Semantically Conditioned LSTM-based Natural Language Generation for Spoken Dialogue Systems

Tsung-Hsien Wen, Milica Gasic, Nikola Mrksic, Pei-Hao Su, David Vandyke, and Steve Young
Outline

- Intro
- Semantic Conditioned LSTM
- Deep Architecture
- Experiments
  - Setup
  - Corpus-based Evaluation
  - Human Evaluation
- Conclusion
Outline

- **Intro**
- Semantic Conditioned LSTM
- Deep Architecture
- Experiments
  - Setup
  - Corpus-based Evaluation
  - Human Evaluation
- Conclusion
Spoken Dialogue System

Diagram:
- Speech Recognition
- Language Understanding
- Dialogue Manager
- Language Generation
- Speech Synthesis
- Knowledge Base
- Web

Diagram illustrates the components of a Spoken Dialogue System, including speech recognition, language understanding, dialogue manager, language generation, speech synthesis, knowledge base, and web.
Given a meaning representation, map it into natural language utterances.

What do we care about?
- adequacy, fluency, readability, variation (Stent et al 2005)
Traditional approaches to NLG
Outline

- Intro
- **Semantic Conditioned LSTM**
- Deep Architecture
- Experiments
  - Setup
  - Corpus-based Evaluation
  - Human Evaluation
- Conclusion
Recurrent Generation Model

Inform(name=Seven_Days, food=Chinese)

\[ \begin{bmatrix} 0, 0, 1, 0, 0, \ldots, 1, 0, 0, \ldots, 1, 0, 0, 0, 0, 0, \ldots \end{bmatrix} \]

dialog act 1-hot representation

\[ \begin{array}{c}
\text{delexicalisation} \\
\text{SLOT_NAME} \\
\text{Seven Days} \\
\text{serves} \\
\text{serves} \\
\text{SLOT_FOOD} \\
\text{Chinese} \\
. \\
. \\
</s> \\
</s> \\
</s> \\
</s> \\
</s>
\]

RNNLM (Mikolov et al 2010)
SC-LSTM

- **Original LSTM cell**
  
  \[
  \begin{align*}
  i_t &= \sigma(W_{wi}w_t + W_{hi}h_{t-1}) \\
  f_t &= \sigma(W_{wf}w_t + W_{hf}h_{t-1}) \\
  o_t &= \sigma(W_{wo}w_t + W_{ho}h_{t-1}) \\
  \hat{c}_t &= \tanh(W_{wc}w_t + W_{hc}h_{t-1}) \\
  c_t &= f_t \odot c_{t-1} + i_t \odot \hat{c}_t \\
  h_t &= o_t \odot \tanh(c_t)
  \end{align*}
  \]

- **DA cell**
  
  \[
  \begin{align*}
  r_t &= \sigma(W_{wr}w_t + W_{hr}h_{t-1}) \\
  d_t &= r_t \odot d_{t-1}
  \end{align*}
  \]

- **Modify C_t**
  
  \[
  c_t = f_t \odot c_{t-1} + i_t \odot \hat{c}_t + \tanh(W_{dc}d_t)
  \]

- **Inform**

  \[
  Inform(name=Seven_Days, food=Chinese)
  \]

  \[
  0, 0, 1, 0, 0, ..., 1, 0, 0, ..., 1, 0, 0, ...
  \]

  *(dialog act 1-hot representation)*

  *(Hochreiter and Schmidhuber, 1997)*
Visualization

The graph shows the feature values for various attributes. The x-axis represents different attributes such as 'there', 'are', '6', 'korean', 'restaurant', 'that', 'allow', 'kid', 'at', 'various', 'price', 'range', 'the', 'castro_inn', 'near', 'haight', 'has', 'internet', 'and', 'accepts', 'credit', 'card', etc. The y-axis represents the feature value, ranging from 0 to 1.

