽然数据量多，但是大部分数据往往**没有标注**
- 收集标注数据，或者从头开始构建每一个模型，**代价高昂且费时**



文本



图片及视频



音频



行为

- 对已有标签的数据和模型进行**重用**成为了可能
 - 传统机器学习方法通常假定这些数据服从**相同分布**，不再适用

如何基于已有的不同分布数据，快速构建模型，实现数据标定，是一个重要问题

1 迁移学习简介

迁移学习是机器学习领域用于解决**标记数据难获取**这一基础问题的重要手段

■ 迁移学习

- 通过减小源域(辅助领域)到目标域的**分布差异**，进行**知识迁移**，从而实现数据标定。



■ 核心思想

- 找到不同任务之间的**相关性**
- “举一反三”、“照猫画虎”，但不要“东施效颦”（负迁移）



1 迁移学习应用场景

■ 应用前景广阔

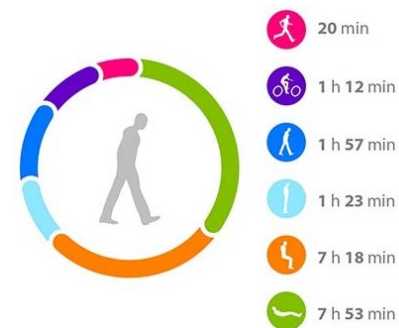
- 模式识别、计算机视觉、语音识别、自然语言处理、数据挖掘...



语料匮乏条件下不同语言的相互翻译学习



不同视角、不同背景、不同光照 的图像识别



不同用户、不同设备、不同位置
的行为识别



不同领域、不同背景下的
文本翻译、舆情分析



不同用户、不同接口、不同情境 的人机交互



不同场景、不同设备、不同时间的 室内定位

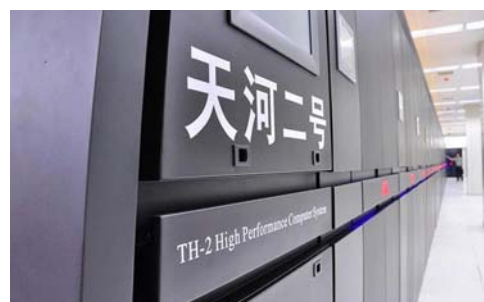
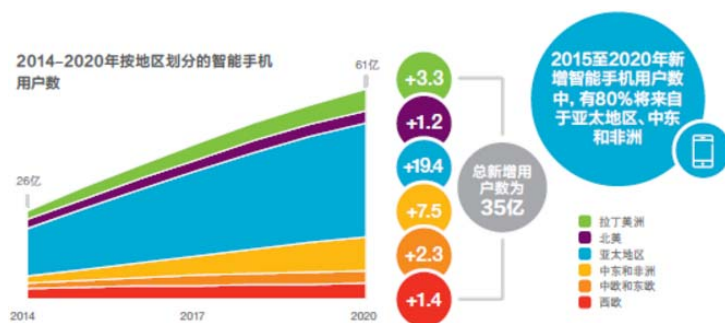
1 迁移学习简介：为什么需要迁移学习

- **数据为王，计算是核心**
 - 数据爆炸的时代！
 - 计算机更强大了！
- **但是**
 - 大数据、大计算能力只是有钱人的游戏

Google amazon



Microsoft



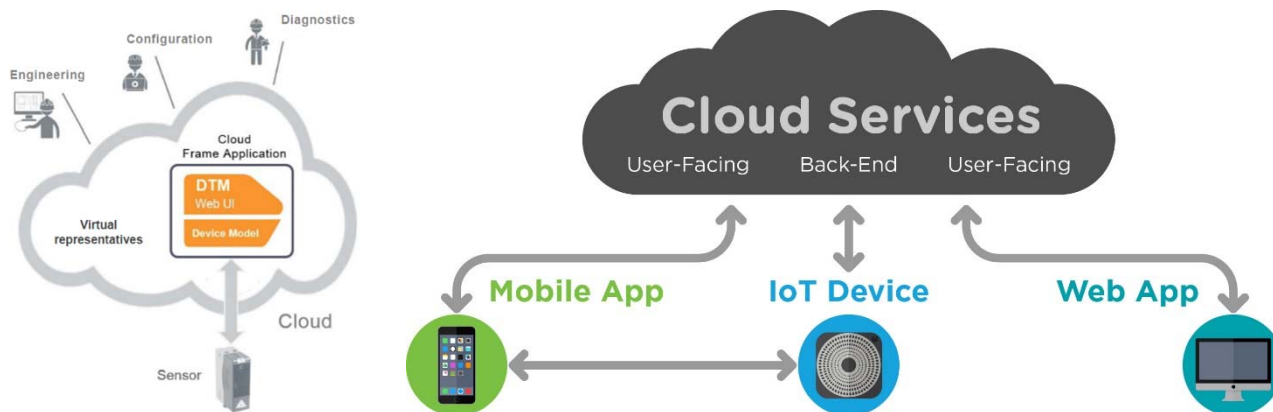
Alibaba Group

Baidu 百度

普通研究人员无法获取足够标定数据，并且没有足够的计算资源

1 迁移学习简介：为什么需要迁移学习

- **云+端的模型被普遍应用**
 - 通常需要对设备、环境、用户作具体优化
 - 个性化适配通常很复杂、很耗时
 - 对于不同用户，需要不同的隐私处理方式



如何针对新用户、新设备、新环境，快速构建模型？

1 迁移学习简介：为什么需要迁移学习

- 特定的机器学习应用

- 推荐系统中的冷启动问题：没有数据，如何作推荐？



Cold Start Problem

| | 3 | | ? |
|--|---|---|---|
| | 2 | 5 | ? |
| | | 3 | ? |

| | 3 | | 4 |
|--|---|---|---|
| | 2 | 5 | |
| | ? | ? | ? |

PANDORA®



amazon®

JD. 京东.COM

没有足够的用户数据，如何构建模型？

1 迁移学习简介：为什么需要迁移学习

- **为什么需要迁移学习**

- 数据的角度
 - 收集数据很困难
 - 为数据打标签很耗时
 - 训练一对一的模型很繁琐
- 模型的角度
 - 个性化模型很复杂
 - 云+端的模型需要作具体化适配
- 应用的角度
 - 冷启动问题：没有足够用户数据，推荐系统无法工作

因此，迁移学习是必要的

1 迁移学习简介：迁移学习方法

■ 常见的迁移学习方法分类

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)

