Joint Transfer and Batch-mode Active Learning

ICML 2013

Introduction

参数分类模型 $\max_{\theta} P(X, Y|\theta) = P(X|\theta)P(Y|X, \theta)$

迁移学习问题 source $S \sim (X_S, Y_S)$ $P_S(X) \neq P_T(X)$
target $T \sim (X_T, Y_T)$

$P_S(X, Y|\theta) \neq P_T(X, Y|\theta)$无法直接利用源域数据训练目标域模型

$P_S(Y|X) = P_T(Y|X)$ 样本迁移、主动学习 $P_{S \cup L \cup Q}(X) = P_{U-Q}(X)$
$T = U \cup L$

MMD (Maximum Mean Discrepancy)

$$MMD[F,p,q] \coloneqq \sup_{f\in F}(\mathbf{E}_x[f(x)] - \mathbf{E}_y[f(y)])$$

$$MMD_b[F,X,Y] \coloneqq \sup_{f\in F}\left(\frac{1}{m}\sum_{i=1}^{m} f(x_i) - \frac{1}{n}\sum_{i=1}^{n} f(y_i)\right)$$

设F为RKHS上单位球,$f(x) = \langle f, \phi(x)\rangle_H$, $\langle \phi(x), \phi(y)\rangle = k(x,y)$

$\mathbf{E}_x f = \langle f, \mu_p\rangle_H$, $\mathbf{E}_y f = \langle f, \mu_q\rangle_H, \mu \in H$

$$\begin{aligned} MMD^2[F,p,q] &= \left[\sup_{\|f\|_H\leq 1}\left(\mathbf{E}_x[f(x)] - \mathbf{E}_y[f(y)]\right)\right]^2 \\ &= \left[\sup_{\|f\|_H\leq 1}\langle \mu_p - \mu_q, f\rangle_H\right]^2 \\ &= \|\mu_p - \mu_q\|_H^2 \\ &= \langle \mu_p, \mu_p\rangle_H + \langle \mu_q, \mu_q\rangle_H - 2\langle \mu_p, \mu_q\rangle_H \end{aligned}$$

MMD (Maximum Mean Discrepancy)

$$MMD^2[F,p,q] = \mathbf{E}_{x,x'}[k(x,x')] - 2\mathbf{E}_{x,y}[k(x,y)] + \mathbf{E}_{y,y'}[k(y,y')]$$

$$MMD_u^2[F,X,Y]$$
$$= \frac{1}{m(m-1)}\sum_{i=1}^{m}\sum_{j\neq i}^{m} k(x_i,x_j) + \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} k(y_i,y_j) - \frac{2}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} k(x_i,y_j)$$

设H为定义在紧致度量空间X上且有连续核函数k的$RKHS$，则当F为H上单位球时，$MMD[F,p,q] = 0, iff\ p = q$。

证明：A Kernel Two-Sample Test

MMD (Maximum Mean Discrepancy)

Choice of kernel

$$K := \left\{ k : k = \sum_{u=1}^{d} \beta_u k_u \, , \sum_{u=1}^{d} \beta_u = D, \beta_u \geq 0, \forall u \in \{1, \dots, d\} \right\}$$

$$H_0 : p = q \qquad Type\ I\ error \leq \epsilon \qquad k^* = \underset{k \in K}{\operatorname{argmin}}\ Type\ II\ error$$

