**INTRODUCTION**

- Building the Natural Language Understanding (NLU) module of goal-oriented Spoken Dialogue Systems (SDS) often involves: the definition of intents (and entities), creating domain-specific and task-relevant datasets, annotating the data with intents/entities, iterative training and evaluation of NLU models, and repeating this tedious process for new/updated use-cases.
- This work explores the potential benefits of data augmentation with paraphrase generation for improving the NLU models trained on limited task-specific datasets.
- Our experiments show that paraphrasing using small seed data with model-in-the-loop (MITL) data augmentation strategies helps boost the performance of the Intent Recognition (IR) task.
- We also investigate extracting entities for potentially further data expansion.

**MOTIVATION**

- Intelligent conversational agent helping the kids learn 'tens and ones' concepts, along with practicing simple counting, addition, and subtraction operations.
- SDS is a crucial building block for handling efficient task-oriented conversation with younger children in play-based learning settings.

**MULTIMODAL SDS**

**METHODOLOGY**

**Natural Language Understanding (NLU)**

- Our NLU and DM models are built on top of the Rasa open-source framework [1].
- The former baseline Intent Classifier in Rasa was inspired by the StarSpace, where embeddings are trained by maximizing the similarity between intents & utterances.
- TF+BERT: In prev. work [2], we enriched this former baseline Rasa NLU architecture by adapting Transformer networks and incorporating pre-trained BERT embeddings.
- DIET+ConveRT: In this work, we adapted the Transformer-based multi-task DIET architecture [3] and incorporated pre-trained ConveRT embeddings [4] to improve Intent Classification performance (and explore the Entity Recognition capabilities).

**Paraphrase Generation**

- We develop a data augmentation module by training a paraphrasing model to generate paraphrased samples from the original seed utterances to augment the NLU training data.
- We adapted the BART seq2seq model [5] that we fine-tuned on the back-translated English sentence pairs from the ParaNMT-SGM, PAWS, and MSRP corpora.
- We examine the data augmentation with certain heuristics:
  - baseline (aug3/aug5/10): augment with 3/5/10x paraphrased samples
  - incShort: paraphrasing only for the low-sample intents or minority classes (<50)
  - excShort: excluding the intents with samples having shorter utterance lengths
- We also investigate model-in-the-loop (MITL) data augmentation techniques using the initial NLU models trained in small datasets.
  - success: augment only the paraphrased utterances with successful predictions
  - success_cont90: success with the confidence level thresholds (>0.9)
  - all_cont90: checking the confidence level thresholds only (>0.9)

**EXPERIMENTS AND RESULTS**

**Datasets**

- Experiments conducted on Kid Space NLU datasets having utterances from math learning experiences (Planting and Watering activities) designed for 5-to-8 years-old kids [6].
- Some intents are highly generic across usages and activities (e.g., affirm, deny, next_step, out_of_scope, goodbye), whereas the rest are highly domain-dependent and task-specific (e.g., intro重要举措, answer_flowers/water, ask_number, counting).

**Intent Recognition (IR)**

<table>
<thead>
<tr>
<th>Model/Dataset</th>
<th>Planting</th>
<th>Watering</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF-BERT (Baseline)</td>
<td>90.15</td>
<td>92.41</td>
</tr>
<tr>
<td>DIET+ConveRT</td>
<td>95.59</td>
<td>97.83</td>
</tr>
<tr>
<td>Performance Gain:</td>
<td>+5.04</td>
<td>+5.42</td>
</tr>
</tbody>
</table>

**Data Augmentation**

- Data size vs. NLU performance for original and augmented datasets.

**Entity Recognition**

- Entity Recognizers with:
  - SpaCy pre-trained NER tags (generic)
  - Manual annotation for domain-specific entities
- Next: ConceptNet Relatedness Entity Expansion

**CONCLUSION**

- Focused on improving the NLU module of the task-oriented SDS pipeline with limited datasets.
- Shown that paraphrasing with model-in-the-loop (MITL) strategies using small seed data is a promising approach yielding higher F1-scores for Intent Recognition on task-specific datasets.
- Explored Entity Extraction (for further data expansion and enrichment) to improve the NLU module of our multimodal SDS.
- We believe these results are highly encouraging in our quest to make dialogue systems more robust and generalizable to new intents with limited data resources.

**SELECTED REFERENCES**