Multi-source Domain Adaptation

kongyiji

January 2021

1 Problem Formulation

$$\min \frac{\{\text{domain shift }\}_{\text{marginal}} + \{\text{domain shift}\}_{\text{conditional}} + \{\text{dist}\}_{\text{intra}}}{\{\text{dist}\}_{\text{inter}}} \tag{1}$$

1.1 Ditribution Matching and Landmark Selection

The model considers both marginal and conditional distribution discrepancy, denoted as E_{MG} and E_{CD},which can be formulated as follows:

$$\begin{array}{l} \min_{A,B} E_{\mathrm{MG}}\left(\alpha, \beta, X_{\mathrm{s}}, X_{\mathrm{t}}, A, B\right) + E_{\mathrm{CD}}\left(\alpha, \beta, X_{\mathrm{s}}, X_{\mathrm{t}}, A, B\right) \\ \text{s.t. } \ \{\alpha_{ui}^{c}, \beta_{i}^{c}\} \in [0,1], \frac{{\alpha_{\mathrm{u}}^{\mathrm{c}}}^{\mathrm{T}} \mathbf{1}_{n_{\mathrm{s}}^{\mathrm{u}}}}{n_{\mathrm{s}}^{\mathrm{u}c}} = \delta_{\mathrm{s}}^{\mathrm{u}}, \frac{{\beta^{c}}^{\mathrm{T}} \mathbf{1}_{n_{\mathrm{t}}^{c}}}{n_{\mathrm{t}}^{c}} = \delta_{\mathrm{t}} \end{array} \tag{2}$$

where $\alpha_{\mathrm{u}} = \left[\alpha_{\mathrm{u}}^{1}; \cdots; \alpha_{\mathrm{u}}^{c}; \cdots; \alpha_{\mathrm{u}}^{C}\right] \in R^{n_{\mathrm{s}}^{\mathrm{u}}}$ are the weights of samples in source domain $X_{\mathrm{s}}^{\mathrm{u}} \in R^{d_{\mathrm{s}}^{\mathrm{u}} \times n_{\mathrm{s}}^{\mathrm{u}}}$,$\alpha = [\alpha_1; \cdots; \alpha_{\mathrm{u}}] \in R^{n_s}$ are the weights of samples in source domain $X_{\mathrm{s}} = \left[X_{\mathrm{s}}^{1}; X_{\mathrm{s}}^{2}; \ldots; X_{\mathrm{s}}^{\mathrm{N_s}}\right] \in R^{d_s \times n_s}$. $\beta = \left[\beta^{1}; \cdots; \beta^{c}; \cdots; \beta^{C}\right] \in R^{n_t}$ are the weights of data in the source domain and the target domain, respectively, $\alpha_{\mathrm{u}}^{c} = \left[\alpha_{\mathrm{u}1}^{\;c}; \cdots; \alpha_{\mathrm{u}n_{\mathrm{s}}^{\mathrm{u}c}}^{\;c}\right], \beta^{c} = \left[\beta_1^{c}; \cdots; \beta_{n^c}^{c}\right], \mathbf{1}_{n_{\mathrm{s}}^{\mathrm{u}c}} \in R^{n_{\mathrm{s}}^{\mathrm{u}c}}$ and $\mathbf{1}_{n_{\mathrm{t}}^{c}} \in R^{n_{\mathrm{t}}^{c}}$ are column vectors with all ones. $\delta_{\mathrm{s}}^{\mathrm{u}}, \delta_t \in [0,1]$ controls the ratio of landmarks in the whole source or target domain samples. The constraints on α and β keep them from trivial solutions such as one-hot vectors that only align one sample from source and one sample from target. Then, E_{MG} and E_{CD} in Eq.(2) further can be calculated by:

$$E_{\mathrm{MG}} = \sum_{u=1}^{\mathrm{N_s}} \left\| \frac{1}{\delta_{\mathrm{s}}^{\mathrm{u}} n_{\mathrm{s}}^{\mathrm{u}}} \sum_{i=1}^{n_{\mathrm{s}}^{\mathrm{u}}} \alpha_{\mathrm{ui}} A^{\mathrm{T}} x_{\mathrm{s}}^{\mathrm{u}i} - \frac{1}{\delta_{\mathrm{t}} n_{\mathrm{t}}} \sum_{j=1}^{n_{\mathrm{t}}} \beta_{\mathrm{j}} B^{\mathrm{T}} x_{\mathrm{t}}^{j} \right\|^2$$

