Sentence-Level Content Planning and Style Specification for Neural Text Generation

Xinyu Hua and Lu Wang
EMNLP 2019

Northeastern University
Motivation

• Text generation in early days

(Joyce [Rambow and Korelsky, 1992])

(FOG [Goldber et al, 1994])
Motivation

• A commonly used architecture [Reiter and Dale, 2000]

Goal

Text Planner

Sentence Planner

Linguistic Realizer

Surface Text
• A commonly used architecture [Reiter and Dale, 2000]
Motivation

- Text generation nowadays

![Diagram](image)
• Text generation nowadays

Training: maximizing log-likelihood of next token
Inference: autoregressive decoding
Is this how humans write?
A cognitive model of writing [Hayes and Flower, 1983]

**GENERATING**: “retrieve information relevant to the writing task from long-term memory”.
Motivation

• A cognitive model of writing [Hayes and Flower, 1983]

**GENERATING:** “retrieve information relevant to the writing task from long-term memory”.

**ORGANIZING:** “select the most useful of the materials retrieved by the GENERATING process and to organize them into a writing plan.”
Motivation

• A cognitive model of writing [Hayes and Flower, 1983]

**GENERATING:** “retrieve information relevant to the writing task from long-term memory”.

**ORGANIZING:** “select the most useful of the materials retrieved by the GENERATING process and to organize them into a writing plan.”

**EDITING:** “evaluate material with respect to the writing goals.”
Motivation

• A cognitive model of writing [Hayes and Flower, 1983]

**GENERATING:** “retrieve information relevant to the writing task from long-term memory”.

**ORGANIZING:** “select the most useful of the materials retrieved by the GENERATING process and to organize them into a writing plan.”

- Neural NLG models are great, but they are handling multiple distinct tasks at once.

- Our goal is to separate planning and realization, for better control over the output.
Roadmap

• Motivation
• Tasks
• Model
• Data
• Evaluation
• Conclusion
Roadmap

• Motivation
• Tasks
• Model
• Data
• Evaluation
• Conclusion
Tasks

Subjective
Open-ended
High-entropy

Objective
Factoid
Low-entropy
Tasks

Argument generation

Paper abstract generation

Wikipedia paragraph generation

Subjective
Open-ended
High-entropy

Objective
Factoid
Low-entropy
Argument generation

Wikipedia paragraph generation

Paper abstract generation

AGENDA

Title: Semantic Embeddings from Hashtags

We describe a convolutional neural network that learns feature representations for …

Title: Artificial intelligence

In computer science, artificial intelligence (AI), sometimes called machine intelligence…

cut financial aid

allies

economic benefits

Foreign aid allows for allies in places that are economically advantageous.

Short textual posts

hashtags

Hashtag prediction

…

Style:

simple

normal

computer science

perceive

artificial

machines
Argument generation

Cut financial aid allies economic benefits...

Foreign aid allows for allies in places that are economically advantageous.

Paper abstract generation

Title: Semantic Embeddings from Hashtags

We describe a convolutional neural network that learns feature representations for...

Wikipedia paragraph generation

Title: Artificial intelligence

In computer science, artificial intelligence (AI), sometimes called machine intelligence...
**Argument generation**

* r/changemyview
  * I think the United States should cut off foreign aid completely. CMV
  * 2011 saw 49.5B in spending on foreign aid. Why aren't US tax dollars being spent to help US allies?

- cut financial aid
- allies
- economic benefits

**Paper abstract generation**

**AGENDA**

- **Title:** Semantic Embeddings from Hashtags
- **Short textual posts**
- **hashtags**
- **Hashtag prediction**

**Wikipedia paragraph generation**

**Title:** Artificial intelligence

**Style:**
- simple
- normal

- computer science
- perceive
- artificial
- machines

**In computer science, artificial intelligence (AI), sometimes called machine intelligence...**
Roadmap

• Motivation
• Tasks
• Model
• Data
• Evaluation
• Conclusion
• Example: argument generation
US should cut off foreign aid completely!
US should cut off foreign aid completely!

2011 saw 49.5B in spending on foreign aid. Why is the US government taking money from citizens and spending it on others?
US should cut off foreign aid completely!

2011 saw 49.5B in spending on foreign aid. Why is the US government taking money from citizens and spending it on others?
US should cut off foreign aid completely!

- cut financial aid
- make homosexuality a crime
- uganda
- ...
- political bargaining chip
US should cut off foreign aid completely!

Sentence 1: [political bargaining chip]
Sentence 2: [cut financial aid; uganda]
Sentence 3: [make homosexuality a crime]
Sentence 4: [NULL]

- cut financial aid
- make homosexuality a crime
- uganda
- ...
- political bargaining chip
US should cut off foreign aid completely!

Sentence 1: It can be a useful political bargaining chip.
US should cut off foreign aid completely!

