Supplementary Material - PropMix: Hard Sample Filtering and Proportional MixUp for Learning with Noisy Labels

Filipe R. Cordeiro[1]
filipe.rolim@ufrpe.br
Vasileios Belagiannis[2]
vasileios.belagiannis@uni-ulm.de
Ian Reid[3]
ian.reid@adelaide.edu.au
Gustavo Carneiro[3]
gustavo.carneiro@adelaide.edu.au

[1] Universidade Federal Rural de Pernambuco Recife, Brazil
[2] Universität Ulm Ulm, Germany
[3] University of Adelaide Adelaide, Australia

1 PropMix Algorithm

SOTA noise-robust classifiers [1, 2, 3] are formed by an ensemble of two classifiers, where the classifier structure is the same, but their parameters are denoted by $\theta(1), \theta(2) \in \Theta$. The training for $\theta(1)$ influences $\theta(2)$ and vice-versa, where this can be achieved by co-training [1, 2] or student-teacher [3] approaches. Our training relies on co-training. The estimation of class prediction for label estimation and evaluation are given by the average outputs of the models. The pseudo-code for the training of PropMix is shown in Algorithm 1.

References

[1] Junnan Li, Richard Socher, and Steven C. H. Hoi. DivideMix: Learning with Noisy Labels as Semi-supervised Learning. *arXiv e-prints*, art. arXiv:2002.07394, February 2020.

[2] Sheng Liu, Jonathan Niles-Weed, Narges Razavian, and Carlos Fernandez-Granda. Early-learning regularization prevents memorization of noisy labels. *arXiv preprint arXiv:2007.00151*, 2020.

[3] Duc Tam Nguyen, Chaithanya Kumar Mummadi, Thi Phuong Nhung Ngo, Thi Hoai Phuong Nguyen, Laura Beggel, and Thomas Brox. Self: Learning to filter noisy labels with self-ensembling. *arXiv preprint arXiv:1910.01842*, 2019.

Algorithm 1: PropMix (PM)

$\mathcal{D}$, number of epochs E, clean sample threshold τ, hard sample threshold τ'
// Self-supervised pre-training
$f_\phi(\mathbf{x})$,$\{\mathcal{N}_{\mathbf{x}_i}\}_{i=1}^{|\mathcal{D}|}$ = PreTrain($\mathcal{D}$)
// Warm Up
$p_\theta(\mathbf{y}|\mathbf{x})$ = WarmUp($\mathcal{D}$,$f_\phi(\mathbf{x})$)
while $e < E$ **do**
  // Estimate sets of clean and noisy samples
  for $i = \{1,...,|\mathcal{D}|\}$ **do**
    Estimate $p(\text{clean}|\ell_i,\gamma)$, with $\ell_i = -\mathbf{y}_i^\top \log p_\theta(.|\mathbf{x}_i)$
  end
  $\mathcal{X}$,$\mathcal{U}$=FormCleanNoisySets($\{p(\text{clean}|\ell_i,\gamma)\}_{i=1}^{|\mathcal{D}|},\tau$)
  // Estimate sets of hard noisy and easy noisy samples
  for $i = \{1,...,|\mathcal{U}|\}$ **do**
    Estimate $p\,(\text{hard}|\mathbf{y}_i^*(c),\gamma)$, with $\mathbf{y}_i^*(c) =_{c\in\mathcal{Y}} p\,(c|\mathbf{x}_i)$
  end
  $\mathcal{U}_H$,$\mathcal{U}_E$=FormHardEasySets($\{p\,(\text{hard}|\mathbf{y}_i^*(c),\gamma)\}_{i=1}^{|\mathcal{U}|},\tau'$)
  for b=*1 to* B **do**
    $\{\hat{\mathbf{x}}_{b,m}\}_{m=1}^M$ = DataAugment($\mathbf{x}_b \in \mathcal{X}, M$)
    $\{\hat{\mathbf{u}}_{b,m}\}_{m=1}^M$ = DataAugment($\mathbf{u}_b \in \mathcal{U}_E, M$)
    $\mathbf{p}_b = \frac{1}{2M}\sum_{m=1,k=1}^{M,2} f(\hat{\mathbf{x}}_{b,m};\theta(k))$
    $\mathbf{q}_b = \frac{1}{2M}\sum_{m=1,k=1}^{M,2} f(\hat{\mathbf{u}}_{b,m};\theta(k))$
    $\hat{\mathbf{y}}_b = \text{TempShrp}(w_b\mathbf{y}_b + (1-w_b)\mathbf{p}_b;T)$
    $\hat{\mathbf{q}}_b = \text{TempShrp}(\mathbf{q}_b;T)$
  end
  $\hat{\mathcal{X}} = \{\hat{\mathbf{x}}_{b,m},\hat{\mathbf{y}}_b\}_{m=1}^M$, $\hat{\mathcal{U}}_E = \{\hat{\mathbf{u}}_{b,m},\hat{\mathbf{q}}_b\}_{m=1}^M$
  $\hat{\mathcal{D}}$ = ProportionalMixUp($\hat{\mathcal{X}}$,$\hat{\mathcal{U}}_E$)
  Update $\theta(k1)$, $\theta(k2)$ with $\mathcal{L}$ from (5)
end