$$= \textstyle\sum_{u=1}^{\mathrm{N_s}} \left\| \frac{1}{\delta_{\mathrm{s}}^{\mathrm{u}} n_{\mathrm{s}}^{\mathrm{u}}} A^{\mathrm{T}} \left[\begin{array}{c} \mathrm{x}_{\mathrm{u}1} \mathrm{x}_{\mathrm{u}2} \cdots \mathrm{x}_{n_{\mathrm{s}}^{\mathrm{u}}} \end{array}\right]_{1 \times n_{\mathrm{s}}^{\mathrm{u}}} \left[\begin{array}{c} \alpha_{\mathrm{u}1} \\ \alpha_{\mathrm{u}2} \\ \vdots \\ \alpha_{n_{\mathrm{s}}^{\mathrm{u}}} \end{array}\right]_{n_{\mathrm{s}}^{\mathrm{u}} \times 1} - \frac{1}{\delta_t n_t} B^{\mathrm{T}} \left[\begin{array}{c} \mathrm{x}_1 \mathrm{x}_2 \cdots \mathrm{x}_{n_t} \end{array}\right]_{1 \times n_t} \left[\begin{array}{c} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{n_t} \end{array}\right]_{n_t \times 1} \right\|$$

$$= \textstyle\sum_{u=1}^{\mathrm{N_s}} \operatorname{tr}\left(\frac{1}{{\delta_{\mathrm{s}}^{\mathrm{u}}}^2 {n_{\mathrm{s}}^{\mathrm{u}}}^2} A^{\mathrm{T}} X_{\mathrm{s}}^{\mathrm{u}} \alpha_{\mathrm{u}} \left(A^{\mathrm{T}} X_{\mathrm{s}}^{\mathrm{u}} \alpha_{\mathrm{u}}\right)^{\mathrm{T}} + \frac{1}{\delta_{\mathrm{t}}^2 n_t^2} B^{\mathrm{T}} X_t \beta \left(B^{\mathrm{T}} X_t \beta\right)^{\mathrm{T}} - \frac{1}{\delta_{\mathrm{s}}^{\mathrm{u}} \delta_{\mathrm{t}} n_{\mathrm{s}}^{\mathrm{u}} n_t} A^{\mathrm{T}} X_{\mathrm{s}}^{\mathrm{u}} \alpha_{\mathrm{u}} \left(B^{\mathrm{T}} X_t \beta\right)^{\mathrm{T}} \right.$$
$$\left. - \frac{1}{\delta_{\mathrm{s}}^{\mathrm{u}} \delta_{\mathrm{t}} n_{\mathrm{s}}^{\mathrm{u}} n_t} B^{\mathrm{T}} X_t \beta \left(A^{\mathrm{T}} X_{\mathrm{s}}^{\mathrm{u}} \alpha_{\mathrm{u}}\right)^{\mathrm{T}} \right)$$

$$=\sum_{u=1}^{\mathrm{N_s}} \operatorname{tr}\left(\frac{1}{\delta_{\mathrm{s}}^{\mathrm{u}\,2} n_{\mathrm{s}}^{\mathrm{u}\,2}} A^{\mathrm{T}} X_{\mathrm{s}}^{\mathrm{u}} \alpha_{\mathrm{u}} \alpha_{\mathrm{u}}^{\mathrm{T}} X_{\mathrm{s}}^{\mathrm{u}\,\mathrm{T}} A+\frac{1}{\delta_{\mathrm{t}} n_{t}^{2}} B^{\mathrm{T}} X_{t} \beta \beta^{\mathrm{T}} X_{t}^{\mathrm{T}} B-\frac{1}{\delta_{\mathrm{s}}^{\mathrm{u}} \delta_{\mathrm{t}} n_{\mathrm{s}}^{\mathrm{u}} n_{t}} A^{\mathrm{T}} X_{\mathrm{s}}^{\mathrm{u}} \alpha_{\mathrm{u}} \beta^{\mathrm{T}} X_{t}^{\mathrm{T}}-\frac{1}{\delta_{\mathrm{s}}^{\mathrm{u}} \delta_{\mathrm{t}} n_{\mathrm{s}}^{\mathrm{u}} n_{t}} B^{\mathrm{T}} X_{t} \beta \alpha_{\mathrm{u}}^{\mathrm{T}} X_{\mathrm{s}}^{\mathrm{u}\,\mathrm{T}}\right)$$

