Discovering Latent Domains for Multisource Domain Adaptation: Supplemental Material

Judy Hoffman[1,2], Kate Saenko[1,2,3], Brian Kulis[4], and Trevor Darrell[1,2]

UC Berkeley EECS[1], ICSI[2], Berkeley, California
Harvard University[3], Cambridge, Massachusetts
Ohio State University[4], Columbus, Ohio
jhoffman,saenko,trevor@eecs.berkeley.edu
kulis@cse.ohio-state.edu

We present here a proof that our clustering algorithm converges to a local optimum under the assumption that S, the number of global clusters, is small. We will do this by showing that after every iteration our objective function is non-increasing, and since the objective function is lower bounded (by zero) this proves that the algorithm will converge to a local optimum. Since each iteration of our clustering algorithm has four parts, we will individually prove that each part does not increase the objective function.

Let us define the objective of our clustering optimization in terms of our four variables:

$$J(\boldsymbol{Z}^L, \mu, \boldsymbol{Z}^G, \boldsymbol{m}) = \sum_{i=1}^{n}\sum_{j=1}^{J} \boldsymbol{Z}^L_{ij}(x_i - \mu_j)^2 + \sum_{j=1}^{J}\sum_{k=1}^{S} \boldsymbol{Z}^G_{jk}(\mu_j - \boldsymbol{m}_k)^2 \qquad (1)$$

Similarly, we let the value of our four variables after iteration t be denoted as: $(\boldsymbol{Z}^L_{ij})^{(t)}, \mu_j(t), \boldsymbol{Z}^G_{jk}(t), \boldsymbol{m}_k(t)$.

Claim. $1 J(\boldsymbol{Z}^L(t), \mu(t), \boldsymbol{Z}^G(t), \boldsymbol{m}(t)) \geq J(\boldsymbol{Z}^{L^{(t+1)}}, \mu(t), \boldsymbol{Z}^G(t), \boldsymbol{m}(t))$
Proof: To compute the local assignment variables during iteration $t+1$ we set

$$\boldsymbol{Z}^{L^{(t+1)}}_{ij} = \begin{cases} 1 \text{ if } \mu_j \text{ is the closest cluster center to point } x_i \\ 0 \text{ otherwise} \end{cases} \qquad (2)$$

Therefore, we can say that $\sum_{i=1}^{n}\sum_{j=1}^{J} \boldsymbol{Z}^L_{ij}(t)(x_i - \mu_j(t))^2 \geq \sum_{i=1}^{n}\sum_{j=1}^{J} \boldsymbol{Z}^{L^{(t+1)}}_{ij}(x_i - \mu_j(t))^2$, where equality only holds if all data points x_i were already assigned to their closest local cluster center μ_j. Since the second term of the objective function remains unchanged by this update, our claim is proven. □

Claim. $2 J(\boldsymbol{Z}^{L^{(t+1)}}, \mu^{(t)}, \boldsymbol{Z}^{G^{(t)}}, \boldsymbol{m}^{(t)}) \geq J(\boldsymbol{Z}^{L^{(t+1)}}, \mu^{(t+1)}, \boldsymbol{Z}^{G^{(t)}}, \boldsymbol{m}^{(t)})$
Proof: In this second stage we want to show that updating the local means does

not increase the total objective function. The local means update rule is derived by setting the derivative of the objective function with respect to μ to zero.

$$\begin{aligned}
\frac{d}{d\mu_j} J(\boldsymbol{Z}^{L^{(t+1)}}, \mu_j, \boldsymbol{Z}^{G^{(t)}}, \boldsymbol{m}^{(t)}) &= \frac{d}{d\mu_j} \sum_{i=1}^{n} \sum_{j=1}^{J} \boldsymbol{Z}_{ij}^{L^{(t+1)}} (x_i - \mu_j)^2 && (3) \\
&+ \sum_{j=1}^{J} \sum_{k=1}^{S} \boldsymbol{Z}_{jk}^{G^{(t)}} (\mu_j - \boldsymbol{m}_k^{(t)})^2 \\
&= -2 \sum_{i=1}^{n} \boldsymbol{Z}_{ij}^{L^{(t+1)}} (x_i - \mu_j) && (4) \\
&+2 \sum_{j=1}^{S} \boldsymbol{Z}_{jk}^{G^{(t)}} (\mu_j - \boldsymbol{m}_k) \\
&= 0 && (5)
\end{aligned}$$

