Controlling Text Generation

Alexander Rush
(based on work by Sam Wiseman,
Sebastian Gehrmann, and Yuntian Deng)

HarvardNLP
1 Review: Neural Machine Translation

2 Past Work: Some Progress on Text Generation
   - Summary
   - Image-to-Markup
   - Data-to-Text

3 Current Research: Uphill Battles in Generation
   - Generation Analysis
   - Controlling Generation
1) Preface: End-to-End Models for NLP
Example: Neural Machine Translation (Sutskever et al., 2014)

Over the line!
Example: Neural Machine Translation (Sutskever et al., 2014)
Example: Neural Machine Translation (Sutskever et al., 2014)

Over the line ! <s>
Example: Neural Machine Translation (Sutskever et al., 2014)
Example: Neural Machine Translation (Sutskever et al., 2014)
Example: Neural Machine Translation (Sutskever et al., 2014)
Example: Neural Machine Translation (Sutskever et al., 2014)
Example: Neural Machine Translation (Sutskever et al., 2014)
Model 1: Seq2Seq Model

Encoder \((\text{enc}(x))\):

\[
h^x_m \leftarrow \text{RNN}(h^x_{m-1}, x_m)
\]

Context:

\[
c_n = h^x_M
\]

Decoder \((\text{dec}(c_n))\):

\[
h_n \leftarrow \text{RNN}(h_{n-1}, w_n)
\]

Prediction:

\[
p(w_{n+1} \mid w_{1:n}, x_{1:M}) = \text{softmax}(W[h_n, c_n])
\]
Attention-based Neural Machine Translation (Bahdanau et al., 2015)
Attention-based Neural Machine Translation (Bahdanau et al., 2015)
Attention-based Neural Machine Translation (Bahdanau et al., 2015)
Attention-based Neural Machine Translation (Bahdanau et al., 2015)
Attention-based Neural Machine Translation (Bahdanau et al., 2015)
Attention-based Neural Machine Translation (Bahdanau et al., 2015)
Attention-based Neural Machine Translation (Bahdanau et al., 2015)
Attention-based Neural Machine Translation (Bahdanau et al., 2015)
Attention-based Neural Machine Translation (Bahdanau et al., 2015)
Attention-based Neural Machine Translation (Bahdanau et al., 2015)
Model 2: Seq2Seq+Attention Model

**Encoder (\(enc(x)\)):**

\[
\mathbf{h}_m^x \leftarrow \text{RNN}(\mathbf{h}_{m-1}^x, x_m)
\]

**Attention**

\[
p_{\text{att}}(m) \leftarrow \text{softmax}([\mathbf{h}_1^x; \ldots; \mathbf{h}_M^x]^{\top} \mathbf{h}_n)
\]

\[
\mathbf{c}_n \leftarrow \mathbb{E}_{m \sim p_{\text{att}}} [\mathbf{h}_m^x] = \sum_{m=1}^{M} p_{\text{att}}(m) \mathbf{h}_m^x
\]

**Decoder (\(dec(c_n)\))**

\[
\mathbf{h}_n \leftarrow \text{RNN}(\mathbf{h}_{n-1}, w_n)
\]

**Prediction**

\[
p(w_{n+1}|w_{1:n}, x_{1:M}) = \text{softmax}(\mathbf{W}[\mathbf{h}_n, \mathbf{c}_n])
\]
Attention-based Neural Machine Translation (Bahdanau et al., 2015)
Model 3: Seq2Seq+Attention+Copy Model (Gulcehre et al, 2016) ...

Encoder \( (enc(x)) \):
\[
\mathbf{h}_m^x \leftarrow \text{RNN}(\mathbf{h}_{m-1}^x, x_m)
\]

Attention
\[
p_{\text{att}}(m) \leftarrow \text{softmax}([\mathbf{h}_1^x; \ldots; \mathbf{h}_M^x]^{\top} \mathbf{h}_n)
\]
\[
\mathbf{c}_n \leftarrow \mathbb{E}_{m \sim p_{\text{att}}} [\mathbf{h}_m^x] = \sum_{m=1}^{M} p_{\text{att}}(m) \mathbf{h}_m^x
\]

Decoder \( (dec(c_n)) \):
\[
\mathbf{h}_n \leftarrow \text{RNN}(\mathbf{h}_{n-1}, w_n)
\]

Prediction:
\[
p_{\text{gen}} = \sigma(U[\mathbf{h}_n, \mathbf{c}_n])
\]
\[
p(w_{n+1}|w_{1:n}, x_{1:M}) = p_{\text{gen}} \times \text{softmax}(W[\mathbf{h}_n, \mathbf{c}_n])
\]
\[
+ (1 - p_{\text{gen}}) \times \mathbb{E}_{m \sim p_{\text{att}}} [\mathbf{1}(w_{n+1} = x_m)]
\]
Model 3: Seq2Seq+Attention+Copy Model (Gulcehre et al, 2016) ...