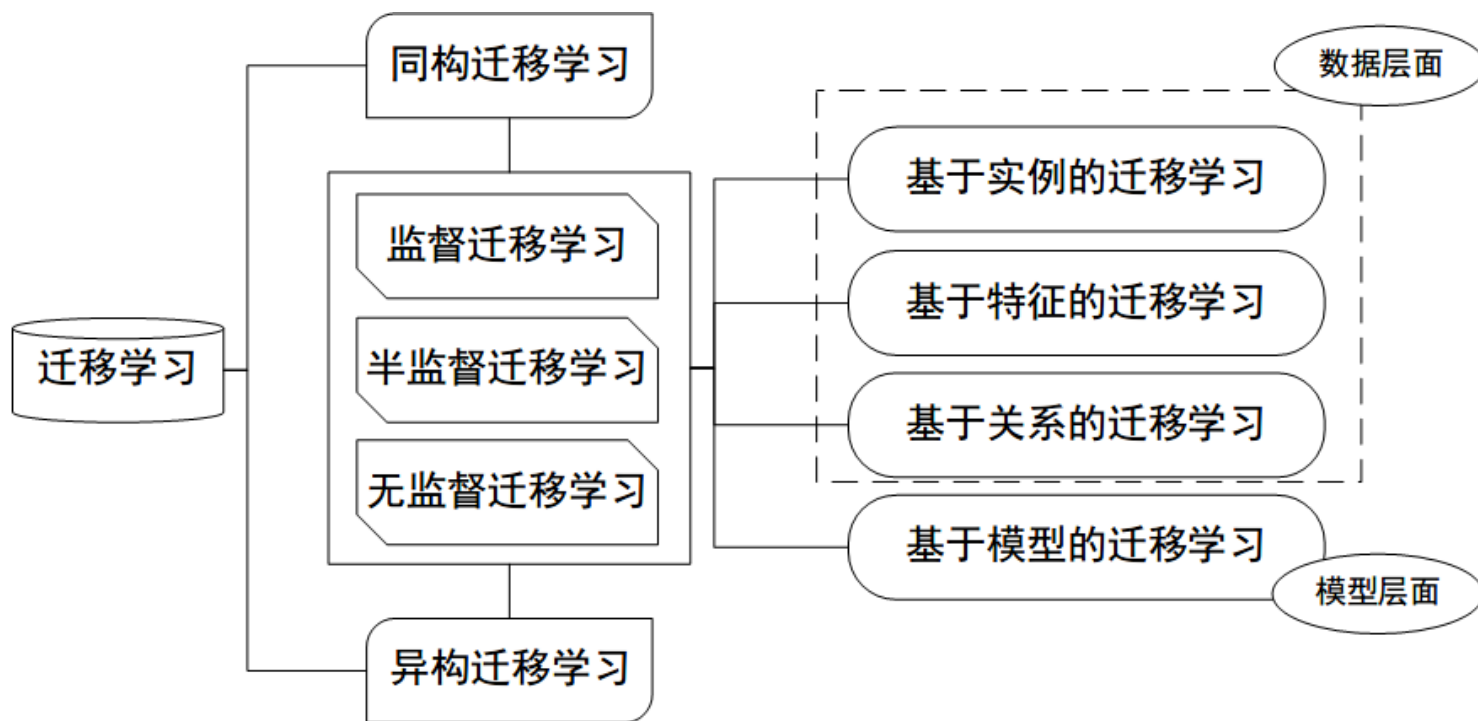
- 利用源域和目标域的参数共享模型

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

1 迁移学习简介：迁移学习方法研究领域

■ 常见的迁移学习研究领域与方法分类



目录

CONTENTS

- 1 迁移学习简介
- 2 领域自适应问题
- 3 领域自适应方法
- 4 最新研究成果
- 5 参考资料

2 领域分布自适应：形式化

■ 迁移学习基本概念

- **域(Domain)**: 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
- **任务(Task)**: 由目标函数和学习结果组成, 是学习的结果

■ 形式化

- 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 $\mathcal{T}_S \neq \mathcal{T}_T$

2 领域自适应：形式化

■ 领域自适应问题

- **Domain Adaptation (DA)**; cross-domain learning; 同构迁移学习
- 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

$$\mathcal{D}_S \neq \mathcal{D}_T: P_S(X) \neq P_T(X)$$

- 计算机视觉中的一个重要问题
 - 每年发表大量相关论文: CVPR、ICCV、ICML、NIPS、IJCAI、AAAI

Training



Testing



Caltech-256



Amazon



DSLR



Webcam



2 领域自适应：形式化

- **领域自适应问题**

- 按照目标域有无标签
 - 目标域全部有标签：supervised DA
 - 目标域有一些标签：semi-supervised DA
 - 目标域全没有标签：unsupervised DA
- **Unsupervised DA**最有挑战性，是我们的关注点

目录

CONTENTS

- 1 迁移学习简介
- 2 领域自适应问题
- 3 领域自适应方法
- 4 最新研究成果
- 5 参考资料

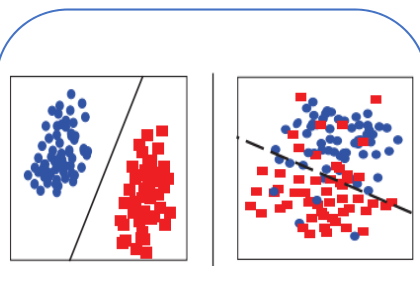
3 领域自适应：方法概览

■ 基本假设

- 数据分布角度：源域和目标域的**概率分布相似**
 - **最小化**概率分布距离
- 特征选择角度：源域和目标域共享着**某些特征**
 - **选择**出这部分公共特征
- 特征变换角度：源域和目标域共享**某些子空间**
 - 把两个域**变换**到相同的子空间

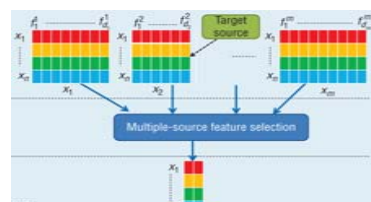
■ 解决思路

数据分布



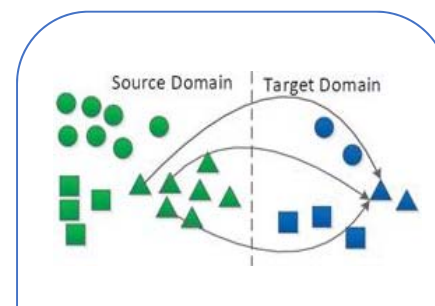
概率分布适配法
(Distribution Adaptation)