Sentence 1: It can be a useful political bargaining chip.

Sentence 2: US threatened to cut off financial aid to Uganda.
US should cut off foreign aid completely!

Sentence 1: It can be a useful political bargaining chip.
Sentence 2: US threatened to cut off financial aid to Uganda.
Sentence 3: Because it planned to criminalize homosexuality.
Sentence 4: [NULL]
**US should cut off foreign aid completely!**

Sentence 1: It can be a useful political bargaining chip.
Sentence 2: US threatened to cut off financial aid to Uganda.
Sentence 3: Because it planned to criminalize homosexuality.
Sentence 4: Please consider changing your mind!
US should cut off foreign aid completely!

2011 saw 49.5B in spending on foreign aid. Why is the US government taking money from citizens and spending it on others?

CLAIM
- cut financial aid
- make homosexuality a crime
- Uganda

FUNCTIONAL
- US threatened to cut off financial aid to Uganda.

PREMISE
- should
- cut
- off

PHRASE ENCODER

INPUT ENCODER

PLANNER

REALIZER
US should cut off foreign aid completely!

2011 saw 49.5B in spending on foreign aid. Why is the US government taking money from citizens and spending it on others?

Cut financial aid
Make homosexuality a crime
Uganda
Political bargaining chip

US threatened to cut off financial aid to Uganda.
US should cut off foreign aid completely!

2011 saw 49.5B in spending on foreign aid. Why is the US government taking money from citizens and spending it on others?

US threatened to cut off financial aid to Uganda.

CLAIM

PREMISE

PREMISE

FUNCTIONAL

US threatened to cut off financial aid to Uganda.

Phrase encoder

Input encoder

Planner

Realizer

cut financial aid
make homosexuality a crime
uganda
political bargaining chip
US should cut off foreign aid completely!

2011 saw 49.5B in spending on foreign aid. Why is the US government taking money from citizens and spending it on others?

US threatened to cut off financial aid to Uganda.

Input encoder

Phrases:
- cut financial aid
- make homosexuality a crime
- uganda
- political bargaining chip

Sum:
\[
\sum w_e
\]

Phase encoder:
- \( h_1 \)
- \( h_2 \)
- \( h_3 \)
- \( h_M \)

Output:
- CLAIM
- PREMISE
- PREMISE
- FUNCTIONAL

Planner

Realizer

US threatened to cut off financial aid to Uganda.
Sentence 1: [political bargaining chip]
Sentence 1: [political bargaining chip]

\[ s_j = LSTM(s_{j-1}, m_j) \]
Sentence 1: [political bargaining chip]

\[ s_j = LSTM(s_{j-1}, m_j) \]

\[ m_j = \sum_{sel(k)} h_k^e \]
Sentence 1: [political bargaining chip]

\[ \text{softmax}(w^T_s \tanh(W^s[m_j; s_j])) \]
Sentence 1: [political bargaining chip]
Sentence 1: [political bargaining chip]

\[ \text{softmax}(w_s^T \tanh(W^s[m_j; s_j])) \]

Style specification

CLAIM: “I believe foreign aid is a useful bargaining chip.”

PREMISE: “In 2014, the US cuts aid to Uganda over anti-gay law.”
Sentence 1: [political bargaining chip]

**FUNCTIONAL**: “Please change your mind!”

**PREMISE**: “In 2014, the US cuts aid to Uganda over anti-gay law.”

**CLAIM**: “I believe foreign aid is a useful bargaining chip.”

\[
softmax(W_s^T \tanh(W_s[m_j; s_j]))
\]

Style specification
Sentence 1: [political bargaining chip]

Sentence 2: [cut financial aid; uganda]

Keyphrase selection

$$P(v_{j+1,k} = 1 | v_{1:j}) = \sigma(w_v^T s_j + q_j w^c h_k^e)$$

select k-th phrase in (j+1)-th sentence

Content Planning
Sentence 1: [political bargaining chip]

Sentence 2: [cut financial aid; uganda]
• Content selection decoding

Sentence 1: [political bargaining chip]
Sentence 2: [cut financial aid; uganda]
Sentence 3: [make homosexuality a crime]
Sentence 4: [NULL]
US should cut off foreign aid completely!

2011 saw 49.5B in spending on foreign aid. Why is the US government taking money from citizens and spending it on others?

"...

Cut financial aid
Make homosexuality a crime
Uganda
Political bargaining chip

US threatened to cut off financial aid to Uganda."
US should cut off foreign aid completely!

2011 saw 49.5B in spending on foreign aid. Why is the US government taking money from citizens and spending it on others?