具体过程：Optimal kernel choice for large-scale two-sample tests

Joint Optimization Framework

$$P_{S \cup L \cup Q}(X) = P_{U-Q}(X)$$

$$\min_{\alpha,\beta} \| \frac{1}{n_s + n_l + b} \left(\sum_{i \in S} \beta_i \phi(x_i) + \sum_{i \in L} \phi(x_i) + \sum_{i \in U} \alpha_i \phi(x_i) \right) - \frac{1}{n_u - b} \sum_{i \in U} (1 - \alpha_i) \phi(x_i) \|_H^2$$

$$s.t.\ \alpha_i \in \{0,1\}, \beta_i \in [0,1], \alpha^\top \mathbf{1} = b$$

其中 $\phi: X \to H$

Joint Optimization Framework

$$\begin{aligned}\min \quad & \frac{1}{2}\alpha^T K_{u,u}\alpha + \frac{1}{2}\beta^T K_{s,s}\beta + \beta^T K_{s,u}\alpha \\ & - k_{u,u}^T\alpha - k_{s,u}^T\beta + k_{u,l}^T\alpha + k_{s,l}^T\beta + \text{const.} \\ \text{s.t.} \quad & \alpha_i \in \{0,1\}, \beta_i \in [0,1], \alpha^T\mathbf{1} = b.\end{aligned}$$

$$c = \frac{n_l+n_s+n_u}{n_u-b}$$

$$\begin{aligned} K_{s,s} &= \frac{1}{c^2}G(1:n_s, 1:n_s), \\ K_{u,u} &= G(n_s+1:n_s+1+n_u, n_s+1:n_s+1+n_u), \\ K_{s,u} &= \frac{1}{c}G(1:n_s, n_s+1:n_s+1+n_u), \end{aligned}$$

$$k_{u,u}(i) = \frac{n_l+n_s+b}{c^2(n_u-b)}\sum_{j=1}^{n_u} K_{u,u}(i,j), \qquad k_{s,l}(i) = \frac{1}{c^2}\sum_{j=1}^{n_l} G(i, n_s+n_u+j),$$

$$k_{s,u}(i) = \frac{n_l+n_s+b}{c^2(n_u-b)}\sum_{j=1}^{n_u} K_{s,u}(i,j), \qquad k_{u,l}(i) = \frac{1}{c}\sum_{j=1}^{n_l} G(i+n_s, n_s+n_u+j).$$

Joint Optimization Framework

$$\min_{X: X_i \in [0,1], X^\top B = b} 0.5 X^\top H X + f^\top X \quad \text{where}$$

$$X = \begin{pmatrix} \beta \\ \alpha \end{pmatrix}, H = \begin{pmatrix} K_{s,s} & K_{s,u} \\ K_{s,u}^\top & K_{u,u} \end{pmatrix}, f = \begin{pmatrix} k_{s,l} - k_{s,u} \\ k_{u,l} - k_{u,u} \end{pmatrix}, B = \begin{pmatrix} O \\ I \end{pmatrix},$$

$$I = \mathbf{1}_{n_u \times 1}, O = \mathbf{0}_{n_s \times 1}.$$

Joint Optimization Framework

Algorithm 1 JO-TAL

Input: S: source domain data; L: set of labeled target domain data; U: set of unlabeled target domain data; b: batch size; β_{new} : source weights (for iteration nos. > 1);

Output: β_{new}: source weights (updated), Q: target query set;

Compute H and f as explained in Section 2.1 and 2.2.

