



2021.09.01 - 2021.11.30

KOR_STT EMR-PROJECT

김홍범

CONTACT
billkim418@naver.com

THINKING INFINITY
MAKE BETTER WORLD

CONTENTS

Thinking Infinity

About ERROR

01

PATH ERROR

라이브러리 실행을 위한 파이썬 경로설정

02

AUDIO OPEN ERROR

오디오 파일이 열리지 않는 문제

03

AUDIO NONE ERROR

오디오 파일 매칭이 되지 않는 문제

04

ZERO DIVISION ERROR

오디오 경로 설정 오류로 인한 변수 선언 문제

BACKGROUND

ERROR

KoSpeechPATH 설정

01. Solution

1. kospeech/models 안의 BeamDecoderRNN 경로 변경

-> from kospeech.models.beam_search import BeamDecoderRNN

2. local 패키지 로드 경로 추가하기

vim .profile

아래 문구 추가하기

-> export PYTHONPATH="\${PYTHONPATH}\$경로(kospeech 경로)"

source .profile

BACKGROUND

ERROR

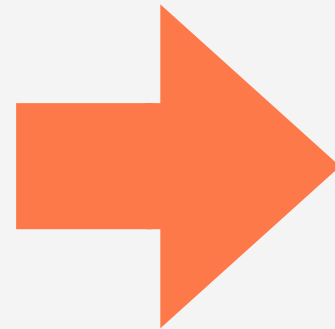
Audio file open error : pcm

01. Solution

데이터셋 : 한국어음성 데이터(kspon speech)

코덱(포맷) : 16kHz/16bits, headerless (little endian) linear PCM

문제 : 음성 데이터 재생이 불가능했음



기존의 오디오 프로그램(팟 플레이어, 곰 플레이어)가 아닌 audicty, 음성 편집 프로그램을 사용

BACKGROUND

ERROR

ValueError: not enough values to unpack
(expected 3, got 0)

01. image

```
Error executing job with overrides: ['model=ds2', 'train=ds2_train', 'train.dataset_path=/home/gitlabmaster/anaconda3/envs/STT_KOR/kospeech/data/kspon']
Traceback (most recent call last):
  File "./bin/main.py", line 162, in main
    last_model_checkpoint = train(config)
  File "./bin/main.py", line 89, in train
    epoch_time_step, trainset_list, validset = split_dataset(config, config.train.transcripts_path, vocab)
  File "/home/gitlabmaster/anaconda3/envs/STT_KOR/kospeech/kospeech/data/data_loader.py", line 359, in split_dataset
    validset = SpectrogramDataset(
  File "/home/gitlabmaster/anaconda3/envs/STT_KOR/kospeech/kospeech/data/data_loader.py", line 73, in __init__
    self.shuffle()
  File "/home/gitlabmaster/anaconda3/envs/STT_KOR/kospeech/kospeech/data/data_loader.py", line 126, in shuffle
    self.audio_paths, self.transcripts, self.augment_methods = zip(*tmp)
ValueError: not enough values to unpack (expected 3, got 0)
```

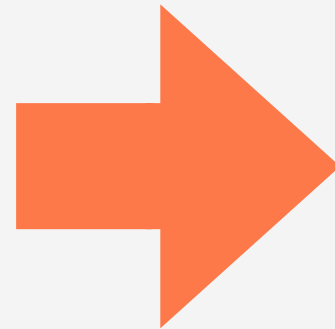
BACKGROUND

ERROR

ValueError : not enough values to unpack
(expected 3, got 0)

02. Cause

원인 : shuffle 함수를 구현 함에 있어서 zip 함수가 3가지 인자들을 호출하지 못하고 있음



-> spec Augmentation의 오류
-> train, valid set 선언 문제(data_loader.py)

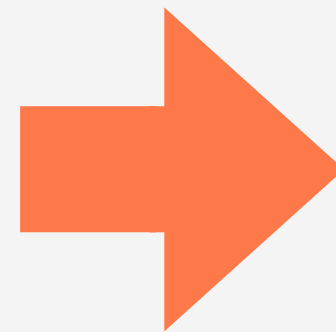
BACKGROUND

ERROR

ValueError : not enough values to unpack
(expected 3, got 0)

03. Solution

```
if config.train.dataset == 'kspon':  
    train_num = 8 #620000 # train 데이터 갯수  
    valid_num = 2 #2545 # validation 데이터 갯수  
elif config.train.dataset == 'libri':  
    train_num = 281241  
    valid_num = 5567
```



데이터셋 전체를 쓰지 않고 부분만을 학습하므로 train_num부분의 수치를 변경해줘야만 한다.

