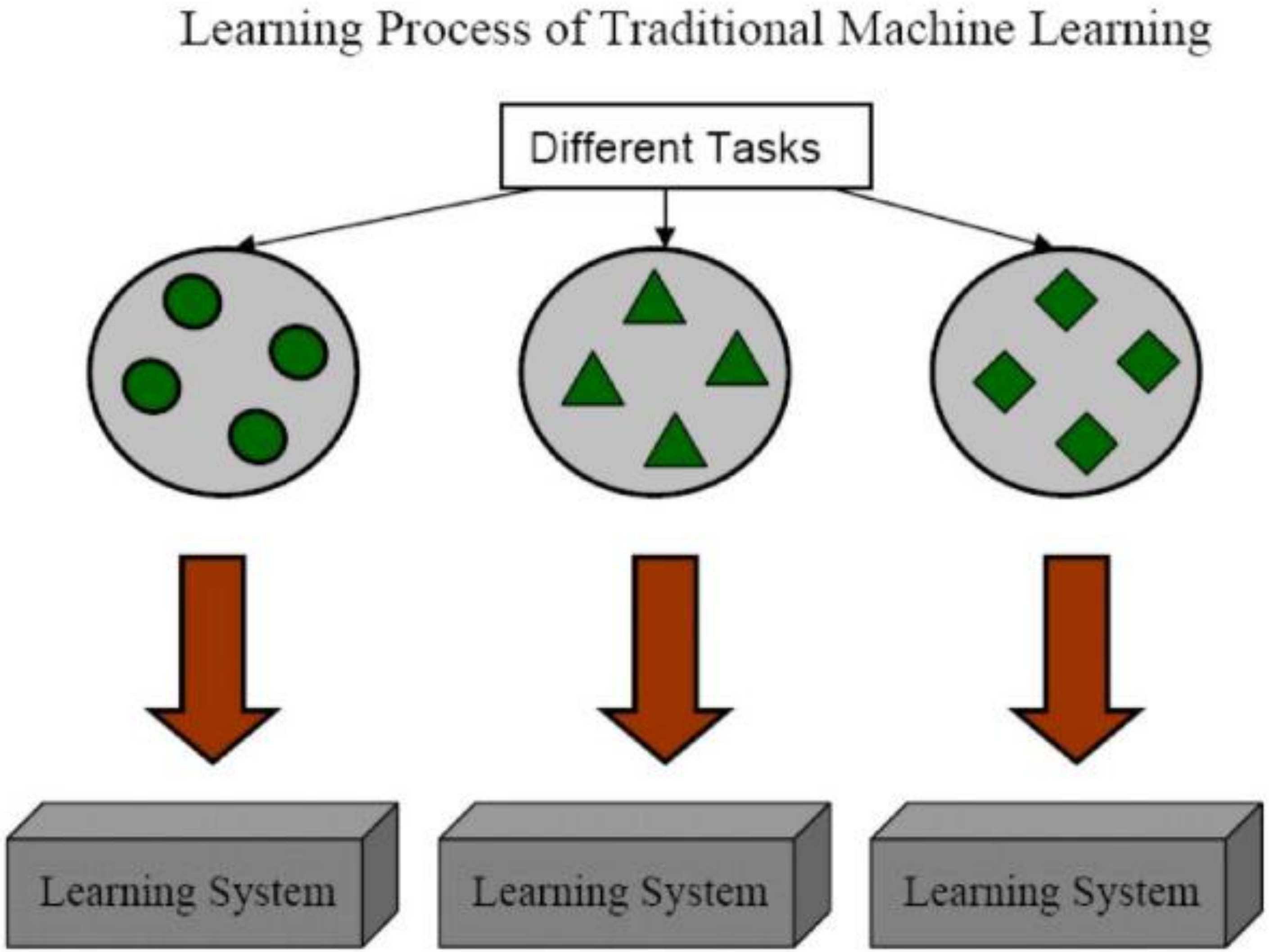
# Multi-target Unsupervised Domain Adaptation without Exactly Shared Categories