特征选择



特征选择法
(Feature Selection)

特征变换

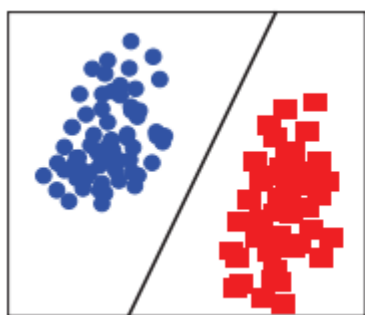


子空间学习法
(Subspace Learning)

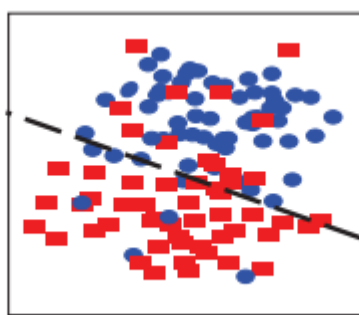
3 领域自适应：概率分布适配法

- 概率分布适配法 (Distribution Adaptation)

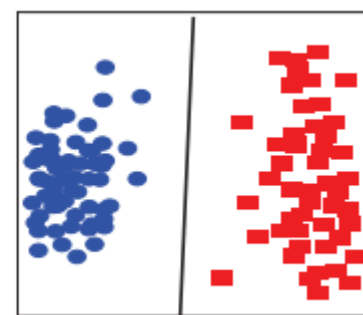
- 边缘分布适配 (Marginal distribution adaptation)
 - 假设: $P(\mathbf{X}_s) \neq P(\mathbf{X}_t)$
- 条件分布适配 (Conditional distribution adaptation)
 - 假设: $P(y_s|\mathbf{X}_s) \neq P(y_t|\mathbf{X}_t)$
- 联合分布适配 (Joint distribution adaptation)
 - 假设: $P(\mathbf{X}_s, y_s) \neq P(\mathbf{X}_t, y_t)$



源域数据



目标域数据(1)
优先考虑边缘分布



目标域数据(2)
优先考虑条件分布

3 领域自适应：概率分布适配法

■ 边缘分布适配 (1)

- 迁移成分分析 (Transfer Component Analysis, TCA) [Pan, TNN-11]

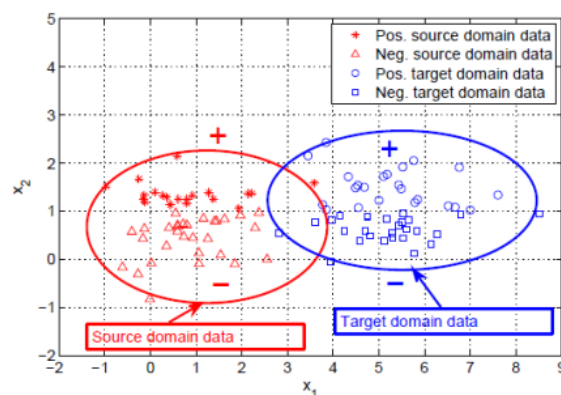
- 优化目标:

$$\min_{\varphi} \text{Dist}(\varphi(\mathbf{X}_S), \varphi(\mathbf{X}_T)) + \lambda \Omega(\varphi)$$

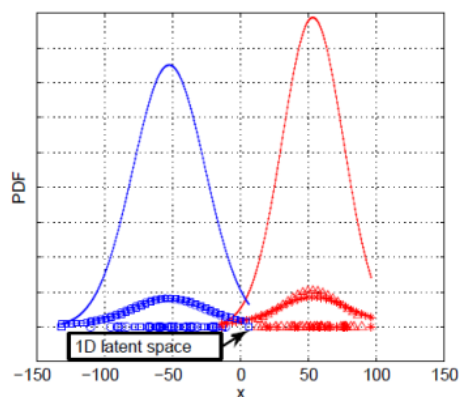
$$\text{s.t. constraints on } \varphi(\mathbf{X}_S) \text{ and } \varphi(\mathbf{X}_T)$$

- 最大均值差异(Maximum Mean Discrepancy, MMD)

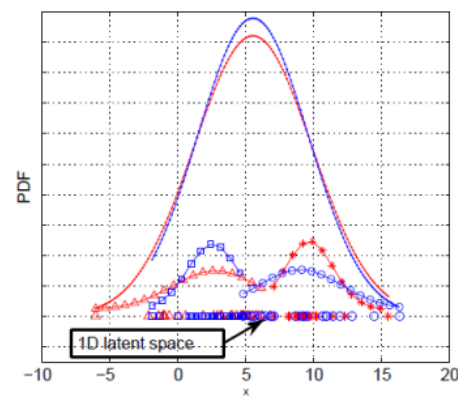
$$\text{Dist}(P(X_S), P(X_T)) = \left\| \frac{1}{n_S} \sum_{i=1}^{n_S} \Phi(x_{S_i}) - \frac{1}{n_T} \sum_{j=1}^{n_T} \Phi(x_{T_j}) \right\|_{\mathcal{H}}$$



Original feature space



PCA



TCA

3 领域自适应：概率分布适配法

■ 边缘分布适配 (2)

■ 迁移成分分析 (TCA)方法的一些扩展

- Adapting Component Analysis (ACA) [Dorri, ICDM-12] $\text{maximize } \frac{\text{tr}(HK_X HL_\Phi)}{\text{tr}(HL_M HL_\Phi)}$
 - 最小化MMD，同时维持迁移过程中目标域的结构
- Domain Transfer Multiple Kernel Learning (DTMKL) [Duan, PAMI-12]
 - 多核MMD
$$k = \sum_{m=1}^M d_m k_m$$
- Deep Domain Confusion (DDC) [Tzeng, arXiv-14]
 - 把MMD加入到神经网络中
- Deep Adaptation Networks (DAN) [Long, ICML-15]
 - 把MKK-MMD加入到神经网络中
- Distribution-Matching Embedding (DME) [Baktashmotlagh, JMLR-16]
 - 先计算变换矩阵，再进行映射
- Central Moment Discrepancy (CMD) [Zellinger, ICLR-17]
 - 不只是一阶的MMD，推广到了k阶

3 领域自适应：概率分布适配法

■ 条件分布适配

- Domain Adaptation of Conditional Probability Models via Feature Subsetting [Satpal, PKDD-07]

