Towards Practical Natural Language Generation (NLG) System in Task-Oriented Dialog

Jinfeng Rao, Facebook AI

Joint work with Anusha, Kartikeya, Michael, Rajen, Anuj

Berkeley NLP Seminar
09/30/2019
Overview
“Do I have any new messages?”

“Your latest chat is with Tom, and you have an unread message from your chat with Jimmy from September 1st. Which one would you like me to read?”
Natural language generation

• Naturalness is the primary metric we want to optimize
• We want our system to be natural in context
  • Dialog context
  • User context, including memory
  • Multimodal context
Who we think we are

NLG

- Seq2Seq
- Encoder-Decoder
- Image Captioning
- Pun/Story Generation
- Dialog
- IC
- REG
- LM
- NMT
- Machine Translation
- Text Summarization
- Referring Expression Generation

Superb! 😊
Who we actually are

A General Architecture View of Task-Oriented Dialog System
Who we actually are

A General Architecture View of Task-Oriented Dialog System
Templates

{date_time} in {location}, it’ll be {condition} with a high of {temp_high} and a low of {temp_low} F.

Tomorrow in Palo Alto, it’ll be cloudy with a high of 64 and a low of 41 F.

March 6th in Golden Gate Bridge, it’ll be rain with a high of 64 and a low of 41 F.

- **Strengths**
  - Simple to bootstrap
  - Robust and interpretable

- **Drawbacks**
  - Hard to maintain -- lots of templates to handle every case
  - Surface realization rely on argument generators
  - Unable to generalize across domains
Why Model-based generation?

- **Strengths**
  - Flexible surface form realization
  - Easy to maintain
  - Reusable across domains
  - Easy to condition on wide feature set (e.g. contextual NLG)

- **Drawbacks**
  - Adds latency
  - Could generate ungrammatical or incorrect responses
But still… Some Open Questions

- Response Quality
- Controllability
- Evaluation
- Model Robustness
- Cross-domain/language generalizability
Generate-Filter-Rank

Templates
Models
Generators

Correctness filter
Grammar filter
Filters

Ranker
Related Papers

● Constrained Decoding with Compositional Meaning Representations for Neural NLG in Task-oriented Dialog, **ACL 2019**

● Tree-to-Sequence Model for Neural NLG in Task-oriented Dialog, **INLG 2019**.
Background
E2E NLG Dataset

~50K (meaning representation, sentence) pairs for the restaurants domain

**MR:** name[JJs’ Pub] rating[5 out of 5] familyFriendly[no] eatType[restaurant] near[Crowne Plaza Hotel]

**Sentence:** JJ’s pub is not a family friendly restaurant. It has a high customer rating of 5 out of 5. You can find it near the Crowne Plaza Hotel.
E2E Dataset and Shortcomings

● Provided crowdworkers with MR and asked them to write responses

● Lots of diversity in dataset, e.g. argument grouping, contrast, rich language
  ○ [But no way to control these aspects]

● **However**, models trained on the data lack this diversity
  ○ e.g. contrast only occurs in 0.4% of model generations
E2E NLG Dataset

**MR:** name[JJs’ Pub]  rating[5 out of 5]  
familyFriendly[no]  eatType[restaurant]  
near[Crowne Plaza Hotel]

**S1:** JJ’s pub is not a family friendly restaurant. It has a high customer rating of 5 out of 5. You can find it near the Crowne Plaza Hotel.

**S2:** JJ’s Pub is not family friendly, but it has a high customer rating of 5 out of 5. It is a restaurant near the Crowne Plaza Hotel.
Why control discourse?

Increased **perceived naturalness** when user models influence discourse

What can you tell me about JJ’s Pub? I’d like to eat English food.

JJ’s pub is highly rated, *but it serves English food.*

5/5 rating and English food are usually contrasted

-discourse-unaware system

(Lemon et al. 2004; Carenin and Moore, 2006; Walker et al. 2007; White et al. 2010; Demberg et al. 2011)
Why control discourse?