-> train num : train 데이터 숫자

-> valid_num : 검증 데이터 숫자

Train + valid -> 총 데이터 숫자

Ex) $8 + 2 = 10$

BACKGROUND**ERROR**

Audio is None error

01. Image

```
[2021-10-07 10:29:10,924][kospeech.utils][INFO] - Audio is None : $/home/gitlabmaster/anaconda3/envs/STT_KOR/kospeech/data/kspon/KsponSpeech_01/KsponSpeech_0001/KsponSpeech_000006.pcm
```

02. solution

오디오 signal 파일을 생성하지 못했는데 그 이유는 경로 설정이 잘못 되어 있었기 때문이다.
python ./bin/main.py model=ds2 train=ds2_train train.dataset_path=/home/gitlabmaster/anaconda3/envs/STT_KOR/kospeech/data/kspon/

BACKGROUND

ERROR

ZeroDivisionError : float division by zero

01. Image

```
Error executing job with overrides: ['model=ds2', 'train=ds2_train', 'train.dataset_path=/home/gitlabmaster/anaconda3/envs/STT_KOR/kospeech/data/kspon']
Traceback (most recent call last):
  File "./bin/main.py", line 162, in main
    last_model_checkpoint = train(config)
  File "./bin/main.py", line 122, in train
    model = trainer.train(
  File "/home/gitlabmaster/anaconda3/envs/STT_KOR/kospeech/kospeech/trainer/supervised_trainer.py", line 164, in train
    model, train_loss, train_cer = self._train_epoches(
  File "/home/gitlabmaster/anaconda3/envs/STT_KOR/kospeech/kospeech/trainer/supervised_trainer.py", line 326, in _train_epoches
    return model, epoch_loss_total / total_num, cer      #total_num error occurred
ZeroDivisionError: float division by zero
```

BACKGROUND**ERROR**

ZeroDivisionError : float division by zero

02. cause

1. total_num = 0으로 설정되어 있음
2. total_num += int(input_lengths.sum()) 에서 input_lengnth를 불러오지 못함

03. Solution

1. 오디오 경로 설정 필요 - 오류 혹은 안되어 있을 가능성이 있음
2. input_lengths 생성 후 error 해결

BACKGROUND

ERROR

TypeError: __init__() got an unexpected keyword argument 'require'

01. Image

```
Traceback (most recent call last):
  File "./bin/inference.py", line 50, in <module>
    parser.add_argument('--model_path', type=str, require=True)
  File "/home/gitlabmaster/anaconda3/lib/python3.8/argparse.py", line 1368, in add_argument
    action = action_class(**kwargs)
TypeError: __init__() got an unexpected keyword argument 'require'
```

BACKGROUND**ERROR**

TypeError: __init__() got an unexpected keyword argument 'require'

02. Solution

Spelling error

require -> required로 수정해주면 실행이 된다.

BACKGROUND

ERROR

type error : input audio type is must .pcm

01. Image

```
(STT_KOR) gitlabmaster@rollex:~/anaconda3/envs/STT_KOR/data/test_data$ file 20210906_01.wav
20210906_01.wav: RIFF (little-endian) data, WAVE audio, IMA ADPCM, mono 8000 Hz
```

02. Solution

.wav 타입을 .pcm 파일 타입으로 변경해주어야만 한다.