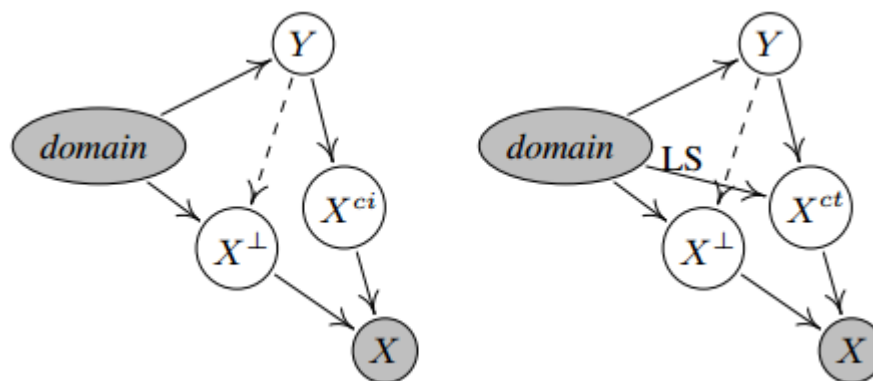
- 条件随机场+分布适配
- 优化目标：

$$\operatorname{argmax}_{\mathbf{w}, S} \sum_{(\mathbf{x}, \mathbf{y}) \in D} \sum_{k \in S} w_k f_k(\mathbf{x}, \mathbf{y}) - \log z_{\mathbf{w}}(\mathbf{x})$$

such that $\operatorname{dist}(\mathcal{D}, \mathcal{D}' | S, D, D') \leq \epsilon$.

- Conditional Transferrable Components (CTC) [Gong, ICML-15]

- 定义条件转移成分，对其进行建模



3 领域自适应：概率分布适配法

■ 联合分布适配 (1)

- 联合分布适配 (Joint Distribution Adaptation, JDA) [Long, ICCV-13]

- 直接继承于TCA，但是加入了条件分布适配
- 优化目标：

$$D(\mathcal{D}_s, \mathcal{D}_t) \approx D(P(\mathbf{x}_s), P(\mathbf{x}_t)) \\ + D(P(y_s|\mathbf{x}_s), P(y_t|\mathbf{x}_t))$$

- 问题：如何获得估计条件分布？
 - 充分统计量：用类条件概率近似条件概率
 - 用一个弱分类器生成目标域的初始软标签
- 最终优化形式

$$\min_{\mathbf{A}^T \mathbf{K} \mathbf{H} \mathbf{K}^T \mathbf{A} = \mathbf{I}} \sum_{c=0}^C \text{tr}(\mathbf{A}^T \mathbf{K} \mathbf{M}_c \mathbf{K}^T \mathbf{A}) + \lambda \|\mathbf{A}\|_F^2$$

- 联合分布适配的结果普遍优于比单独适配边缘或条件分布

3 领域自适应：概率分布适配法

■ 联合分布适配 (2)

- 联合分布适配(JDA)方法的一些扩展
 - Adaptation Regularization (ARTL) [Long, TKDE-14]
 - 分类器学习+联合分布适配
 - Visual Domain Adaptation (VDA) [Tahmoresnezhad, KIS-17]
 - 加入类内距、类间距
 - Joint Geometrical and Statistical Alignment (JGSA) [Zhang, CVPR-17]
 - 加入类内距、类间距、标签适配
 - [Hsu, TIP-16]: 加入结构不变性控制
 - [Hsu, AVSS-15]: 目标域选择
 - Joint Adaptation Networks (JAN) [Long, ICML-17]
 - 提出JMMD度量, 在深度网络中进行联合分布适配

3 领域自适应：概率分布适配法

■ 联合分布适配 (3)

- 平衡分布适配 (Balanced Distribution Adaptation, BDA) [Wang, ICDM-2017]

- 仅仅适配条件分布和边缘分布就够了吗?
 - 联合分布适配的问题：两种分布同等重要
 - 真实环境：两种分布**不一定**同等重要
- 加入**平衡因子**动态衡量两种分布的重要性

$$D(\mathcal{D}_s, \mathcal{D}_t) \approx (1 - \underbrace{\mu}_{\text{平衡因子}}) D(P(\mathbf{x}_s), P(\mathbf{x}_t)) + \mu D(P(y_s|\mathbf{x}_s), P(y_t|\mathbf{x}_t)) \quad \mu \in [0, 1]$$

- 当 $\mu \rightarrow 0$, 表示边缘分布更占优, 应该优先适配
- 当 $\mu \rightarrow 1$, 表示条件分布更占优, 应该优先适配
- 最终表示形式

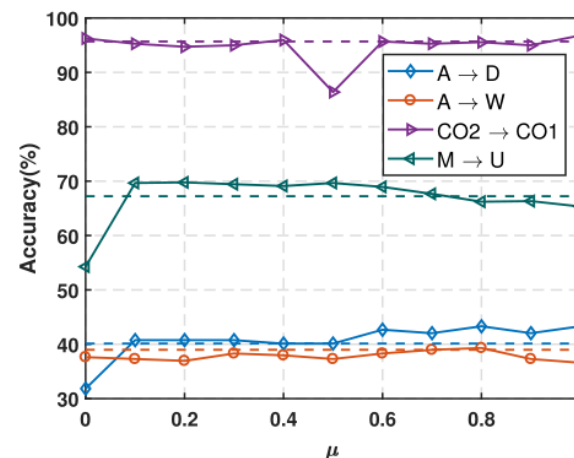
$$\begin{aligned} \min \quad & \text{tr} \left(\mathbf{A}^\top \mathbf{X} \left((1 - \mu) \mathbf{M}_0 + \mu \sum_{c=1}^C \mathbf{M}_c \right) \mathbf{X}^\top \mathbf{A} \right) + \lambda \|\mathbf{A}\|_F^2 \\ \text{s.t.} \quad & \mathbf{A}^\top \mathbf{X} \mathbf{H} \mathbf{X}^\top \mathbf{A} = \mathbf{I}, \quad 0 \leq \mu \leq 1 \end{aligned}$$

3 领域自适应：概率分布适配法

■ 联合分布适配 (4)