Prior work has shown increases in **perceived naturalness** when user models influence discourse.

(Lemon et al. 2004; Carenin and Moore, 2006; Walker et al. 2007; White et al. 2010; Demberg et al. 2011)
Reed et al. (2018)

- Add input tokens to indicate target discourse structure
  - contrast
  - # of sentences

- Greatly improved accuracy of contrast expression and # sentences

- **But:** no way to control which slots gets contrasted!

How can we control the expression of discourse relations?
Compositional Meaning Representations
Proposed Representation Components

- **Arguments** e.g. `date_time`, `family_friendly`
  - Entities or slots to be mentioned
  - Can be nested (e.g. `date_time contains week_day`)

- **Dialog acts** e.g. `INFORM`, `RECOMMEND`
  - Contain arguments

- **Discourse relations** e.g. `JUSTIFY`, `CONTRAST`
  - Relationship between dialog acts / discourse relations

(Mann and Thompson, 1988; Rambow et al. 2001; Reiter and Dale, 2000; Walker et al. 2007)
## Dialog Acts

<table>
<thead>
<tr>
<th>DG:INFORM</th>
<th>Response with requested information</th>
</tr>
</thead>
<tbody>
<tr>
<td>DG:ACK</td>
<td>Acknowledgement of a request</td>
</tr>
<tr>
<td>DG:CONFIRM</td>
<td>Confirmation before performing an action</td>
</tr>
<tr>
<td>DG:PROMPT</td>
<td>Asks the user if they want to perform a task, usually as a follow-up</td>
</tr>
<tr>
<td>DG:ERROR</td>
<td>When the assistant can't perform an action, usually due to an invalid argument</td>
</tr>
<tr>
<td>DG:REQUEST</td>
<td>When more information is needed to perform an action</td>
</tr>
<tr>
<td>DG:DISAMB</td>
<td>When arguments need to be disambiguated</td>
</tr>
<tr>
<td>DG:SELECT</td>
<td>When the user needs to choose from a list</td>
</tr>
</tbody>
</table>

## Discourse Acts

<table>
<thead>
<tr>
<th>DS:JOIN</th>
<th>Combines dialog acts, usually with a conjunction like “and”</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS:CONTRAST</td>
<td>Combines dialog acts that differ in some way, usually with a conjunction like “but”</td>
</tr>
<tr>
<td>DS:JUSTIFY</td>
<td>Used when one dialog act provides justification for another dialog act</td>
</tr>
</tbody>
</table>
Putting it all together

- Tree-structured representation
- Allows arbitrary nesting of relations and acts

```
[JOIN
  [CONTRAST
    [INFORM [name JJs’ Pub] [rating 5 out of 5] ]
    [INFORM [familyFriendly no] ]
  ]
  [INFORM [eatType restaurant] [near Crowne Plaza Hotel] ]
]
```
# Flat MR vs. Tree-structured MR

<table>
<thead>
<tr>
<th>Flat MR</th>
<th>Our MR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>INFORM</strong> [ chance[likely], wind_summary[heavy], date_time[ week_day[Saturday] colloquial[morning] ] ]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Annotated Reference</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[INFORM] It’ll be [condition sunny] throughout [date_time_range colloquial[this weekend]].</td>
<td><strong>CONTRAST</strong> [INFORM] The high will be in the [avg_high 60s]],</td>
</tr>
<tr>
<td>[INFORM] but expect temperatures to drop as low as [avg_low 43 degrees] by [date_time [week_day Sun colloquial evening] ]].</td>
<td>[INFORM] There’s also [chance a chance of] [wind_summary strong winds] on [date_time [week_day Saturday] colloquial morning]]</td>
</tr>
</tbody>
</table>
Data
Dataset Creation

User query crowdsourcing → Compositional MR generation → Response generation and annotation → Response quality evaluation

Rule-based

33K examples for weather domain

https://github.com/facebookresearch/TreeNLG
Q: Is it going to rain in Miami tomorrow?

