Bypassing the Transport Plan: Dynamic Reweighting for Out-of-Distribution Detection with Optimal Transport

Supplementary Material

6. Proof of Proposition 3.1

Proposition 3.1 *(Dynamic Reweighting with Exact SemiUOT Solver). Given SemiUOT with KL-Divergence shown in Equation* (2)*, we can rewrite its dual form as below:*

$$\begin{aligned} &\min_{\boldsymbol{u},\boldsymbol{v},\boldsymbol{s},\boldsymbol{\zeta}} \mathcal{J} = \tau_a \left\langle \boldsymbol{a}, \exp\left(-\frac{\boldsymbol{u}+\boldsymbol{\zeta}}{\tau_a}\right)\right\rangle - \left\langle \boldsymbol{v}-\boldsymbol{\zeta}, \boldsymbol{b}\right\rangle, \\ &s.t.\ u_i + v_j + s_{ij} = C_{ij}, s_{ij} \geq 0, \end{aligned} \tag{13}$$

where $\boldsymbol{u}$, $\boldsymbol{v}$, s, and $\boldsymbol{\zeta}$ are dual variables. Equation (13) *can be rewritten as classical OT:*

$$\begin{aligned} &\min_{\boldsymbol{\pi}\geq 0} \mathcal{J}_{\text{S}} = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle, \\ &s.t.\ \boldsymbol{\pi}\mathbf{1}_N = \boldsymbol{a} \odot \exp\left(-\frac{\boldsymbol{u}^* + \boldsymbol{\zeta}^*}{\tau_a}\right), \boldsymbol{\pi}^{\text{T}}\mathbf{1}_M = \boldsymbol{b}, \end{aligned} \tag{14}$$

where $\boldsymbol{u}^*$ and $\boldsymbol{\zeta}^*$ are the optimal value of $\boldsymbol{u}$ and $\boldsymbol{\zeta}$.

Proof. We can rewrite the SemiUOT problem as below:

$$\begin{aligned} &\min_{\boldsymbol{\pi}\geq 0} J = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle + \tau_a \text{KL}\left(\boldsymbol{\pi}\mathbf{1}_N \| \boldsymbol{a}\right) \\ &s.t.\ \text{(Constraint)} : \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b}, \text{(Optional)} : \boldsymbol{\pi}\mathbf{1}_N = \boldsymbol{\alpha}. \end{aligned}$$

Note that we do not need to know the exact value of $\boldsymbol{\alpha}$ beforehand. We adopt this optional constraint only for simplifying the following deduction. The Lagrange multipliers of Semi-UOT with KL-Divergence is given as:

$$\begin{aligned} &\max_{\boldsymbol{s}\geq 0,\boldsymbol{u},\boldsymbol{v},\boldsymbol{\zeta}} \min_{\boldsymbol{\pi}\geq 0} \mathcal{J} = \\ &\tau_a \text{KL}\left(\boldsymbol{\pi}\mathbf{1}_N \| \boldsymbol{a}\right) + \langle \boldsymbol{u}+\boldsymbol{\zeta}, \boldsymbol{\pi}\mathbf{1}_N \rangle + \langle \boldsymbol{v}-\boldsymbol{\zeta}, \boldsymbol{b} \rangle + \mathscr{C}_{\text{SUOT}} \end{aligned}$$

where $\mathscr{C}_{\text{SUOT}} = \sum_{i,j}(C_{ij} - u_i - v_j - s_{ij})\pi_{ij} = \langle \boldsymbol{C} - \boldsymbol{u} \otimes \mathbf{1}_N^\top - \mathbf{1}_M \otimes \boldsymbol{v}^\top - \boldsymbol{s}, \boldsymbol{\pi} \rangle$ and $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{\zeta}$ are dual variables. By taking the differentiation on π_{ij} we have:

$$\begin{aligned} &\frac{\partial \mathcal{J}}{\partial \pi_{ij}} \\ &= \left[\tau_a \log \frac{\sum_{j=1}^N \pi_{ij}}{a_i} + u_i + \boldsymbol{\zeta}\right] + (C_{ij} - u_i - v_j - s_{ij}) \\ &= C_{ij} + \tau_a \log \frac{\sum_{j=1}^N \pi_{ij}}{a_i} + \boldsymbol{\zeta} - v_j - s_{ij} = 0. \end{aligned}$$