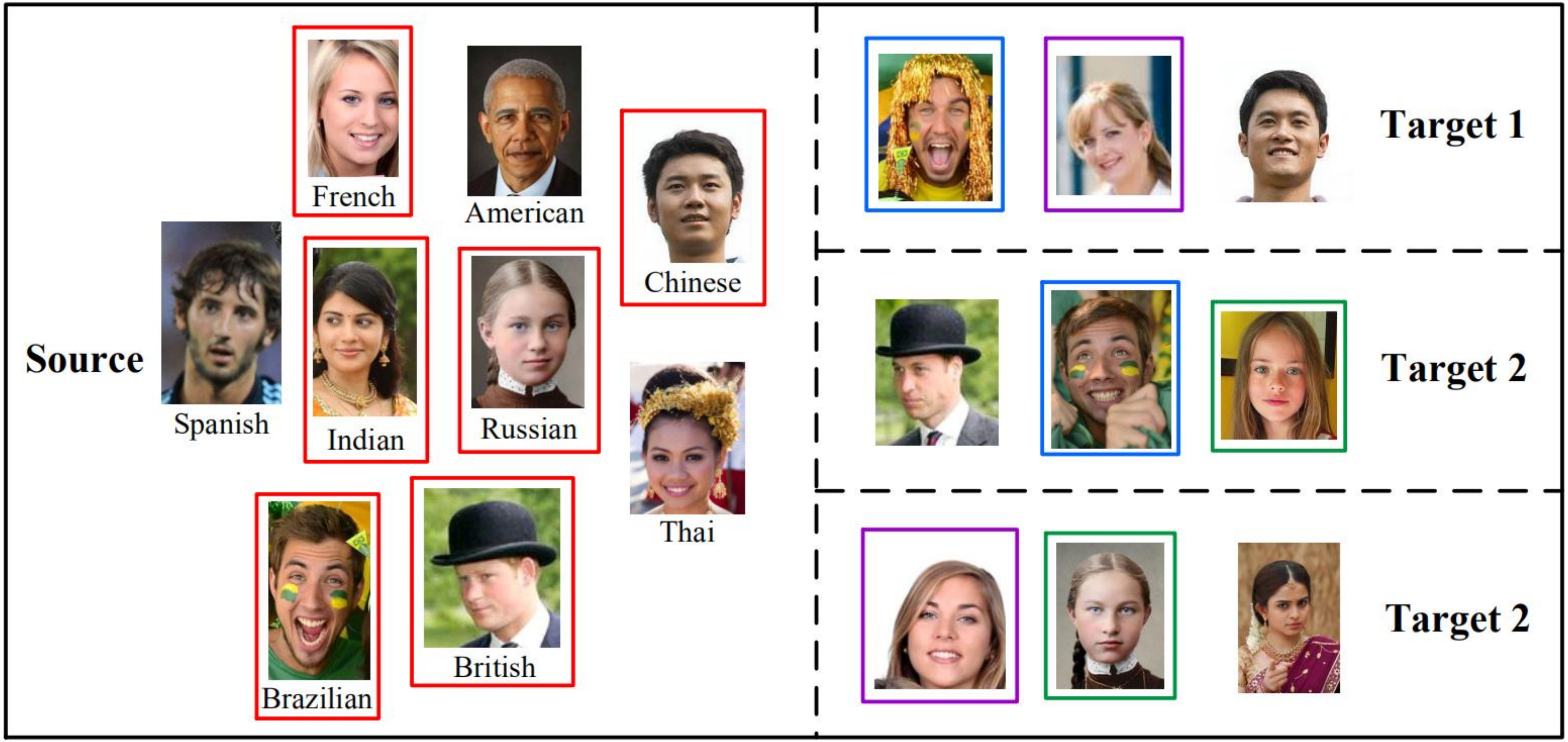
arXiv:1809.00852

# Introduction




源域和目标域往往属于同一类任务，但是分布不同。

# Introduction




(c) One Source Domain and Multiple Target Domains

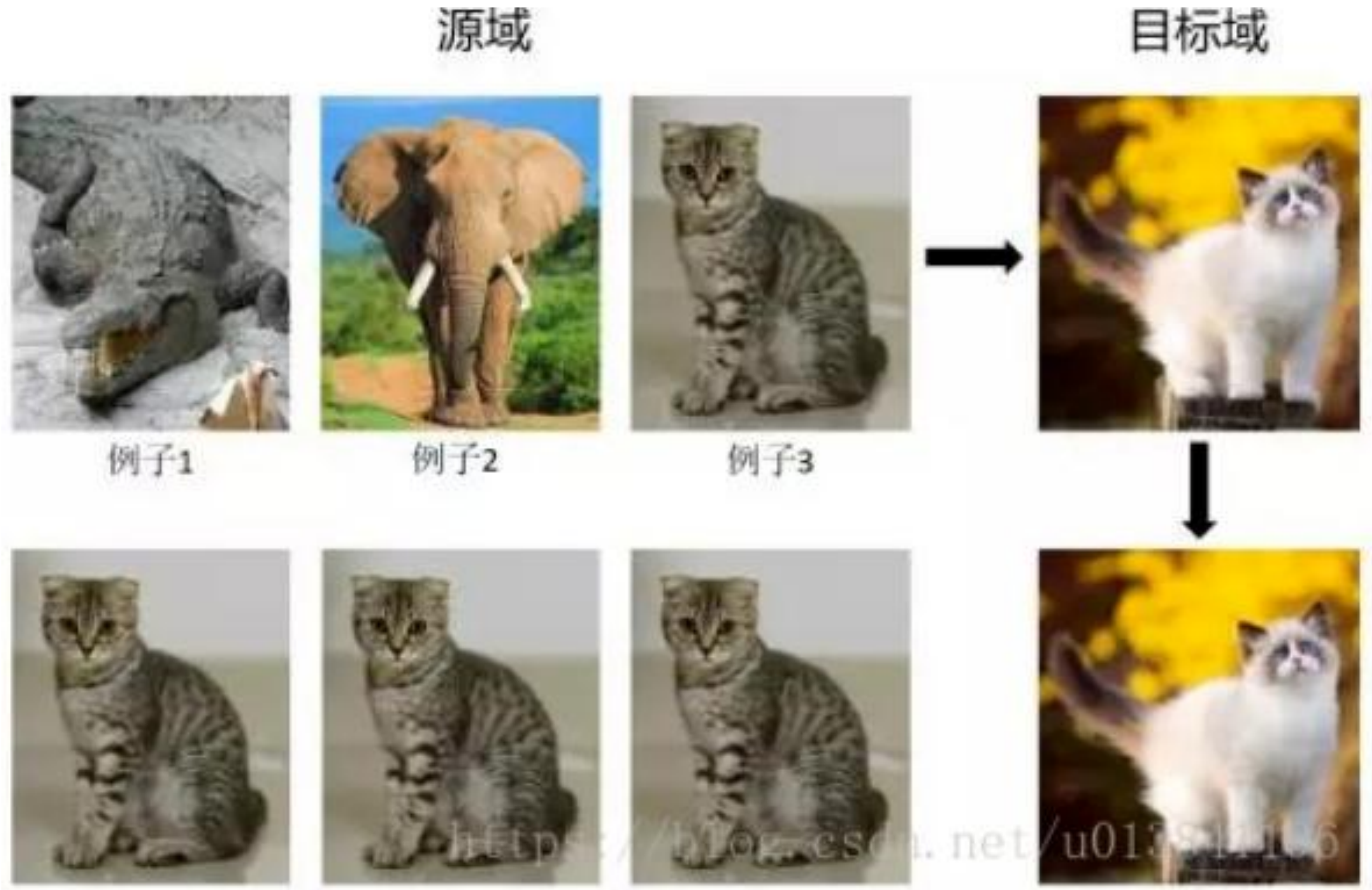
# Introduction

Obviously, there exist some implicit connections among three target domains, which is different from the previously mentioned scenarios.

Since mSmT is essentially the combination of mS1T and 1SmT, this paper mainly focuses on 1SmT.

