DLHLP 2021 Fall HW1 E2E ASR

Team A

報告者：吳冠緯、梁正

HW1 E2E ASR

1. Briefly talking about E2E ASR – CTC
2. Processing steps
3. Trying to improve performence and the results (1)
4. Trying to improve performance and the results (2)

HW1 E2E ASR

1. Briefly talking about E2E ASR – CTC
2. Processing steps
3. Trying to improve performence and the results (1)
4. Trying to improve performance and the results (2)

1. Briefly talking about E2E ASR – CTC

- E2E ASR

- CTC

1. Briefly talking about E2E ASR – CTC

- E2E ASR

- CTC

E2E ASR

- Input voice signal -> continuous acoustic features

- Go through ASR model

- Output text message -> discrete text tokens

E2E ASR

- Raw waveform -> feature extractor -> acoustic features

- Acoustic features -> ASR model -> probability distributions

- Probability distributions <- CTC loss -> text token sequence

1. Briefly talking about E2E ASR – CTC

- E2E ASR

- CTC

CTC

- Ignoring post-processing -> an E2E model

- Acoustic features -> encoder (uni-directional RNN for on-line streaming speech recognition) -> linear classifier -> token distribution (vocabulary size)

- Another special NULL token

CTC

- Ignoring down sampling -> input T acoustic features -> output T tokens

- Merging duplicate tokens & removing NULL token

CTC

- Paired training data ? -> alignment !

- Linear classifier as a decoder -> only attend on one vector and each output is decided independently ? Unknowing what has done before -> encorder may help !

HW1 E2E ASR

1. Briefly talking about E2E ASR – CTC
2. Processing steps
3. Trying to improve performence and the results (1)
4. Trying to improve performance and the results (2)

2. Processing steps

- Toolkit installation and data preprocessing

- Modify config files

- Training and testing

2. Processing steps

- Toolkit installation and data preprocessing

- Modify config files

- Training and testing

Toolkit installation and data preprocessing

- Toolkit -> MiniASR

- Download code & install dependencies

- Download data -> training set - Libri-light fine-tuning set / development and testing set - LibriSpeech

2. Processing steps

- Toolkit installation and data preprocessing

- Modify config files

- Training and testing

Modify config files

- Feature extractor -> fbank / MFCC / spectrogram

- Model architecture -> GRU / LSTM / RNN

- SpecAugment -> w/ or w/o

2. Processing steps

- Toolkit installation and data preprocessing

- Modify config files

- Training and testing

Training and testing

- Character / Word error rate (CER / WER)

- Report and compare CER in this HW

HW1 E2E ASR

1. Briefly talking about E2E ASR – CTC
2. Processing steps
3. Trying to improve performence and the results (1)
4. Trying to improve performance and the results (2)

Trying to improve performance and the results (1)

- Data augmentation

- Encoder's module

- Input features

Trying to improve performance and the results (1)

- Data augmentation

- Encoder's module

- Input features

Data augmentation

- w/ vs. w/o SpecAugment

```
Character errors
| #Snt      #Tok    | Sub    Del    Ins    Err    SErr  |
| 2620      281530  | 16.9   8.3    5.0    30.2   99.9  |

Word errors
| #Snt      #Tok    | Sub    Del    Ins    Err    SErr  |
| 2620      52576   | 60.9   5.5    5.5    72.0   99.9  |
```

```
Character errors
| #Snt      #Tok    | Sub    Del    Ins    Err    SErr  |
| 2620      281530  | 13.9   7.7    3.9    25.5   99.8  |

Word errors
| #Snt      #Tok    | Sub    Del    Ins    Err    SErr  |
| 2620      52576   | 54.7   4.7    5.1    64.6   99.8  |
```

Trying to improve performance and the results (1)

- Data augmentation

- Encoder's module

- Input features

Encoder's module

- GRU vs. LSTM

```
Character errors
| #Snt      #Tok    | Sub    Del    Ins    Err    SErr   |
| 2620      281530  | 16.9   8.3    5.0    30.2   99.9   |

Word errors
| #Snt      #Tok    | Sub    Del    Ins    Err    SErr   |
| 2620      52576   | 60.9   5.5    5.5    72.0   99.9   |
```

```
Character errors
| #Snt      #Tok    | Sub    Del    Ins    Err    SErr   |
| 2620      281530  | 13.4   6.4    4.7    24.5   99.6   |

Word errors
| #Snt      #Tok    | Sub    Del    Ins    Err    SErr   |
| 2620      52576   | 52.0   4.9    4.7    61.7   99.6   |
```

Encoder's module

- GRU vs. RNN

```
Character errors
| #Snt      #Tok     | Sub    Del    Ins    Err    SErr   |
| 2620      281530   | 16.9   8.3    5.0    30.2   99.9   |

Word errors
| #Snt      #Tok     | Sub    Del    Ins    Err    SErr   |
| 2620      52576    | 60.9   5.5    5.5    72.0   99.9   |
```

```
Character errors
| #Snt      #Tok     | Sub    Del    Ins    Err    SErr   |
| 2620      281530   | 16.2   12.6   2.8    31.7   100.0  |

Word errors
| #Snt      #Tok     | Sub    Del    Ins    Err    SErr   |
| 2620      52576    | 65.2   5.9    5.7    76.8   100.0  |
```

Trying to improve performance and the results (1)

- Data augmentation

- Encoder's module

- Input features

Input features

• fbank vs. MFCC

```
Character errors
| #Snt       #Tok     | Sub     Del     Ins     Err     SErr    |
| 2620       281530   | 16.9    8.3     5.0     30.2    99.9    |

Word errors
| #Snt       #Tok     | Sub     Del     Ins     Err     SErr    |
| 2620       52576    | 60.9    5.5     5.5     72.0    99.9    |
```

```
Character errors
| #Snt       #Tok     | Sub     Del     Ins     Err     SErr    |
| 2620       281530   | 12.8    8.0     3.5     24.3    99.9    |

Word errors
| #Snt       #Tok     | Sub     Del     Ins     Err     SErr    |
| 2620       52576    | 54.2    4.7     4.7     63.6    99.9    |
```

Input features

- **fbank** vs. spectrogram

```
Character errors
| #Snt      #Tok      | Sub    Del    Ins    Err    SErr   |
| 2620      281530    | 16.9   8.3    5.0    30.2   99.9   |

Word errors
| #Snt      #Tok      | Sub    Del    Ins    Err    SErr   |
| 2620      52576     | 60.9   5.5    5.5    72.0   99.9   |
```

```
Character errors
| #Snt      #Tok      | Sub    Del    Ins    Err    SErr   |
| 2620      281530    | 16.9   9.7    4.4    31.0   100.0  |

Word errors
| #Snt      #Tok      | Sub    Del    Ins    Err    SErr   |
| 2620      52576     | 62.5   5.6    5.1    73.3   100.0  |
```

HW1 E2E ASR

1. Briefly talking about E2E ASR – CTC
2. Processing steps
3. Trying to improve performance and the results (1)
4. Trying to improve performance and the results (2)