

## 음성 중첩 데이터 분류 1등 코드

Torch.cuda.amp : AMP(Automatic Mixed Precision) : mixed precision training이란 처리 속도를 높이기 위한 16bit floating point 연산과 정확도 유지를 위한 32bit floating point 연산을 섞어 학습하는 방법이다. (8배의 연산 처리량&2배의 메모리 처리량 효과) 최적화되어 배치를 늘릴 수 있기 때문에 학습 속도도 빨라지고 배치 뿐만 아니라 모델 최적화도 이루어지기 때문에 속도가 증가함.

Multi-resolution mel-spectrogram: 사람의 귀를 칼라맵인 스펙트로그램에 반영하는 것. 고주파로 갈수록 사람이 구분하는 주파수 간격이 넓어지는 원리를 이용하여 필터를 이용하여 스케일 단위를 변환하는 것.

Hop, window size : 데이터 만들 수 있음. Ex) 10초 짜리 데이터 window size는 1초, hop size는 0.5초이면 20개의 데이터가 나옴. 데이터수=데이터초/(window size\*hop size), window size는 푸리에 변환하는 단위

Tqdm: 연속적인 작업을 수행할 때 진행률을 알고 싶을 때.

Frequencymasking: 주파수와 시간 축으로 이루어진 spectrogram의 주파수 축을 따라 일정 영역을 0으로 마스킹

Timemasking : 시간 축을 따라 일정영역을 0으로 마스킹

데이터 augmentation : 데이터를 부풀려서 모델의 성능을 향상시키는 기법으로 이미지 인식 분야에서 많이 쓰이고 좌우 반전, 사진의 일부 발체, 밝기 조절 등을 하여 한정된 데이터를 조금씩 변형시켜 새로운 데이터처럼 활용하는 방법. Augmentation을 하면 대부분 성능이 향상되고 방법이 간단하고 패턴이 정해져있기 때문에 단순하다. 오디오에 대해서는 noise를 추가해주거나 음성 신호를 좌/우 shift하고 빈 공간은 0으로 채우는 방법, 음높이/주파수를 랜덤하게 변경, 속도를 빠르거나 느리게 변경하는 방법으로 데이터를 변형시킴.

Pitch shifting : 소리의 원래 피치를 높이거나 낮추는 기술. 미리 지정된 음악 간격(조 옮김)만큼 피치를 높이거나 낮춤.

Random crop : 원본 이미지의 무작위 하위 집합을 만드는 데이터 확대 기술.

Mixup: 두 데이터의 이미지와 label을 각각 weighted linear interpolation하여 새로운 sample을 생성하는 기법

KL-Divergence: 상대적 엔트로피

AMP(Automatic Mixed Precision) : mixed precision training이란 처리 속도를 높이기 위한 16bit floating point 연산과 정확도 유지를 위한 32bit floating point 연산을 섞어 학습하는 방법이다. (8배의 연산 처리량&2배의 메모리 처리량 효과) 최적화되어 배치를 늘릴 수 있기 때문에 학습 속도도 빨라지고 배치 뿐만 아니라 모델 최적화도 이루어지기 때문에 속도가 증가함.