

제로샷·원샷·퓨샷

간단하게

• Zero-shot learning (ZSL)

- 모델이 학습과정에서 본적 없는 새로운 클래스를 인식할 수 있도록 하는 학습 방법이다.

↳ 모델이 클래스간의 관계나 속성을 통해 일반화하는 능력을 활용한다.

(예) CLIP같은 VLM은 이미지와 자연어를 같이 학습했기 때문에 여러 '강아지'에 대해 학습하지 않아도

강아지 사진과 '잔디밭 위의 강아지' 문장의 유사도를 비교해서 '강아지'를 예측할 수 있다.

• One-shot learning (OSL)

- 모델에게 특정 클래스를 인식할 수 있게 단 한 예시 이미지를 보여주고 이후에 특정 클래스를 분류할 수 있게 된다.

(예) 강아지 사진과 '이 동물은 강아지야' 즉게 되면 모델은 '강아지'와 유사한 것을 찾아낸다.

↳ 내부적으로는 즉각 한 예시 이미지와 새로운 이미지의 임베딩 벡터간 거리를 비교해서 가까우면 '강아지'라고 예측한다.

• Few-shot learning (FSL)

- 딱히 설명 할게 없다. 그저 OSL에서 예시 개수만 좀 더 늘어난 것이 끝이다.

• OSL과 FSL은 어떻게 사용하는지는 아래를 보면 이해할 수 있다.

OSL, FSL과 일반적인 지도학습 차이

• 일반적인 과잉학습

- 데이터를 노아서 학습하고 테스트 진행

• OSL과 FSL

~ 이것들은 사용방법이 또 2개로 나뉜다.

① 파인튜닝 방식 (대표: MAML, ANIL)

• Support set (예시 이미지)를 학습하기 위해 준비한다.

• 모델 전체가 아닌 일부 계층만 학습한다.

(예) 다지 막 볼클레이어 (FC, head) 또는 Adapter, Prompt tuning 계층만 학습한다.

• 처음 모델은 LR를 작게 기존의 지도 학습보다 적게 학습이 진행될다.

• 이후에 워크에서 준비한 데이터와 레이어 부분에 조금 더 큰 LR로 몇번만 학습한다.

② 유사도 비교 방식 (대표: Siamese Network, clip, blip)

• input 이미지와 Support 이미지들 예측할 때 준다

• 두 이미지간의 유사도를 비교 판단 한다.

• 두가지 방식을 다할수있는 애들도 있다. (clip, blip)