Encoder ($enc(x)$):

$$h^x_m \leftarrow \text{RNN}(h^x_{m-1}, x_m)$$

Attention

$$p_{\text{att}}(m) \leftarrow \text{softmax}([h^x_1; \ldots; h^x_M] \top h_n)$$

$$c_n \leftarrow \mathbb{E}_{m \sim p_{\text{att}}} [h^x_m] = \sum_{m=1}^{M} p_{\text{att}}(m) h^x_m$$

Decoder ($dec(c_n)$):

$$h_n \leftarrow \text{RNN}(h_{n-1}, w_n)$$

Prediction:

$$p_{\text{gen}} = \sigma(U[h_n, c_n])$$

$$p(w_{n+1}|w_{1:n}, x_{1:M}) = p_{\text{gen}} \times \text{softmax}(W[h_n, c_n])$$

$$+ (1 - p_{\text{gen}}) \times \mathbb{E}_{m \sim p_{\text{att}}} [1(w_{n+1} = x_m)]$$
Model 3: Seq2Seq+Attention+Copy Model (Gulcehre et al, 2016) ...

Encoder ($enc(x)$):

$$h^x_m \leftarrow \text{RNN}(h^x_{m-1}, x_m)$$

Attention

$$p_{\text{att}}(m) \leftarrow \text{softmax}([h^x_1; \ldots; h^x_M] \top h_n)$$

$$c_n \leftarrow \mathbb{E}_{m \sim p_{\text{att}}}[h^x_m] = \sum_{m=1}^M p_{\text{att}}(m)h^x_m$$

Decoder ($dec(c_n)$):

$$h_n \leftarrow \text{RNN}(h_{n-1}, w_n)$$

Prediction:

$$p_{\text{gen}} = \sigma(U[h_n, c_n])$$

$$p(w_{n+1}|w_{1:n}, x_{1:M}) = p_{\text{gen}} \times \text{softmax}(W[h_n, c_n])$$

$$+ (1 - p_{\text{gen}}) \times \mathbb{E}_{m \sim p_{\text{att}}}[\mathbf{1}(w_{n+1} = x_m)]$$
Model 3: Seq2Seq+Attention+Copy Model (Gulcehre et al, 2016) ...

Encoder \((\text{enc}(x))\):

\[
\mathbf{h}_{m}^{x} \leftarrow \text{RNN}(\mathbf{h}_{m-1}^{x}, x_{m})
\]

Attention

\[
p_{\text{att}}(m) \leftarrow \text{softmax}([\mathbf{h}_{1}^{x}; \ldots; \mathbf{h}_{M}^{x}]^\top \mathbf{h}_{n})
\]

\[
\mathbf{c}_{n} \leftarrow \mathbb{E}_{m \sim p_{\text{att}}} [\mathbf{h}_{m}^{x}] = \sum_{m=1}^{M} p_{\text{att}}(m) \mathbf{h}_{m}^{x}
\]

Decoder \((\text{dec}(c_{n}))\):

\[
\mathbf{h}_{n} \leftarrow \text{RNN}(\mathbf{h}_{n-1}, w_{n})
\]

Prediction:

\[
p_{\text{gen}} = \sigma(\mathbf{U}[\mathbf{h}_{n}, \mathbf{c}_{n}])
\]

\[
p(w_{n+1}|w_{1:n}, x_{1:M}) = p_{\text{gen}} \times \text{softmax}(\mathbf{W}[\mathbf{h}_{n}, \mathbf{c}_{n}])
\]

\[
+ (1 - p_{\text{gen}}) \times \mathbb{E}_{m \sim p_{\text{att}}} [\mathbf{1}(w_{n+1} = x_{m})]
\]
Applications From HarvardNLP: OpenNMT

Home

OpenNMT is a industrial-strength, open-source (MIT) neural machine translation system utilizing the Torch/PyTorch mathematical toolkit.

OpenNMT is used as provided in production by major translation providers. The system is designed to be simple to use and easy to extend, while maintaining efficiency and state-of-the-art translation accuracy.
Applications From HarvardNLP: Seq2Seq-Vis

Innovations in communication have increased the contact among different cultures and shortened distances, changing our ways and the society we live in.

The day before yesterday we arrived at "unk" where we enjoy sightseeing in "unk" touristic town "unk".

He was born in "unk", "unk" (Austria). Johann "unk" was born 1810 in "unk" (Oberösterreich) geboren. Im Alter von 5 Jahren bekam er seinen ersten "unk".

What we have an effect is the Commission which, when it comes to coordination, draws up draft recommendations.

Many shopping opportunities and restaurants are in the immediate vicinity.

Apple Hostsels of Philadelphia is located in a newly renovated 19th Century building and offers all the modern comforts you would expect.

Elections do occur in the Arab world, and they vary in frequency and significance.