S: [DS_CONTRAST [DG_NO ] [DG_INFORM [date_time tomorrow ] [location Miami ] [condition sunny ] ] ]

[DS_CONTRAST [DG_NO No, ] [DG_INFORM it’ll be [condition sunny ] [date_time tomorrow ] in [location Miami ] ] ].
Modified E2E Dataset

- Used the Berkeley neural parser to automatically infer CONTRAST and JOIN relations in the E2E challenge dataset
- Used Slug2Slug token tagger and HarvardNLP latent segmentation model to annotate responses
Approach
Model Overview

Standard Seq2Seq model with attention

**Input:** Linearized tree structure

```
[CONTRAST [INFORM [condition rain ] [date_time [week_day Saturday ] ] ]
  [INFORM [cloud_coverage sunny ] [date_time [week_day Sunday ] ] ] ]
```

**Output:** Response with tree structure preserved

```
[CONTRAST [INFORM There’ll be [condition rain ] on [date_time [week_day Saturday ] ] ], but [INFORM [date_time [week_day Sunday ] ] will be [cloud_coverage sunny ] ] ].
```
Semantic Correspondence

Leverage correspondence between input and output structures

MR: [CONTRAST [INFORM [condition rain] [date_time [week_day Saturday]]] [INFORM [cloud_coverage sunny] [date_time [week_day Sunday]]]]

S: [JOIN [INFORM There’ll be [condition rain] on [date_time [week_day Saturday]]], and [INFORM [date_time [week_day Sunday]] will be [cloud_coverage sunny]].]
Semantic Correspondence

Leverage correspondence between input and output structures

(CONTRAST [INFORM [condition rain] [date_time [week_day Saturday]]] [INFORM [cloud_coverage sunny] [date_time [week_day Sunday]]])

JOIN [INFORM There’ll be [condition rain] on [date_time [week_day Saturday]], and [INFORM [date_time [week_day Sunday]] will be [cloud_coverage sunny]]].

Tree accuracy metric for semantic correctness
Key Model Consideration

- Lexicalization vs. delexicalization
  - i.e., “It will be [condition sunny ]” vs. “It will be [condition _condition_ ]”
  - Free surface form realization vs vocabulary sparsity

- Coarse-grained vs. fine-grained
  - Coarse: [INFORM [ARG_CONDITION sunny] [DATE today] ].
  - Fine: [INFORM [ARG_CONDITION_SUNNY sunny] [DATE today] ].
  - Fine-grained arguments provides us better correctness control, but also increase the vocab space.
Constrained Decoding

Idea: During beam search, invalidate tokens that would result in responses that don’t match the input structure (based on the nonterminals in the output)

MR: [INFORM [condition_not rain] [date_time [colloquial today ] ] ]

<table>
<thead>
<tr>
<th></th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>[condition]</td>
<td>0.35</td>
</tr>
<tr>
<td>not</td>
<td>0.28</td>
</tr>
<tr>
<td>[condition_not]</td>
<td>0.19</td>
</tr>
<tr>
<td>[location]</td>
<td>0.07</td>
</tr>
<tr>
<td>be</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Constrained Decoding (contd.)

Mask tokens that invalidate constraints:

- validity of child node
- coverage of all children before closing a node
- allow aggregation of similar attributes

MR: [INFORM [location Menlo Park ] [condition sunny] ]
[INFORM [location Menlo Park ] [temp 72 ] ]

S: [INFORM [location Menlo Park ] will be [condition sunny] ] ] [INFORM with a temperature of [temp 72 ] ].
Tree2Seq
Why Tree2Seq?

Would combining a better learning and decoding algorithm leveraging tree structures lead to better effectiveness?

