



# [논문 리뷰] Pseudo - Label [2013]

✨ Labeled data와 Unlabel data를 동시에 학습하는 신경망을 제안합니다.

예측 확률이 가장 높은 class를 가진 UnLabeled data( Pseudo-Labels )은 True labels 을 갖습니다.

이것은 Entropy Regularization과 동일한 효과를 갖습니다.



The proposed Network는 Labeled된 data와 Unlabeled된 data를 동시에 학습합니다.

## Pseudo-Label Method

Pseudo-Label은 unlabeled data에 관한 target classes입니다.

they were true labels. We just pick up the class which has maximum predicted probability for each unlabeled sample.

$$y'_i = \begin{cases} 1 & \text{if } i = \operatorname{argmax}_{i'} f_{i'}(x) \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

Pseudo-Label은 fine-tuning된 상황에서 Dropout과 함께 사용됩니다.

가중치가 변할 때마다 계속해서 계산되는 Psuedo-Labels은 지도 학습의 손실함수와 동일하게 사용됩니다.

Labeled data와 UnLabeled data의 총 개수가 꽤 차이가 있고, Neural Net의 학습에서의 균형은 상당히 중요하기 때문에, 전반적인 손실 함수는 아래와 같이 사용합니다.

$$L = \frac{1}{n} \sum_{m=1}^n \sum_{i=1}^C L(y_i^m, f_i^m) + \alpha(t) \frac{1}{n'} \sum_{m=1}^{n'} \sum_{i=1}^C L(y_i'^m, f_i'^m),$$

만약, 균형 계수  $a(t)$ 가 상당히 크면, Label data의 학습에 상당히 방해가 되고, 균형 계수  $a(t)$ 가 작으면, UnLabel data의 사용에 대한 benefit을 사용 불가능합니다.

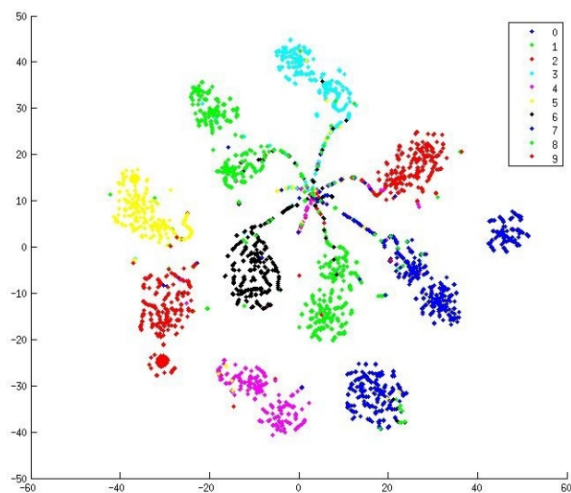
균형 계수  $a(t)$ 를 천천히 증가시킨다면,  
pool local minima를 피하기 위한 optimization process을 돕습니다.

$$\alpha(t) = \begin{cases} 0 & t < T_1 \\ \frac{t-T_1}{T_2-T_1} \alpha_f & T_1 \leq t < T_2 \\ \alpha_f & T_2 \leq t \end{cases} \quad (16)$$

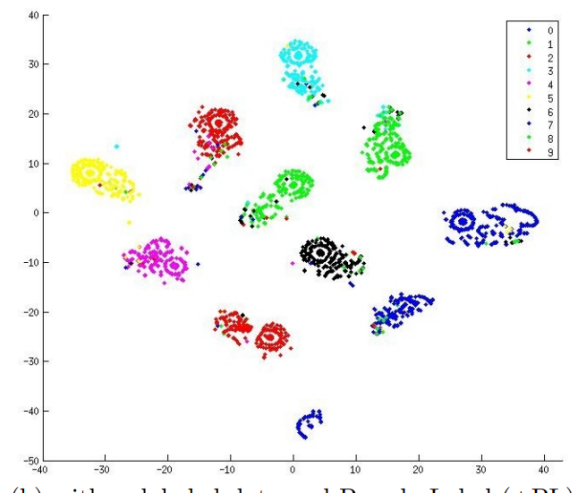
with  $\alpha_f = 3$ ,  $T_1 = 100$ ,  $T_2 = 600$  without pre-training,  
 $T_1 = 200$ ,  $T_2 = 800$  with DAE.

## Experiments

✨ 좌측의 경우, Pseudo Label을 사용하지 않았을 때,  
우측의 경우 Pseudo Label을 사용했을 경우입니다. [T-SNE 적용 예시]



(a) without unlabeled data (dropNN)



(b) with unlabeled data and Pseudo-Label (+PL)

✨ MNIST에서의 성능 예시입니다.

METHOD	100	600	1000	3000
NN	25.81	11.44	10.7	6.04
SVM	23.44	8.85	7.77	4.21
CNN	22.98	7.68	6.45	3.35
TSVM	16.81	6.16	5.38	3.45
DBN-RNCA	-	8.7	-	3.3
EMBEDNN	16.86	5.97	5.73	3.59
CAE	13.47	6.3	4.77	3.22
MTC	12.03	5.13	3.64	<b>2.57</b>
DROPNN	21.89	8.57	6.59	3.72
+PL	16.15	5.03	4.30	2.80
+PL+DAE	<b>10.49</b>	<b>4.01</b>	<b>3.46</b>	2.69

## Conclusion

단순하고, 효율적인 방식인 Semi-Supervised Learning을 연구를 통해 보여줬습니다.

Complex training scheme와 Computationally expensive similarity matrix 없이, SOTA를 달성하였습니다,

 <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.664.3543&rep=rep1&type=pdf>