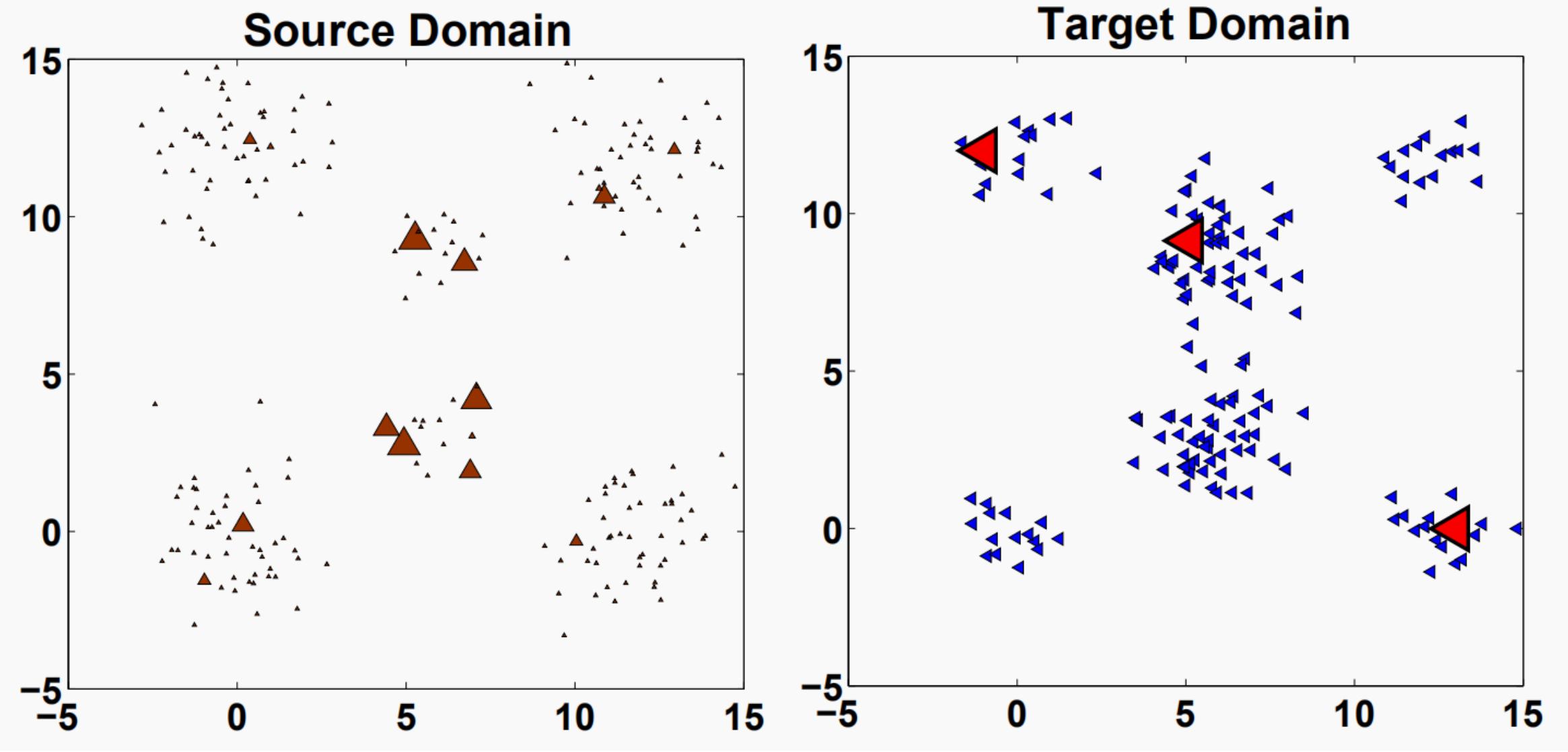
Compute β and α by solving (4).

$Q \rightarrow$ top b instances of U, sorted in descending order of α.

Update L, U : $L \leftarrow L \cup Q$, $U \leftarrow U \backslash Q$.

Update β_{new} : $\beta_{new} \leftarrow \beta_{new} + \beta$ (iteration nos. > 1).