I think it is of interest to take part in the second European Conference scheduled for 5 October, given that issues of mutual concern will be discussed there.
2) Applications: End-to-End Natural Language Generation

Natural language generation is the process of deliberately constructing a natural language text in order to meet specified communicative goals. - MacDonald (1987)
Common Practice Templated Generation / Intents

```
"types": [
    [
        "name": "articleType",
        "values": [
            {
                "name": {
                    "value": "a"
                }
            },
            {
                "name": {
                    "value": "an"
                }
            },
            {
                "name": {
                    "value": "the"
                }
            }
        ]
    ],
    [
        "name": "atTheType",
        "values": [
            {
                "name": {
                    "value": "at the"
                }
            },
            {
                "name": {
                    "value": "on the"
                }
            },
            {
                "name": {
                    "value": "around the"
                }
            }
        ]
    ],
    [
        "name": "sizeType",
        "values": [
            {
                "name": {
                    "value": "large",
                    "synonyms": ["huge", "trick", "gigantic", "eat me out of house", "scary big", "ginormous", "ride", "waist height"
                ]
            },
            {
                "name": {
                    "value": "medium",
                    "synonyms": ["bigger than a cat", "on the bed", "up to my knees", "average"
                ]
            },
            {
                "name": {
                    "value": "small",
                    "synonyms": ["little", "take on an airplane"
                ]
            },
            {
                "name": {
                    "value": "tiny",
                    "synonyms": ["cheap to feed", "teacup", "pocket", "yipppy"
                ]
            }
        ]
    ]
```
End-to-End Generation

- Neural MT has inspired interest in generation with E2E models.
- Differs significantly from much past work in NLG.
- Three areas we have worked on:
  - Summarization
  - Image-to-Markup
  - Data-to-Text
- Many others, e.g. image/video captioning, chatbots, dialogue response generation.
mexico city, mexico - cnn - heavy rains and flooding have forced hundreds of thousands of people from homes in southern mexico 's state of tabasco over the past four days , with nearly as many trapped by the rising waters , state officials said thursday . officials say about 300,000 people are still trapped by the worst flooding in the region for 50 years . the grijalva river pushed over its banks through the state capital of villahermosa on thursday , forcing government workers to evacuate and leaving up to 80 percent of the city flooded , gov. andres granier 's office told cnn . about 700,000 people have seen their homes flooded , with about 300,000 of those still trapped there , granier 's office reported . one death had been blamed on the floods , which followed weeks of heavy rain in the largely swampy state . tabasco borders guatemala to the south and the gulf of mexico to the north . . . .
mexico city, mexico — heavy rains and flooding have forced hundreds of thousands of people from homes in southern mexico’s state of tabasco over the past four days, with nearly as many trapped by the rising waters, state officials said thursday. officials say about 300,000 people are still trapped by the worst flooding in the region for 50 years. the grijalva river pushed over its banks through the state capital of villahermosa on thursday, forcing government workers to evacuate and leaving up to 80 percent of the city flooded, gov. andres granier’s office told cnn. about 700,000 people have seen their homes flooded, with about 300,000 of those still trapped there, granier’s office reported. one death had been blamed on the floods, which followed weeks of heavy rain in the largely swampy state. tabasco borders guatemala to the south and the gulf of mexico to the north. . . .
Abstractive E2E Sentence Summarization  (Rush et al, 2015)

Input (First Sentence)

*Russian Defense Minister Ivanov called Sunday for the creation of a joint front for combating global terrorism.*

Output (Title)

*Russia calls for joint front against terrorism.*
Abstractive E2E Sentence Summarization  (Rush et al, 2015)

Input (First Sentence)

Russian Defense Minister Ivanov called Sunday for the creation of a joint front for combating global terrorism.

Output (Title)

Russia calls for joint front against terrorism.
Abstractive E2E Sentence Summarization  (Rush et al, 2015)

Input (First Sentence)

*Russian Defense Minister Ivanov called Sunday for the creation of a joint front for combating global terrorism.*

Output (Title)

*Russia calls for joint front against terrorism.*
GERMANY IMPLEMENTS TEMPORARY BORDER CHECKS TO LIMIT MIGRANTS

BY GEIR MOULSON AND SHAWN POGATCHNIK
ASSOCIATED PRESS

BERLIN (AP) -- Germany introduced temporary border controls Sunday to stem the tide of thousands of refugees streaming across its frontier, sending a clear message to its European partners that it needs more help with an influx that is straining its ability to cope.

Germany is a preferred destination for many people fleeing Syria's civil war and other troubled nations in the migration crisis that has bitterly divided Europe. They have braved dangerous sea crossings in flimsy
Minimalist blogging platform Posterous drew its last breath earlier this week.

The service, a favorite among mobile bloggers who liked to post on the go, officially shut down four years after it was originally created, one year after it was purchased by Twitter, and two months after it informed users that it was closing.

That came on the heels of other closures, and announcements thereof. Google Reader, the popular RSS tool, will shut down July 1 and EveryBlock, the hyper-local news site, was shuttered in February.

It's a reality of the Internet that sites are constantly starting up, shutting down or getting acquired. But that doesn't make the loss of a beloved site any less upsetting or inconvenient for its faithful fans.

To preserve your sanity, and your data, here are a few tips for handling the death of a favorite website or service.