# Introduction

Instance based TL

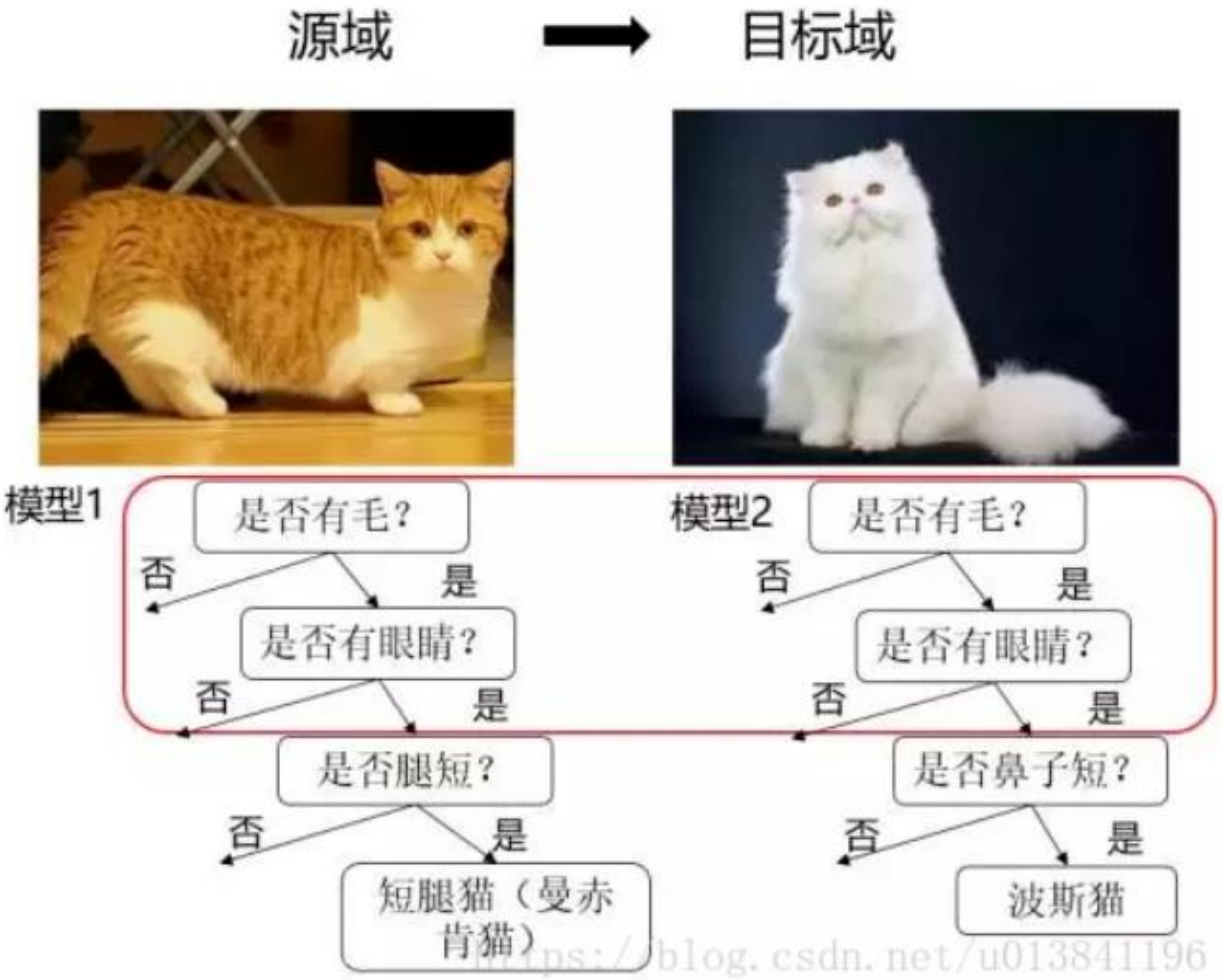


# Introduction

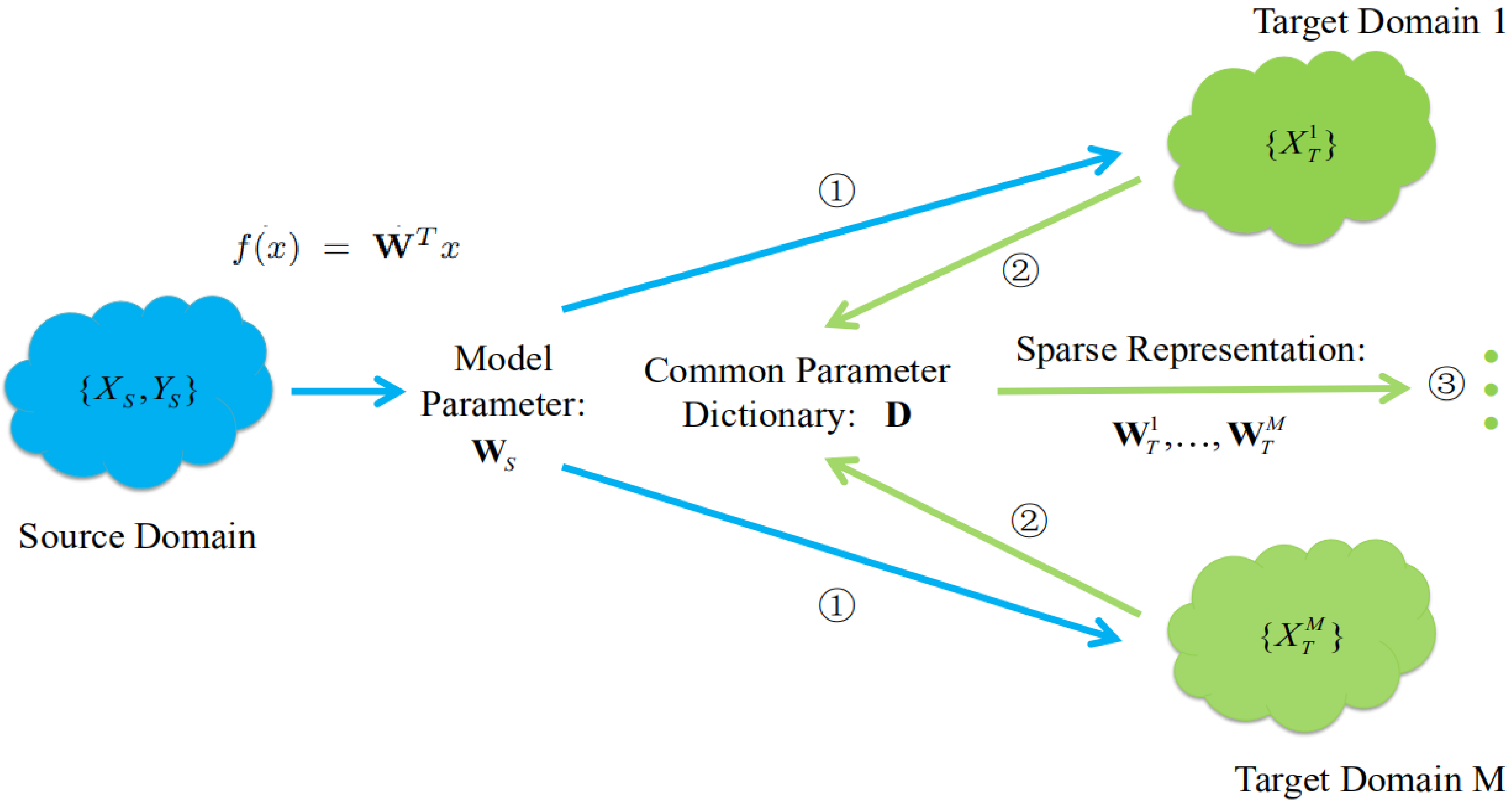
Feature based TL

# Introduction

Parameter based TL



# Framework



Target Domain 1
$\{X_T^1\}$
①
$f(x) = \mathbf{W}^T x$
②
$\{X_S, Y_S\}$
Model
Parameter:
$\mathbf{W}_S$
Common Parameter
Dictionary: $\mathbf{D}$
Sparse Representation:
③
$\mathbf{W}_T^1, \ldots, \mathbf{W}_T^M$
Source Domain
②
①
$\{X_T^M\}$
Target Domain M

# Framework

For the j-th target domain,we adopt SLMC to train the target model $\boldsymbol{W}_T^j$

**Soft Larget Margin Clustering**

$$\min_{u_{ki}^j, \boldsymbol{W}_T^j} \frac{1}{2}||\boldsymbol{W}_T^j||_F^2 + \frac{\lambda}{2}\sum_{k=1}^{C_T^j}\sum_{i=1}^{n_t^j}(u_{ki}^j)^2||(\boldsymbol{W}_T^j)^T x_i^j - l_k^j||_2^2$$

$$s.t. \sum_{k=1}^{C_t^j} u_{ki}^j = 1, 0 \leq u_{ki}^j \leq 1, \forall k = 1 \cdots C_T^j, i = 1 \cdots n_t^j$$