■ 平衡分布适配 (BDA): 平衡因子的重要性

- 对于不同的任务，边缘分布和条件分布并不是同等重要，因此，BDA方法可以**有效衡量**这两个分布的权重，从而达到最好的结果



■ 平衡分布适配 (BDA): 平衡因子的求解与估计

- 目前尚无精确的估计方法; 我们采用A-distance来进行估计
 - 求解源域和目标域整体的A-distance
 - 对目标域聚类，计算源域和目标域每个类的A-distance
 - 计算上述两个距离的比值，则为平衡因子

3 领域自适应：概率分布适配法

■ 概率分布适配：总结

■ 方法

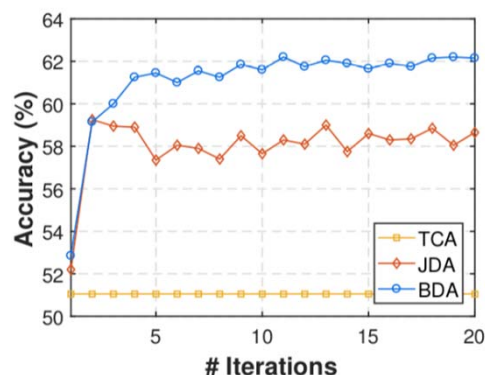
- 基础：大多数方法基于MMD距离进行优化求解
- 分别进行边缘 / 条件 / 联合概率适配
- 效果：平衡 (BDA) > 联合 (JDA) > 边缘 (TCA) > 条件

■ 使用

- 数据整体差异性大 (相似度较低)，边缘分布更重要
- 数据整体差异性小 (协方差漂移)，条件分布更重要

■ 最新成果

- 深度学习+分布适配往往有更好的效果 (DDC、DAN、JAN)



BDA、JDA、TCA精度比较

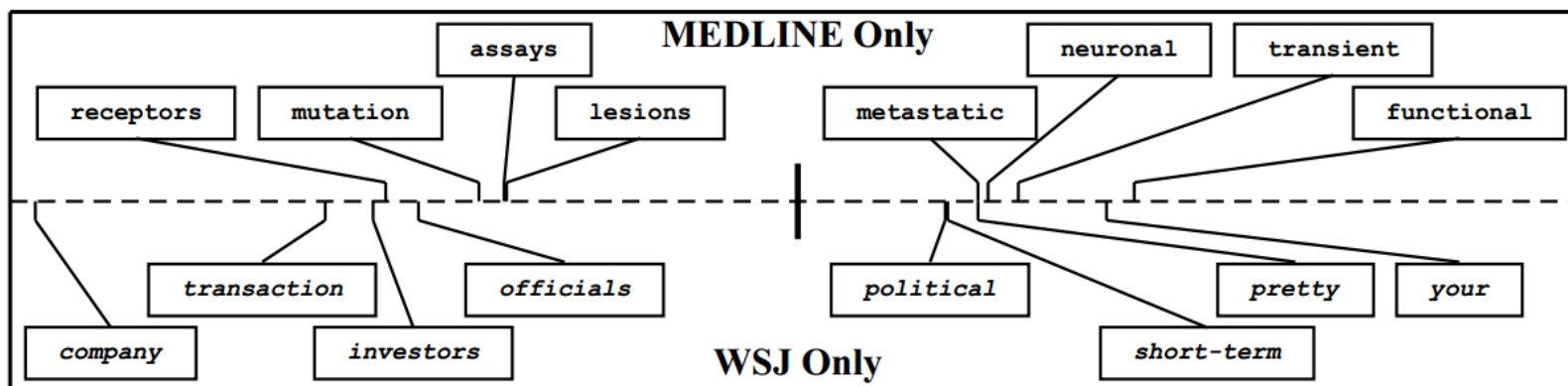
| Method | A → W | D → W | W → D | A → D | D → A | W → A | Avg |
|-----------------------------------|----------|----------|----------|----------|----------|----------|------|
| AlexNet (Krizhevsky et al., 2012) | 61.6±0.5 | 95.4±0.3 | 99.0±0.2 | 63.8±0.5 | 51.1±0.6 | 49.8±0.4 | 70.1 |
| TCA (Pan et al., 2011) | 61.0±0.0 | 93.2±0.0 | 95.2±0.0 | 60.8±0.0 | 51.6±0.0 | 50.9±0.0 | 68.8 |
| GFK (Gong et al., 2012) | 60.4±0.0 | 95.6±0.0 | 95.0±0.0 | 60.6±0.0 | 52.4±0.0 | 48.1±0.0 | 68.7 |
| DDC (Tzeng et al., 2014) | 61.8±0.4 | 95.0±0.5 | 98.5±0.4 | 64.4±0.3 | 52.1±0.6 | 52.2±0.4 | 70.6 |
| DAN (Long et al., 2015) | 68.5±0.5 | 96.0±0.3 | 99.0±0.3 | 67.0±0.4 | 54.0±0.5 | 53.1±0.5 | 72.9 |
| RTN (Long et al., 2016) | 73.3±0.3 | 96.8±0.2 | 99.6±0.1 | 71.0±0.2 | 50.5±0.3 | 51.0±0.1 | 73.7 |
| RevGrad (Ganin & Lempitsky, 2015) | 73.0±0.5 | 96.4±0.3 | 99.2±0.3 | 72.3±0.3 | 53.4±0.4 | 51.2±0.5 | 74.3 |
| JAN (ours) | 74.9±0.3 | 96.6±0.2 | 99.5±0.2 | 71.8±0.2 | 58.3±0.3 | 55.0±0.4 | 76.0 |
| JAN-A (ours) | 75.2±0.4 | 96.6±0.2 | 99.6±0.1 | 72.8±0.3 | 57.5±0.2 | 56.3±0.2 | 76.3 |

DDC、DAN、JAN与其他方法结果比较

3 领域自适应：特征选择法

■ 特征选择法 (Feature Selection)

- 从源域和目标域中选择提取共享的特征，建立统一模型
- Structural Correspondence Learning (SCL) [Blitzer, ECML-06]
 - 寻找Pivot feature，将源域和目标域进行对齐



3 领域自适应：特征选择法

■ 特征选择法其他扩展

- Joint feature selection and subspace learning [Gu, IJCAI-11]

- 特征选择/变换+子空间学习

- 优化目标：
$$\min_{\mathbf{A}} \|\mathbf{A}\|_{2,1} + \mu \text{tr}(\mathbf{A}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{A})$$
$$\text{s.t. } \mathbf{A}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{A} = \mathbf{I},$$

- Transfer Joint Matching (TJM) [Long, CVPR-14]

