Context-aware Neural-based Dialog Act Classification On Automatically Generated Transcriptions

**Motivation**
Explore the effect of training and testing a context-aware neural-based dialog act (DA) classifier on transcriptions generated from two different automatic speech recognition (ASR) systems, so that the DA classification is taken into a more realistic scenario.

**Dialog Act Classification Model**
Our two-fold model consist of:
- Convolutional neural networks (CNNs) for utterance representation.
- Conditional random fields (CRFs) for sequence labeling.

**Automatic Speech Recognition**
Two types of ASR architectures:
- Hybrid Time Delay Neural Network and Hidden Markov Model (TDNN/HMM) trained with lattice-free maximum mutual information.
- Joint CTC-Attention End-to-End (E2E): shared-encoder representation trained by both Connectionist Temporal Classification (CTC) and attention model using the following combined training loss:
  \[ L = \alpha L_{ctc} + (1 - \alpha) L_{att} \]

**Experimental Results**
- **Experiments varying the context**
- **Experiments on MRDA**
- **Experiments on SwDA**

**Conclusion**
- We explored dialog act classification on automatic transcriptions by means of CNNs and CRFs.
- Although the WERs from both ASR systems are comparable, the End-to-End ASR system might be more suitable for dialog act classification.
- Punctuation yields central cues for the task. Therefore, it should be integrated into the ASR output in future works.

**Experimental Setup**
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dataset</strong></td>
<td><strong>C</strong></td>
<td><strong>Train</strong></td>
</tr>
<tr>
<td>MRDA</td>
<td>5</td>
<td>12k</td>
</tr>
<tr>
<td>SwDA</td>
<td>42</td>
<td>20k</td>
</tr>
</tbody>
</table>

**CNN hyperparameters**
- **Hyperparameter**  
  - **Value**
  - **Train/Val/Test**
  - **# of utts.**

**Best ASR performance in terms of WER (%)**

- **Training data**
- **Accuracy**
- **Context**
- **60 65 70 74 78 82 86 90 94 98**
- **0 1 2 3**

- **MRDA**
- **SwDA**

---

**References**