Test



Competing Methods

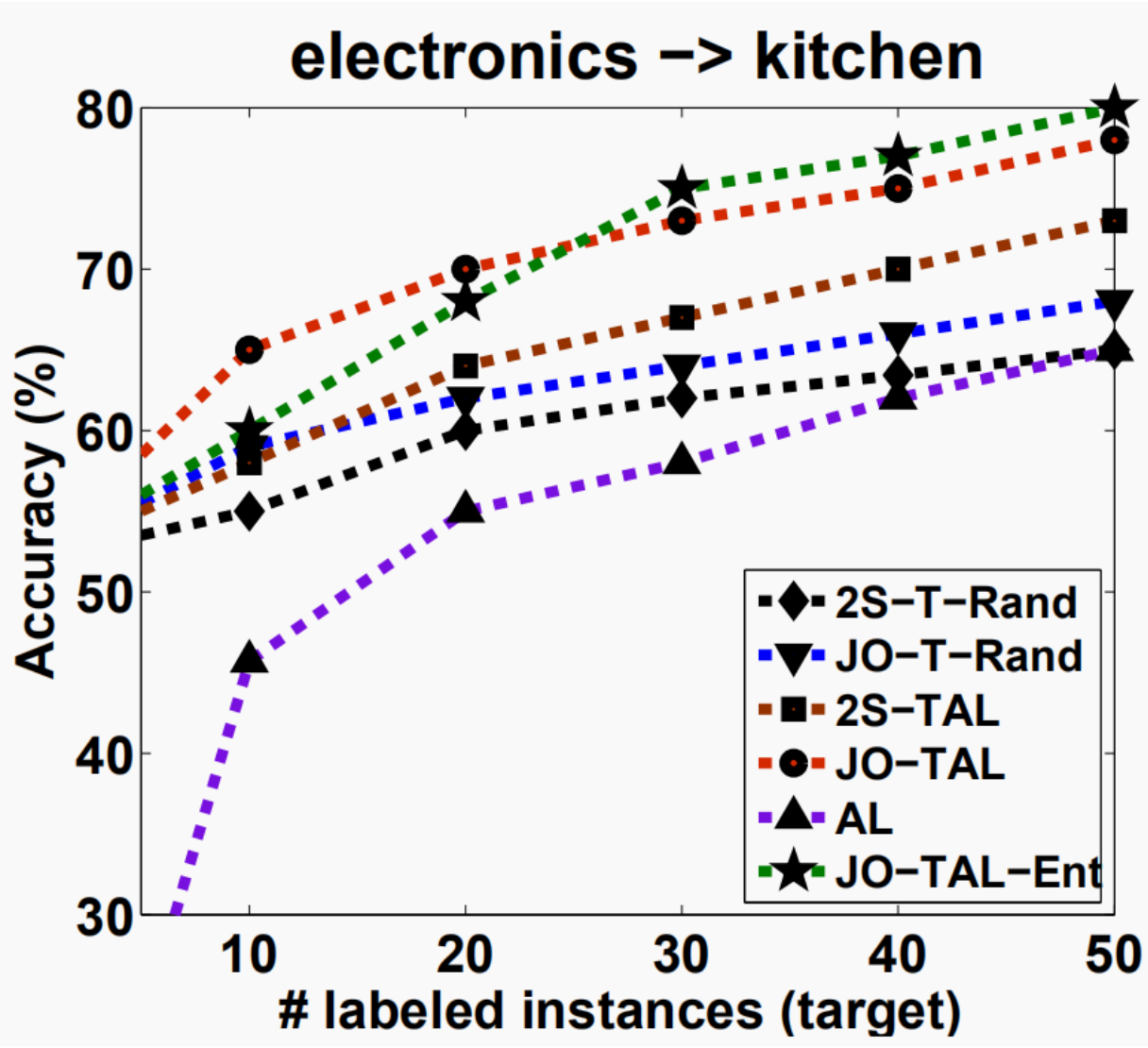
2S-TAL: 先样本迁移，再主动学习

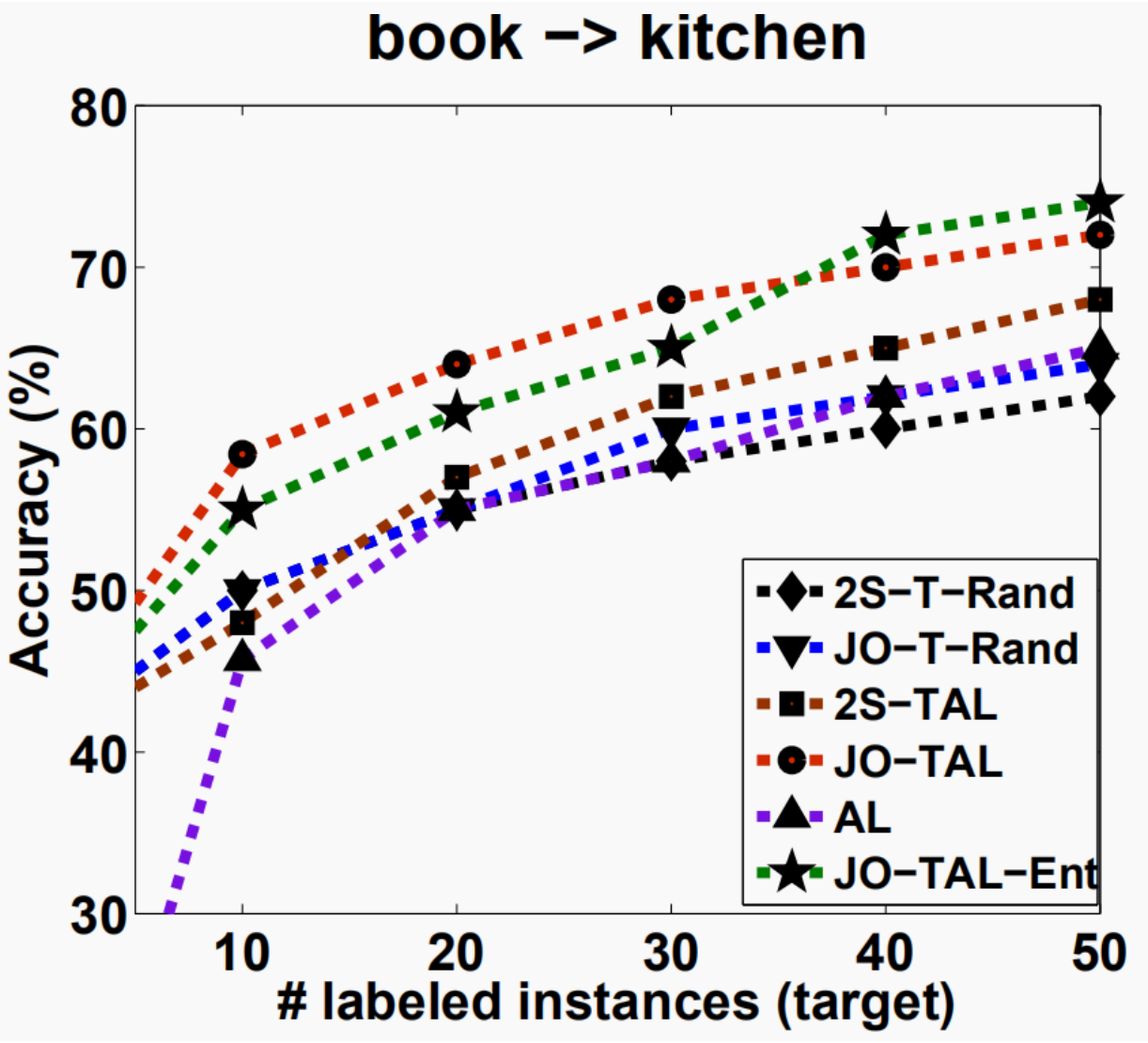
JO-T-Rand:先随机query，再样本迁移，减小$MMD(S \cup Q \cup L, U - Q)$

2S-T-Rand:先样本迁移，再随机主动学习

AL:主动学习，减小$MMD(Q \cup L, U - Q)$

Experiment


electronics -> kitchen
Accuracy (%)
80
70
60
50
40
30
10
20
30
40
50
labeled instances (target)
2S-T-Rand
JO-T-Rand
2S-TAL
JO-TAL
AL
JO-TAL-Ent


book -> kitchen
Accuracy (%)
80
70
60
50
40
30
10
20
30
40
50
labeled instances (target)
2S-T-Rand
JO-T-Rand
2S-TAL
JO-TAL
AL
JO-TAL-Ent

Thanks