Us should cut off

cut financial aid
make homosexuality a crime
uganda
political bargaining chip

US threatened to cut off financial aid to Uganda.
• Surface realization

Sentence 1: [political bargaining chip]

Sentence 2: [cut financial aid; uganda]

Sentence 3: [make homosexuality a crime]

Sentence 4: [NULL]

US threatened to cut off financial aid to Uganda.
Surface realization

Sentence 1: [political bargaining chip]
Sentence 2: [cut financial aid; uganda]
Sentence 3: [make homosexuality a crime]
Sentence 4: [NULL]

\[ z_t = LSTM(z_{t-1}, \tanh(W^{ws} s_{J(t)} + W^{ww} y_{t-1})) \]

US threatened to cut off financial aid to Uganda.
Surface realization

Sentence 1: [political bargaining chip]

Sentence 2: [cut financial aid; uganda]

Sentence 3: [make homosexuality a crime]

Sentence 4: [NULL]

\[ z_t = LSTM(z_{t-1}, \tanh(W^{ws}s_J(t) + W^{ww}y_{t-1})) \]

Content control
Surface realization

Sentence 1: [political bargaining chip]

Sentence 2: [cut financial aid; uganda]

Sentence 3: [make homosexuality a crime]

Sentence 4: [NULL]

Output layer

\[ P(y_t | y_{1:t-1}) = \text{softmax}(\tanh(W^o [z_t; c_t; t_J(t)])) \]

US threatened to cut off financial aid to Uganda.
Model

• Training objective

\[ L(\theta) = \sum -\log P(y^*|X; \theta) + \gamma L_{style}(\theta) + \eta L_{sel}(\theta) \]
• Training objective

\[ L(\theta) = \sum \log P(y^* | X; \theta) + \gamma L_{style}(\theta) + \eta L_{sel}(\theta) \]

- Token level cross-entropy
- Style cross-entropy
- Selection, binary cross-entropy
Roadmap

- Motivation
- Tasks
- Model
- Data
- Evaluation
- Conclusion
Data

- Argument generation
  - Data: r/changemyview
  - Input statement: OP post
  - Output argument: “delta” awarded and well-upvoted replies
• Argument generation
  • Data: r/changemyview
  • Input statement: OP post
  • Output argument: “delta” awarded and well-upvoted replies

I think the United States should cut off foreign aid completely. CMV

2011 saw 49.5B in spending on foreign aid. Why aren't US tax dollars being spent to help US citizens in need instead of foreign citizens in need? 2007 saw over 70B in private companies/individuals foreign aid, which I am completely on board with.

Why is the US government taking money from citizens and spending it on (mostly) Africa when the US has its own problems? Let the private investors help those outside of the borders.
Data

• Argument generation
  • Data: r/changemyview
  • Input statement: OP post
  • Output argument: “delta” awarded and well-upvoted replies

Foreign aid allows for allies in places that are economically advantageous.

Because if the US government did, then really bad shit would happen, in short.
Argument generation

- **Data:** ![Reddit](r/changemyview)
- Input statement: OP post
- Output argument: “delta” awarded and well-upvoted replies

Foreign aid allows for allies in places that are economically advantageous.

Because if the US government did, then really bad shit would happen, in short.

I think the United States should cut off foreign aid completely. CMV

2011 saw 49.5B in spending on foreign aid. Why aren't US tax dollars being spent to help US citizens in need instead of foreign citizens in need? 2007 saw over 70B in private companies/individuals foreign aid, which I am completely on board with.
Data

• Argument generation
  • Data: r/changemyview
  • Input statement: OP post
  • Output argument: “delta” awarded and well-upvoted replies
  • Keyphrases: noun/verb phrases from retrieved passages
• Argument generation
  • Data: r/changemyview
  • Input statement: OP post
  • Output argument: “delta” awarded and well-upvoted replies
  • Keyphrases: noun/verb phrases from retrieved passages

Foreign aid allows for allies in places that are economically advantageous.

The New York Times
Its investment (...) foreign aid helps advance peace and stability ... and strengthening allies with military and economic assistance.

Reuters
Almost 50 percent of U.S. international assistance goes to (...) allies in the campaigns (...)

The Economist Journal
You will never look at global economic development the same way again...
• Argument generation
  • Data: r/changemyview
  • Input statement: OP post
  • Output argument: “delta” awarded and well-upvoted replies
• Keyphrases: noun/verb phrases from retrieved passages
  • Is at most 10 words
  • Contains at least one non-stopwords
  • Is a topic signature [Lin and Hovy, 2000] or a Wikipedia title
• Argument generation
  • Data: r/changemyview
  • Input statement: OP post
  • Output argument: “delta” awarded and well-upvoted replies
  • Keyphrases: noun/verb phrases from retrieved passages
  • Sentence style: labeled by pattern matching and rules
Data

• Abstract generation for scientific papers
  • AGENDA dataset [Koncel-Kedziorski, Bekal, Luan, Lapata, and Hajishirzi, 2019]
Abstract generation for scientific papers