Then we can obtain the results:

$$\begin{aligned} &\begin{cases} \sum_{j=1}^N \pi_{ij} = a_i \exp\left(-\frac{u_i + \boldsymbol{\zeta}}{\tau_a}\right) \\ \sum_{i=1}^M \pi_{ij} = b_j \end{cases} \\ &\Rightarrow C_{ij} - u_i - v_j - s_{ij} = 0, \quad s_{ij} \geq 0. \end{aligned}$$

Thus SemiUOT can be can be regarded as classic optimal transport problem:

$$\begin{aligned} &\min_{\boldsymbol{\pi}\geq 0} \mathcal{J}_{\text{S}} = \langle \boldsymbol{C}, \boldsymbol{\pi} \rangle \\ &s.t. \boldsymbol{\pi}\mathbf{1}_N = \boldsymbol{a} \odot \exp\left(-\frac{\boldsymbol{u}^* + \boldsymbol{\zeta}^*}{\tau_a}\right), \boldsymbol{\pi}^\top \mathbf{1}_M = \boldsymbol{b}. \end{aligned}$$

7. Proof of Proposition 3.2

Proposition 3.2 *(Accelerating dynamic reweighting with approximation). We consider a smooth approximation of SemiUOT to replace* $\inf(\cdot)$ *as* $\inf_{k\in[M]}[C_{kj} - u_k] \approx -\epsilon \log[\sum_{k=1}^M e^{\frac{u_k - C_{kj}}{\epsilon}}]$. *Note that $\epsilon > 0$ denotes the balanced hyperparameter among precision and smoothness of the function. Smaller ϵ (e.g., ϵ approaches to 0) could lead to more accurate while less smooth solutions. Then we can obtain the proposed Approximate SemiUOT Equation as $\widehat{\mathcal{J}}_{\text{S}}$ by replacing* $\inf(\cdot)$ *with the smoothness term for* $\widehat{\mathrm{u}}$,

$$\begin{aligned} \min_{\boldsymbol{u},\boldsymbol{\zeta}} \widehat{\mathcal{J}}_{\text{S}} = \tau_a \exp\left(-\frac{\boldsymbol{\zeta}}{\tau_a}\right) \sum_{i=1}^M a_i \exp\left(-\frac{u_i}{\tau_a}\right) + \\ \sum_{j=1}^N \left[\epsilon \log\left[\sum_{k=1}^M \exp\left(\frac{u_k - C_{kj}}{\epsilon}\right)\right] + \boldsymbol{\zeta}\right] b_j. \end{aligned} \tag{15}$$

Then, u_i can be iteratively updated as follows:

$$\begin{aligned} u_i^{(l+1)} = \mathcal{T}(u_i^{(l)}) = \frac{\tau_a \epsilon}{\tau_a + \epsilon} \log\left(a_i \exp\left(-\frac{\boldsymbol{\zeta}}{\tau_a}\right)\right) - \\ \frac{\tau_a \epsilon}{\tau_a + \epsilon} \log\left[\sum_{j=1}^N \left[\frac{\exp\left(-\frac{C_{ij}}{\epsilon}\right)}{\sum_{k=1}^M \exp\left(\frac{u_k^{(l)} - C_{kj}}{\epsilon}\right)}\right] b_j\right]. \end{aligned} \tag{16}$$

Meanwhile, $\boldsymbol{\zeta}$ can eventually be computed as:

$$\boldsymbol{\zeta} = \tau_a \left[\log\left(\sum_{i=1}^M a_i \exp\left(-\frac{u_i}{\tau_a}\right)\right) - \log\left(\sum_{j=1}^N b_j\right)\right]. \tag{17}$$

Proof. Although $\mathcal{J}_{\text{S}}$ is convex and has unique solutions, the presence of $\inf(\cdot)$ renders it a non-smooth function, leading to inefficient optimization. To further accelerate the optimization process, we consider to make a smooth approximation on replacing $\inf(\cdot)$ as $\inf_{k\in[M]}[C_{kj} - u_k] \approx -\epsilon \log[\sum_{k=1}^M e^{\frac{u_k - C_{kj}}{\epsilon}}]$. Note that $\epsilon > 0$ denotes the balanced hyper parameters among the accuracy and function smoothness. Smaller ϵ (e.g., ϵ approaches to 0) could lead to more accurate while less smooth solutions. Then we can obtain the proposed *Approximate SemiUOT Equation* as $\widehat{\mathcal{J}}_{\mathcal{S}}$ by replacing $\inf(\cdot)$ with the smoothness term for $\widehat{f}$ as below:

