**Introduction**
- Natural Language Generation (NLG) is the final step in task-oriented dialogue pipeline
- Its readability and informativeness directly impact users’ perception of the system
- Existing methods struggle to achieve high fluency and variation
- We utilize the language model as an unconditioned auxiliary task to generate natural responses

**Formulation**
- Meaning Representations (MR) are slot-value pairs offering information category and content
  - E.g. {(area, south), (name, The Phoenix)}
- Dialogue acts (DA) indicate types of system actions
  - E.g. inform/request/confirm
- Response is the desired natural language output
  - E.g. The Phoenix is a restaurant in the south of city.

**Model**
- Seq2seq model generates response with maximum likelihood conditioned on input $I$
  \[
p(w_1, \ldots, w_n | I) = \prod_{t=1}^{n} p(w_t | w_1, \ldots, w_{t-1}; I)
\]
- Language model as unconditioned probability
  \[
p^{LM}(w_1, \ldots, w_n) = \prod_{t=1}^{n} p^{LM}(w_t | w_1, \ldots, w_{t-1})
\]
- Both tasks share decoder parameters and use a linear combination for loss function
  \[
  \mathcal{L}(\theta) = \mathcal{L}^{NLG}(\theta) + \alpha \mathcal{L}^{LM}(\theta)
  \]

**Generated Examples**
- MR: Name[The Cricketers], eatType[restaurant], food[Chinese], priceRange[high], customer rating[1 out of 5], area[riverside], familyFriendly[no], near[All Bar One]
- Answer: The high priced Chinese restaurant The Cricketers, is based in the riverside area near to the All Bar One. It is not child friendly and has a customer rating of 1 out of 5.
- w/o LM: The Cricketers is a restaurant providing Chinese food in the high price range. It is located in the riverside. It is near All Bar One. Its customer rating is 1 out of 5.
- NLG-LM: The Cricketers is a Chinese restaurant in the riverside area near All Bar One. It is not children friendly and has a high price range and a customer rating of 1 out of 5.

**Experiments**
- **Datasets**
  - E2E-NLG dataset with 51.2K samples
  - RNNLG dataset, including TV, Laptop, Hotel and Restaurant, with 14.1K, 26.5K, 8.7K and 8.5K samples
- **Use official evaluation scripts from E2E-NLG and RNNLG**

**Result (BLEU-4)**

<table>
<thead>
<tr>
<th>Model</th>
<th>E</th>
<th>TV</th>
<th>L</th>
<th>H</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGen</td>
<td>0.659</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Slug</td>
<td>0.662</td>
<td>0.529</td>
<td>0.524</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>SCLSTM</td>
<td>0.527</td>
<td>0.512</td>
<td>0.848</td>
<td>0.752</td>
<td>/</td>
</tr>
<tr>
<td>RALSTM</td>
<td>0.541</td>
<td>0.525</td>
<td>0.898</td>
<td>0.779</td>
<td>/</td>
</tr>
</tbody>
</table>

**Running Time Comparison**
- **NLG-LM**
  - w/o LM: 618.87s
  - Per-batch: 500.64s

**Conclusion**
- Propose multi-task learning model NLG-LM for response generation in task-oriented dialogues
- Incorporate language model as an unconditioned auxiliary task
- Achieve state-of-the-art results on five public datasets