**Title:** Semantic Embeddings from Hashtags

**Entities:**
- short textual posts
- document recommendation task
- hashtag prediction task
- ...
- convolutional neural network

**Abstract:** We describe a convolutional neural network that learns feature representations for short textual posts using hashtags as a supervised signal. The proposed approach is …
Data

- Wikipedia paragraph generation
  - First paragraphs of Wikipedia articles
Data

• Wikipedia paragraph generation
  • First paragraphs of Wikipedia articles
  • Keyphrase as noun/verb chunks with content word(s)
Data

• Wikipedia paragraph generation
  • First paragraphs of Wikipedia articles
  • Keyphrase as noun/verb chunks with content word(s)

In computer science, artificial intelligence (AI), sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans. Leading AI textbooks define the field as the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals.[1] Colloquially, the term "artificial intelligence" is often used to describe machines (or computers) that mimic "cognitive" functions that humans associate with the human mind, such as "learning" and "problem solving". [2]

• computer science
• artificial intelligence
• machine intelligence
• …
• perceives its environment
Data

• Wikipedia paragraph generation
  • First paragraphs of Wikipedia articles
  • Keyphrase as noun/verb chunks with content word(s)
  • Sentence style as sentence complexities, proxied by length
Data

- Wikipedia paragraph generation
  - First paragraphs of Wikipedia articles
  - Keyphrase as noun/verb chunks with content word(s)
  - Sentence style as sentence complexities, proxied by length
- Global style of simple vs. normal

In computer science, (...) any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals (...) that mimic "cognitive" functions that humans...

Artificial intelligence is the ability of a computer program or a machine to think and learn. (...) which tries to make computers "smart". (...) John McCarthy came up with the name (...)
Data

• Wikipedia paragraph generation
  • First paragraphs of Wikipedia articles
  • Keyphrase as noun/verb chunks with content
  • Sentence style as sentence complexities, proxied by length
• Global style of simple vs. normal

In computer science, (...) any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals (...) that mimic "cognitive" functions that humans...

Artificial intelligence is the ability of a computer program or a machine to think and learn. (...) which tries to make computers "smart". (...) John McCarthy came up with the name (...) Model needs to capture the interplay between style and content.
In computer science, (...) any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals (...) that mimic "cognitive" functions that humans...

Artificial intelligence is the ability of a computer program or a machine to think and learn. (...) which tries to make computers "smart". (...) John McCarthy came up with the name (...) Model needs to capture the interplay between style and content. We append one extra bit to the planner’s input as global style control.
Roadmap

• Motivation
• Tasks
• Model
• Data
• Evaluation
• Conclusion
Experiment

• Comparisons for argument generation
  • RETRIEVAL: returns the highest ranked passage as output
  • SEQ2SEQ: encodes input prompt and keyphrase as tokens
  • Our 2018 ArgGen model (H&W 2018): generates keyphrases as auxiliary task
Experiment

• Comparisons for argument generation
• Comparisons for abstract generation
  • GraphWriter [Koncel-Kedziorski et al, 2019]: state-of-the-art model with graph transformer that handles both entities and relations
  • Seq2Seq: encodes input prompt and keyphrase as tokens
Experiment

• Comparisons for argument generation
• Comparisons for abstract generation

• Comparisons for Wikipedia paragraph generation
  • RETRIEVAL: returns the most similar paragraph from training set
  • SEQ2SEQ: encodes input prompt and keyphrase as tokens
  • LOGREGSEL: predicts whether a phrase should be included, given global style
Automatic Evaluation

- Argument generation

<table>
<thead>
<tr>
<th>Method</th>
<th>BLEU-2</th>
<th>ROUGE-L</th>
<th>METEOR</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>RETRIEVAL</td>
<td>7.81</td>
<td>15.68</td>
<td>10.59</td>
<td>150.0</td>
</tr>
<tr>
<td>SEQ2SEQ</td>
<td>3.64</td>
<td>19.00</td>
<td>9.85</td>
<td>51.7</td>
</tr>
<tr>
<td>H&amp;W (2018)</td>
<td>5.73</td>
<td>14.44</td>
<td>3.82</td>
<td>36.5</td>
</tr>
</tbody>
</table>
Automatic Evaluation

- Argument generation

<table>
<thead>
<tr>
<th></th>
<th>BLEU-2</th>
<th>ROUGE-L</th>
<th>METEOR</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>RETRIEVAL</td>
<td>7.81</td>
<td>15.68</td>
<td>10.59</td>
<td>150.0</td>
</tr>
<tr>
<td>SEQ2SEQ</td>
<td>3.64</td>
<td>19.00</td>
<td>9.85</td>
<td>51.7</td>
</tr>
<tr>
<td>H&amp;W (2018)</td>
<td>5.73</td>
<td>14.44</td>
<td>3.82</td>
<td>36.5</td>
</tr>
<tr>
<td>Ours</td>
<td>13.19</td>
<td>20.15</td>
<td>10.42</td>
<td>65.5</td>
</tr>
</tbody>
</table>
## Automatic Evaluation

