

MixMatch: A Holistic Approach to Semi-Supervised Learning (NIPS 2019)

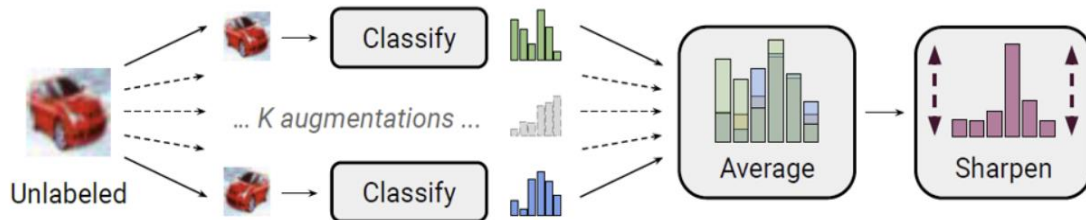


Figure 1: Diagram of the label guessing process used in MixMatch. Stochastic data augmentation is applied to an unlabeled image K times, and each augmented image is fed through the classifier. Then, the average of these K predictions is “sharpened” by adjusting the distribution’s temperature. See algorithm [1](#) for a full description.

MixMatch 의 주요 contribution 은 다음과 같다.

1. Consistency Regularization

하나의 이미지에 대해서 augmentation 을 여러 개 취하는 방법이다. 이렇게 얻어진 각기 다른 이미지들은 사실 같은 이미지로부터 파생된 것이다. 때문에 각 이미지들에 대한 예측 분포 또한 서로 유사해야 함이 당연하다. 즉, 특정 데이터를 약간 변형하여도 “일관되게” 예측해야 한다는 정의이다.

2. Entropy Minimization

unlabeled data 로부터 얻은 예측 entropy 를 최소화 하는 것으로, 예측에 확신을 부여하는 작업이다. MixMatch 논문에서는 Sharpening 기법을 사용한다. temperature T 값을 활용하여 그림과 같이 예측 분포에 대한 entropy 를 줄여준다. 해당 작업까지 끝난 예측 분포는 pseudo label 로 작용한다. 이때 one-hot vector 가 아닌 확률 분포이다.

3. Mixup

labeled data 와 unlabeled data 간의 Mixup 을 적용한다. 이는 모델의 일반화 성능을 향상 시키고, unlabeled data 에 대한 불확실한 예측에 대해 Label

Smoothing 효과가 발생되어 모델이 unlabeled data 에 대해 너무 큰 확신을 가지지 않게 한다.

Algorithm

1. Input: labeled data 의 batch X , unlabeled data 의 batch U , temperature T 값, number of augmentations K 값
2. Batch size 만큼 labeled data 와 unlabeled data 의 각 sample 을 순환한다. 이때 labeled data 에 대해 augmentation 을 1 번 진행하고, unlabeled data 에 대해 augmentation 을 K 번 진행한다.
3. K 번 augmentation 을 적용한 unlabeled data 에 대해 평균 값을 내고 sharpening 을 적용하여 pseudo label 을 형성한다.
4. 각 augmentation 을 적용한 data 들을 새로운 batch (각각 X' , U')로 정의하고, Concat 하고 섞어서 새로운 batch W 를 생성한다.
5. labeled data 는 $\text{MixUp}(X', W)$ 을 통해서 unlabeled data 는 $\text{MixUp}(U', W)$ 를 통해서 진행한다.

Remixmatch: Semi-supervised learning with distribution matching and augmentation anchoring

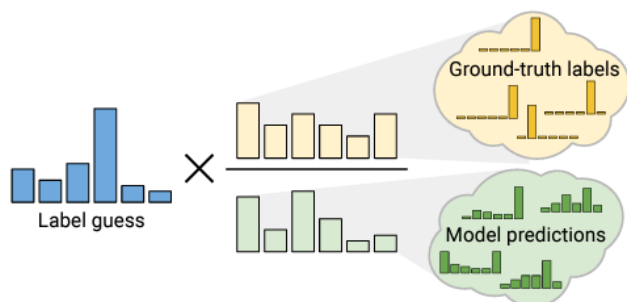


Figure 1: Distribution alignment. Guessed label distributions are adjusted according to the ratio of the empirical ground-truth class distribution divided by the average model predictions on unlabeled data.