- MMD分布适配+源域样本选择

- 优化目标：
$$\min_{\mathbf{A}^T \mathbf{K} \mathbf{H} \mathbf{K}^T \mathbf{A} = \mathbf{I}} \text{tr}(\mathbf{A}^T \mathbf{K} \mathbf{M} \mathbf{K}^T \mathbf{A}) + \lambda (\|\mathbf{A}_s\|_{2,1} + \|\mathbf{A}_t\|_F^2)$$

- Feature Selection and Structure Preservation (FSSL) [Li, IJCAI-16]

- 特征选择+信息不变性

- 优化目标：

$$\min_{\mathbf{P}, \mathbf{Z}, \mathbf{E}} \|\mathbf{P}\|_{2,1} + \frac{\lambda}{2} \text{tr}(\mathbf{P}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{P}) + \frac{\beta}{2} \|\mathbf{Z}\|_F^2 + \gamma \|\mathbf{E}\|_1$$
$$\text{s.t. } \mathbf{P}^T \mathbf{X} = \mathbf{P}^T \mathbf{X}_s \mathbf{Z} + \mathbf{E}, \mathbf{P}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{P} = \mathbf{I},$$

3 领域自适应：特征选择法

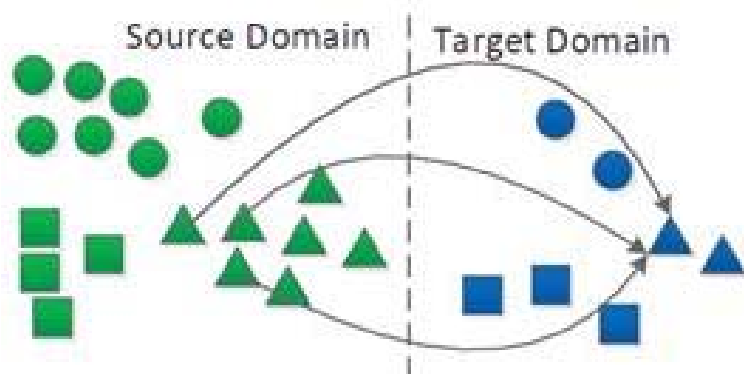
- **特征选择法：总结**

- 从源域和目标域中选择提取共享的特征，建立统一模型
- 通常与分布适配进行结合
- 选择特征通常利用稀疏矩阵

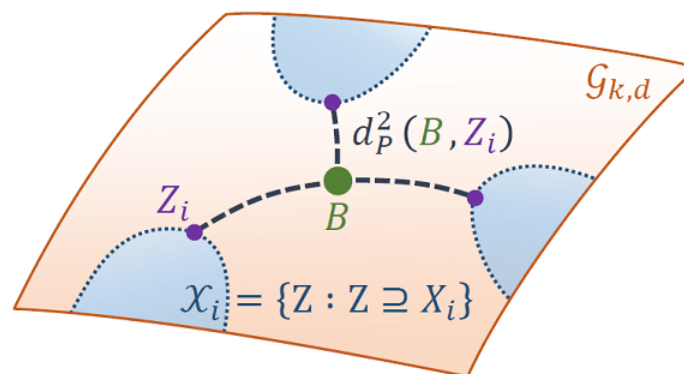
3 领域自适应：子空间学习法

■ 子空间学习法 (Subspace Learning)

- 将源域和目标域变换到相同的子空间，然后建立统一的模型
 - 统计特征变换 (Statistical Feature Transformation)
 - 将源域和目标域的一些统计特征进行变换对齐
 - 流形学习 (Manifold Learning)
 - 在流形空间中进行子空间变换



统计特征变换



流形学习

3 领域自适应：子空间学习法

■ 统计特征变换 (1)

- 子空间对齐法 (Subspace Alignment, SA) [Fernando, ICCV-13]

- 直接寻求一个线性变换, 把source变换到target空间中
- 优化目标: $F(M) = \|X_S M - X_T\|_F^2$

$$M^* = \operatorname{argmin}_M (F(M))$$

- 直接获得线性变换的闭式解:

$$F(M) = \|X_S' X_S M - X_S' X_T\|_F^2 = \|M - X_S' X_T\|_F^2.$$

- 子空间分布对齐法 (Subspace Distribution Alignment, SDA) [Sun, BMVC-15]

- 子空间对齐+概率分布适配

$$M_s = S_s T_{TS} A_{TS} S_t^T = S_s (S_s^T S_t) (E_s^{-\frac{1}{2}} E_t^{\frac{1}{2}}) S_t^T$$

- 空间对齐法: 方法简洁, 计算高效

3 领域自适应：子空间学习法

■ 统计特征变换 (2)

- 关联对齐法 (CORrelation Alignment, CORAL) [Sun, AAAI-15]

- 最小化源域和目标域的二阶统计特征

- 优化目标：
$$\min_A \|C_{\hat{S}} - C_T\|_F^2$$
$$= \min_A \|A^T C_S A - C_T\|_F^2$$

- 形式简单，求解高效

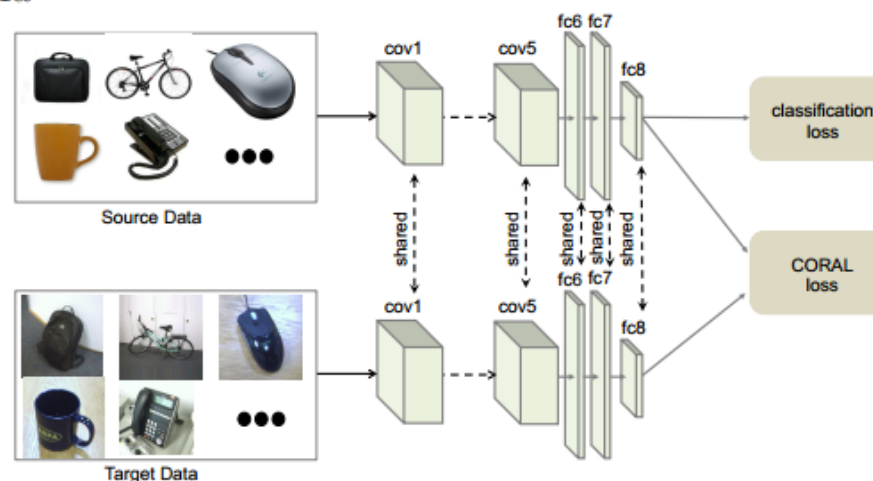
- 深度关联对齐 (Deep-CORAL) [Sun, ECCV-16]

