Unsupervised ASR

Team A
吳冠緯 梁正

Outline

- Problem 1 result

- Problem 2 result

Problem 1 result

ckpt \ WER	LibriSpeech	WikiText	Image Caption
203	87.13272	86.88652	85.62535
686	27.38089	87.8359	81.00591
1012	23.72802	87.23213	78.71259
best	22.72398	88.01132	81.46864
last	25.14743	87.12439	78.21447

Probelem 1 result

Librispeech (1686 epoch)

```
REF: N AO R W AA Z DH IH S IH G Z AE K T L IY DH AH SH EY P DH AH TH IH NG T UH K T UW DH AH K AA
HYP: N AO W AA Z IH S IH K S AE K L IY DH AH SH EY P DH AH TH IH NG K T UH K T UW DH AH K AA N SH
```

Wikitext (1686 epoch)

```
REF: N AO R W AA Z DH IH S IH G Z AE K T L IY DH AH SH EY P DH AH TH IH NG T UH K T UW DH AH K AA
HYP: M T OW W AA Z AH N Y AH S IH T AH Y UW Z P IH K S IY Y UW F ER S T L AO S K ER Y UW S EH N Y
```

Image Caption (1686 epoch)

```
REF: N AO R W AA Z DH IH S IH G Z AE K T L IY DH AH SH EY P DH AH TH IH NG T UH K T UW DH AH K AA N SH AH
HYP: Y AA V ER HH ER P AH M P UW M D IH NG AH Z SH AE S HH AH HH AW N D S T EY JH S T IH N AH M AA N SH AH
```

Problem 2 result

- learning rate
- size of kernel, dimension, depth
- loss weight

Original

- **last training epoch :**

  REF: N AO R W AA Z DH IH S IH G Z AE K T L IY DH AH SH

  HYP: B AW N AO W AA Z IH S IH K S AE K L IY DH AH SH EY

  PER(Phoneme Error Rate) = 23.059

- **evaluation :**

  WER(Word Error Rate) = 22.191

Problem 2 result

- **learning rate**

- size of kernel, dimension, depth

- loss weight

learning rate

generator_lr : 0.0004 -> 0.0003

- **last training epoch :**

  REF: N AO R W AA Z DH IH S IH G Z AE K T

  HYP: N AO W AA Z IH S IH G Z AE K L IY DH

  PER = 22.841 ⇓

- **evaluation :**

  WER = 21.756 ⇓

learning rate

generator_lr : 0.0004 -> 0.0005

- **last training epoch :**

  REF: N AO R W AA Z DH IH S IH G Z AE K T

  HYP: N AO W AA Z IH S IH G Z AE K L IY DH(nearly same)

  PER = 23.781 ⇑

- **evaluation :**

  WER = 23.560 ⇑

learning rate

discriminator_lr : 0.0005 -> 0.0004

- **last training epoch :**

  REF: N AO R W AA Z DH IH S IH G Z AE K T

  HYP: N AO W AA Z IH S IH G Z AE K L IY DH(nearly same)

  PER = 23.570 ⇑

- **evaluation :**

  WER = 21.633 ⇓

learning rate

discriminator_lr : 0.0005 -> 0.0006

- **last training epoch :**

  REF: N AO R W AA Z DH IH S IH G Z AE K T

  HYP: AY N AO W AA Z IH S IH G Z AE K L IY

  PER = 24.424 ⇑

- **evaluation :**

  WER = 23.138 ⇑

Problem 2 result

- learning rate
- **size of kernel, dimension, depth**
- loss weight

size of kernel

discriminator_kernel : 6 -> 4

- **last training epoch :**

  REF: N AO R W AA Z DH IH S IH G Z AE K T

  HYP: N AO W AA Z DH S IH JH AE K L IY DH AH

  PER = 22.362 ⇓

- **evaluation :**

  WER = 22.302 ⇑

size of kernel

discriminator_kernel : 6 -> 8

- **last training epoch :**

  REF: N AO R W AA Z DH IH S IH G Z AE K T

  HYP: SH D N R W AA Z IH S IH G Z AE K L

  PER = 25.994 ⇑

- **evaluation :**

  WER = 22.875 ⇑

dimension

discriminator_dim : 384 -> 192

- **last training epoch :**

  REF: N AO R W AA Z DH IH S IH G Z AE K T

  HYP: HH IY N AO W AA Z IH S IH G Z AE K L

- **evaluation :**

  WER = 23.17838 ⇑

dimension

discriminator_dim : 384 -> 768

- **last training epoch :**

  REF: N AO R W AA Z DH IH S IH G Z AE K T

  HYP: HH AY D AE N IH R AH N D L EY D T ER

- **evaluation :**

  WER = 88.47509 ⇑⇑⇑

depth

discriminator_depth : 2 -> 3

- **last training epoch :**

  REF: N AO R W AA Z DH IH S IH G Z AE K T

  HYP: S AH W EH T AA R T IH NG T ER IH NG K

  PER = 84.9997 ⇑⇑⇑

- **evaluation :**

  WER = 78.925 ⇑⇑⇑

Problem 2 result

- learning rate
- size of kernel, dimension, depth
- **loss weight**

loss weight

discriminator_weight_decay : 0.0001 -> 0.001

- **last training epoch :**

  REF: N AO R W AA Z DH IH S IH G Z AE K T

  HYP: N UH N AO W AA Z IH S IH G Z AE K L

- **evaluation :**

  WER = 24.77423 ⇑

loss weight

generator_weight_decay : 0 -> 0.001

- **last training epoch :**

  REF: N AO R W AA Z DH IH S IH G Z AE K T

  HYP: DH EH W AO L OW L IH Z IY Z AE N ER Y

- **evaluation :**

  WER = 23.56198 ⇑

Summary

- **generator_lr** change from 0.0004 to **0.0003** is a good choice

- **discriminator_depth** change from 2 to **3** can't be trained well, **discriminator_dim** change from 384 to **768** can't be trained well, too

- other changes of parameters that we had tried didn't influence much, in the change of learning rate, the HYP parts were nearly same