**Pay attention to warnings**

Most sites won't shutter without giving their users official notice. To avoid being caught off-guard, read any updates, e-mails, blog posts or tweets from the company warning of major changes or sharing goodbyes.
E2E Generation Challenge:
Talk about the Environment (Multimodal)
E2E Generation Challenge:
Talk about the Environment (Multimodal)

Dow and S&P 500 close out week at all-time highs ...
Image-to-Latex Dataset (Deng et al, 2017)
Coarse-to-Fine Attention

\[ r = \frac{\sqrt{Q_3}}{l} \sin \left( \frac{l}{\sqrt{Q_3}} u \right). \]
Coarse-to-Fine Attention

\[ r = \frac{\sqrt{Q_3}}{l} \sin \left( \frac{l}{\sqrt{Q_3}} u \right), \]
Coarse-to-Fine Attention

\[ r = \frac{\sqrt{Q_3}}{l} \sin \left( \frac{l}{\sqrt{Q_3}} u \right) , \]
Coarse-to-Fine Attention

\[ r = \frac{\sqrt{Q_3}}{l} \sin \left( \frac{l}{\sqrt{Q_3} u} \right), \]
Coarse-to-Fine Attention

\[ r = \frac{\sqrt{Q_3}}{l} \sin \left( \frac{l}{\sqrt{Q_3} u} \right), \]
Coarse-to-Fine Attention

\[ r = \sqrt{Q_3} \cdot \sin \left( \frac{l}{\sqrt{Q_3} u} \right), \]
Baseline Results

![Bar chart showing exact match accuracy for different models: CTC (7.60%), COMMERCIAL Model (15.60%), CAPTION (53.53%), IM2TEX (77.46%).]
Handwritten Formulas

- Synthetic handwritten formulas by using handwritten characters [Kirsch, 2010] as font, used for pretraining
- Finetune and evaluate on CROHME 13 and 14 (8K training set)
**Handwritten Formulas**

- Synthetic handwritten formulas by using handwritten characters [Kirsch, 2010] as font, used for pretraining
- Finetune and evaluate on CROHME 13 and 14 (8K training set)

  CROHME 13 (*uses private in-domain handwritten training data)
The Atlanta Hawks defeated the Miami Heat, 103 - 95, at Philips Arena on Wednesday. Atlanta...
The Atlanta Hawks defeated the Miami Heat, 103 - 95, at Philips Arena on Wednesday. Atlanta...
A coffee shop located on the riverside called The Golden Palace, has a 5 out of 5 customer rating. Its price range are fairly cheap for its excellent Fast food.
Harvard System (Gehrmann et al, 2018)

Details:

- Copy mechanism
- Coverage and length beam search
- Built in OpenNMT-py
- Transformer and LSTM model
- Diverse Ensembling (Lee et al, 2016)
Results: 2018 E2E Challenge (Novikova et al, 2017)

- Total of 60 submissions by 16 institutions with about 1/3 of these submissions coming from industry.

- Harvard systems finished first in ROUGE, CIDEr, and second in METEOR

- Best system: 2pts ROUGE improvement of baseline model.

- Human evaluation results more mixed, all systems similar.
How can we maintain coherence through long-form text outputs?

How can we discover complex relationships in source input?

When is textual improvisation allowed versus literal mappings?
3) Uphill Battles in E2E Generation

1. Challenges in Data-to-Document Generation
   (with Sam Wiseman)

2. Controllable Generation
The Atlanta Hawks defeated the Miami Heat, 103 - 95, at Philips Arena on Wednesday. Atlanta was in desperate need of a win and they were able to take care of a shorthanded Miami team here. Defense was key for the Hawks, as they held the Heat to 42 percent shooting and forced them to commit 16 turnovers. Atlanta also dominated in the paint, winning the rebounding battle, 47 - 34, and outscoring them in the paint 58 - 26. The Hawks shot 49 percent from the field and assisted on 27 of their 43 made baskets. This was a near wire-to-wire win for the Hawks, as Miami held just one lead in the first five minutes. Miami ( 7 - 15 ) are as beat-up as anyone right now and it's taking a toll on the heavily used starters. Hassan Whiteside really struggled in this game, as he amassed eight points, 12 rebounds and one blocks on 4 - of - 12 shooting ...
### Vocab Statistics

<table>
<thead>
<tr>
<th></th>
<th>RoboCup</th>
<th>WeatherGov</th>
<th>RotoWire</th>
<th>SBNation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocab</td>
<td>409</td>
<td>394</td>
<td>11,331</td>
<td>68,574</td>
</tr>
<tr>
<td>Tokens</td>
<td>11K</td>
<td>0.9M</td>
<td>1.6M</td>
<td>8.8M</td>
</tr>
<tr>
<td>Examples</td>
<td>1,919</td>
<td>22,146</td>
<td>4,853</td>
<td>10,903</td>
</tr>
<tr>
<td>Avg Len</td>
<td>5.7</td>
<td>28.7</td>
<td>337.1</td>
<td>805.4</td>
</tr>
<tr>
<td>Field Types</td>
<td>4</td>
<td>10</td>
<td>39</td>
<td>39</td>
</tr>
<tr>
<td>Avg Records</td>
<td>2.2</td>
<td>191</td>
<td>628</td>
<td>628</td>
</tr>
</tbody>
</table>

### Player Types

<table>
<thead>
<tr>
<th>POSN</th>
<th>MIN</th>
<th>PTS</th>
<th>FGM</th>
<th>FGA</th>
<th>FG-PCT</th>
<th>FG3M</th>
<th>FG3A</th>
<th>FG3-PCT</th>
<th>FTM</th>
<th>FTA</th>
<th>FT-PCT</th>
<th>OREB</th>
<th>DREB</th>
<th>REB</th>
<th>AST</th>
<th>TOV</th>
<th>STL</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLK</td>
<td>PF</td>
<td>NAME1</td>
<td>NAME2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Team Types