-> wav파일일 경우에 다른 언어(영어)로 인식됨

BACKGROUND

ERROR

DeepSpeech2 -> greedy_search가 존재하지 않음

01. Image

```
Traceback (most recent call last):
  File "./bin/inference.py", line 71, in <module>
    y_hats = model.greedy_search(feature.unsqueeze(0), input_length, opt.device)
  File "/home/gitlabmaster/anaconda3/lib/python3.8/site-packages/torch/nn/modules/module.py", line 778, in __getattr__
    raise ModuleAttributeError("'{}' object has no attribute '{}'.format(
torch.nn.modules.module.ModuleAttributeError: 'DeepSpeech2' object has no attribute 'greedy_search'
```

01. Image

함수를 model의 내장함수인 recognize로 수정해주면 된다.

ex) model.recognize()

Training result

KOR STT

Result

위에서는 여러가지의 오류를 통해 모델이 돌아가게끔 수정하였다. 앞으로는 여러 파라미터들을 수정하여 좀 더 나은 방향으로 학습이 되게끔 진행할 것이며, 추후 이 모델을 이용하여 test를 진행할 것이다.

BACKGROUND

ERROR

Train 오류(txt 한글 데이터 인코딩 변경)

01. Image

```
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
  File "/home/gitlabmaster/anaconda3/lib/python3.8/site-packages/numpy/core/memmap.py", line 236, in __new__
    raise ValueError("Size of available data is not a "
ValueError: Size of available data is not a multiple of the data-type size.
```

01. Explanation

find . -type f -execdir iconv -c -f euckr -t utf-8 {} -o {}.tmp \; -execdir mv -f {}.tmp {} \;

위와 같은 조건을 주고 인코딩을 변경 하니 특정 pcm 파일의 signal이 생성 되지 않는 오류가 발생하였다.

BACKGROUND

ERROR

Train 오류(txt 한글 데이터 인코딩 변경)

01. Solution

```
iconv: 0 위치에 잘못된 입력 순서열이 있음
iconv: 2521 위치에 잘못된 입력 순서열이 있음
iconv: 8775 위치에 잘못된 입력 순서열이 있음
iconv: 2808 위치에 잘못된 입력 순서열이 있음
iconv: 4221 위치에 잘못된 입력 순서열이 있음
iconv: 5558 위치에 잘못된 입력 순서열이 있음
```

특정 파일에 관해 파일의 인코딩 혹은 내용이 변경된 것을 확인

-> 기존 코드로 인코딩을 변경할 시에 전혀 관계 없는 pcm 파일도 같이 인코딩이 변경

특정 txt파일만을 변경하는 코드를 새롭게 짜 해결하였다.

```
find . -name '*.txt' -execdir iconv -c -f euckr -t utf-8 {} -o {}.tmp \; -execdir mv -f {}.tmp {} \;
```

BACKGROUND

ERROR

Checkpoint를 이용한 학습 오류 해결

01. Image

```
def get_latest_checkpoint(self):  
    """  
        returns the path to the last saved checkpoint's  
        subdirectory.  
        Precondition: at least one checkpoint has been made (i.e.,  
        latest checkpoint subdirectory exists).  
    """  
    checkpoints_path = sorted(os.listdir(self.LOAD_PATH),  
reverse=True)[1]  
    print("checkpoint path : ", checkpoints_path)  
    sorted_listdir =  
sorted(os.listdir(os.path.join(self.LOAD_PATH, checkpoints_path)),  
reverse=True)  
    print("sorted_listdir : ", sorted_listdir)  
    return os.path.join(self.LOAD_PATH, checkpoints_path,  
sorted_listdir[0])
```

02. Solution

sorted_listdir의 경로가 가장 최신의 모델을 불러오게끔 되어 있지만 특정 그 전 모델을 불러오고 싶은 상황
이므로 이 값을 변경하여 해결하였다.