$$\begin{aligned}\min_{\boldsymbol{u},\boldsymbol{\zeta}} \widehat{\mathcal{J}}_{\mathrm{S}} =& \tau_a \exp\left(-\frac{\boldsymbol{\zeta}}{\tau_a}\right) \sum_{i=1}^{M} a_i \exp\left(-\frac{u_i}{\tau_a}\right) \\ &+ \sum_{j=1}^{N} \left[\epsilon \log\left[\sum_{k=1}^{M} \exp\left(\frac{u_k - C_{kj}}{\epsilon}\right)\right] + \boldsymbol{\zeta}\right] b_j .\end{aligned}$$

Take the differentiation on u_i we can obtain:

$$\begin{aligned}\frac{\partial \widehat{\mathcal{J}}_{\mathrm{S}}}{\partial u_i} =& - a_i \exp\left(-\frac{\boldsymbol{\zeta}}{\tau_a}\right) \exp\left(-\frac{u_i}{\tau_a}\right) \\ &+ \exp\left(\frac{u_i}{\epsilon}\right) \sum_{j=1}^{N} \left[\frac{\exp\left(-\frac{C_{ij}}{\epsilon}\right)}{\sum_{k=1}^{M} \exp\left(\frac{u_k - C_{kj}}{\epsilon}\right)}\right] b_j \\ =& 0.\end{aligned}$$

To solve the above problem, we can obtain:

$$\begin{aligned}u_i^{(l+1)} =& \frac{\tau_a \epsilon}{\tau_a + \epsilon} \log\left(a_i \exp\left(-\frac{\boldsymbol{\zeta}}{\tau_a}\right)\right) \\ &- \frac{\tau_a \epsilon}{\tau_a + \epsilon} \log\left[\sum_{j=1}^{N} \left[\frac{\exp\left(-\frac{C_{ij}}{\epsilon}\right)}{\sum_{k=1}^{M} \exp\left(\frac{u_k^{(l)} - C_{kj}}{\epsilon}\right)}\right] b_j\right] \\ =& \mathcal{T}(u_i^{(l)}).\end{aligned}$$

We can adopt Banach theorem to verify the convergence of the algorithm.

$$\begin{aligned}&\frac{\partial \mathcal{T}(u_i^{(l)})}{\partial u_i^{(l)}} = \\ &- \frac{\tau_a \epsilon}{\tau_a + \epsilon} \frac{\frac{\partial}{\partial u_i^{(l)}} \left(\sum_{j=1}^{N} \left[\frac{\exp\left(-\frac{C_{ij}}{\epsilon}\right)}{\sum_{k=1}^{M} \exp\left(\frac{u_k^{(l)} - C_{kj}}{\epsilon}\right)}\right] b_j\right)}{\sum_{j=1}^{N} \left[\frac{\exp\left(-\frac{C_{ij}}{\epsilon}\right)}{\sum_{k=1}^{M} \exp\left(\frac{u_k^{(l)} - C_{kj}}{\epsilon}\right)}\right] b_j} = \\ &\frac{\tau_a}{\tau_a + \epsilon} \underbrace{\frac{\sum_{j=1}^{N} \left[\frac{b_j \exp\left(-\frac{C_{ij}}{\epsilon}\right) \exp\left(\frac{u_i^{(l)} - C_{ij}}{\epsilon}\right)}{\sum_{k=1}^{M} \exp\left(\frac{u_k^{(l)} - C_{kj}}{\epsilon}\right) \sum_{k=1}^{M} \exp\left(\frac{u_k^{(l)} - C_{kj}}{\epsilon}\right)}\right]}{\sum_{j=1}^{N} \left[\frac{\exp\left(-\frac{C_{ij}}{\epsilon}\right)}{\sum_{k=1}^{M} \exp\left(\frac{u_k^{(l)} - C_{kj}}{\epsilon}\right)}\right] b_j}}_{\leq 1} \\ &\leq 1.\end{aligned}$$