Tree-structured MR:

```
INFORM [ condition[sunny], date.time.range[ colloquial[this weekend ] ] ]
CONTRAST [ ]
  INFORM [ avg.high[60s] date.time[ colloquial this weekend ] ]
INFORM [ chance[likely], wind.summary[heavy], date.time[ week.day[Saturday] colloquial[morning] ] ]
```
Tree-to-sequence model

- Dialog and discourse acts are inherently tree-structured
- A structure-aware encoder may learn semantics better
Tree-to-sequence Model

- N-ary Tree-based encoder (Tai et al. ACL’15)

\[ h_k^p = f_{\text{tree}}(\{h_k^{c1}, ..., h_k^{cN}\}) \]

- Tree inputs are hard to parallelize
  - Developed an iterative bottom-to-up traversal algorithm to support batch forward/backward
  - Achieved 5-10x speedup in training and inference
Tree-to-sequence Model

- Structured-aware Decoder
  - Initialize decoder state as root hidden state
    \[ s_1 = h_{\text{root}} \]

- Incorporate tree structures into decoding

\[
\hat{s}_j = \tanh(W_d \cdot [s_j; d_j; s_{j-1}] + b_d)
\]

\[
P(y_i|y_{<i}, x) = (W_s \cdot \hat{s}_j + b_s)
\]
Experiments and Results
Models

- **S2S-FLAT**
  - Input and output contain a flat list of arguments

- **S2S-TOKEN**
  - Inspired by Reed et al. 2018, input is similar to Flat but has tokens indicating # of Contrast and Join relations

- **S2S-TREE**
  - Input and output are linearized tree representations

- **S2S-CONSTR**
  - TREE with constrained decoding

- **T2S**
  - Tree2seq model with linearized tree representations

- **T2S-CONSTR**
  - Tree2seq + constrained decoding
Metrics

- Automatic
  - BLEU
  - Tree accuracy
- Human evaluation
  - Grammaticality
  - Semantic correctness
  - Disc. correctness (semantic correctness measured on challenging subset)
Results

<table>
<thead>
<tr>
<th>Model</th>
<th>BLEU</th>
<th>TreeAcc</th>
<th>E2E Gram</th>
<th>Corr</th>
<th>Disc</th>
<th>BLEU</th>
<th>TreeAcc</th>
<th>E2E Gram</th>
<th>Corr</th>
<th>Disc</th>
</tr>
</thead>
<tbody>
<tr>
<td>S2S-FLAT</td>
<td>0.6360</td>
<td>-</td>
<td>94.03</td>
<td>63.85</td>
<td>30.87</td>
<td>0.7455</td>
<td>-</td>
<td>98.77</td>
<td>77.09</td>
<td>79.04</td>
</tr>
<tr>
<td>S2S-TOKEN</td>
<td>0.7441⁺</td>
<td>-</td>
<td>92.29</td>
<td>69.02⁺</td>
<td>42.29⁺</td>
<td>0.7493⁺</td>
<td>-</td>
<td>96.7</td>
<td>81.56⁺</td>
<td>83.93⁺</td>
</tr>
<tr>
<td>S2S-TREE</td>
<td>0.7458⁺</td>
<td>94.86</td>
<td>93.59</td>
<td>83.85⁺</td>
<td>54.35⁺</td>
<td>0.7612⁺</td>
<td>92.5</td>
<td>95.26</td>
<td>87.61⁺</td>
<td>85.97⁺</td>
</tr>
<tr>
<td>S2S-CONSTR</td>
<td>0.7469⁺</td>
<td>99.25</td>
<td>94.33</td>
<td>85.89⁺</td>
<td>66.09⁺</td>
<td>0.7660⁺</td>
<td>96.92</td>
<td>95.30</td>
<td>91.82⁺</td>
<td>93.44⁺</td>
</tr>
</tbody>
</table>

Tree-based representations result in responses that are more consistent with desired discourse structure.