BACKGROUND

Default

1000개의 문장을 이용하여 학습

01. Parameters

trainer

num_epoch : 70

batch_size : 1

save_result_every : 1000

checkpoin_every : 5000

print_every : 10

device

num_worker : 4

use_cuda : True

optim

optimizer: 'adam'

init_lr: 1e-06

final_lr: 1e-06

peak_lr: 1e-04

init_lr_scale: 0.01

final_lr_scale: 0.05

max_grad_norm: 400

BACKGROUND

Result

training set : 1000

epochs : 70

batch size : 1

num_worker : 4

01.Validation set

targets	predicts
아 맞다. 우리 대학교 가야 되지.	그 그 그리 그.
아 너가 입어서 이쁘진 않았다고?	아 하 하다 거..
주말마다 나와	주말마다 나와

1000개의 문장만 학습시키다보니 predicts 예측 결과가 매우 좋지 않게 나온 것을 확인할 수 있었고, 추후 이를 보완하기 위해 시간이 걸리더라도 ai hub에서 제공되는 모든 데이터셋을 이용해 학습을 진행해보았다.

BACKGROUND

Default

40만개의 모든 데이터를 학습시킴
batch size : 1 -> 시간이 많이 소모 되었다.

01.Validation set(32 epochs)

targets	predicts
뭐 먹었어?	뭐 먹었어?
아 절반 정도	아 잘 반정도?
아마 배에 있는 시간이 있으니까 좀 그게 수당이 있는 나오는 거 같애.	아 에에 있는 시간이어 있었으니까 좀 그게 수당이 있는 나오는 거 같애.

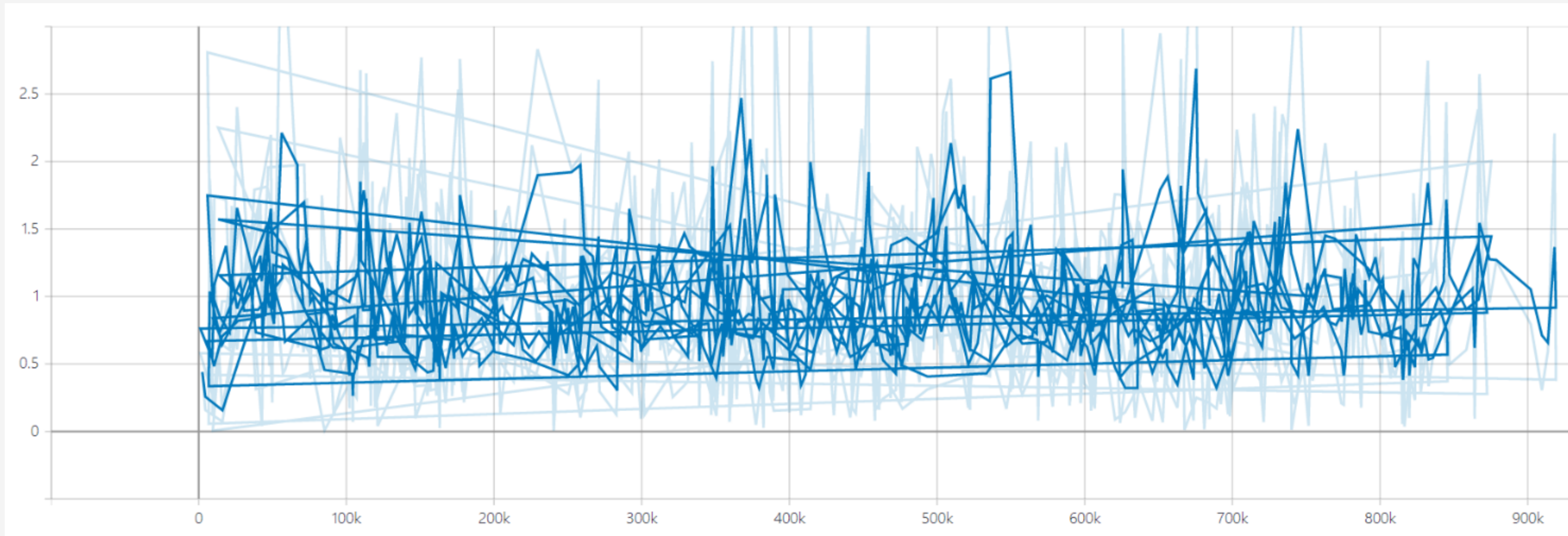
많은 데이터 셋을 가지고 학습을 진행하다보니 1000개의 데이터셋으로 학습시켰을 때와 비교하여 월등히 나
은 결과가 도출된 것을 볼 수 있었다.

