Introduction

- Dialogue State Tracking (DST) is an important step in task-oriented dialogue systems.
- Slot filling based on the current user utterance and dialogue history.
- We explore pre-training strategies and data augmentation methods to better leverage the power of large-scale, pre-trained language models (like BERT and T5) for DST.
- Multi-Domain Wizard-of-Oz (MultiWoZ) Dataset [1]
  - Multi-turn dialog utterances with labeled slot-value pairs.
  - Evaluation: Joint Goal Accuracy, Slot F1, Slot Accuracy.

Data Augmentation for Dialogue Understanding

- LMs perform better when provided with huge amounts of training data [3].
- Why Data Augmentation?
  - Additional training data without any human annotation effort.
  - Provides diversity for better generalizability of NLP models.
  - Noisy annotations in MultiWoZ dataset (multiple versions).
- Rule-Based Augmentation: Entity Replacement, Crop and Rotate using dependency parse trees, Sequential augmentation to increase complexity.
- Deep Learning Techniques:
  - Paraphrasing using Pegasus, Reverse translation with English-Spanish NMT.
  - Data augmentation significantly improves performance on DST.
  - High levels of augmentation can at times hurt the performance too.

Multi-phase Adaptive Pre-training

- Significant domain mismatch between large-scale text corpus (for language modelling) and dialogue datasets.
  - Domain adaptive pre-training on open-domain dialogues.
  - Task adaptive pre-training on target dataset.
- Span-level objectives to reason across multiple turns and span-selection where entities are contiguous sequences of words.

<table>
<thead>
<tr>
<th>Model</th>
<th>Objective</th>
<th>MultiWoZ 2.0</th>
<th>MultiWoZ 2.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>BERT</td>
<td>Span</td>
<td>Base</td>
<td>51.2 75.2 92.6</td>
</tr>
<tr>
<td></td>
<td>Prediction</td>
<td>+ DAPT</td>
<td>52.7 75.1 92.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ TAPT</td>
<td>56.7 75.3 92.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ DAPT + TAPT</td>
<td>51.7 75.1 92.6</td>
</tr>
<tr>
<td>T5</td>
<td>Span</td>
<td>Base</td>
<td>50.2 89.6 96.3</td>
</tr>
<tr>
<td></td>
<td>Correction</td>
<td>+ DAPT</td>
<td>50.2 89.4 96.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ TAPT</td>
<td>51.6 89.9 96.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ DAPT + TAPT</td>
<td>51.5 89.6 96.3</td>
</tr>
</tbody>
</table>

Domain Analysis

- Combining best pre-training method and data augmentation techniques.
- Consistent gains in goal accuracy across all domains in MultiWoZ.

Conclusion & Future Work

- Incorporating language structure of dialogues through span-level pre-training and additional domain data through augmentation methods is helpful for DST.
- Potential future direction would be to study the impact of these techniques on end-to-end dialogue systems including generation.

References

[2] Gururangan et al., Don’t stop pre-training: Adapt language models to domains and tasks, ACL 2020