$u_{ki}^j \in [0,1]$ the soft membership of $x_i^j$ to the $k$th cluster

$\{l_1^j ... l_{C_T^j}^j\}$ the given encodings for the $C_T^j$ cluster respectively, and each $l_k^j = [0, ..., 0, 1, 0, ..., 0]^T \in \mathbb{R}^{C_T^j}$ corresponding to the $k$-th cluster

# Framework

*Learning from Source Domain to Target Domain*

$$\min_{\mathbf{W}_T^j, \mathbf{V}_j} \|\mathbf{W}_S - \mathbf{W}_T^j \mathbf{V}^j\|_F^2 + \eta \|\mathbf{V}^j\|_{2,1}$$

**dictionary of source model parameter** ($\mathbf{W}_T^j$)

**sparse representation of $W_S$** ($\mathbf{V}^j$)

$$\boldsymbol{W}_S \in \mathbb{R}^{d \times C_S}, \boldsymbol{W}_T^j \in \mathbb{R}^{d \times C_T^j}, \boldsymbol{V}^j \in \mathbb{R}^{C_T^j \times C_S}$$

$$\|\mathbf{M}\|_{2,1} = \sum_{i=1}^{n} \|\mathbf{m}^i\|_2$$

# Framework

*Learning among Multiple target Domains*

$$\min_{\mathbf{D},\mathbf{V}_T^j} \|\mathbf{W}_T^j - \mathbf{D}\mathbf{V}_T^j\|_F^2 + \eta\|\mathbf{V}_T^j\|_{2,1}$$