BACKGROUND

Visualization

Tensorboard를 이용하여 학습의 진행에 따른 loss 및 cer의 변화 정도를 살펴보았다.

01.Result_step_loss



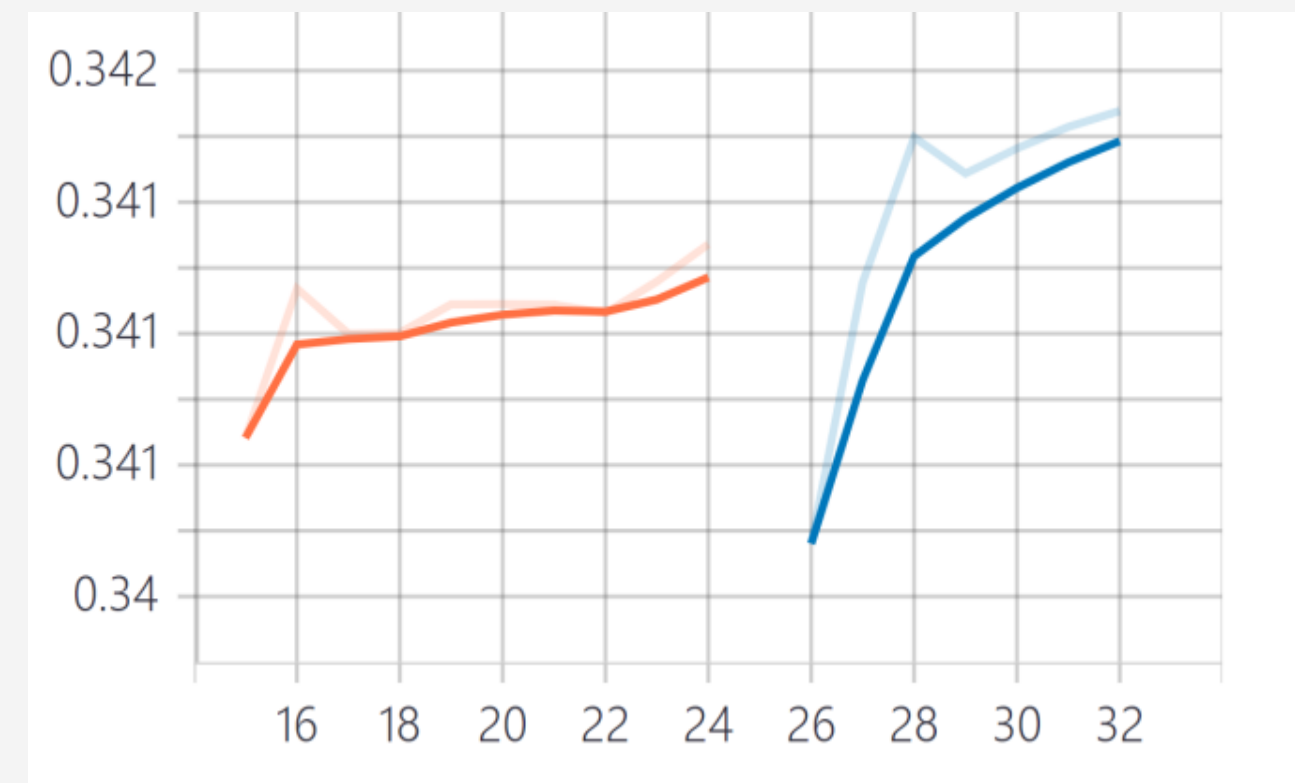
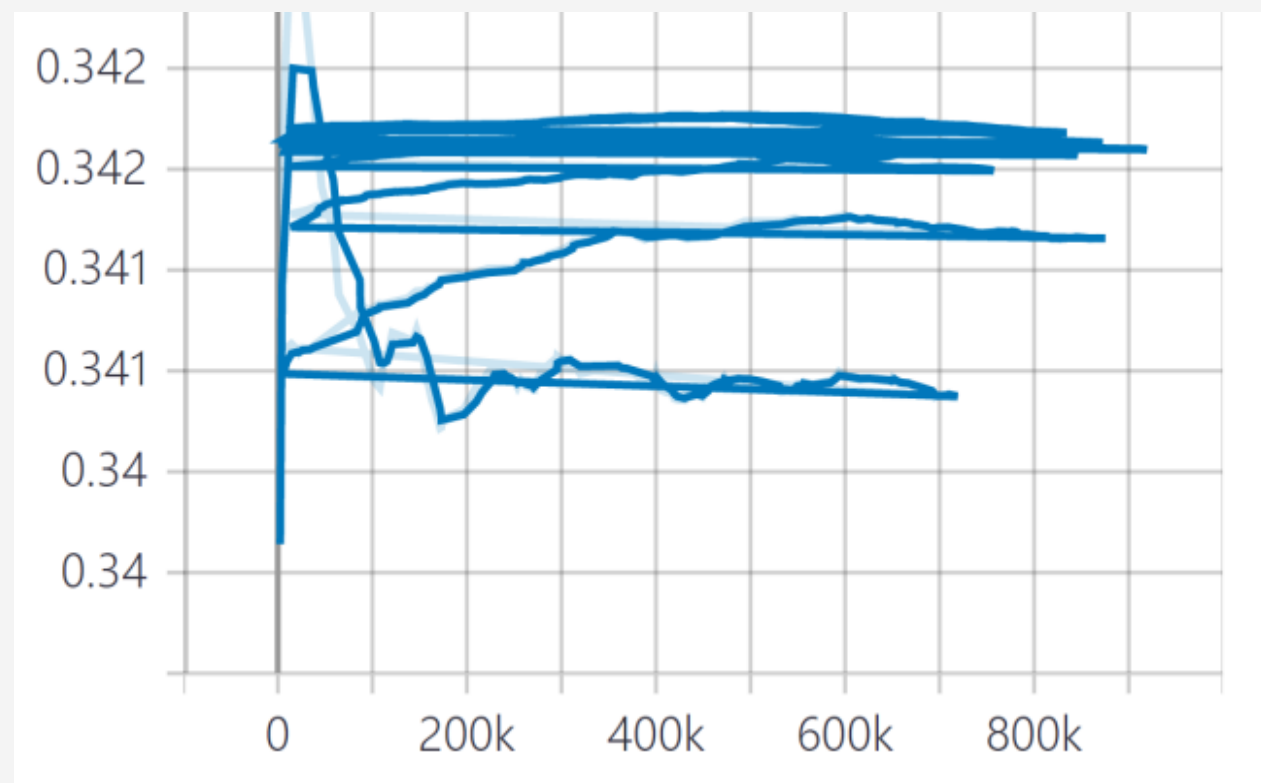
step별 loss를 시각화해 보았다. loss값은 변동폭이 심함을 알 수 있었다.