$$= \operatorname{tr}\left(A^{T}\left(\sum_{u=1}^{\mathrm{N_s}} X_{\mathrm{s}}^{\mathrm{u}} H_{sm}^{\mathrm{u}} X_{\mathrm{s}}^{\mathrm{u}\,T}\right) A+B^{T}\left(\sum_{u=1}^{\mathrm{N_s}} X_{t} H_{tm} X_{t}^{T}\right) B-A^{T}\left(\sum_{u=1}^{\mathrm{N_s}} X_{\mathrm{s}}^{\mathrm{u}} H_{stm}^{\mathrm{u}} X_{t}^{T}\right) B\right.$$

$$\left.-B^{T}\left(\sum_{u=1}^{\mathrm{N_s}} X_{t} H_{stm}^{\mathrm{u}}{}^{T} X_{\mathrm{s}}^{\mathrm{u}\,T}\right) A\right)$$

$$\mathrm{E_{CD}} = \sum_{u=1}^{\mathrm{N_s}} \sum_{c=1}^{C} \left\| \frac{1}{\delta_{\mathrm{s}}^{\mathrm{u}\,c} n_{\mathrm{s}}^{\mathrm{u}\,c}} \sum_{i=1}^{n_{\mathrm{s}}^{\mathrm{u}\,c}} \alpha_{\mathrm{ui}}^{\mathrm{c}} A^{\mathrm{T}} x_{\mathrm{s}}^{i,c} - \frac{1}{\delta_{\mathrm{t}}^{c} n_{\mathrm{t}}^{c}} \sum_{j=1}^{n_{\mathrm{t}}^{c}} \beta_{\mathrm{i}}^{\mathrm{c}} B^{\mathrm{T}} x_{\mathrm{t}}^{j,c} \right\|^{2}$$

$$=\sum_{c=1}^{C} \operatorname{tr}\left(A^{T}\left(\sum_{u=1}^{\mathrm{N_s}} X_{\mathrm{s}}^{\mathrm{u}} H_{sc}^{\mathrm{u}} X_{\mathrm{s}}^{\mathrm{u}\,T}\right) A+B^{T}\left(\sum_{u=1}^{\mathrm{N_s}} X_{t} H_{tc} X_{t}^{T}\right) B-A^{T}\left(\sum_{u=1}^{\mathrm{N_s}} X_{\mathrm{s}}^{\mathrm{u}} H_{stc}^{\mathrm{u}} X_{t}^{T}\right) B\right.$$

$$\left.-B^{T}\left(\sum_{u=1}^{\mathrm{N_s}} X_{t} H_{stc}^{\mathrm{u}}{}^{T} X_{\mathrm{s}}^{\mathrm{u}\,T}\right) A\right)$$

where

$$H_{sm}^{\mathrm{u}} = \frac{1}{\delta_{\mathrm{us}}^{2} n_{\mathrm{s}}^{\mathrm{u}\,2}} \alpha_{\mathrm{u}} \cdot \alpha_{\mathrm{u}}^{T}, H_{tm} = \frac{1}{\delta_{\mathrm{t}}^{2} n_{t}^{2}} \beta \cdot \beta^{T}, H_{stm}^{\mathrm{u}} = \frac{1}{\delta_{\mathrm{us}} \delta_{\mathrm{t}} n_{\mathrm{s}}^{\mathrm{u}} n_{t}} \alpha_{\mathrm{u}} \cdot \beta^{T},$$

$$H_{sc}^{\mathrm{u}\ \mathrm{c}} = \frac{1}{\delta_{\mathrm{us}}^{2} n_{\mathrm{s}}^{\mathrm{u} c 2}} \alpha_{\mathrm{u}}^{\mathrm{c}} \cdot \alpha_{\mathrm{u}}^{c^{T}}, H_{tc}^{\mathrm{c}} = \frac{1}{\delta_{\mathrm{us}}^{2} n_{t}^{c 2}} \beta^{\mathrm{c}} \cdot \beta^{\mathrm{c}^{T}}, H_{stc}^{\mathrm{u}\ \mathrm{c}} = \frac{1}{\delta_{\mathrm{us}} \delta_{\mathrm{t}} n_{\mathrm{s}}^{\mathrm{u} c} n_{t}^{\mathrm{c}}} \alpha_{\mathrm{u}}^{\mathrm{c}} \cdot \beta^{\mathrm{c}^{T}},$$