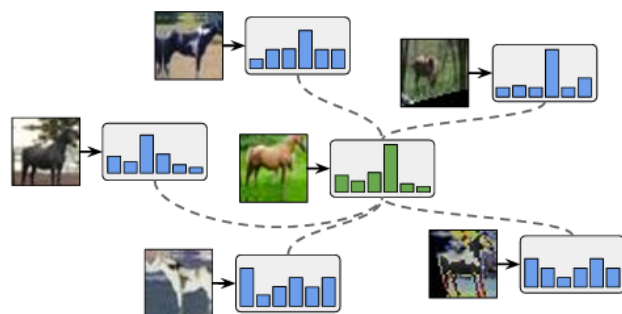


Figure 2: Augmentation anchoring. We use the prediction for a weakly augmented image (green, middle) as the target for predictions on strong augmentations of the same image (blue).

RemixMatch 는 MixMatch 의 후속 논문으로, 전반적으로 MixMatch 의 방법론을 토대로 하고 있다.

여기서 몇 가지 추가적인 방법론이 제안된다. Figure 1 을 보면 label guess 단계에서 ground truth labels 의 분포(이때는 labeled data 의 target 값을 의미) 대비 model predictions(unlabeled data sample 에 대한 예측)분포를 곱하게 된다. 이는 결국 위 두 개의 분포를 일치하겠다는 의미이다.

여기서 **상호 의존도**의 개념을 짚고 넘어가야 한다.

$$\begin{aligned}\mathcal{I}(y; x) &= \iint p(y, x) \log \frac{p(y, x)}{p(y)p(x)} dy dx \\ &= \mathcal{H}(\mathbb{E}_x[p_{\text{model}}(y|x; \theta)]) - \mathbb{E}_x[\mathcal{H}(p_{\text{model}}(y|x; \theta))]\end{aligned}$$

상호 의존도는 입력과 출력이 서로 얼마나 의존적인지 측정하는 지표이다. unlabeled data 의 관점에서 바라보았을 때, 입력 x 는 image, 출력 y 는 pseudo label 에 해당된다.

상호 의존도가 크다는 말은 y 에 대해 x 가 많은 정보를 제공한다는 말이고, 즉 unlabeled data에 대해 얼마나 잘 label을 예측했는지 나타내는 것이다.

결과적으로 값을 키워야 하기 때문에, 첫 번째 entropy 값을 키우고, 두 번째 entropy 값을 줄이는 방식으로 이를 해결할 수 있다. 기존 MixMatch에서 사용했듯 후자의 경우는 entropy minimization으로 sharpening 기법을 사용하여 이를 해결했다.

RemixMatch도 마찬가지로 sharpening을 사용하여 이를 해결했지만, 여기서 그치지 않고, 첫 번째 entropy를 극대화 하는 방식도 추가를 하였다. 이를 위해서는 모델이 예측한 분포가 평평하도록, 모든 class에 대한 예측이 동일한 값을 가져야 한다. 하지만 이 때 발생하는 문제점이 모델이 예측한 label의 분포가 실제 정답 분포와 다르다면 학습에 안 좋은 영향을 미친다고 한다. 따라서 앞서 언급한 새로운 label guess 방법으로 Distribution Alignment를 도입하게 된다.

RemixMatch에서는 한 가지 더 추가적인 방법론을 제시한다. 기존 MixMatch에서는 K 개의 Augmentation을 적용하여 Consistency Regularization을 해결하였다. 하지만 RemixMatch에서는 weak Augmentation, strong augmentation 단 두 개만 사용하여 이를 해결하였다.

전체적인 알고리즘은 다음과 같다.

Algorithm

1. Input은 MixMatch와 동일하게 적용된다.
2. Batch size를 돌면서 labeled data와 unlabeled data에 대해 Augmentation을 진행한다. labeled data는 strong, unlabeled data는 weak, strong 각각 적용한다.
3. unlabeled data의 weak augmentation을 거친 예측 값을 실제 label 분포로 distribution alignment 한다.
4. 이후 최종적으로 sharpening을 거치고 pseudo label이 생성된다.

Mixup의 과정은 MixMatch와 동일하지만, 한 가지 추가된 점이 있다면 strong augmented data의 batch를 별도로 내어 추가적인 loss를 계산한다는 것이다.