**COSE361\_assignment2\_2022100124권서은**

1. **Definition of your translation task**

목소리 변환을 구현하려고 하였습니다.

1. **Description of the source code that you implemented or downloaded**
2. preprocess-vcc2018.py

이는 전처리를 구현하기 위한 코드입니다.

speaker\_list에 있는 화자를 다 도는 loop문을 실행합니다. 이부분은 corpus\_dir 내부의 모든 wav파일의 목록을 flist.txt 파일에 저장합니다.

텍스트, 스크린샷, 소프트웨어이(가) 표시된 사진

자동 생성된 설명

이 부분은 flist.txt에 저장했던 리스트에 있는 음성파일들을 로드하고 전처리 하는 부분입니다. dtype(데이터타입)이 train인지, 아닌지에 따라서 다르게 전처리해 저장합니다.

1. train\_base.py

각 에포크마다 학습 데이터를 통해 모델을 학습시키고 손실을 계산합니다. 또한, 검증 데이터를 활용하여 모델을 평가하고 손실을 기록합니다.

1. train\_further.py

역시 각 에포크마다의 학습 데이터를 통해 모델을 학습시키고, 손실을 계산합니다.

1. convert.py

이는 녹음 파일을 목소리 화자별로 전환하는 로직을 구현합니다.

1. calculate.py

해당 파일에서는, total mcd와 msd, m2m, f2m별 mcd와 msd를 구하는 로직을 펼치고 있습니다.

1. **Description of the training/validation/test data that you used**

/corpus 폴더 내부에서 test와 train 폴더를 나누어 우선 test data와 train data를 넣었습니다.

스크린샷, 텍스트, 디자인이(가) 표시된 사진

자동 생성된 설명 다음은 dataset의 폴더 구조입니다. train data는 훈련 모델을 학습시키는 데 사용되는 데이터입니다. 따라서 train data에서 제 목소리 (여자)를 source와 target으로 주었고, 다른 남성의 목소리 역시 source와 target으로 주었습니다. 이번 과제에서 사용할 source는 제 목소리이고, target은 이미 주어진 남성의 목소리이나, 우선 일관성을 위해서 다음과 같이 구현하였습니다. 제 목소리 (여자) 주어진 남성의 목소리는 동일한 내용을 다루고 있습니다. 이후 이들은 전처리되어 이후 학습에 활용됩니다.

텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

자동 생성된 설명다음은 validation data입니다. 따로 폴더를 두지 않았고, 전처리 과정에서 dev\_list로 train data의 일부분을 넣어 validation dataset을 설정했습니다. validation data란 훈련 중 모델의 성능을 평가하는 데 사용되며, train data와는 다른 내용을 가져야합니다.

test data는 학습 이후, 최종적으로 모델의 성능을 평가하기 위해 사용되는 데이터입니다. 역시 train data와 validation data와 겹치지 않는 내용을 담고 있으며, 모델의 최종 일반화 성능을 측정하는 데 사용됩니다. Test data는 test 폴더에 구현하였으며, 역시 VCC2SF1~4는 제 목소리이고, VCC2SM1~4는 제공된 다른 성별 (남성)의 목소리입니다.

1. **Description of the compiling, training, validation, and testing procedures**

**텍스트, 스크린샷이(가) 표시된 사진

자동 생성된 설명**다음은 전처리 과정입니다. 앞서 나온 preprocess-vcc2018.py의 speaker\_list의 화자의 데이터를 전처리합니다.

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

이는 training, validation, testing 과정입니다. 여기서 생성되는 vae\_exp\_0520.txt는 다음과 같은 내용을 포함합니다.

텍스트, 스크린샷, 폰트, 문서이(가) 표시된 사진

자동 생성된 설명 이부분은 training과 validation part입니다. 각 epoch마다 발생하는 train loss와 dev loss를 나타내고 있습니다.