After some algebra operations, Eq(2) can be written as the following equivalent equation:

$$\begin{aligned} & A^{T} M_{ss} A + B^{T} M_{tt} B - A^{T} M_{st} B - B^{T} M_{ts} A \\ =& A^{T}\left(\sum_{u=1}^{\mathrm{N_s}} X_{\mathrm{s}}^{\mathrm{u}}\left(H_{sm}^{\mathrm{u}}+H_{sc}^{\mathrm{u}}\right) X_{\mathrm{s}}^{\mathrm{u}\,T}\right) A+B^{T}\left(\sum_{u=1}^{\mathrm{N_s}} X_{t}\left(H_{tm}+H_{tc}\right) X_{t}^{T}\right) B \\ & -A^{T}\left(\sum_{u=1}^{\mathrm{N_s}} X_{\mathrm{s}}^{\mathrm{u}}\left(H_{stm}^{\mathrm{u}}+H_{stc}^{\mathrm{u}}\right) X_{t}^{T}\right) B-B^{T}\left(\sum_{u=1}^{\mathrm{N_s}} X_{t}\left(H_{tsm}^{\mathrm{u}}+H_{tsc}^{\mathrm{u}}\right)^{T} X_{\mathrm{s}}^{\mathrm{u}\,T}\right) A \end{aligned} \tag{3}$$

$$\begin{aligned} M_{ss} &= \left(\sum_{u=1}^{\mathrm{N_s}} X_{\mathrm{s}}^{\mathrm{u}}\left(H_{sm}^{\mathrm{u}}+H_{sc}^{\mathrm{u}}\right) X_{\mathrm{s}}^{\mathrm{u}\,T}\right) \\ M_{tt} &= \left(\sum_{u=1}^{\mathrm{N_s}} X_{t}\left(H_{tm}+H_{tc}\right) X_{t}^{T}\right) \\ M_{st} &= -\left(\sum_{u=1}^{\mathrm{N_s}} X_{\mathrm{s}}^{\mathrm{u}}\left(H_{stm}^{\mathrm{u}}+H_{stc}^{\mathrm{u}}\right) X_{t}^{T}\right) \\ M_{ts} &= M_{st}^{T} \end{aligned} \tag{4}$$

At last, the above equation(3), can be further transformed to its matrix form as follows:

$$\operatorname{Tr}\left(\left[\begin{array}{ll} A^{\mathrm{T}} & B^{\mathrm{T}} \end{array}\right]\left[\begin{array}{ll} M_{\mathrm{ss}} & M_{\mathrm{st}} \\ M_{\mathrm{ts}} & M_{\mathrm{tt}} \end{array}\right]\left[\begin{array}{l} A \\ B \end{array}\right]\right) \tag{5}$$

1.2 Structure Preservation

(a) Construct the intrinsic weight matrix W_{w} : For each sample x, connect the nearest neighbor pair v and x if v has the same label information with x. (b) Construct the penalty weight matrix W_{b} : For each domain, connect the k -nearest vertex pairs where samples in each pair belong to different classes.

