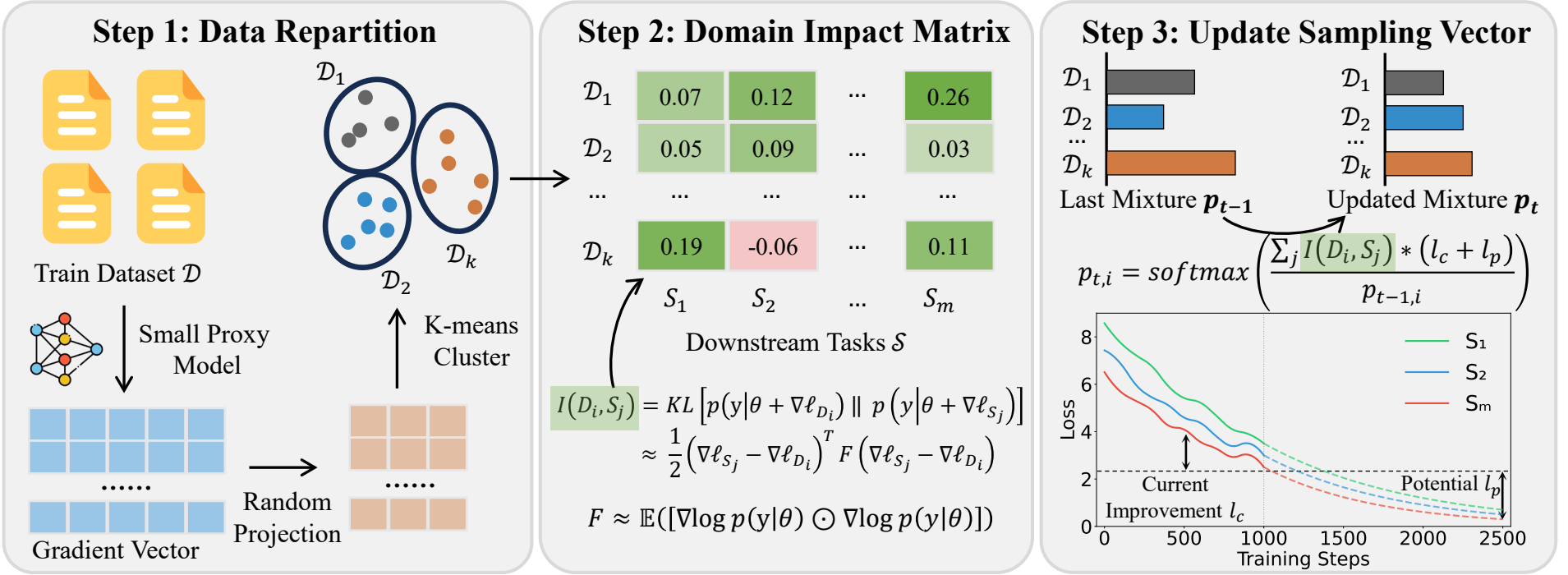
Step 1: Data Repartition
Train Dataset $\mathcal{D}$
Small Proxy Model
Gradient Vector
Random Projection
K-means Cluster
$\mathcal{D}_1$
$\mathcal{D}_2$
$\mathcal{D}_k$
Step 2: Domain Impact Matrix
0.07 0.12 ... 0.26
0.05 0.09 ... 0.03
0.19 -0.06 ... 0.11
S_1 S_2 ... S_m
Downstream Tasks $\mathcal{S}$
$I(D_i, S_j) = KL\left[p(y|\theta + \nabla\ell_{D_i}) \parallel p\left(y\middle|\theta + \nabla\ell_{S_j}\right)\right]$
$\approx \frac{1}{2}\left(\nabla\ell_{S_j} - \nabla\ell_{D_i}\right)^T F \left(\nabla\ell_{S_j} - \nabla\ell_{D_i}\right)$
$F \approx \mathbb{E}([\nabla \log p(y|\theta) \odot \nabla \log p(y|\theta)])$
Step 3: Update Sampling Vector
Last Mixture p_{t-1}
Updated Mixture p_t
$p_{t,i} = softmax\left(\frac{\sum_j I(D_i, S_j) * (l_c + l_p)}{p_{t-1,i}}\right)$
Loss
Training Steps
Current Improvement l_c
Potential l_p
S_1
S_2
S_m