Call-Sign Extraction Task

A Grammar-based regular expression regex was proposed to describe the input as analytic formal language in terms of a set of rules for recognizing strings in the language. Two forms of call-sign annotations were identified as follow:

- **Full call sign term** is a sequence of terms starting by Airline code and followed by NATO or Numbers.
- **Short call sign term** is a sequence of terms contains only NATO or Numbers and is referred to the last Full call sign which contains the same NATO or Numbers after the Airline code.

Motivations and Goals

**Motivations**

- **Automatic Speech Recognition** is one of the active research field in speech processing community and has been improved tremendously thanks to the introduction of deep neural networks.
- **ASR in ATC Environment** is a challenging task because of the speech signal variability (e.g., speaker accents, recording conditions, etc.) and the highly domain-specific vocabulary.
- **Call sign extraction** is another challenging task which requires the detection of the aircraft identifier (short & full forms).

**Goals**

- Explore robust features dedicated for noisy environment.
- Take advantage of the recent advances in deep learning techniques.
- Evaluate the different speech recognition approaches (HMM-DNN based ASR, and CTC based ASR).

Preprocessing

Preprocessing aims at making the audio-transcriptions suitable for training the speech recognition system. It consists on normalizing the original text based on the following steps:

- Correct the errors of spelling.
- Remove all signs of ambiguities (e.g., ‘huh’ in call sign).
- Handle the specific call sign (e.g., ‘thousand’, ‘nineteen’ that doesn’t exist in NATO Codification).

About 250 errors were identified through a manual audio examination and the correction was proposed based on the American English spelling.

Audio Transcription Task

Modeling the variability in speech, to reduce the negative environmental influences, and modeling the temporal dynamics in speech, to capture the long term dependencies between acoustic events, require an acoustic model which can effectively deal with these speech characteristics.

Variability in speech signal

- **Feature representations**: MFCC, PLP, FDLP
- **Speaker adaptation**: fMLLR, iVector
- **Data augmentation**: Speed perturbation, VTLP

Temporal dynamics in speech

- **Acoustic models**, which can learn the long term dependencies based on short-term feature representations
- **Deep neural network**, and its variants: DNN, TDNN, LSTM

Experimental Results

Results of audio transcription task - WER

<table>
<thead>
<tr>
<th>Table: Feature representations</th>
<th>Table: Speaker adaptation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFCC, PLP, FDLP</td>
<td>without fMLLR, iVector</td>
</tr>
<tr>
<td>Dev. Set: 10.52, 11.04, 12.63</td>
<td>Dev. Set: 10.52, 13.58, 12.27</td>
</tr>
<tr>
<td>Test Set: 14.19, 11.21, 13.78</td>
<td>Test Set: 14.19, 15.56</td>
</tr>
</tbody>
</table>

ASR approaches

<table>
<thead>
<tr>
<th></th>
<th>SGMM-HMM</th>
<th>DNN-HMM</th>
<th>TDNN-HMM</th>
<th>LSTM-CTC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dev. Set</td>
<td>14.94</td>
<td>10.52</td>
<td>11.21</td>
<td>13.78</td>
</tr>
<tr>
<td>Test Set</td>
<td>18.50</td>
<td>14.19</td>
<td>13.10</td>
<td>13.56</td>
</tr>
</tbody>
</table>

Results of call-sign extraction task - F1 Score

<table>
<thead>
<tr>
<th>Table: Performance of Call Sign Extraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>regex seq2seq</td>
</tr>
<tr>
<td>Dev. Set (Original Annotation): 93.78, 60.59</td>
</tr>
<tr>
<td>Test Set: 75.33</td>
</tr>
</tbody>
</table>

Important Results

- Speaker adaptation has negative impact on the performance
- HMM-based ASR outperforms the recent advanced approach (CTC-based)
- Increasing the amount of training data leads to more robust acoustic model
- Final evaluation results: Transcript WER: 13.06%, callsign F1: 73.64

Material

- KALDI Toolkit: https://github.com/kaldi-asr/kaldi
- EESN Toolkit: https://github.com/srvk/eesen
- FDLPP Feature: https://github.com/iiscleap/FeatureExtractionUsingFDLP