$$
\begin{aligned}
\min \sum_{u=1}^{\mathrm{N_s}} \frac{\operatorname{Tr}\left(A^{\mathrm{T}} X_{\mathrm{s}}^{\mathrm{u}} L_{\mathrm{b}}^{\mathrm{us}} X_{\mathrm{s}}^{\mathrm{u}\mathrm{T}} A\right)}{\operatorname{Tr}\left(A^{\mathrm{T}} X_{\mathrm{s}}^{\mathrm{u}} L_{\mathrm{w}}^{\mathrm{us}} X_{\mathrm{s}}^{\mathrm{u}\mathrm{T}} A\right)} &= \min \frac{\operatorname{Tr}\left(A^{\mathrm{T}} S_{\mathrm{w}}^{\mathrm{s}} A\right)}{\operatorname{Tr}\left(A^{\mathrm{T}} S_{\mathrm{b}}^{\mathrm{s}} A\right)} \\
\min \frac{\operatorname{Tr}\left(B^{\mathrm{T}} X_{\mathrm{t}} L_{\mathrm{w}}^{\mathrm{t}} X_{\mathrm{t}}^{\mathrm{T}} B\right)}{\operatorname{Tr}\left(B^{\mathrm{T}} X_{\mathrm{t}} L_{\mathrm{b}}^{\mathrm{t}} X_{\mathrm{t}}^{\mathrm{T}} B\right)} &= \min \frac{\operatorname{Tr}\left(B^{\mathrm{T}} S_{\mathrm{w}}^{\mathrm{t}} B\right)}{\operatorname{Tr}\left(B^{\mathrm{T}} S_{\mathrm{b}}^{\mathrm{t}} B\right)}
\end{aligned}
\tag{6}
$$

where

$$
\begin{gathered}
S_{\mathrm{b}}^{\mathrm{s}} = \sum_{u=1}^{\mathrm{N_s}} X_{\mathrm{s}}^{\mathrm{u}} L_{\mathrm{b}}^{\mathrm{us}} X_{\mathrm{s}}^{\mathrm{u}\mathrm{T}}, \quad S_{\mathrm{w}}^{\mathrm{s}} = \sum_{u=1}^{\mathrm{N_s}} X_{\mathrm{s}}^{\mathrm{u}} L_{\mathrm{w}}^{\mathrm{us}} X_{\mathrm{s}}^{\mathrm{u}\mathrm{T}} \\
S_{\mathrm{b}}^{\mathrm{t}} = X_{\mathrm{t}} L_{\mathrm{b}}^{\mathrm{t}} X_{\mathrm{t}}^{\mathrm{T}}, \quad S_{\mathrm{w}}^{\mathrm{t}} = X_{\mathrm{t}} L_{\mathrm{w}}^{\mathrm{t}} X_{\mathrm{t}}^{\mathrm{T}}
\end{gathered}
$$

$$
\min_{A,B} \frac{\operatorname{Tr}\left(\left[A^{\mathrm{T}} B^{\mathrm{T}}\right]\left[\begin{array}{cc} M_{\mathrm{ss}}+\gamma S_{\mathrm{w}}^{\mathrm{s}} & M_{\mathrm{st}} \\ M_{\mathrm{ts}} & M_{\mathrm{tt}}+\gamma S_{\mathrm{w}}^{\mathrm{t}}+\mu I \end{array}\right]\left[\begin{array}{c} A \\ B \end{array}\right]\right)}{\operatorname{Tr}\left(\left[\begin{array}{cc} A^{\mathrm{T}} & B^{\mathrm{T}} \end{array}\right]\left[\begin{array}{cc} \gamma S_{\mathrm{b}}^{\mathrm{s}} & \mathbf{0} \\ \mathbf{0} & \gamma S_{\mathrm{b}}^{\mathrm{t}}+\mu S_{\mathrm{h}}^{\mathrm{t}} \end{array}\right]\left[\begin{array}{c} A \\ B \end{array}\right]\right)}
\tag{7}
$$

where

$$
S_{\mathrm{h}}^{\mathrm{t}} = X_{\mathrm{t}}\left(I_{\mathrm{t}} - \frac{1}{n_{\mathrm{t}}} \mathbf{1}_{n_{\mathrm{t}}} \mathbf{1}_{n_{\mathrm{t}}}^{\mathrm{T}}\right) X_{\mathrm{t}}^{\mathrm{T}}
$$

is the covariance matrix of the target domain, to avoid projecting features into irrelevant dimensions, we encourage the variances of target domain is maximized in the respective subspaces.Tr $\left(B^{T} B\right)$ constraint is further imposed small to control the scale of B.γ and μ are trade-off parameters for the locality preserving term and the target variance term, respectively.