- Argument generation

<table>
<thead>
<tr>
<th></th>
<th>BLEU-2</th>
<th>ROUGE-L</th>
<th>METEOR</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>RETRIEVAL</td>
<td>7.81</td>
<td>15.68</td>
<td>10.59</td>
<td>150.0</td>
</tr>
<tr>
<td>SEQ2SEQ</td>
<td>3.64</td>
<td>19.00</td>
<td>9.85</td>
<td>51.7</td>
</tr>
<tr>
<td>H&amp;W (2018)</td>
<td>5.73</td>
<td>14.44</td>
<td>3.82</td>
<td>36.5</td>
</tr>
<tr>
<td>Ours</td>
<td>13.19</td>
<td>20.15</td>
<td>10.42</td>
<td>65.5</td>
</tr>
<tr>
<td>w/o Style</td>
<td>12.61</td>
<td>20.28</td>
<td>9.03</td>
<td>62.6</td>
</tr>
</tbody>
</table>
Automatic Evaluation

- Argument generation

<table>
<thead>
<tr>
<th></th>
<th>BLEU-2</th>
<th>ROUGE-L</th>
<th>METEOR</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrieval</td>
<td>7.81</td>
<td>15.68</td>
<td>10.59</td>
<td>150.0</td>
</tr>
<tr>
<td>Seq2Seq</td>
<td>3.64</td>
<td>19.00</td>
<td>9.85</td>
<td>51.7</td>
</tr>
<tr>
<td>H&amp;W (2018)</td>
<td>5.73</td>
<td>14.44</td>
<td>3.82</td>
<td>36.5</td>
</tr>
<tr>
<td>Ours</td>
<td>13.19</td>
<td>20.15</td>
<td>10.42</td>
<td>65.5</td>
</tr>
<tr>
<td>w/o Style</td>
<td>12.61</td>
<td>20.28</td>
<td>9.03</td>
<td>62.6</td>
</tr>
</tbody>
</table>

BLEU, METEOR are improved with Style module.
## Automatic Evaluation

- Argument generation

<table>
<thead>
<tr>
<th></th>
<th>BLEU-2</th>
<th>ROUGE-L</th>
<th>METEOR</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RETRIEVAL</strong></td>
<td>7.81</td>
<td>15.68</td>
<td>10.59</td>
<td>150.0</td>
</tr>
<tr>
<td><strong>SEQ2SEQ</strong></td>
<td>3.64</td>
<td>19.00</td>
<td>9.85</td>
<td>51.7</td>
</tr>
<tr>
<td><strong>H&amp;W (2018)</strong></td>
<td>5.73</td>
<td>14.44</td>
<td>3.82</td>
<td>36.5</td>
</tr>
<tr>
<td><strong>Ours</strong></td>
<td>13.19</td>
<td>20.15</td>
<td>10.42</td>
<td>65.5</td>
</tr>
<tr>
<td><strong>w/o Style</strong></td>
<td>12.61</td>
<td>20.28</td>
<td>9.03</td>
<td>62.6</td>
</tr>
<tr>
<td><strong>w/ Oracle Plan</strong></td>
<td><strong>16.30</strong></td>
<td><strong>20.25</strong></td>
<td><strong>11.61</strong></td>
<td>65.5</td>
</tr>
</tbody>
</table>
Automatic Evaluation

- Argument generation

<table>
<thead>
<tr>
<th></th>
<th>BLEU-2</th>
<th>ROUGE-L</th>
<th>METEOR</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>RETRIEVAL</td>
<td>7.81</td>
<td>15.68</td>
<td>10.59</td>
<td>150.0</td>
</tr>
<tr>
<td>SEQ2SEQ</td>
<td>3.64</td>
<td>19.00</td>
<td>9.85</td>
<td>51.7</td>
</tr>
<tr>
<td>H&amp;W (2018)</td>
<td>5.73</td>
<td>14.44</td>
<td>3.82</td>
<td>36.5</td>
</tr>
<tr>
<td>Ours</td>
<td>13.19</td>
<td>20.15</td>
<td>10.42</td>
<td>65.5</td>
</tr>
<tr>
<td>w/o Style</td>
<td>12.61</td>
<td>20.28</td>
<td>9.03</td>
<td>62.6</td>
</tr>
<tr>
<td>w/ Oracle Plan</td>
<td>16.30</td>
<td>20.25</td>
<td>11.61</td>
<td>65.5</td>
</tr>
</tbody>
</table>