Key attributes and their values include:
- A-inform
- pricerange=don't care
- kids_allowed=yes
- food=VALUE
- type=VALUE
- accepts_cards=yes
- has_internet=yes
- near=VALUE
- name=VALUE

The lines on the graph indicate the feature value trends for each attribute.
Outline

- Intro
- Semantic Conditioned LSTM
- **Deep Architecture**
- Experiments
  - Setup
  - Corpus-based Evaluation
  - Human Evaluation
- Conclusion
Deep Architecture
Outline

- Intro
- Semantic Conditioned LSTM
- Deep Architecture
- Experiments
  - Setup
    - Corpus-based Evaluation
    - Human Evaluation
- Conclusion
Setup

- Data collection:
  - SFX restaurant/hotel domains
# Ontologies

<table>
<thead>
<tr>
<th></th>
<th>SF Restaurant</th>
<th>SF Hotel</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>act type</strong></td>
<td>inform, inform_only, reject, confirm, select, request, reqmore, goodbye</td>
<td></td>
</tr>
<tr>
<td><strong>shared</strong></td>
<td>name, type, *pricerange, price, phone, address, postcode, *area, *near</td>
<td></td>
</tr>
<tr>
<td><strong>specific</strong></td>
<td>*food *goodformeal *kids-allowed</td>
<td>*hasinternet *acceptscards *dogs-allowed</td>
</tr>
</tbody>
</table>

**bold**=binary slots, *=slots can take “don’t care” value
Setup

- Data collection:
  - SFX restaurant/hotel domains
  - Workers recruited from Amazon MT.
  - Asked to generate system responses given a DA.
  - Result in ~5.1K utterances, 228/164 distinct acts.

- Training: BPTT, L2 reg, SGD w/ early stopping.
  - train/valid/test: 3/1/1, data up-sampling
**Generated Examples**

| inform_only(name='red victorian bed breakfast',accepts_credit_cards='yes',near='haight',has_internet='yes') |
| red victorian bed breakfast is the only hotel near haight and accepts credit cards and has internet. |
| red victorian bed breakfast is the only hotel near haight and has internet and accepts credit cards. |
| red victorian bed breakfast is the only hotel near haight that accept credit cards and offers internet. |
| the red victorian bed breakfast has internet and near haight, it does accept credit cards. |
| the red victorian bed breakfast is the only hotel near haight that accepts credit cards, and offers internet. |
Generated Examples

```plaintext
inform_no_match(area=tenderloin)
there are no restaurants in the tenderloin area .
there are 0 restaurants in the tenderloin area .
unfortunately there are 0 restaurants in the tenderloin area .
i could not find any restaurants in tenderloin .
there are no restaurants in tenderloin .
```
Generated Examples

?request(food)
what type of food are you looking for ?
what type of food would you like ?
what kind of food are you looking for ?
what type of food are you interested in ?
what kind of food would you like ?
Outline

- Intro
- Semantic Conditioned LSTM
- Deep Architecture
- Experiments
  - Setup
  - Corpus-based Evaluation
  - Human Evaluation
- Conclusion
Corpus-based Evaluation

- Test set: ~1K utterances each domain
- Metrics: BLEU-4 (against multiple references), ERR(slot error rates)
- Averaged over 5 random initialised networks.
- Over-gen 20, evaluate on top-5
- Baselines:
  - handcrafted generator (hdc)
  - kNN example-based generator (kNN)
  - class-based LM generator (classlm, O&R 2000)
  - rnn-based generator (rnn, Wen et al 2015)
Corpus-based Evaluation

Selection scheme: 5/20
Corpus-based Evaluation

Selection scheme: 5/20

ERR (%)

Model

Restaurant
Hotel

hdc
knn
classlm
rnn
sc-lstm
+deep

Selection scheme: 5/20
Outline

- Intro
- Semantic Conditioned LSTM
- Deep Architecture
- Experiments
  - Setup
  - Corpus-based Evaluation
  - Human Evaluation
- Conclusion
Human Evaluation

- **Setup**
  - Judges (~60) recruited from Amazon MT.
  - Asked to evaluate two system responses pairwise.
  - Comparing `classlm`, `rnn`, `sc-lstm`, and `+deep`

- **Metrics:**
  - Informativeness, Naturalness (rating out of 3)
  - Preference
## Human Evaluation

<table>
<thead>
<tr>
<th>Method</th>
<th>Informativeness</th>
<th>Naturalness</th>
</tr>
</thead>
<tbody>
<tr>
<td>+deep</td>
<td>2.58</td>
<td>2.51</td>
</tr>
<tr>
<td>sc-lstm</td>
<td><strong>2.59</strong></td>
<td>2.50</td>
</tr>
<tr>
<td>rnn</td>
<td>2.53</td>
<td>2.42*</td>
</tr>
<tr>
<td>classlm</td>
<td><strong>2.46</strong>**</td>
<td>2.45</td>
</tr>
</tbody>
</table>