<table>
<thead>
<tr>
<th>PTS-QTR1</th>
<th>PTS-QTR2</th>
<th>PTS-QTR3</th>
<th>PTS-QTR4</th>
<th>PTS</th>
<th>FG-PCT</th>
<th>FG3-PCT</th>
<th>FT-PCT</th>
<th>REB</th>
</tr>
</thead>
<tbody>
<tr>
<td>AST</td>
<td>TOV</td>
<td>WINS</td>
<td>LOSSES</td>
<td>CITY</td>
<td>NAME</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **RoboCup**
- **WeatherGov**
- **RotoWire**
- **SBNation**
Content Encoding with Cell Embeddings

- \( \{r_1, \ldots, r_S\} \)

- \( r.t = \text{POINTS} \), and such that entity \( r.e = (\text{Tyler Johnson}) \) and value \( r.m = 27 \) (Liang et al, 2009)

- \( s_j = E(r_j) \) for \( j \in \{1, \ldots S\} \)

<table>
<thead>
<tr>
<th>PLAYER</th>
<th>AS</th>
<th>RB</th>
<th>PT</th>
<th>FG</th>
<th>FGA</th>
<th>CITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyler Johnson</td>
<td>5</td>
<td>2</td>
<td>8</td>
<td>16</td>
<td></td>
<td>Miami</td>
</tr>
<tr>
<td>Dwight Howard</td>
<td>11</td>
<td>17</td>
<td>23</td>
<td>9</td>
<td>11</td>
<td>Atlanta</td>
</tr>
<tr>
<td>Paul Millsap</td>
<td>2</td>
<td>9</td>
<td>21</td>
<td>8</td>
<td>12</td>
<td>Atlanta</td>
</tr>
<tr>
<td>Goran Dragic</td>
<td>4</td>
<td>2</td>
<td>8</td>
<td>17</td>
<td></td>
<td>Miami</td>
</tr>
<tr>
<td>Wayne Ellington</td>
<td>2</td>
<td>3</td>
<td>19</td>
<td>7</td>
<td>15</td>
<td>Miami</td>
</tr>
<tr>
<td>Dennis Schroder</td>
<td>7</td>
<td>4</td>
<td>8</td>
<td>15</td>
<td></td>
<td>Atlanta</td>
</tr>
<tr>
<td>Rodney McGruder</td>
<td>5</td>
<td>5</td>
<td>11</td>
<td>3</td>
<td>8</td>
<td>Miami</td>
</tr>
</tbody>
</table>

...
Additional Extensions: Machine Translation

Architectural changes focusing on source selection.

- Copy Attention, Pointer Networks, and Reconstruction Models
The ⟨team1⟩ (⟨wins1⟩-⟨losses1⟩) defeated the ⟨team2⟩ (⟨wins2⟩-⟨losses2⟩) ⟨pts1⟩-⟨pts2⟩.

(6×)

⟨player⟩ scored ⟨pts⟩ points (⟨fgm⟩-⟨fga⟩ FG, ⟨tpm⟩-⟨tpa⟩ 3PT, ⟨ftm⟩-⟨fta⟩ FT) to go with ⟨reb⟩ rebounds.

The ⟨team1⟩ next game will be at home against the Dallas Mavericks, while the ⟨team2⟩ will travel to play the Bulls.
<table>
<thead>
<tr>
<th>Beam</th>
<th>Model</th>
<th>Development</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>PPL</td>
</tr>
<tr>
<td>Template</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>1</td>
<td>Joint Copy</td>
<td>7.46</td>
</tr>
<tr>
<td></td>
<td>Joint Copy + Rec</td>
<td>7.25</td>
</tr>
<tr>
<td></td>
<td>Joint Copy + Rec + TVD</td>
<td>7.22</td>
</tr>
<tr>
<td></td>
<td>Conditional Copy</td>
<td>7.44</td>
</tr>
<tr>
<td>5</td>
<td>Joint Copy</td>
<td>7.46</td>
</tr>
<tr>
<td></td>
<td>Joint Copy + Rec</td>
<td>7.25</td>
</tr>
<tr>
<td></td>
<td>Joint Copy + Rec + TVD</td>
<td>7.22</td>
</tr>
<tr>
<td></td>
<td>Conditional Copy</td>
<td>7.44</td>
</tr>
</tbody>
</table>
The Utah Jazz (38 - 26) defeated the Houston Rockets (38 - 26) 117 - 91 on Wednesday at Energy Solutions Arena in Salt Lake City. The Jazz got out to a quick start in this one, out-scoring the Rockets 31 - 15 in the first quarter alone. Along with the quick start, the Rockets were the superior shooters in this game, going 54 percent from the field and 43 percent from the three-point line, while the Jazz went 38 percent from the floor and a meager 19 percent from deep. The Rockets were able to out-rebound the Rockets 49 - 49, giving them just enough of an advantage to secure the victory in front of their home crowd. The Jazz were led by the duo of Derrick Favors and James Harden. Favors went 2 - for - 6 from the field and 0 - for - 1 from the three-point line to score a game-high of 15 points, while also adding four rebounds and four assists.
The Utah Jazz (38 - 26) defeated the Houston Rockets (38 - 26) 117 - 91 on Wednesday at Energy Solutions Arena in Salt Lake City. The Jazz got out to a quick start in this one, out-scoring the Rockets 31 - 15 in the first quarter alone. Along with the quick start, the Rockets were the superior shooters in this game, going 54 percent from the field and 43 percent from the three-point line, while the Jazz went 38 percent from the floor and a meager 19 percent from deep. The Rockets were able to out-rebound the Rockets 49 - 49, giving them just enough of an advantage to secure the victory in front of their home crowd. The Jazz were led by the duo of Derrick Favors and James Harden. Favors went 2 - for - 6 from the field and 0 - for - 1 from the three-point line to score a game-high of 15 points, while also adding four rebounds and four assists ....
Generations are fluent and accurate...