Oracle plan setup shows good planning leads to improved overall quality.
### Automatic Evaluation

- **Abstract generation**

<table>
<thead>
<tr>
<th>Model</th>
<th>BLEU-2</th>
<th>ROUGE-L</th>
<th>METEOR</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>GraphWriter</td>
<td>29.95</td>
<td>28.56</td>
<td>19.90</td>
<td>130.1</td>
</tr>
<tr>
<td>Seq2Seq</td>
<td>18.13</td>
<td>21.03</td>
<td>13.95</td>
<td>134.8</td>
</tr>
<tr>
<td>Ours</td>
<td>20.32</td>
<td>23.30</td>
<td>15.95</td>
<td>128.3</td>
</tr>
<tr>
<td>w/ Oracle Plan</td>
<td>25.03</td>
<td>26.18</td>
<td>19.21</td>
<td>125.8</td>
</tr>
</tbody>
</table>
## Automatic Evaluation

- Abstract generation

<table>
<thead>
<tr>
<th></th>
<th>BLEU-2</th>
<th>ROUGE-L</th>
<th>METEOR</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GRAPHWRITER</strong></td>
<td>29.95</td>
<td>28.56</td>
<td>19.90</td>
<td>130.1</td>
</tr>
<tr>
<td><strong>SEQ2SEQ</strong></td>
<td>18.13</td>
<td>21.03</td>
<td>13.95</td>
<td>134.8</td>
</tr>
<tr>
<td><strong>Ours</strong></td>
<td>20.32</td>
<td>23.30</td>
<td>15.95</td>
<td>128.3</td>
</tr>
<tr>
<td>w/ Oracle Plan</td>
<td>25.03</td>
<td>26.18</td>
<td>19.21</td>
<td>125.8</td>
</tr>
</tbody>
</table>

With oracle plan, our system is competitive to the relation-aware SotA model.
## Automatic Evaluation

- Wikipedia paragraph generation (Normal)

<table>
<thead>
<tr>
<th></th>
<th>BLEU-2</th>
<th>ROUGE-L</th>
<th>METEOR</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>RETRIEVAL</td>
<td>20.10</td>
<td>28.60</td>
<td>12.23</td>
<td>44.5</td>
</tr>
<tr>
<td>SEQ2SEQ</td>
<td>22.62</td>
<td>27.49</td>
<td>14.74</td>
<td>52.9</td>
</tr>
<tr>
<td>LOGREGSel</td>
<td>29.28</td>
<td>28.65</td>
<td>27.76</td>
<td>34.4</td>
</tr>
<tr>
<td>Ours</td>
<td>33.76</td>
<td>40.08</td>
<td>25.70</td>
<td>65.4</td>
</tr>
</tbody>
</table>
### Automatic Evaluation

- Wikipedia paragraph generation (Normalized)

<table>
<thead>
<tr>
<th></th>
<th>BLEU-2</th>
<th>ROUGE-L</th>
<th>METEOR</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrieval</td>
<td>20.10</td>
<td>28.60</td>
<td>12.23</td>
<td>44.5</td>
</tr>
<tr>
<td>Seq2Seq</td>
<td>22.62</td>
<td>27.49</td>
<td>14.74</td>
<td>52.9</td>
</tr>
<tr>
<td>LogRegSel</td>
<td>29.28</td>
<td>28.65</td>
<td>27.76</td>
<td>34.4</td>
</tr>
<tr>
<td>Ours</td>
<td>33.76</td>
<td>40.08</td>
<td>25.70</td>
<td>65.4</td>
</tr>
</tbody>
</table>

Our model yields much better BLEU, ROUGE than comparisons.
Automatic Evaluation

- Wikipedia paragraph generation (Normal)

<table>
<thead>
<tr>
<th></th>
<th>BLEU-2</th>
<th>ROUGE-L</th>
<th>METEOR</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrieval</td>
<td>20.10</td>
<td>28.60</td>
<td>12.23</td>
<td>44.5</td>
</tr>
<tr>
<td>Seq2Seq</td>
<td>22.62</td>
<td>27.49</td>
<td>14.74</td>
<td>52.9</td>
</tr>
<tr>
<td>LogRegSel</td>
<td>29.28</td>
<td>28.65</td>
<td>27.76</td>
<td>34.4</td>
</tr>
<tr>
<td>Ours</td>
<td>33.76</td>
<td>40.08</td>
<td>25.70</td>
<td>65.4</td>
</tr>
<tr>
<td>w/ Oracle Plan</td>
<td>37.70</td>
<td>45.41</td>
<td>31.65</td>
<td>79.8</td>
</tr>
</tbody>
</table>
## Automatic Evaluation

- Wikipedia paragraph generation (Simple)