- 在深度网络中加入CORAL

- CORAL loss: $\ell_{CORAL} = \frac{1}{4d^2} \|C_S - C_T\|_F^2$

$$C_S = \frac{1}{n_S - 1} (D_S^T D_S - \frac{1}{n_S} (\mathbf{1}^T D_S)^T (\mathbf{1}^T D_S))$$

$$C_T = \frac{1}{n_T - 1} (D_T^T D_T - \frac{1}{n_T} (\mathbf{1}^T D_T)^T (\mathbf{1}^T D_T))$$

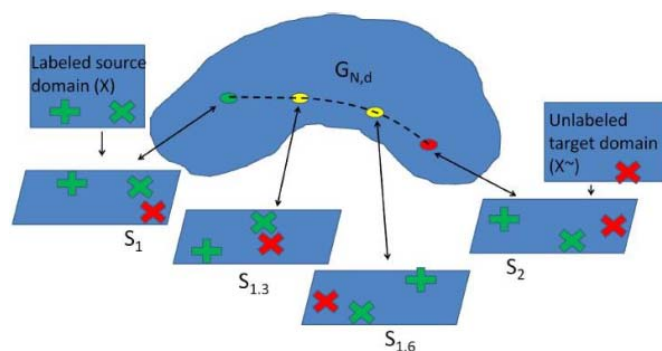


3 领域自适应：子空间学习法

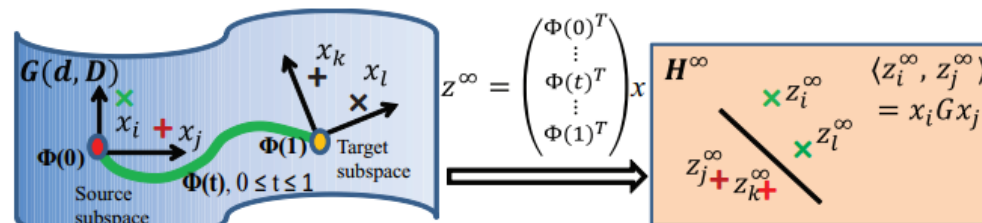
■ 流形学习 (1)

- 采样测地线流方法 (Sample Geodesic Flow, SGF) [Gopalan, ICCV-11]
 - 把领域自适应的问题看成一个增量式“行走”问题
 - 从源域走到目标域就完成了自适应过程
 - 在流形空间中采样有限个点，构建一个测地线流
- 测地线流式核方法 (Geodesic Flow Kernel, GFK) [Gong, CVPR-12]
 - 继承了SGF方法，采样无穷个点
 - 转化成Grassmann流形中的核学习，构建了GFK
 - 优化目标：

$$\langle z_i^\infty, z_j^\infty \rangle = \int_0^1 (\Phi(t)^T x_i)^T (\Phi(t)^T x_j) dt = x_i^T G x_j$$



SGF方法



GFK方法

3 领域自适应：子空间学习法

■ 流形学习 (2)

- 域不变映射 (Domain-Invariant Projection, DIP) [Baktashmotlagh, CVPR-13]
 - 直接度量分布距离是不好的：原始空间特征扭曲
 - 仅作流形子空间学习：无法刻画分布距离
 - 解决方案：流形映射+分布度量
- 统计流形法 (Statistical Manifold) [Baktashmotlagh, CVPR-14]
 - 在统计流形(黎曼流形)上进行分布度量
 - 用Fisher-Rao distance (Hellinger distance)进行度量

$$\min_{\alpha} \frac{1}{\sum_{i=1}^{n_s} \alpha_i} \sum_{i=1}^{n_s} \alpha_i \left(\sqrt{\hat{T}(x_i^s)} - \sqrt{1 - \hat{T}(x_i^s)} \right)^2 + \frac{1}{n_t} \sum_{i=1}^{n_t} \left(\sqrt{\hat{T}(x_i^t)} - \sqrt{1 - \hat{T}(x_i^t)} \right)^2$$

3 领域自适应：子空间学习法

- **子空间学习法：总结**
 - 主要包括统计特征对齐和流形学习方法两大类
 - 和分布适配结合效果更好
 - 趋势：与神经网络结合

目录

CONTENTS

1

迁移学习简介

2

领域自适应问题

3

领域自适应方法

4

最新研究成果

5

参考资料

4 最新研究成果

■ 领域自适应的最新研究成果(1)

- 与深度学习进行结合
 - Deep Adaptation Networks (DAN) [Long, ICML-15]
 - 深度网络+MMD距离最小化
 - Joint Adaptation Networks (JAN) [Long, ICML-17]
 - 深度网络+联合分布距离最小化
 - Simultaneous feature and task transfer [Tzeng, ICCV-15]
 - 特征和任务同时进行迁移
 - Deep Hashing Network (DHN) [CVPR-17]
 - 在深度网络中同时学习域适应和深度Hash特征
 - Label Efficient Learning of Transferable Representations across Domains and Tasks [Luo, NIPS-17]
 - 在深度网络中进行任务迁移

4 最新研究成果

■ 领域自适应的最新研究成果(2)

- 与对抗学习进行结合
 - Domain-adversarial neural network [Ganin, JMLR-16]
 - 深度网络中加入对抗
 - Adversarial Discriminative Domain Adaptation (ADDA) [Tzeng, arXiv-17]
 - 对抗+判别
- 开放世界领域自适应
 - Open set domain adaptation [Busto, ICCV-17]
 - 当源域和目标域只共享一部分类别时如何迁移?
- 与张量 (Tensor)表示相结合
 - When DA Meets tensor representation [Lu, ICCV-17]
 - 用tensor的思想来做领域自适应
- 与增量学习结合
 - Learning to Transfer (L2T) [Wei, arXiv-17]
 - 提取已有的迁移学习经验, 应用于新任务

目录

CONTENTS

1

迁移学习简介

2

领域自适应问题

3

领域自适应方法

4

最新研究成果

5

参考资料

5 参考资料

■ 参考资料

- 迁移学习综述文章
 - A survey on Transfer Learning [Pan and Yang, TKDE-10]
- (可能是有史以来)最全的**迁移学习资料库**, (文章/资料/代码/数据)
 - <https://github.com/jindongwang/transferlearning>
- 迁移学习视频教程
 - <https://www.youtube.com/watch?v=qD6iD4TFsdQ>
- 知乎专栏 “机器有颗玻璃心” 中《小王爱迁移》系列
 - <https://zhuanlan.zhihu.com/p/27336930>
 - 用浅显易懂的语言深入讲解经典+最新的迁移学习文章
- 迁移学习与领域自适应论文分享与笔记
 - Paperweekly: <http://www.paperweekly.site/collections/231/papers>
- 迁移学习与领域自适应公开**数据集**
 - <https://github.com/jindongwang/transferlearning/blob/master/doc/dataset.md>