- Along with the quick start, the Rockets were the superior shooters in this game, going 54 percent from the field and 43 percent from the three-point line

... but also have issues in the content and discourse

- The Rockets were able to out-rebound the Rockets (incorrect discourse)
- The Jazz were led by the duo of Derrick Favors and James Harden (wrong player selected)
- to score a game-high (non-factual) of 15 points
Question 2: How can we quantify the issues in generation?

Criteria:

1. Relation Generation: Referring expressions should be easy trace.

2. Content Selection: Relevant content should be generated.

3. Content Ordering: Discourse structure should be consistent.

Observation: NLU is currently a lot easier than NLG.
Extractive Evaluation

Use information extraction system for generations (details in paper)

Criteria:

1. **Relation Generation:** Referring expressions should be easy trace.
   - Precision and count of identified data points.

2. **Content Selection:** Relevant content should be generated.
   - F-score on generated data points.

3. **Content Ordering:** Discourse structure should be consistent.
   - Damerau-Levenshtein distance between ordered elements.
## Higher-Level Properties

<table>
<thead>
<tr>
<th>Beam</th>
<th>Model</th>
<th>RG</th>
<th>CS</th>
<th>CO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>P%</td>
<td>P%</td>
<td>R%</td>
</tr>
<tr>
<td></td>
<td>Template</td>
<td>99.35</td>
<td>45.17</td>
<td>24.85</td>
</tr>
<tr>
<td>B=1</td>
<td>Joint Copy</td>
<td>47.55</td>
<td>20.53</td>
<td>22.49</td>
</tr>
<tr>
<td></td>
<td>Conditional Copy</td>
<td>68.94</td>
<td>25.15</td>
<td>22.94</td>
</tr>
<tr>
<td>B=5</td>
<td>Joint Copy</td>
<td>47.00</td>
<td>16.52</td>
<td>26.08</td>
</tr>
<tr>
<td></td>
<td>Conditional Copy</td>
<td>71.07</td>
<td>21.90</td>
<td>27.27</td>
</tr>
</tbody>
</table>
Next Steps: Robust Generation

- Current systems target fluency, very hard to check for accuracy.
- Unlike NMT, NLG work needs high-precision outputs, very hard with black box models.
- Research focus?: learning template-based generation systems.
3) Uphill Battles in Generation

1. Challenges in Data-to-Document Generation

2. Controllable Generation with Neural Templates

(Work in Progress with Sam Wiseman)
Project Aim

Can we build a neural generation system that is:

1. Interpretable in its content selection.
2. Easily controllable in terms of style and form.

Tension between end-to-end neural approach and desired modularity.
Project Aim

Can we build a neural generation system that is:

1. Interpretable in its content selection.

2. Easily controllable in terms of style and form.

Tension between end-to-end neural approach and desired modularity.
Standard Copy Generation Approach

Step 1: Encode the Source

Fitzbillies, ty[coffee shop], pr[< £20], food[Chinese], cust[3/5], area[city centre]

Step 2: Generate with Copy Decoder

Fitzbillies is a coffee shop providing Chinese food in the moderate price range. It is located in the city centre. Its customer rating is 3 out of 5.
(Neural) Template Generation Approach

Step 1: Encode the Source
Fitzbillies, ty [coffee shop], pr [< £20], food [Chinese], cust [3/5], area [city centre]

Step 2: Select a Template

The ___ is a ___ expensive ___ providing ___ offering ___ food ___ in the ___

high moderate less than average ___ price range ___ It is located in the located near ___ ___

Its customer rating is ___ out of ___.