Constrained decoding improves semantic correctness further.
Observations

- Constrained decoding can still result in incorrect generations
  - Model generates surface forms without generating the non-terminal
  - [restaurant <name>] is not family friendly and serves [cuisine Indian] food.
- Occasional (<1%) stuttering due to constrained decoding
- Unnatural phrasings
  - Yes, you should wear [ATTIRE an umbrella]
Generalization Study

Trained two models:
1) On E2E flat data
2) On E2E flat data + all Weather data

Gradually add more compositional E2E examples to training
Tree2Seq Results

- No big differences in BLEU and Grammaticality score
- Tree2Seq + constrained decoding achieved the best tree accuracy
- Human correctness highly correlates to tree accuracy
Model Effectiveness

Tree Acc by % data

Weather Data Efficiency Study

Tree Acc by # DialogActs

Weather Tree Accuracy Distribution
## Model Samples

<table>
<thead>
<tr>
<th>Column</th>
<th>G/T/C</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR Input</td>
<td>-</td>
<td><img src="image1.png" alt="Image" /></td>
</tr>
<tr>
<td>Annotated Reference</td>
<td>-</td>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td>S2S</td>
<td>1/0/0</td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td>T2S</td>
<td>1/0/0</td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
<tr>
<td>T2S-Constr</td>
<td>1/1</td>
<td><img src="image5.png" alt="Image" /></td>
</tr>
</tbody>
</table>

- G/T/C stands for grammatical/tree accuracy/correct (human annotation).
Grammaticality Filter
## Correctness filter motivation

<table>
<thead>
<tr>
<th>Input</th>
<th>[INFORM [CONDITION sunny ] [DATE_TIME today] ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct Response</td>
<td>It will be sunny today.</td>
</tr>
<tr>
<td>Slot Drop</td>
<td>It will be sunny.</td>
</tr>
<tr>
<td>Hallucination</td>
<td>It will be cloudy today.</td>
</tr>
<tr>
<td>Added information</td>
<td>It will be sunny and 45 F today.</td>
</tr>
</tbody>
</table>
Correctness filter

Idea: Use tree structure predicted by model to automatically detect incorrect responses

<table>
<thead>
<tr>
<th>Input</th>
<th>[INFORM [CONDITION sunny ] [DATE_TIME today] ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct Response with non-terminals</td>
<td>[INFORM It will be [CONDITION sunny ] [DATE_TIME today] ] .</td>
</tr>
<tr>
<td>Correct Response with fine-grained non-terminals</td>
<td>[INFORM It will be [CONDITION_SUNNY sunny ] [DATE_TIME today] ] .</td>
</tr>
<tr>
<td>Response with Hallucination and fine-grained non-terminals</td>
<td>[INFORM It will be [CONDITION_CLOUDY cloudy ] [DATE_TIME today] ] .</td>
</tr>
</tbody>
</table>
Grammaticality filter

Expect *raining* today in Miami.

- DocNN: Biased for precision, 0.85 recall at 0.98 precision
- BERT: 0.99 precision at 0.90 recall
- Future work:
  - Develop a domain-agnostic grammaticality filter
  - Distill BERT to smaller model for product serving.

Expect *raining* today in Miami.
Conclusions

- Shown that tree structured representations can greatly improve controllability of generated text
- Introduced a simple constrained decoding technique that is only applied to the decoder
- Introduced a tree-to-sequence model for better structure-aware learning
- Released a conversational NLG corpus for the weather domain
- Introduced grammatical filtering approach

Check out our dataset and code:

https://github.com/facebookresearch/TreeNLG
Future Work

- Add constraints into training (trainable beam search)
- Add LM/reranking approach for better grammaticality and correctness
- Adversarial approaches to improve model robustness
- Add multi-modal contexts into generation
- Personalized NLG model
- Cross-domain and cross-lingual generalization
- Data Augmentation to leverage public data for domain bootstrapping
Thank you!

https://github.com/facebookresearch/TreeNLG