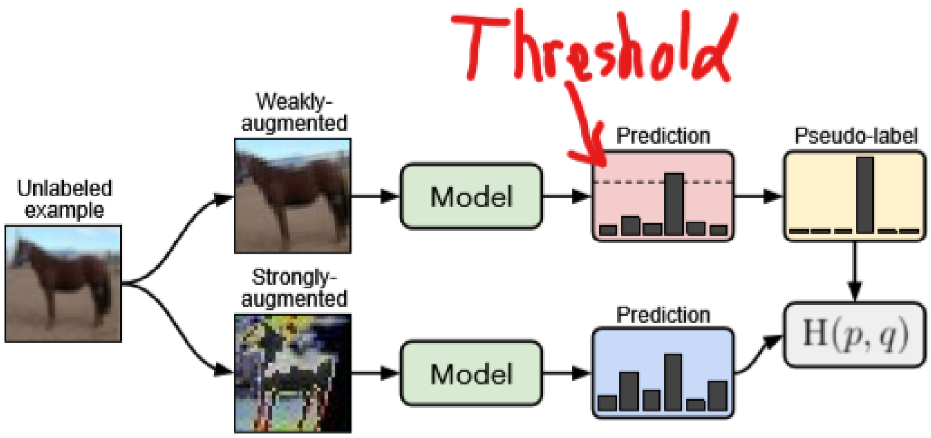
Notes on FixMatch Original Paper

Abstract

1. FixMatch combines two methods, **consistency regularization** and **pseudo-labeling**
2. Steps:
   -> weakly augmented unlabeled images
   -> retain pseudo-labels if the $confidence\ level > threshold$
   -> strongly augmented unlabeled images, predict its label



Introduction

1. FixMatch leverages **CutOut**, **CTAugment**, **RandAugment** for strong augmentation.

FixMatch

Some notations:

1. A batch of B labeled examples, where x_b are training samples and p_b are one-hot labels
   $\mathcal{X} = \{(x_b, p_b) : b \in (1, \ldots, B)\}$
2. A batch of μB unlabeled examples. μ determines the ratio of $\mathcal{X}$ and $\mathcal{U}$.
   $\mathcal{U} = \{u_b : b \in (1, \ldots, \mu B)\}$
3. Predicted class distribution, $p_m(y|x)$
4. Cross entropy between P and Q, $H(p, q)$
5. Strong augmentation: $\mathcal{A}(\cdot)$. Weak augmentation: $\alpha(\cdot)$

Background

Pseudo-labeling refers to the use of “hard” labels (argmax of the model’s output) and only retaining artificial labels whose largest class probability fall above a predefined threshold

The use of a hard label makes pseudo-labeling closely related to entropy minimization, where the model’s predictions are encouraged to be **low-entropy** (i.e., high-confidence) on unlabeled data.

Algorithm

Supervised loss ℓ_s, cross entropy on weakly augmented labeled examples:

$$\ell_s = \frac{1}{B}\sum_{b=1}^{B} H(p_b, p_m(y \mid \alpha(x_b)))$$

Unsupervised loss ℓ_u, perform weak augmentation on image -> model prediction distribution, $q_b = p_m(y|\alpha(u_b))$ -> pseudo-label, $\hat{q}_b$(if exceed threshold). Next, compute cross entropy loss with the prediction distribution from strongly augmented image.

$$\ell_u = \frac{1}{\mu B}\sum_{b=1}^{\mu B} 1(\max(q_b) \geq \tau) H(\hat{q}_b, p_m(y \mid \mathcal{A}(u_b)))$$

$$Total\ loss = \ell_s + \lambda_u \ell_u$$

I guess the ideology of pseudo-label with threshold is to *label the image* if the model is *highly confident* on its predictions.

Augmentation

Weak augmentation: Flip and shift

Strong augmentation: *Cutout* followed by *RandAugment* or**CTAugment* where both RandAugment and CTAugment are variants of AutoAugment (Reinforce Method).

Related work

FixMatch resembles **UDA** and **ReMixMatch**.

Sharpening: UDA and ReMixMatch “sharpen” the artificial label not using *pseudo-label*. The *thresholded pseudo-labeling* of FixMatch has a similar effect to sharpening.

Weight Annealing: ReMixMatch anneals the weight while FixMatch doesn’t. *thresholded pseudo-labeling* in Fixmatch has similar effect.

Therefore, FixMatch considered as *simplified* version of UDA and ReMixMatch.
(pseudo-labeling and consistency regularization to replace sharpening, training signal annealing from UDA, distribution alignment and the rotation loss from ReMixMatch, etc.)

Ablation study

Sharpening and Thresholding

- Pseudo-labeling – 1 hyperparameter
- Sharpening+thresholding – 2 hyperparameter

In summary, swapping *pseudo-labeling* for *sharpening and thresholding* would introduce a new hyperparameter while achieving no better performance.

Augmentation Strategy

- Both CutOut and CTAugment are required to obtain the best performance
- Model diverged when replace weak augmentation for label guessing instead of strong augmentation, therefore, **pseudo-label** needs to be generated **using weakly augmented data**.

Ratio of unlabeled data

Using **large amount of unlabeled data** decreases the error rate significantly.

Scaling the learning rate η linearly with the batch size (a technique for large-batch supervised training) was effective for FixMatch, especially when μ is small.

Weight Decay

Tuning the weight decay is exceptionally important for low-label regimes.

Conclusion

Strategies that help FixMatch wins the game:

1. Pseudo-labeling with threshold
2. Augmentation (CutOut+ CTAugment or RandAugment)

Questions

1. What is the purpose of decaying learning rate?
2. Pseudo-labeling with threshold to prevent overfitting due to small amount of labeled data?

Reference

[FixMatch original paper]