$$\boldsymbol{D} \in \mathbb{R}^{r \times C_T^j}$$

$\mathbf{D}$ common dictionary of multiple target model parameters

# Framework

*Unified Objective Function*

$$\min_{u_{ki}^j, \boldsymbol{W}_T^j, \boldsymbol{D}, \boldsymbol{V}^j, \boldsymbol{V}_T^j} \sum_{j=1}^{M} \left\{ \frac{1}{2}||\boldsymbol{W}_T^j||_F^2 + \frac{\lambda}{2} \sum_{k=1}^{C_T^j} \sum_{i=1}^{n_t^j} (u_{ki}^j)^2 ||(\boldsymbol{W}_T^j)^T x_i^j - l_k^j||_2^2 + \frac{\beta}{2}||\boldsymbol{W}_S - \boldsymbol{W}_T^j \boldsymbol{V}^j||_F^2 \right.$$

$$\left. + \frac{\gamma}{2}||\boldsymbol{W}_T^j - \boldsymbol{D}\boldsymbol{V}_T^j||_F^2 + \eta(||\boldsymbol{V}^j||_{2,1} + ||\boldsymbol{V}_T^j||_{2,1}) \right\}$$

$$s.t. \sum_{k=1}^{C_t^j} u_{ki}^j = 1, 0 \leq u_{ki}^j \leq 1, \forall k = 1 \cdots C_T^j, i = 1 \cdots n_t^j, j = 1 \cdots M$$

not-convex w.r.t joint$(u_{ki}^j, \boldsymbol{W}_T^j, \boldsymbol{D}, \boldsymbol{V}^j, \boldsymbol{V}_T^j)$but block convex

# Optimization

Tseng P. Convergence of a block coordinate descent method for non-differentiable minimization. Journal of optimization theory and ap-plications, 2001, 109(3): 475-494.

**An iterative strategy:**

(1)updating $u_{ki}^{j}$ with fixed $\boldsymbol{W}_T^j, \boldsymbol{D}, \boldsymbol{V}^j, \boldsymbol{V}_T^j$;

(2)updating $\boldsymbol{W}_T^j$ and $\boldsymbol{D}$ with fixed $u_{ki}^{j}, \boldsymbol{V}^j, \boldsymbol{V}_T^j$;

(3)updating $\boldsymbol{V}^j$ and $\boldsymbol{V}_T^j$ with fixed $\boldsymbol{W}_T^j, \boldsymbol{D}, u_{ki}^{j}$;

# Optimization

$$u_{ki}^{j} = \frac{\|f^{j}(x_i^j) - l_k\|_2^{-2}}{\sum_{r=1}^{C_T^j} \|(f^{j}(x_i^j) - l_r\|_2^{-2}} \quad (8)$$

$$\left(\mathbf{I} + \lambda \sum_{k=1}^{C_T^j} \mathbf{X}^j \hat{\mathbf{U}}_k (\mathbf{X}^j)^T + \gamma \mathbf{I}\right) \mathbf{W}_T^j + \mathbf{W}_T^j \left(\beta \mathbf{V}^j (\mathbf{V}^j)^T\right)$$
$$= \lambda \sum_{k=1}^{C_T^j} \mathbf{X}^j \hat{\mathbf{U}}_k (\mathbf{L}_k^j)^T + \beta \mathbf{W}_S (\mathbf{V}^j)^T + \gamma \mathbf{D} \mathbf{V}_T^j \quad (11)$$

$\mathbf{AX} + \mathbf{XB} = \mathbf{Q}$ can directly be solved by python function: *scipy.linalg._solvers.solve_sylvester*$(\mathbf{A}, \mathbf{B}, \mathbf{Q})$.

# Optimization

$$\mathbf{D} = \sum_{j=1}^{M} \mathbf{W}_T^j (\mathbf{V}_T^j)^T \left( \sum_{j=1}^{M} \mathbf{V}_T^j (\mathbf{V}_T^j)^T \right)^{-1} \quad (13)$$

$$\mathbf{V}^j = \left( \beta (\mathbf{W}_T^j)^T \mathbf{W}_T^j + 2\eta \mathbf{M}^j \right)^{-1} \beta (\mathbf{W}_T^j)^T \mathbf{W}_S \quad (17)$$

$$\mathbf{V}_T^j = \left( \gamma \mathbf{D}^T \mathbf{D} + 2\eta \mathbf{M}_T^j \right)^{-1} \gamma \mathbf{D}^T \mathbf{W}_T^j \quad (18)$$