BACKGROUND

Visualization

Tensorboard를 이용하여 학습의 진행에 따른 loss 및 cer의 변화 정도를 살펴보았다.

01.Result_step_loss



왼쪽은 step별 cer 값인데 0.341 근처에 수렴함을 볼 수 있었고, 전체적인 epoch의 진행에 따른 cer 역시 0.341 근처로 수렴함을 시각화를 통해 확인할 수 있었다.

BACKGROUND

Conclusion

학습을 진행하며 loss, cer을 확인한것까지는 좋았으나 전체적인 DeepSpeech2의 단점을 파악하고 이를 개선하지 못한 부분이 아쉽다.

처음 진행해본 음성인식 프로젝트에서 리눅스를 활용하고 PyTorch로 학습까지 진행해보았다.

기본적인 Training까지의 과정은 매끄러웠으나, 기술 스택(PyTorch)의 기본기 부족과 논문의 구현과정에서 시간을 너무 소모하였다.

따라서, 다음 번 단계로는 DeepSpeech2의 기술적 한계를 파악하고 이를 극복한 다른 라이브러리를 사용할 것이다. 만약 최신 구현된 방법이 없다면 로직을 개선하는 방식으로 진행해 볼 것이다.



20XX. 08 - 20XX. 09

Thinking Infinity

THANK YOU.

CONTACT

hongbum.kim@sobetec.com

MAKE BETTER WORLD