2 Problem Optimization

1) Optimizing the space mappings A and B : To optimize Eq. (7), we write $[A; B]$ as P. Thus, the objective function can be rewritten as:

$$
\min_{P} \frac{\operatorname{Tr}\left(P^{\mathrm{T}}\left[\begin{array}{cc} M_{\mathrm{ss}}+\gamma S_{\mathrm{w}}^{\mathrm{s}} & M_{\mathrm{st}} \\ M_{\mathrm{ts}} & M_{\mathrm{tt}}+\gamma S_{\mathrm{w}}^{\mathrm{t}}+\mu I \end{array}\right] P\right)}{\operatorname{Tr}\left(P^{\mathrm{T}}\left[\begin{array}{cc} \gamma S_{\mathrm{b}}^{\mathrm{s}} & \mathbf{0} \\ \mathbf{0} & \gamma_{\mathrm{b}}^{\mathrm{t}}+\mu S_{\mathrm{h}}^{\mathrm{t}} \end{array}\right] P\right)}
\tag{8}
$$

We can reformulate Eq. (8) as:

$$
\begin{aligned}
&\max_{P} \operatorname{Tr}\left(P^{\mathrm{T}}\left[\begin{array}{cc} \gamma S_{\mathrm{b}}^{\mathrm{s}} & \mathbf{0} \\ \mathbf{0} & \gamma S_{\mathrm{b}}^{\mathrm{t}}+\mu S_{\mathrm{h}}^{\mathrm{t}} \end{array}\right] P\right) \\
&\text{s.t. } \operatorname{Tr}\left(P^{\mathrm{T}}\left[\begin{array}{cc} M_{\mathrm{ss}}+\gamma S_{\mathrm{w}}^{\mathrm{s}} & M_{\mathrm{st}} \\ M_{\mathrm{ts}} & M_{\mathrm{tt}}+\gamma S_{\mathrm{w}}^{\mathrm{t}}+\mu I \end{array}\right] P\right) = 1
\end{aligned}
\tag{9}
$$

According to the constraincd optimization theory, we introduce a Lagrange multiplier Φ, and get the Lagrange function for Eq. (19) as follows:

$$
\begin{aligned}
\mathcal{L} = & \operatorname{Tr}\left(P^{\mathrm{T}}\left[\begin{array}{cc} \gamma S_{\mathrm{b}}^{\mathrm{s}} & 0 \\ 0 & \gamma S_{\mathrm{b}}^{\mathrm{u}}+\mu S_{\mathrm{h}}^{\mathrm{u}} \end{array}\right] P\right) \\
& -\operatorname{Tr}\left(\left(P^{\mathrm{T}}\left[\begin{array}{cc} M_{\mathrm{ss}}+\gamma S_{\mathrm{w}}^{\mathrm{s}} & M_{\mathrm{su}} \\ M_{\mathrm{s}}^{\mathrm{u}} & M_{\mathrm{uu}}+\gamma S_{\mathrm{w}}^{\mathrm{u}}+\mu I \end{array}\right] P-I\right) \Phi\right)
\end{aligned} \tag{10}
$$

where $\Phi = \operatorname{diag}(\phi_1, \cdots, \phi_d)$ and $(\phi_1, \cdots, \phi_d)$ are the d largest eigenvalues of the following eigendecomposition problem:

$$
\left[\begin{array}{cc} \gamma S_{\mathrm{b}}^{\mathrm{s}} & \mathbf{0} \\ \mathbf{0} & \gamma S_{\mathrm{b}}^{\mathrm{u}}+\mu S_{\mathrm{h}}^{\mathrm{u}} \end{array}\right] P=\left[\begin{array}{cc} M_{\mathrm{ss}}+\gamma S_{\mathrm{w}}^{\mathrm{s}} & M_{\mathrm{su}} \\ M_{\mathrm{s}}^{\mathrm{u}} & M_{\mathrm{uu}}+\gamma S_{\mathrm{w}}^{\mathrm{u}}+\mu I \end{array}\right] P \Phi \tag{11}
$$

As a result, P consists of the corresponding d largest eigenvectors of Eq. (11). At last, the subspaces spanned by A and B can be obtained easily once the transformation matrix P is obtained.