Step 3: Fill-in Each Segment

∥ Fitzbillies ∥ is a ∥ coffee shop ∥ providing ∥ Chinese ∥ food ∥ in the ∥

moderate ∥ price range ∥ . ∥ It is ∥ located in the ∥ city centre ∥ . ∥
(Neural) Template Generation Approach

Step 1: Encode the Source
Fitzbillies, ty[coffee shop], pr[< £20], food[Chinese], cust[3/5], area[city centre]

Step 2: Select a Template

| The ___ | is an expensive | ___ | providing | ___ | food | ___ | in the ___ |
| high | moderate | price range | . | It is | located in the | near | ___ | . |
| less than average | . | Its customer rating is | Their customer rating is | Customers have rated it | ___ out of ___ | . |

Step 3: Fill-in Each Segment
|| Fitzbillies || is a || coffee shop || providing || Chinese || food || in the || moderate || price range || . || It is || located in the || city centre || . ||
(Neural) Template Generation Approach

Step 1: Encode the Source
Fitzbillies, ty [coffee shop], pr [< £20], food [Chinese], cust [3/5], area [city centre]

Step 2: Select a Template

| The ____ | is a ____ | is an ____ | providing ____ | ____ | food ____ | in the ____ |
| high ____ | moderate ____ | price ____ | offering ____ | ____ | price range ____ | ... |
| __ | ____ | Its customer rating is ____ |
| ____ | ____ | Their customer rating is ____ |
| ____ | ____ | Customers have rated it ____ |

Step 3: Fill-in Each Segment

| Fitzbillies | is a ____ | coffee shop ____ | providing ____ | Chinese ____ | food ____ | in the ____ |
| moderate ____ | price range ____ | ... | It is ____ | located in the ____ | city centre ____ | ... |
(Neural) Template Generation Approach

Step 1: Encode the Source
Fitzbillies, ty[coffee shop], pr[< £20], food[Chinese], cust[3/5], area[city centre]

Step 2: Select a Template

| The ___ | is a ___ | ___ | providing ___ | ___ | food ___ | in the ___ |
| high | moderate | less than average | price range | . | It is located in the ___ |
| ___ | ___ | ___ | ___ | ___ | . |
| Its customer rating is ___ | Their customer rating is ___ |
| Customers have rated it ___ out of ___ | . |

Step 3: Fill-in Each Segment

∥ Fitzbillies ∥ is a ∥ coffee shop ∥ providing ∥ Chinese ∥ food ∥ in the ∥ moderate ∥ price range ∥ . ∥ It is ∥ located in the ∥ city centre ∥ . ∥
(Neural) Template Generation Approach

Step 1: Encode the Source
Fitzbillies, ty[coffee shop], pr[< £20], food[Chinese], cust[3/5], area[city centre]

Step 2: Select a Template

\[
\begin{array}{|c|c|c|c|c|}
\hline
\text{The} & \text{is a} & \text{is an} & \text{serving} & \text{food} \\
\hline
\text{...} & \text{...} & \text{...} & \text{...} & \text{...} \\
\text{high} & \text{moderate} & \text{less than average} & \text{price range} & \text{located in the} \\
\text{...} & \text{...} & \text{...} & \text{...} & \text{...} \\
\text{Its customer rating is} & \text{Their customer rating is} & \text{Customers have rated it} & \text{...} & \text{...} \\
\hline
\end{array}
\]

Step 3: Fill-in Each Segment

\[
\begin{array}{|c|c|c|c|c|}
\hline
\text{Fitzbillies} & \text{is a} & \text{coffee shop} & \text{serving} & \text{Chinese} \\
\hline
\text{...} & \text{...} & \text{...} & \text{...} & \text{...} \\
\text{moderate} & \text{price range} & \text{...} & \text{...} & \text{...} \\
\text{...} & \text{...} & \text{...} & \text{...} & \text{...} \\
\text{It is} & \text{located in the} & \text{city centre} & \text{...} & \text{...} \\
\hline
\end{array}
\]
(Neural) Template Generation Approach

Step 1: Encode the Source
Fitzbillies, ty[coffee shop], pr[< £20], food[Chinese], cust[3/5], area[city centre]

Step 2: Select a Template

The ___ is a ___ expensive ___ providing ___ offering ___ food cuisine foods in the ___ high moderate less than average price price range . It is ___ located in the ___ located near near ___ . Its customer rating is Their customer rating is Customers have rated it ___ out of ___ .

Step 3: Fill-in Each Segment

∥ Fitzbillies ∥ is a ∥ coffee shop ∥ providing ∥ Chinese ∥ food ∥ in the ∥ moderate ∥ price range ∥ . ∥ It is ∥ located in the ∥ city centre ∥ .
Criteria

1. Interpretable in its content selection.

   Decisions are localized to a segment of the template.

2. Easily controllable in terms of style and form.

   Alternative realizations through different templates.

However: templates feel much less “end-to-end”.
How can we learn them from data?
Criteria

1. Interpretable in its content selection.

*Decisions are localized to a segment of the template.*

2. Easily controllable in terms of style and form.

*Alternative realizations through different templates.*

However: templates feel much less “end-to-end”.
How can we learn them from data?
Technical Methodology: Hidden Semi-Markov Model

- HMM: discrete latent states with single emissions (e.g. words).
- HSMM: discrete states produce multiple emissions (e.g. phrases).
- Parameterized with transition, emission, and length distributions.
Technical Methodology: Neural Hidden Semi-Markov Model

- Employ HSMM as a conditional latent variable language model, $p(y_1, \ldots, y_T, z \mid x)$.

- Transition Distribution: NN between states.

- Emission Distribution: Seq2Seq+Copy-Attention, one per state $k$. 
Technical Methodology: Learning Templates

- Fit model by maximizing log-marginal likelihood on training data.

\[
\max_\theta \sum_j \ln \sum_z p(y^{(j)}, z \mid x^{(j)}; \theta)
\]

Details: Pre-score segmentations, HSMM forward algorithm for sum, backprop with autograd, all inference is exact.