* $p < 0.05$ ** $p < 0.005$
## Human Evaluation

<table>
<thead>
<tr>
<th>Pref. %</th>
<th>classlm</th>
<th>rnn</th>
<th>sc-lstm</th>
<th>+deep</th>
</tr>
</thead>
<tbody>
<tr>
<td>classlm</td>
<td>-</td>
<td>46.0</td>
<td>40.9**</td>
<td>37.7**</td>
</tr>
<tr>
<td>rnn</td>
<td>54.0</td>
<td>-</td>
<td>43.0</td>
<td>35.7*</td>
</tr>
<tr>
<td>sc-lstm</td>
<td>59.1*</td>
<td>57</td>
<td>-</td>
<td>47.6</td>
</tr>
<tr>
<td>+deep</td>
<td>62.3**</td>
<td>64.3**</td>
<td>52.4</td>
<td>-</td>
</tr>
</tbody>
</table>

* $p < 0.05$ ** $p < 0.005$
Outline

- Intro
- Semantic Conditioned LSTM
- Deep Architecture
- Experiments
  - Setup
  - Corpus-based Evaluation
  - Human Evaluation
- Conclusion
Conclusion

- Train NLG N2N using LSTM.
- Learn LM & slot gating control signal jointly.
- Deep architecture helps.
- Corpus-based/Human evaluation.
- Achieve best performance.
- Potential for open domain SDS.
Papers


Selected References

Thank you! Questions?

Tsung-Hsien Wen is supported by a studentship funded by Toshiba Research Europe Ltd, Cambridge Research Laboratory
SC-LSTM

- **Cost function**
  \[
  F(\theta) = \sum_t p_t^T \log(y_t) + \|d_T\| + \sum_{t=0}^{T-1} \eta \xi \|d_{t+1} - d_t\|
  \]

- **1\textsuperscript{st} term**: Log-likelihood
- **2\textsuperscript{nd} term**: make sure rendering all the information needed
- **3\textsuperscript{rd} term**: close only one gate each time step.

(Hochreiter and Schmidhuber, 1997)
Intuition behind the 3\textsuperscript{rd} term

\[ \eta = 0.01, \xi = 100 \]
Traditional pipeline approach

Inform(
    name=Z_House,
    price=cheap
)

Dialogue Act

Tree-like template

Utterance

Sentence Planning

Surface Realisation

Z House is a cheap restaurant.
Problems

- Scalability
  - Grammars are handcrafted.
  - Require expert knowledge.
Problems

- Boring
  - Frequent repetition of outputs.
  - Non-colloquial, awkward utterances.

Seven Days is a nice restaurant in the expensive price range, in the north part of the town, if you don’t care about what food they serve.
SC-LSTM

- **Original LSTM cell**
  \[
  i_t = \sigma(W_{wi}w_t + W_{hi}h_{t-1}) \\
  f_t = \sigma(W_{wf}w_t + W_{hf}h_{t-1}) \\
  o_t = \sigma(W_{wo}w_t + W_{ho}h_{t-1}) \\
  \hat{c}_t = \tanh(W_{wc}w_t + W_{hc}h_{t-1}) \\
  c_t = f_t \odot c_{t-1} + i_t \odot \hat{c}_t \\
  h_t = o_t \odot \tanh(c_t)
  \]

- **DA cell**
  \[
  r_t = \sigma(W_{wr}w_t + W_{hr}h_{t-1}) \\
  d_t = r_t \odot d_{t-1}
  \]

- **Modify C\_t**
  \[
  c_t = f_t \odot c_{t-1} + i_t \odot \hat{c}_t + \tanh(W_{dc}d_t)
  \]

(Hochreiter and Schmidhuber, 1997)
Deep Architecture

- Techniques applied
  - Skip connection
    (Graves et al 2013)
  - RNNN dropout
    (Srivastava et al 2014)

- Gating Equation is modified from

$$\mathbf{r}_t = \sigma(\mathbf{W}_{wr}\mathbf{w}_t + \alpha\mathbf{W}_{hr}\mathbf{h}_{t-1})$$

- To

$$\mathbf{r}_t = \sigma(\mathbf{W}_{wr}\mathbf{w}_t + \sum_l \alpha_l \mathbf{W}_{hr}^l\mathbf{h}_{t-1}^l)$$