# Optimization

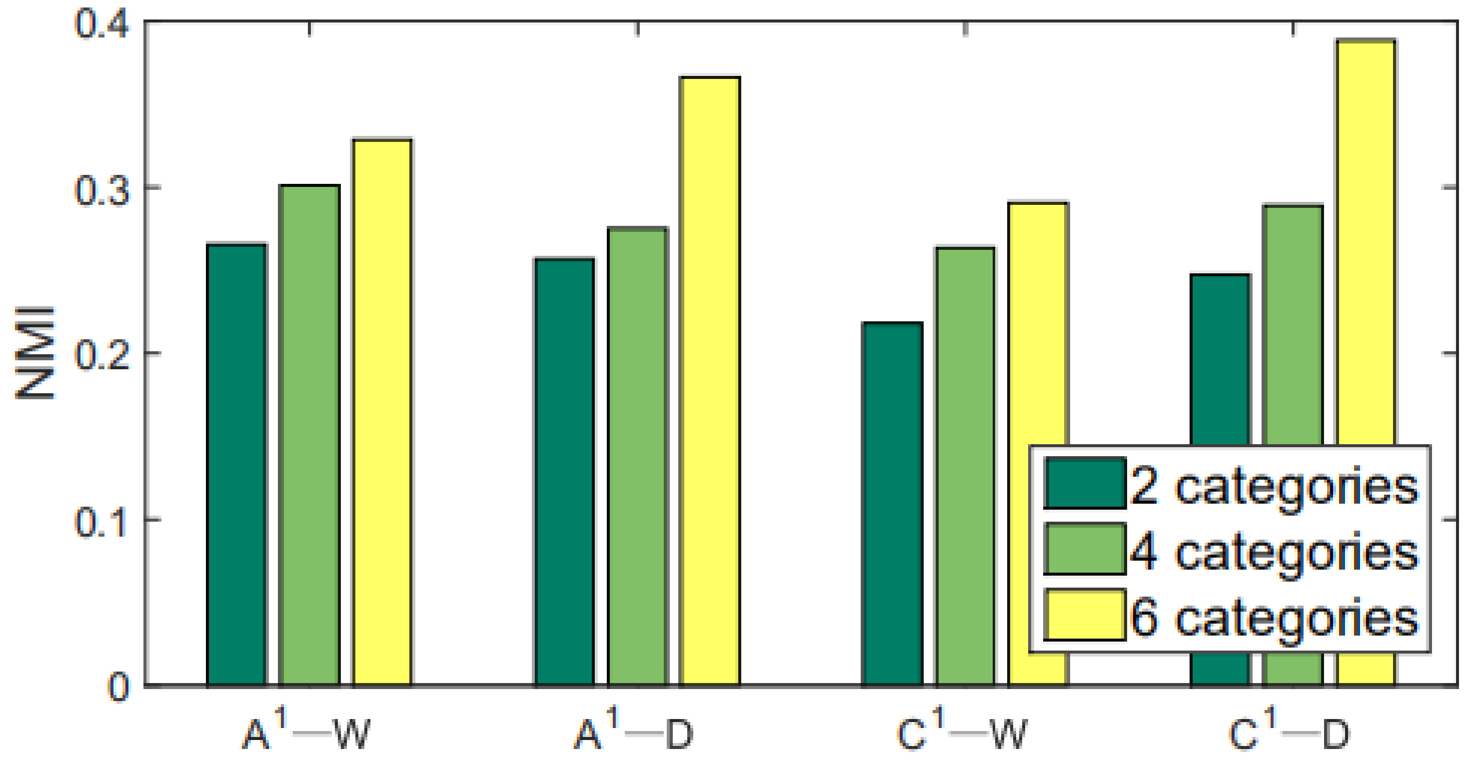
**Theorem 1:** The objective function value shown in Eq.(4) monotonically decreases until convergence by applying the proposed algorithm.

# Experiments

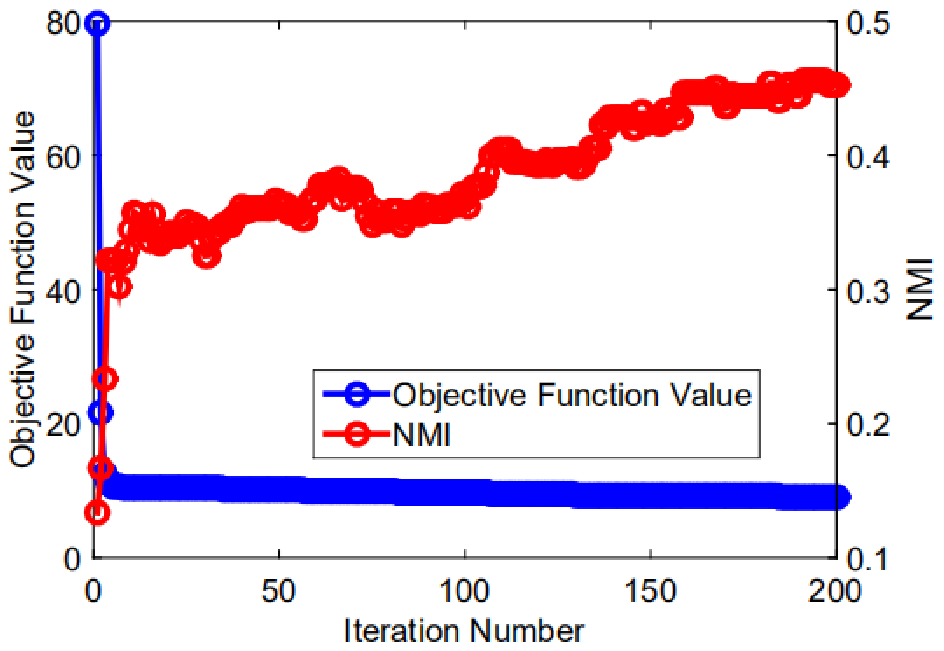
TABLE 5
The Result of Two Representative Scenarios on Three Target Domains