- Compute argmax segmentations to find common templates.

\[
z^{(j)} = \arg \max_z p(y^{(j)}, z \mid x^{(j)}; \theta)
\]
Technical Methodology: Learning Templates

- Fit model by maximizing log-marginal likelihood on training data.

\[
\max_\theta \sum_j \ln \sum_z p(y^{(j)}, z \mid x^{(j)}; \theta)
\]

Details: Pre-score segmentations, HSMM forward algorithm for sum, backprop with autograd, all inference is exact.

- Compute argmax segmentations to find common templates.

\[
z^{(j)} = \arg\max_z p(y^{(j)}, z \mid x^{(j)}; \theta)
\]
Neural Template

The ___ is a ___ providing serving food ___ in the ___
is an expensive offering ___ cuisine ___.

... high price ... located in the ___.

... moderate price range ... located near ___.

... less than average ... nearby ... Its customer rating is

... price range ... located near ___.

... average ... nearby ... Their customer rating is

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.

... its customer rating is ... located near ___.

... price range ... located near ___.
Experimental Setup

- Two datasets, E2E challenge and WikiBio

- Training with 35 and 65 state models, each 1x300 LSTMs.

- Extract 100 most common templates for each.

- Vocabulary limited to non-copy-able words.

- Generation with beam search with a pre-selected template.
Frederick Parker-Rhodes (21 March 1914 - 21 November 1987) was an English linguist, plant pathologist, computer scientist, mathematician, mystic, and mycologist.
# E2E Challenge

<table>
<thead>
<tr>
<th></th>
<th>BLEU</th>
<th>NIST</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Val</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substitution</td>
<td>43.71</td>
<td>6.72</td>
</tr>
<tr>
<td>Neural Template</td>
<td>66.06</td>
<td>7.93</td>
</tr>
<tr>
<td>Full Neural Model</td>
<td>69.25</td>
<td>8.48</td>
</tr>
<tr>
<td><strong>Test</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substitution</td>
<td>43.78</td>
<td>6.88</td>
</tr>
<tr>
<td>Neural Template</td>
<td>56.72</td>
<td>7.63</td>
</tr>
<tr>
<td>Full Neural Model</td>
<td>65.93</td>
<td>8.59</td>
</tr>
<tr>
<td>Model</td>
<td>BLEU</td>
<td>NIST</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>Conditional KN-LM</td>
<td>19.8</td>
<td>5.19</td>
</tr>
<tr>
<td>NNLM (field)</td>
<td>33.4</td>
<td>7.52</td>
</tr>
<tr>
<td>NNLM (field &amp; word)</td>
<td>34.7</td>
<td>7.98</td>
</tr>
<tr>
<td>Neural Template</td>
<td>33.8</td>
<td>7.51</td>
</tr>
</tbody>
</table>

Custom KN and NNLM Baselines from LeBret et al (2016)
Interpretability

kenny warren

**name:** kenny warren, **birth date:** 1 april 1946,
**birth name:** kenneth warren deutscher, **birth place:** brooklyn, new york,
**occupation:** ventriloquist, comedian, author,
**notable work:** book - the revival of ventriloquism in america

<table>
<thead>
<tr>
<th>No.</th>
<th>Name and Birth Details</th>
<th>Occupation and Notable Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>kenny warren deutscher (april 1, 1946)</td>
<td>American ventriloquist.</td>
</tr>
<tr>
<td>2</td>
<td>kenny warren deutscher (april 1, 1946, brooklyn,)</td>
<td>American ventriloquist.</td>
</tr>
<tr>
<td>3</td>
<td>kenny warren deutscher (april 1, 1946)</td>
<td>American ventriloquist, best known for his the revival of ventriloquism.</td>
</tr>
<tr>
<td>4</td>
<td>“kenny” warren</td>
<td>American ventriloquist.</td>
</tr>
<tr>
<td>5</td>
<td>kenneth warren “kenny” warren (born april 1, 1946)</td>
<td>American ventriloquist, and author.</td>
</tr>
</tbody>
</table>
## Controllability

### The Golden Palace

<table>
<thead>
<tr>
<th>name</th>
<th>type</th>
<th>food</th>
<th>priceRange</th>
<th>custRating</th>
<th>area</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Golden Palace</td>
<td>coffee shop</td>
<td>Chinese</td>
<td>cheap</td>
<td>5 out of 5</td>
<td>city centre</td>
</tr>
</tbody>
</table>

1. The Golden Palace is a coffee shop located in the city centre.
2. In the city centre is a cheap Chinese coffee shop called The Golden Palace.
3. The Golden Palace that serves Chinese food in the cheap price range. It is located in the city centre. Its customer rating is 5 out of 5.
4. The Golden Palace is a Chinese coffee shop.
5. The Golden Palace is a Chinese coffee shop with a customer rating of 5 out of 5.
Conclusion: Challenges in Text Generation

- End-to-end models are a remarkable step forward.

- Still significant challenges when we go beyond the sentence.

- Interpretability and controllability are non-trivial issues.

- Scaling probabilistic inference and structured prediction is getting easier, and presents an interesting step forward.