

5 / 23(화)

Semi-supervised

→ 지도 학습에서 우리는 각 예제가 해당 대상 또는 출력 값과 쌍을 이루는 레이블이 지정된 데이터 셋을 가지고 있다. 목표는 레이블이 지정된 예제를 기반으로 입력 기능에서 올바른 출력으로의 매핑을 학습하는 것이다.

→ 비지도 학습에서는 레이블이 지정되지 않은 데이터 세트가 있으며, 목표는 데이터 내에서 기본 패턴, 구조 또는 관계를 찾는 것이다.

→ 준지도 학습은 지도 학습과 비지도 학습 사이에 있는 학습이다. 위의 두 가지 접근법의 요소를 결합한다. 레이블이 지정된 예제와 레이블이 지정되지 않은 예제가 모두 포함된 데이터 세트를 활용한다. 레이블이 지정된 예제는 올바른 출력에 대한 명시적인 정보를 제공하는 반면 레이블이 지정되지 않은 예제는 관련된 목표 값을 가지고 있지 않다. 준지도 학습은 레이블이 지정되지 않은 추가 데이터를 활용함으로써 모델이 더 강력한 표현을 학습하거나 데이터에서 더 의미 있는 패턴을 캡처하여 성능을 향상시킬 수 있다.

→ 이 아이디어는 레이블이 지정되지 않은 데이터가 더 나은 일반화 및 의사 결정에 도움이 될 수 있다는 추가 정보를 제공한다는 것이다. 준지도 학습 알고리즘은 일반적으로 레이블이 지정된 예제에서 레이블이 지정되지 않은 예제로 레이블 정보를 의미 있는 방식으로 전파하는 것을 목표로 한다.

Pseudo

유사 레이블링은 준지도 학습에 일반적으로 사용되는 기술이다. 레이블이 지정된 데이터에 대해 훈련된 모델의 예측을 기반으로 레이블이 지정되지 않은 데이터 지점에 레이블을 할당하는 작업이 포함된다. 레이블이 지정되지 않은 데이터에 대해 “의사 레이블”, “가짜 레이블” 이라고 한다. 유사 레이블은 레이블이 지정된 데이터 셋을 보강하는데 사용되며, 추가 교육을 위해 레이블이 지정된 데이터와 유사 레이블이 지정된 데이터를 함께 결합한

다. 이 방식을 통해 모델은 레이블이 지정된 예제와 유사 레이블이 지정된 예제를 모두 학습하여 성능을 향상 시킬 수 있다.

Mixmatch

주어진 레이블 된 데이터로부터 data augmentation을 통해 새 데이터를 생성, 그 후에 레이블이 없는 데이터에 대한 data augmentation을 진행, 이 때 label이 없는 데이터의 pseudo-label을 augmentation된 데이터에 대한 예측 평균값에 sharpening을 하여 사용. 그 후에는 augmentation 된 데이터들을 섞은 후 레이블 된 데이터들끼리 mixup, pseudo-label 된 데이터들끼리 mixup을 하여 학습에 사용할 데이터를 생성(Data augmentation + self-training + sharpening(=entropy minimization) + mixup)

ReMixmatch

Mixmatch + Distribution alignment, Augmentation anchoring