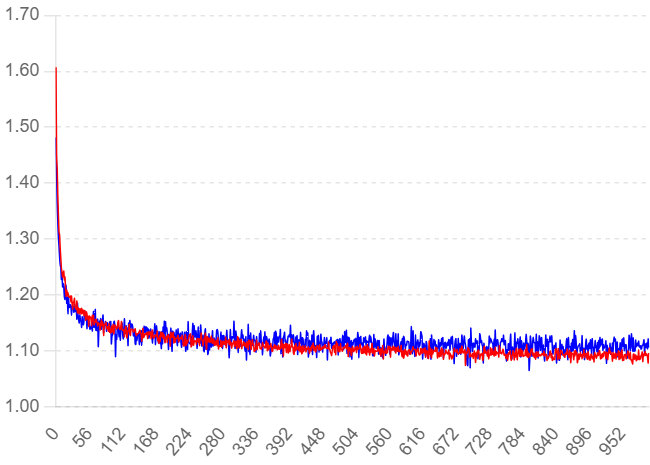
텍스트, 스크린샷, 폰트, 문서이(가) 표시된 사진

자동 생성된 설명 이 부분은 conversion part 입니다. 음성 변환 모델을 사용하여 입력 음성 파일을 변환된 음성 파일로 변환하고 있는 단계입니다. Convert된 파일들은 result>VAE3에서 “VCC2SF1\_to\_VCC2SM2”와 같은 형식으로 저장됩니다.

1. **Resulting graphs of the hyperparameter tuning experiments**

스크린샷, 텍스트, 그래프이(가) 표시된 사진

자동 생성된 설명해당 그래프는 hyperparameter수정 없이 모든 과정을 마쳤을 때 나오는 그래프입니다. 가로축은 epoch이며, 세로축은 loss를 나타냅니다. 붉은색은 Train Loss를 나타내며, 푸른색은 Dev Loss를 나타냅니다. Train loss는 모델이 training data에서 얼마나 잘 학습하였는지 나타내며, Dev Loss는 모델이 validation data에서 얼마나 잘 일반화했는지, 모델 학습 성능을 평가하기 위해 사용됩니다. 이상적인 상태로는 train loss와 dev loss는 비슷한 값을 가져야 하기 때문에, dev loss를 나타내는 파란색이 train loss를 나타내는 빨간색에 근접하도록 수정해야 합니다.

 해당 그래프는 hyperparameter 수정을 거친 뒤 모든 과정을 마쳤을 때 나오는 그래프입니다. 가로축은 epoch이며 세로축은 loss를 나타냅니다. 붉은색은 Train Loss를 나타내며, 푸른색은 Dev Loss를 나타냅니다. 여기서는 rec (reconstruction loss)를 1.5, cyc (cycle consistency loss)를 0.5, kl (kl-divergence loss)를 0.05로 수정했습니다. Rec loss의 가중치를 높여, 모델이 입력 데이터를 재구성하는 데 더 많은 중요성을 부여했습니다. 이는 원본이 입력 데이터를 정확하게 재구성하도록 합니다. Cyc loss의 가중치를 낮춰, 모델의 순환 일관성에 대한 의존성을 줄였습니다. 이후, kl divergence loss의 가중치를 낮춰, 잠재 공간 분포의 정규화 강도를 낮췄습니다.

이렇게 hyperparameter 수정을 거쳐, train loss와 dev loss가 비슷한 값을 가지도록 구성할 수 있었습니다.

1. **A sample set of source (input), target, and converted (output) test data**

train data를 위해 corpus>train 폴더에 이미 제공되어 있던 transcript를 파일별로 읽어 81개의 제 목소리 source 파일을 구현했습니다. (파일명: VCC2SF1, VCC2SF2, VCC2SF3, VCC2SF4). 또한, 동일한 폴더에 동일한 transcript를 다루는 남성의 목소리를 target data로 준비했습니다. (파일명: VCC2TM1, VCC2TM2) 이후 result>VAE3 폴더에 “VCC2SF2\_to\_VCC2SM1”와 같은 파일들이 있는데, 이는 VCC2SF2화자의 목소리에서 VCC2SM1화자의 목소리로 converted된 test data를 의미합니다.

추가로, 첨부하는 파일은 3번째 파라미터를 수정한 뒤 학습을 시키다 과제 임박이 마감하여 학습이 완료되지 못한 상황입니다. 부디 참고 부탁드립니다.