<table>
<thead>
<tr>
<th>Method</th>
<th>BLEU-2</th>
<th>ROUGE-L</th>
<th>METEOR</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RETRIEVAL</strong></td>
<td>21.99</td>
<td>33.44</td>
<td>12.97</td>
<td>34.7</td>
</tr>
<tr>
<td><strong>SEQ2SEQ</strong></td>
<td>21.98</td>
<td>29.36</td>
<td>16.94</td>
<td>52.8</td>
</tr>
<tr>
<td><strong>LOGREGSEL</strong></td>
<td>5.59</td>
<td>23.21</td>
<td>13.27</td>
<td>13.0</td>
</tr>
<tr>
<td><strong>Ours</strong></td>
<td>31.22</td>
<td>40.76</td>
<td>26.76</td>
<td>58.7</td>
</tr>
<tr>
<td><strong>w/ Oracle Plan</strong></td>
<td><strong>34.22</strong></td>
<td><strong>45.48</strong></td>
<td><strong>32.84</strong></td>
<td><strong>70.5</strong></td>
</tr>
</tbody>
</table>
Human Evaluation

• Human assessment
  • Grammaticality (1-5): fluency, free of grammar errors
  • Correctness (1-5): non-contradictory, right stance (argument)
  • Content richness (1-5): coverage of relevant points
Human Evaluation

• Human assessment
  • Grammaticality (1-5): fluency, free of grammar errors
  • Correctness (1-5): non-contradictory, right stance (argument)
  • Content richness (1-5): coverage of relevant points

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HUMAN</strong></td>
<td>4.81</td>
<td>3.90</td>
<td>3.48</td>
</tr>
<tr>
<td>Ours</td>
<td>3.99</td>
<td>2.78</td>
<td>2.61</td>
</tr>
<tr>
<td>w/o Style</td>
<td>3.03</td>
<td>2.26</td>
<td>2.03</td>
</tr>
</tbody>
</table>
Human Evaluation

- Human assessment
  - Grammaticality (1-5): fluency, free of grammar errors
  - Correctness (1-5): non-contradictory, right stance (argument)
  - Content richness (1-5): coverage of relevant points

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HUMAN</strong></td>
<td>4.84</td>
<td>4.73</td>
<td>4.49</td>
</tr>
<tr>
<td>Ours</td>
<td>3.38</td>
<td>3.24</td>
<td>3.43</td>
</tr>
<tr>
<td>w/o Style</td>
<td>2.99</td>
<td>2.89</td>
<td>3.50</td>
</tr>
</tbody>
</table>

Wikipedia paragraph generation (30 samples)
Human Evaluation

• Human assessment
  • Grammaticality (1-5): fluency, free of grammar errors
  • Correctness (1-5): non-contradictory, right stance (argument)
  • Content richness (1-5): coverage of relevant points

Wikipedia paragraph generation (30 samples)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Human</strong></td>
<td>4.84</td>
<td>4.73</td>
<td>4.49</td>
</tr>
<tr>
<td>Ours</td>
<td>3.38</td>
<td>3.24</td>
<td>3.43</td>
</tr>
<tr>
<td>w/o Style</td>
<td>2.99</td>
<td>2.89</td>
<td>3.50</td>
</tr>
</tbody>
</table>

Our system is still far from human, yet style module helps significantly.
• Effect of keyphrase selection
PROMPT: Aborting a fetus has some non-zero negative moral implications.

SEQ2SEQ: I’m not going to try to change your view here, but I don’t want to change your position. I don’t think it’s fair to say that a fetus is not a person. It’s not a matter of consciousness.

OURS: TL;DR: I agree with you, but I think it’s important to note that fetuses are not fully developed. I don’t know if this is the case, but it doesn’t seem to be a compelling argument to me at all, so I’m not going to try to change your view by saying that it should be illegal to kill.

OURS w/o Style: Now, I don’t think it’s fair to say that it’s okay to kill someone. If you don’t want to die, you shouldn’t be able to do that for the sake of the mother and the fetus.
Breaking is an American television soap opera that aired on CBS from September 26, 2007 to April 19, 2009. The series is set in main, New South Wales, in the United Kingdom. It is a spin-off of the ”breaking” series, and was produced by colloquialism productions and broadcast from the television series “ “.

Breaking bad is an American television series set in Albuquerque, New Mexico. It was started by January 2008 and ended on September 2013 after being cancelled shortly thereafter.

Bad breaking is an American television series set in Albuquerque, New Mexico on cable channel from 2007 to 2013.
**Title:** Breaking Bad

**Style:** Simple Wikipedia

**Seq2Seq:** Breaking is an American television soap opera that aired on CBS from September 26, 2007 to April 19, 2009. The series is set in main, New South Wales, in the United Kingdom. It is a spin-off of the "breaking" series, and was produced by colloquialism productions and broadcast from the television series “ “.