Through this we derive the following update rule:

$$\mu_j^{(t+1)} = \frac{\sum_i (\boldsymbol{Z}_{ij}^L)^{(t+1)} x_i + \sum_k \boldsymbol{Z}^{G}{}_{jk}^{(t)} \boldsymbol{m}_k}{\sum_i \boldsymbol{Z}^{L}{}_{ij}^{(t+1)} + \sum_k \boldsymbol{Z}^{G}{}_{jk}^{(t)}} \quad (6)$$

Additionally, since $\frac{d^2}{d\mu_j^2} J(\boldsymbol{Z}^{L^{(t+1)}}, \mu_j, (\boldsymbol{Z}^G)^{(t)}, \boldsymbol{m}^{(t)}) \geq 0$ we know that this is a convex function with respect to μ so using the value for $\mu_j^{(t+1)}$ written above minimizes the objective function. Therefore the original claim is proven. □

Claim. 3$J(\boldsymbol{Z}^{L^{(t+1)}}, \mu^{(t+1)}, \boldsymbol{Z}^{G^{(t)}}, \boldsymbol{m}^{(t)}) \geq J(\boldsymbol{Z}^{L^{(t+1)}}, \mu^{(t+1)}, \boldsymbol{Z}^{G^{(t+1)}}, \boldsymbol{m}^{(t)})$
Proof: Our update rule for $\boldsymbol{Z}^G$ is as follows:

$$\boldsymbol{Z}_{jk}^{G^{(t+1)}} = \begin{cases} 1 \text{ if local cluster center } \mu_j^{(t+1)} \text{ is closest to global cluster center } \boldsymbol{m}_k^{(t)} \\ \quad \text{such that all constraints are satisfied} \\ 0 \text{ otherwise} \end{cases}$$

Note that we can choose the best assignment of local clusters to global clusters via exhaustive search for small S. Therefore, under that assumption this stage will not increase the objective function, since local clusters are assigned to the global clusters according to the matching with the lowest objective achievable under the constraints. □

Claim. 4$J(\boldsymbol{Z}^{L^{(t+1)}}, \mu^{(t+1)}, \boldsymbol{Z}^{G^{(t+1)}}, \boldsymbol{m}^{(t)}) \geq J(\boldsymbol{Z}^{L^{(t+1)}}, \mu^{(t+1)}, \boldsymbol{Z}^{G^{(t+1)}}, \boldsymbol{m}^{(t+1)})$
Proof: To derive our update equation for $\boldsymbol{m}_k$ at time $t+1$ we set the derivative of J with respect to $\boldsymbol{m}$, set it to zero and solve for the new $\boldsymbol{m}$ value.

$$\frac{d}{d\boldsymbol{m}_k} J(\boldsymbol{Z}^{L^{(t+1)}}, \mu^{(t+1)}, \boldsymbol{Z}^{G^{(t+1)}}, \boldsymbol{m}) = -2 \sum_{j=1}^{J} \boldsymbol{Z}_{jk}^{G^{(t+1)}} (\mu_j^{(t+1)} - \boldsymbol{m})) = 0 \quad (7)$$

Solving this equation for $\boldsymbol{m}$ we derive our update rule:

$$\boldsymbol{m}_k^{(t+1)} = \frac{\sum_j \boldsymbol{Z}^{G(t+1)} \mu_j^{(t+1)}}{\sum_j \boldsymbol{Z}^{G(t+1)}} \tag{8}$$

Since the second derivative of J with respect to $\boldsymbol{m}$ is positive we know that by setting $\boldsymbol{m}_k$ is this way we have minimized the function. Therefore the objective is non-decreasing after updating $\boldsymbol{m}$. □

Finally, combining our four claims we can now determine that:

$$\begin{aligned} J(\boldsymbol{Z}^{L^{(t)}}, \mu^{(t)}, \boldsymbol{Z}^{G^{(t)}}, \boldsymbol{m}^{(t)}) &\geq J(\boldsymbol{Z}^{L(t+1)}, \mu^{(t)}, \boldsymbol{Z}^{G(t)}, \boldsymbol{m}^{(t)}) && (9) \\ &\geq J(\boldsymbol{Z}^{L(t+1)}, \mu^{(t+1)}, \boldsymbol{Z}^{G(t)}, \boldsymbol{m}^{(t)}) && (10) \\ &\geq J(\boldsymbol{Z}^{L(t+1)}, \mu^{(t+1)}, \boldsymbol{Z}^{G(t+1)}, \boldsymbol{m}^{(t)}) && (11) \\ &\geq J(\boldsymbol{Z}^{L(t+1)}, \mu^{(t+1)}, \boldsymbol{Z}^{G(t+1)}, \boldsymbol{m}^{(t+1)}) && (12) \end{aligned}$$

Therefore, we have shown that the total objective function is non-increasing after each iteration of the algorithm. □