2) Optimizing sample weights α and β : Regarding A and B as constants, Since $\operatorname{Tr}(AB) = \operatorname{Tr}\left(A^T B^T\right)$ and $\operatorname{Tr}(constant) = constant$.The Eq(3) can be formulated as follows:

$$
\begin{aligned}
& \min_{\alpha^u, \beta} \textstyle\sum_{u=1}^{\mathrm{N_s}} \left(\frac{1}{2} \alpha_{\mathrm{u}}^{\mathrm{T}} K_{\mathrm{ss}}^{\mathrm{u}} \alpha_{\mathrm{u}}-\frac{1}{2} \alpha_{\mathrm{u}}^{\mathrm{T}} K_{\mathrm{st}}^{\mathrm{u}} \beta-\frac{1}{2} \beta^{\mathrm{T}} K_{\mathrm{ts}}^{\mathrm{u}} \alpha_{\mathrm{u}}+\frac{1}{2} \beta^{\mathrm{T}} K_{\mathrm{tt}} \beta\right) \\
& \qquad \text{s.t. } \left\{\alpha_{ui}^c, \beta_i^c\right\} \in[0,1], \frac{{\alpha_{\mathrm{u}}^{c}}^{\mathrm{T}} \mathbf{1}_{n^{c}{}_{\mathrm{s}}^{\mathrm{u}}}}{n_{\mathrm{s}}^{\mathrm{u}c}}=\delta_{\mathrm{s}}^{\mathrm{u}}, \frac{{\beta^{c}}^{\mathrm{T}} \mathbf{1}_{n_{\mathrm{t}}^{c}}}{n_{\mathrm{t}}^{c}}=\delta_{\mathrm{t}}
\end{aligned} \tag{12}
$$

where $\left(K_{\mathrm{ss}}^{\mathrm{u}}\right)_{i,j}$ in $K_{\mathrm{ss}}^{\mathrm{u}} \in R^{n_{us} \times n_{us}}$ is the coefficient associated with $\left(A^{\mathrm{T}} x_{\mathrm{s}}^{\mathrm{u}i}\right)^{\mathrm{T}} A^{\mathrm{T}} x_{\mathrm{s}}^{\mathrm{u}i}$, $\left(K_{\mathrm{st}}^{\mathrm{u}}\right)_{i,j}$ in $K_{\mathrm{st}}^{\mathrm{u}} \in R^{n_{\mathrm{s}}^{\mathrm{u}} \times n_{\mathrm{t}}}$ is the coefficient associated with $\left(A^{\mathrm{T}} x_{\mathrm{s}}^{\mathrm{u}i}\right)^{\mathrm{T}} B^{\mathrm{T}} x_{\mathrm{t}}^{j}$, and $\left(K_{\mathrm{tt}}\right)_{i,j}$ in $K_{\mathrm{tt}} \in R^{n_{\mathrm{t}} \times n_{\mathrm{t}}}$ is the coefficient associated with $\left(B^{\mathrm{T}} x_{\mathrm{t}}^{i}\right)^{\mathrm{T}} B^{\mathrm{T}} x_{\mathrm{t}}^{j}$.

With the above formulation, we can apply Quadratic Programming (QP) solvers to optimize the equivalent problem:

$$
\min_{z_i \in[0,1], V^{\mathrm{T}} \cdot Z=G} \frac{1}{2} Z^{\mathrm{T}} Q Z \tag{13}
$$