**Ours:** Breaking bad is an American television series set in Albuquerque, New Mexico. It was started by January 2008 and ended on September 2013 after being cancelled shortly thereafter.

**Ours w/o Style:** Bad breaking is an American television series set in Albuquerque, New Mexico on cable channel from 2007 to 2013.
**Title**: Breaking Bad  

**Style**: Simple Wikipedia

**SEQ2SEQ**: Breaking is an American television soap opera that aired on CBS from September 26, 2007 to April 19, 2009. The series is set in main, New South Wales, in the United Kingdom. It is a spin-off of the "breaking" series, and was produced by colloquialism productions and broadcast from the television series “ “.

**OURS**: Breaking bad is an American television series set in Albuquerque, New Mexico. It was started by January 2008 and ended on September 2013 after being cancelled shortly thereafter.

**OURS w/o Style**: Bad breaking is an American television series set in Albuquerque, New Mexico on cable channel from 2007 to 2013.
Conclusion

• We propose a generic text generation framework with sentence-level planning and realization.

• We utilize keyphrases as units for content selection, offering flexibility in different domains and tasks.

• Experiments on three tasks show the effectiveness of the model and the association between selection and final output quality.
Thanks!

hua.x@husky.neu.edu

https://github.com/XinyuHua/textgen-emnlp19
**TITLE**: Breaking Bad

**STYLE**: Normal Wikipedia

<table>
<thead>
<tr>
<th>SEQ2SEQ</th>
<th>OURS</th>
<th>OURS w/o Style</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad breaking is an American crime drama television series created by David main. It ran from September 21, 2001 to January 19, 2009 on Fox. The series is set in New York City and follows the adventures of a group of in the fictional town of breaking series, after the events of the breaking series.</td>
<td>Bad breaking is an American neo-western crime drama television series an American neo-western crime drama television series drama series television written and produced by vince vince. The show aired on AMC for five seasons from January 20, 2011 to September 29, 2013.</td>
<td>Bad breaking is an American an American neo-western crime drama television series that was produced by Vince Gilligan. The show ran for over five seasons from January 20 to September 29, 2011.</td>
</tr>
</tbody>
</table>
Moon Jae-in is a South Korean politician and current president of South Korea from 2012 to 2017 and again from 2014 to 2017.
PROMPT: All alleged crimes or infractions involving a police officer should be investigated by a higher-level agency.

SEQ2SEQ: If you look at the history of history, there was a time when the founding fathers didn’t understand the history of the world. They didn’t want to go to war. It wasn’t be cause they didn’t exist.

OURS: Jail sentences for example are a result of negligence. If you don’t like it, you’re going to have to deal with a lot of bad lawyers, and you’ll have to prove you’re guilty of a crime in the first place and you’re not going to be able to sue for damages if you’re in jail.

OURS w/o Style: The biggest problem with this is that this is a huge problem. The problem is that there are a lot of cases where the defendant is innocent, and the defendant has the power to prove that the defendant didn’t want to go to trial, (...)
Backup Slides: Experiment Details

- **Argument generation**
  - Optionally append retrieved passages to input prompt.
  - Truncate input to 500 (OP), 400 (passages), and 70 (# keyphrases).
  - Pre-train encoder and realization decoder with more OP and replies.

- **Wikipedia paragraph generation**
  - Use sum of word embedding for title encoding.
  - Truncate keyphrase to at most 30.

- **Abstract generation**
  - Truncate keyphrase to at most 30.
• Overall
  • Two layers LSTM are used for both decoders.
  • Each layer has 512 dimensions, with dropout probabilities as 0.2.
  • Learning is driven by AdaGrad [Duchi et al., 2011] with 0.15 as learning rate and 0.1 as initial accumulator.
  • Gradient clipped to 2.0.
  • Batch size is 64.
  • Beam search decoding of size 5, with trigram repetitions blocked.
## Backup Slides: Data

### Overall statistics

<table>
<thead>
<tr>
<th></th>
<th>Argument generation</th>
<th>Abstract generation</th>
<th>Wikipedia paragraph generation (Nor/Sim)</th>
</tr>
</thead>
<tbody>
<tr>
<td># Tokens</td>
<td>54.87</td>
<td>141.34</td>
<td>70.57/48.60</td>
</tr>
<tr>
<td># Sentences</td>
<td>2.48</td>
<td>5.59</td>
<td>3.15/3.20</td>
</tr>
<tr>
<td># Keyphrase (candidates)</td>
<td>55.80</td>
<td>12.23</td>
<td>23.56</td>
</tr>
<tr>
<td># Keyphrase (selected)</td>
<td>11.61</td>
<td>12.23</td>
<td>16.01/11.11</td>
</tr>
<tr>
<td># Training data</td>
<td>272,147</td>
<td>38,720</td>
<td>125,136</td>
</tr>
</tbody>
</table>