

# TDAPM: Target Distribution Alignment via Progressive Matching for Multi-Source-Free Domain Adaptation – *Supplementary Materials*

Qing Tian, Linmin Wang, Xiang Liu, Junyu Shen, Yuhui Zheng\*, Jun Wan, Zhen Lei

---

In what follows, we provide a more detailed theoretical derivation for the progressive prediction alignment strategy employed in TDAPM.

## 1. Theoretical Derivations for Progressive Prediction Alignment

We provide a more detailed theoretical derivation for the progressive prediction alignment strategy employed in TDAPM, specifically focusing on bounding the generalization error of the target model. Our analysis draws inspiration from established frameworks in domain adaptation theory, adapting them to the source-free, iterative prediction alignment context.

Let  $P_S$  and  $P_T$  denote the true data distributions of the source and target domains, respectively. In the MSFDA setting, we have access to  $m$  source models  $\{\theta_S^j\}_{j=1}^m$  pre-trained on  $S_j \sim P_{S_j}$ , and unlabeled target data  $T \sim P_T$ . Our goal is to learn a target model  $\theta_T$  that performs well on  $P_T$ .

### 1.1. Generalization Bound for Iterative Adaptation

We define a sequence of target models  $\{\theta_k\}_{k=0}^K$ , where  $\theta_0$  is the initial target model constructed from a weighted ensemble of source models, and  $\theta_k$  is the model at iteration  $k$  after adapting to the target data. Let  $L(\theta, P)$  denote the loss of model  $\theta$  on distribution  $P$ . We are interested in bounding  $L(\theta_K, P_T)$ .

Following a chain rule for generalization error, the error of the final model  $\theta_K$  on the true target distribution  $P_T$  can be related to the error of the initial model and the accumulated shifts during the iterative process. We can express the target risk of  $\theta_K$  as:

$$L(\theta_K, P_T) = L(\theta_0, P_T) + \sum_{k=1}^K [L(\theta_k, P_T) - L(\theta_{k-1}, P_T)] \quad (1)$$

This decomposition is generally difficult to bound directly. Instead, we leverage the concept of distribution discrepancy between successive model predictions.

Let  $P_{\theta_k}$  denote the prediction distribution (output probabilities) of model  $\theta_k$  on target domain samples. The core idea of progressive prediction alignment is to minimize the discrepancy between  $P_{\theta_k}$  and  $P_{\theta_{k-1}}$ . Specifically, TDAPM optimizes  $\theta_k$  such that the pairwise alignment loss  $\mathcal{L}_{\text{align}} = -I(P_{\theta_{k-1}}(x'), P_{\theta_k}(x))$  is minimized. This can be viewed as an attempt to make  $P_{\theta_k}$  “close” to  $P_{\theta_{k-1}}$ .

Let  $D(P_1, P_2)$  be a suitable discrepancy measure between two prediction distributions  $P_1$  and  $P_2$  (e.g., KL divergence). We make the following assumptions:

1. **Bounded Prediction Shift:** The progressive alignment strategy ensures that the discrepancy between consecutive prediction distributions is bounded. That is, for each iteration  $k$ , there exists a  $\pi_k \in [0, \Pi_{\max})$  such that  $D(P_{\theta_k}, P_{\theta_{k-1}}) \leq \pi_k$ . This  $\pi_k$  quantifies the “progress” or “shift” in the prediction distribution from one iteration to the next. In TDAPM, the mutual information maximization loss  $\mathcal{L}_{\text{align}}$  directly aims to keep this shift small.
2. **Regularization Contribution:** Each iteration might include regularization terms, denoted by  $\mathcal{R}(\theta_k)$ , which add to the total loss. We assume these regularization terms contribute  $\lambda_k \mathcal{R}(\theta_k)$  to the error at iteration  $k$ .

---

\*Corresponding author: Yuhui Zheng (email: zhengyh@vip.126.com).

3. **Relationship between Model Change and Risk:** A crucial step is to relate the prediction distribution shift to the change in target risk. While a direct equality  $L(\theta_k, P_{\mathcal{T}}) - L(\theta_{k-1}, P_{\mathcal{T}}) \leq D(P_{\theta_k}, P_{\theta_{k-1}})$  is generally strong, under conditions of sufficient model capacity and smooth loss landscapes, it is reasonable to assume that controlling the prediction distribution shift helps control the risk change. Specifically, we can write:

$$L(\theta_k, P_{\mathcal{T}}) \leq L(\theta_{k-1}, P_{\mathcal{T}}) + \gamma_k D(P_{\theta_k}, P_{\theta_{k-1}}) + \lambda_k \mathcal{R}(\theta_k) \quad (2)$$

where  $\gamma_k$  is a scaling factor that depends on the properties of the loss function and model. For simplicity in the initial theoretical insight, we can consider  $\gamma_k \approx 1$  for a direct relationship.

Applying this iteratively from  $k = 1$  to  $K$ :

$$L(\theta_K, P_{\mathcal{T}}) \leq L(\theta_0, P_{\mathcal{T}}) + \sum_{k=1}^K [\gamma_k D(P_{\theta_k}, P_{\theta_{k-1}}) + \lambda_k \mathcal{R}(\theta_k)] \quad (3)$$

Substituting the bounded prediction shift assumption:

$$L(\theta_K, P_{\mathcal{T}}) \leq L(\theta_0, P_{\mathcal{T}}) + \sum_{k=1}^K [\gamma_k \pi_k + \lambda_k \mathcal{R}(\theta_k)] \quad (4)$$

This is the general form of Eq. (1) in the revised manuscript:

$$L(\theta_K, P_{\mathcal{T}}) \leq L(\theta_0, P_{\mathcal{T}}) + \sum_{k=1}^K (D(P_{\theta_k}, P_{\theta_{k-1}}) + \lambda_k \mathcal{R}(\theta_k)). \quad (5)$$

### 1.2. Simplification and Interpretation

To simplify further and provide a clearer interpretation, let's assume:

- The prediction shift is uniformly bounded:  $D(P_{\theta_k}, P_{\theta_{k-1}}) \leq \pi$  for all  $k$ . This is a reasonable assumption if the iterative steps are small and controlled.
- The regularization strength is constant per iteration:  $\lambda_k \mathcal{R}(\theta_k) = R$  for all  $k$ . This  $R$  encapsulates the contribution of all regularization terms (e.g., diversity loss, domain indicator loss) to the overall generalization error in each step.

Under these simplifying assumptions, the sum becomes:

$$\sum_{k=1}^K (D(P_{\theta_k}, P_{\theta_{k-1}}) + \lambda_k \mathcal{R}(\theta_k)) \leq \sum_{k=1}^K (\pi + R) = K(\pi + R) \quad (6)$$

Substituting this back into the bound:

$$L(\theta_K, P_{\mathcal{T}}) \leq L(\theta_0, P_{\mathcal{T}}) + K(\pi + R). \quad (7)$$

Let  $\alpha_0 = L(\theta_0, P_{\mathcal{T}})$  represent the initial loss on the target distribution before any iterative adaptation. Then, we arrive at my revised manuscript Eq. (2):

$$L(\theta_K, P_{\mathcal{T}}) \leq \alpha_0 + K(\pi + R). \quad (8)$$