$$
\begin{aligned}
& Z=\left(\begin{array}{c} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_u \\ \beta \end{array}\right), Q=\left[\begin{array}{ccccc} K_{\mathrm{ss}}^1 & 0 & \cdots & 0 & -K_{\mathrm{st}}^1 \\ 0 & K_{\mathrm{ss}}^2 & \cdots & 0 & -K_{\mathrm{st}}^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & K_{\mathrm{ss}}^{\mathrm{u}} & -K_{\mathrm{st}}^{\mathrm{u}} \\ -K_{\mathrm{ts}}^1 & -K_{\mathrm{ts}}^2 & \cdots & -K_{\mathrm{ts}}^{\mathrm{u}} & \sum_{u=1}^{N_s} K_{\mathrm{tt}} \end{array}\right] \\
& G \in R^{(Ns+1)C \times 1} \text{ with } (G)_c=\left\{\begin{array}{ll} \delta_{\mathrm{s}}^{\mathrm{u}} n_{\mathrm{s}}^{\mathrm{u}c} & \text{if } (\mathrm{u}-1)\mathrm{C} \leq c \leq \mathrm{uC} \\ \delta_t n_{\mathrm{t}}^{c} & \text{if } \mathrm{c} > \mathrm{Ns} \times \mathrm{C} \end{array}\right. \\
& V=\left[\begin{array}{ccccc} V_{1\mathrm{s}} & \mathbf{0}_{n_{1\mathrm{s}} \times C} & \cdots & \mathbf{0}_{n_{1\mathrm{s}} \times C} & \mathbf{0}_{n_{1\mathrm{s}} \times C} \\ \mathbf{0}_{n_{2\mathrm{s}} \times C} & V_{2\mathrm{s}} & \cdots & \mathbf{0}_{n_{2\mathrm{s}} \times C} & \mathbf{0}_{n_{2\mathrm{s}} \times C} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{n_{\mathrm{s}}^{\mathrm{u}} \times C} & \mathbf{0}_{n_{\mathrm{s}}^{\mathrm{u}} \times C} & \cdots & V_{\mathrm{s}}^{\mathrm{u}} & \mathbf{0}_{n_{\mathrm{s}}^{\mathrm{u}} \times C} \\ \mathbf{0}_{n_{\mathrm{t}} \times C} & \mathbf{0}_{n_{\mathrm{t}} \times C} & \cdots & \mathbf{0}_{n_{\mathrm{t}} \times C} & V_{\mathrm{t}} \end{array}\right] \in R^{(Ns+1)C \times (n_{\mathrm{s}}+n_{\mathrm{t}})} \text{ with} \\
& \left(V_{\mathrm{s}}^{\mathrm{u}}\right)_{ij}=\left\{\begin{array}{ll} 1 & \text{if } x_{\mathrm{s}}^{\mathrm{u}i} \in \text{ class } j \\ 0 & \text{otherwise} \end{array}\right., \text{ and} \\
& \left(V_{\mathrm{t}}\right)_{ij}=\left\{\begin{array}{ll} 1 & \text{if } x_{\mathrm{t}}^{i} \text{ predicted as class } j \\ 0 & \text{otherwise} \end{array}\right.
\end{aligned} \tag{14}
$$

3 Optimization Procedure

Procedure 1 Multi-site Domain Adaptation via Landmark Selection

Input: Source and target domain data: X_{s} and X_{t}; labels for source domain data and pseudo-labels for target domain data: y_{s} and $\hat{y}_t$;Parameters: $\delta_{\mathrm{us}}, \delta_{\mathrm{t}}, d, \mu, \gamma$

Output: optimal Predicted labels y_{u} for target domain unlabeled data

Initialize pseudo labels of target domain unlabeled data $\hat{y}_{\mathrm{t}}$ using certain base classifiers with X_{t} Compute $S^{\mathrm{t}}_{\mathrm{h}}, M_{\mathrm{ss}}, M_{\mathrm{uu}}, M_{\mathrm{st}}, M_{\mathrm{ts}}, S^{\mathrm{s}}_{\mathrm{h}}, S^{\mathrm{s}}_{\mathrm{W}}, S^{\mathrm{t}}_{\mathrm{h}}, S^{\mathrm{u}}_{\mathrm{W}}$;

while not converge **do**

  Solve the generalized eigen-decomposition problem in (11) and select d corresponding eigenvectors of d largest eigenvalues as the transformation P, and obtain transformation A and B;

  Map the original data to respective subspace to get the embeddings: $Z_{\mathrm{s}} = A^{\mathrm{T}} X_{\mathrm{s}}, Z_{\mathrm{t}} = B^{\mathrm{T}} X_{\mathrm{t}}$

  Use base classifiers on $Z_{\mathrm{s}}, Z_{\mathrm{t}}, y_{\mathrm{s}}$ to update pseudo labels in target domain $\hat{y}_{\mathrm{t}}$

  Update landmark weights α, β

  Update $M_{\mathrm{ss}}, M_{\mathrm{tt}}, M_{\mathrm{st}}, M_{\mathrm{ts}}, S^{\mathrm{t}}_{\mathrm{b}}, S^{\mathrm{t}}_{\mathrm{w}}$

end while=0