图：Office+Caltech、USPS+MNIST、ImageNet+VOC、COIL20数据集

参考文献(1)

- [Pan, TNN-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.
- [Dorri, ICDM-12] Dorri F, Ghodsi A. Adapting component analysis[C]//Data Mining (ICDM), 2012 IEEE 12th International Conference on. IEEE, 2012: 846-851.
- [Duan, PAMI-12] Duan L, Tsang I W, Xu D. Domain transfer multiple kernel learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(3): 465-479.
- [Long, ICML-15] Long M, Cao Y, Wang J, et al. Learning transferable features with deep adaptation networks[C]//International Conference on Machine Learning. 2015: 97-105.
- [Baktashmotlagh, JMLR-16] Baktashmotlagh M, Harandi M, Salzmann M. Distribution-matching embedding for visual domain adaptation[J]. The Journal of Machine Learning Research, 2016, 17(1): 3760-3789.
- [Zellinger, ICLR-17] Zellinger W, Grubinger T, Lughofer E, et al. Central moment discrepancy (CMD) for domain-invariant representation learning[J]. arXiv preprint arXiv:1702.08811, 2017.
- [Satpal, PKDD-07] Satpal S, Sarawagi S. Domain adaptation of conditional probability models via feature subsetting[C]//PKDD. 2007, 4702: 224-235.
- [Gong, ICML-15] Gong M, Zhang K, Liu T, et al. Domain adaptation with conditional transferable components[C]//International Conference on Machine Learning. 2016: 2839-2848.
- [Long, ICCV-13] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in ICCV, 2013, pp. 2200–2207.
- [Long, TKDE-14] Long M, Wang J, Ding G, et al. Adaptation regularization: A general framework for transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(5): 1076-1089.
- [Tahmoresnezhad, KIS-17] J. Tahmoresnezhad and S. Hashemi, "Visual domain adaptation via transfer feature learning," Knowl. Inf. Syst., 2016.
- [Zhang, CVPR-17] Zhang J, Li W, Ogunbona P. Joint Geometrical and Statistical Alignment for Visual Domain Adaptation, CVPR 2017.
- [Hsu, AVSS-15] T. Ming Harry Hsu, W. Yu Chen, C.-A. Hou, and H. T. et al., "Unsupervised domain adaptation with imbalanced cross-domain data," in ICCV, 2015, pp. 4121–4129.
- [Hsu, TIP-16] P.-H. Hsiao, F.-J. Chang, and Y.-Y. Lin, "Learning discriminatively reconstructed source data for object recognition with few examples," TIP, vol. 25, no. 8, pp. 3518–3532, 2016.
- [Long, ICML-17] Long M, Wang J, Jordan M I. Deep transfer learning with joint adaptation networks. ICML 2017.
- [Wang, ICDM-17] Wang J, Chen Y, Hao S, Feng W, Shen Z. Balanced Distribution Adaptation for Transfer Learning. ICDM 2017. pp.1129-1134.
- [Blitzer, ECML-06] Blitzer J, McDonald R, Pereira F. Domain adaptation with structural correspondence learning[C]//Proceedings of the 2006 conference on empirical methods in natural language processing. Association for Computational Linguistics, 2006: 120-128.

参考文献(2)

- [Gu, IJCAI-11] Gu Q, Li Z, Han J. Joint feature selection and subspace learning[C]//IJCAI Proceedings-International Joint Conference on Artificial Intelligence. 2011, 22(1): 1294.
- [Long, CVPR-14] Long M, Wang J, Ding G, et al. Transfer joint matching for unsupervised domain adaptation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 1410-1417.
- [Li, IJCAI-16] Li J, Zhao J, Lu K. Joint Feature Selection and Structure Preservation for Domain Adaptation[C]//IJCAI. 2016: 1697-1703.
- [Fernando, ICCV-13] Fernando B, Habrard A, Sebban M, et al. Unsupervised visual domain adaptation using subspace alignment[C]//Proceedings of the IEEE international conference on computer vision. 2013: 2960-2967.
- [Sun, BMVC-15] Sun B, Saenko K. Subspace Distribution Alignment for Unsupervised Domain Adaptation[C]//BMVC. 2015: 24.1-24.10.
- [Sun, AAAI-16] Sun B, Feng J, Saenko K. Return of Frustratingly Easy Domain Adaptation[C]//AAAI. 2016, 6(7): 8.
- [Sun, ECCV-16] Sun B, Saenko K. Deep coral: Correlation alignment for deep domain adaptation[C]//Computer Vision—ECCV 2016 Workshops. Springer International Publishing, 2016: 443-450.
- [Gopalan, ICCV-11] Gopalan R, Li R, Chellappa R. Domain adaptation for object recognition: An unsupervised approach[C]//Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011: 999-1006.
- [Gong, 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.
- [Baktashmotlagh, CVPR-13] Baktashmotlagh M, Harandi M T, Lovell B C, et al. Unsupervised domain adaptation by domain invariant projection[C]//Proceedings of the IEEE International Conference on Computer Vision. 2013: 769-776.
- [Baktashmotlagh, CVPR-14] Baktashmotlagh M, Harandi M T, Lovell B C, et al. Domain adaptation on the statistical manifold[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 2481-2488.
- [Ganin, JMLR-16] Ganin Y, Ustinova E, Ajakan H, et al. Domain-adversarial training of neural networks[J]. Journal of Machine Learning Research, 2016, 17(59): 1-35.
- [Busto, ICCV-17] Panareda Busto P, Gall J. Open Set Domain Adaptation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 754-763.
- [Lu, ICCV-17] Lu H, Zhang L, Cao Z, et al. When unsupervised domain adaptation meets tensor representations. ICCV 2017.
- [Tzeng, arXiv-17] Tzeng E, Hoffman J, Saenko K, et al. Adversarial discriminative domain adaptation[J]. arXiv preprint arXiv:1702.05464, 2017.
- [Wei, arXiv-17] Wei Y, Zhang Y, Yang Q. Learning to Transfer. arXiv 1708.05629, 2017.

谢谢！
请批评指正