A Copy-Augmented Seq2Seq Architecture Gives Good Performance on Task-Oriented Dialogue
Mihail Eric and Christopher Manning

The Problem
• Task-oriented dialogue agents have the potential to revolutionize how we interact with computers

Baseline Models
• Vanilla sequence-to-sequence architecture
• Encode context, decode system response
• Extend this architecture with attention mechanism to provide further signal from encoder context

Results
• Single layer sequence-to-sequence already outperforms other models
• Copy-mechanism adds nice gains in per-response accuracy as well as entity F1

<table>
<thead>
<tr>
<th></th>
<th>Per-Response</th>
<th>Per-Discourse</th>
<th>BLEU</th>
<th>Entity F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq2Seq (1)</td>
<td>46.4</td>
<td>1.5</td>
<td>55.0</td>
<td>69.7</td>
</tr>
<tr>
<td>Seq2Seq (2)</td>
<td>43.5</td>
<td>1.3</td>
<td>54.2</td>
<td>67.3</td>
</tr>
<tr>
<td>Seq2Seq (3)</td>
<td>44.2</td>
<td>1.7</td>
<td>55.4</td>
<td>65.9</td>
</tr>
<tr>
<td>+ Attn</td>
<td>46.0</td>
<td>1.5</td>
<td>56.6</td>
<td>67.1</td>
</tr>
<tr>
<td>+ Copy</td>
<td>47.3</td>
<td>1.3</td>
<td>55.4</td>
<td>71.6</td>
</tr>
<tr>
<td>+ EntType</td>
<td>48.0</td>
<td>1.5</td>
<td>56.0</td>
<td>72.9</td>
</tr>
<tr>
<td>MemNMT</td>
<td>41.1</td>
<td>0.0</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>GREW</td>
<td>48.7</td>
<td>1.4</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>QRN</td>
<td>50.7</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Copy-Augmented Seq2Seq also learns nicely interpretatable attention to context:
• Entities in decoder often map to entities mentioned in encoder and phrases surrounding entities

Copy-Augmented Seq2Seq Architecture
• Copy-mechanisms added to sequence-to-sequence architectures show improvements in many tasks such as code generation, machine translation, and text summarization
• In practice, our copy mechanism is a straightforward attention-based extension

Dataset
Dialogue State-Tracking Challenge 2 (DSTC2)
• Restaurant reservation system
• Number dialogues: 1,618/500/1,117
• Vocabulary: 1,229
• Roughly 100 distinct restaurants → 452 entities in knowledge base (KB)
• API calls to underlying KB also included

Background
Traditional task-oriented dialogue systems are modular:

Copy-Augmented Seq2Seq Architecture
• Copy-mechanisms added to sequence-to-sequence architectures show improvements in many tasks such as code generation, machine translation, and text summarization
• In practice, our copy mechanism is a straightforward attention-based extension

Conclusion
• Recent systems using more complex neural architectures don’t drastically outperform simpler architectures
• Sequence-to-sequence architectures + simple extensions can learn powerful representations for linguistically-rich responses and KB entity incorporation

Can powerful neural embeddings infer implicit representations of dialogue state, without modularly-defined components?