| | $D_S$ | $D_T$ | NMI | Increme-ntal NMI | RI | Increme-ntal NMI |
|---|---|---|---|---|---|---|
| $\bigcap_{j=1}^{M} \mathcal{C}_T^j = 0$ $\wedge \mathcal{C}_T^j \cap \mathcal{C}_T^k = \mathcal{C}_{c0}$ | PIE05 | PIE07[2] | 0.4904 | **5.88** ↑ | 0.8636 | **1.31** ↑ |
| | | PIE09[1] | 0.4906 | **3.62** ↑ | 0.8676 | **1.59** ↑ |
| | | PIE29[3] | 0.5069 | - | 0.878 | - |
| | PIE27 | PIE07[2] | 0.4697 | **2.04** ↑ | 0.8524 | 0.05 ↓ |
| | | PIE09[1] | 0.4372 | **1.48** ↑ | 0.8532 | **1.17** ↑ |
| | | PIE29[3] | 0.4556 | - | 0.8641 | - |
| $\bigcap_{j=1}^{M} \mathcal{C}_T^j = \mathcal{C}_c$ | PIE05 | PIE07[2] | 0.4605 | **2.89** ↑ | 0.8555 | **0.5** ↑ |
| | | PIE09[1] | 0.5248 | **7.04** ↑ | 0.8775 | **2.58** ↑ |
| | | PIE29[4] | 0.437 | - | 0.8508 | - |
| | PIE27 | PIE07[2] | 0.4603 | **1.1** ↑ | 0.8586 | **0.57** ↑ |
| | | PIE09[1] | 0.4529 | **3.05** ↑ | 0.8486 | **0.71** |
| | | PIE29[4] | 0.4814 | - | 0.866 | - |

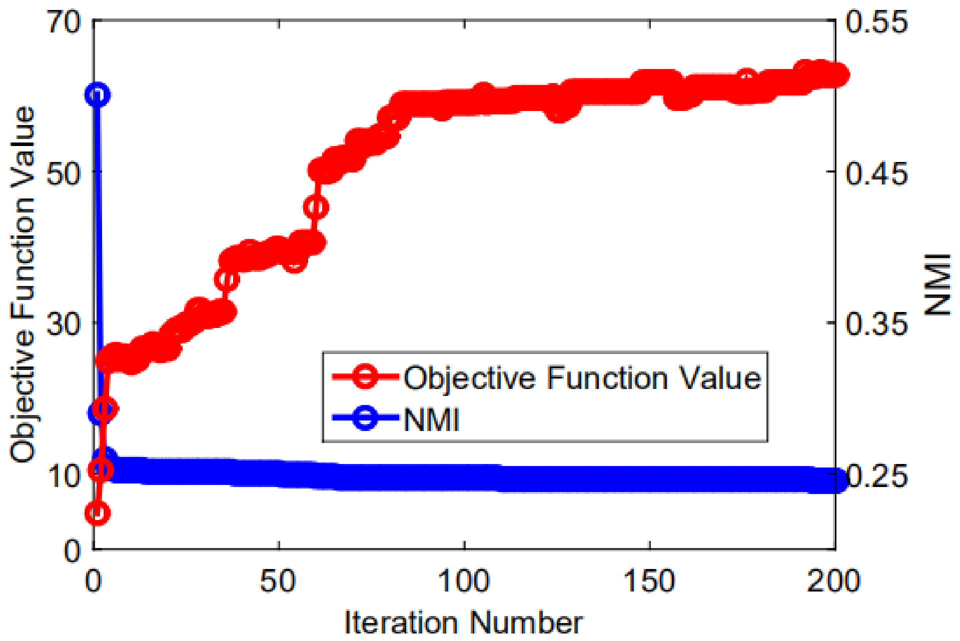
# Experiments



# Experiments




(a)




(b)