8. Proof of Proposition 3.3

Proposition 3.3 *(Approximation error). We consider the analysis between optimal results of* $\boldsymbol{u}$ *and* $\widehat{\boldsymbol{u}}$ *via setting* $\boldsymbol{\zeta} = 0$ *in SemiUOT. Then we define* $E_p(\boldsymbol{u}) = \mathcal{J}_{\mathrm{S}}$ *and* $K_p(\widehat{\boldsymbol{u}}) = \widehat{\mathcal{J}}_{\mathrm{S}}$*. Hence we have the following relationships: (1)* $|K_P(\boldsymbol{u}) - E_P(\boldsymbol{u})| \leq \epsilon \log M$*, (2)* $|E_P(\widehat{\boldsymbol{u}}) - K_P(\widehat{\boldsymbol{u}})| \leq 0$*. Thus we have the following error bound:*

$$|K_P(\boldsymbol{u}) - K_P(\widehat{\boldsymbol{u}})| \leq \epsilon \log M. \tag{18}$$

Therefore $\boldsymbol{u}$ *and* $\widehat{\boldsymbol{u}}$ *will get closer with smaller* ϵ*.*

Proof. By definition of the optimal transport problem, the regularized optimal potential $\widehat{\boldsymbol{u}}$ and the unregularized optimal potential $\boldsymbol{u}$ satisfy:

$$K_P(\widehat{\boldsymbol{u}}) \leq K_P(\boldsymbol{u}), \quad E_P(\boldsymbol{u}) \leq E_P(\widehat{\boldsymbol{u}}). \tag{19}$$

Next, we bound the regularization gap at the unregularized optimum $\boldsymbol{u}$. Specifically,

$$\begin{aligned}\text{error} &= K_P(\boldsymbol{u}) - E_P(\boldsymbol{u}) \\ &= \sum_{j=1}^{N} \varepsilon b_j \log\left[\sum_{k=1}^{M} \exp\left(\frac{u_k - C_{kj}}{\varepsilon}\right)\right] - \\ &\qquad \sum_{j=1}^{N} \sup_{k \in [M]} \left[u_k - C_{kj}\right] b_j \\ &\leq \sum_{j=1}^{N} \varepsilon b_j \left[\log\left(M \exp\left(\sup_{k \in [M]} \left(\frac{u_k - C_{kj}}{\varepsilon}\right)\right)\right)\right. \\ &\qquad \left. - \log\left(\exp\left(\sup_{k \in [M]} \left(\frac{u_k - C_{kj}}{\varepsilon}\right)\right)\right)\right] \\ &= \varepsilon \log M.\end{aligned} \tag{20}$$

Moreover, for any $\boldsymbol{u}$, and in particular for $\widehat{\boldsymbol{u}}$, we have

$$\begin{aligned}&\sum_{j=1}^{N} b_j \sup_{k \in [M]} \left[\hat{u}_k - C_{kj}\right] - \sum_{j=1}^{N} \varepsilon b_j \log \sum_{k=1}^{M} \exp\left(\frac{\hat{u}_k - C_{kj}}{\varepsilon}\right) \\ &\leq 0.\end{aligned} \tag{21}$$

That is,

$$E_P(\widehat{\boldsymbol{u}}) \leq K_P(\widehat{\boldsymbol{u}}). \tag{22}$$

Combining this with Equation (19), we obtain

$$E_P(\boldsymbol{u}) \leq E_P(\widehat{\boldsymbol{u}}) \leq K_P(\widehat{\boldsymbol{u}}) \leq K_P(\boldsymbol{u}) \leq E_P(\boldsymbol{u}) + \varepsilon \log M. \tag{23}$$

Therefore,

$$0 \leq K_P(\widehat{\boldsymbol{u}}) - E_P(\boldsymbol{u}) \leq \varepsilon \log M, \tag{24}$$

which completes the proof.

9. Implementation details

To be fair, we adopt Wide ResNet (WRN) [54] as the backbone, with WRN-28-2 on CIFAR-10 and WRN-28-8 on CIFAR-100, consistent with the baseline. For each batch, the size of $\boldsymbol{X}_u$ is twice that of $\boldsymbol{X}_l$. Since Fixmatch is used in the closed-set semi-supervised classification module of DREW in the experimental part, we set the temperature parameter of pseudo-label to 1 and the threshold of pseudo-label to 0.95. The model is trained with a Nesterov SGD optimizer with 0.9 momentum and 5×10^{-4} weight decay. We implement the cosine annealing learning rate adjustment strategy and set the initial learning rate as 0.03. τ_a in Equation (2) is set to 0.01 when comparing DREW with baselines. Putting all the components together, we set the balance factor of L_x and L_u as 1 while setting that of L_{ood} as 0.01. For the ImageNet-30 dataset, We use ResNet-18 [16] as the backbone.