Controllable Text Generation

Should machines reflect the way humans interact in society?

Shrimai Prabhumoye

Carnegie Mellon University
Language Technologies Institute
We constantly author text!

Style, Content and Structure

Social Media

Personal/Professional documents

Blogs

News/Wiki articles
Style

● Expressed in the choice of *words or phrases* as well as *syntactic structures* used to convey information

● I define style as a group of natural language sentences that belong to a particular class or label.

“Do you have any code that we can look at?”

“Any code? Or do I have to guess at how you did it?”

(Danescu-Niculescu-Mizil, 2013)
Human communication by definition is a process by which individuals exchange information and influence one another through a common system of symbols and signs.

Leverage information in the unstructured form
Alice: The Notebook is hands-down one of my favorite movies EVER! Have you ever seen The Notebook?

Bob: No I have never seen this movie. I am going to try it out now.

Alice: It was a heartwarming story of young love. The main characters are played by Ryan Gosling and Rachel McAdams.

Bob: Ok this sounds nice. I think Ryan is a good actor.

Alice: For all the praise it received, I was surprised to see that it only got a 5.7/10 on Rotten Tomatoes.

Bob: That is interesting. They never get the rating correct.

Alice: The story goes back and forth between present day and the past. Older Ryan is played by James Garner and older Rachel is played by Gena Rowlands. Yeah, Rotten Tomatoes never gets the right ratings. I always like to see the ratings but if I want to see a movie, I will watch it even if it has a bad rating.
**Structure**

**The Lead**: The most important information
May include a “hook” (provocative quote or question)

**The Body**: The crucial information
Argument, Controversy, Story, Issue
Evidence, background, details, logic, etc
Quotes, photos, support, dispute, expand

**The Tail**: extra information
Interesting/Related items like blogs, other editorials
Application

I have a talk to present today!

Best of luck!
Application

I have a talk to present today!

May the force be with you!
Hey, have you watched Avengers Endgame

Yeah, I loved Bradley Cooper in it!
Hey, have you watched *Avengers Endgame*?

Yeah, I loved Robert Downey in it!
Hey, have you watched Avengers Endgame?

Yeah, I loved the movie!
Yeah, I loved the movie!

oh great! Can you tell me the story!

Yeah, Ironman steals the Infinity Stones back from Thanos and uses them to disintegrate Thanos and his army, at the cost of his life. Thor decapitates Thanos. Hulk travels to New York City in 2012 and convinces the Ancient One to give him the Time Stone. Five years later, AntMan escapes from the quantum realm. Ironman builds a time machine to save the world.
Yeah, I loved the movie! oh great! Can you tell me the story!

Yeah, Ironman steals the Infinity Stones back from Thanos and uses them to disintegrate Thanos and his army, at the cost of his life. Thor decapitates Thanos. Five years later, AntMan escapes from the quantum realm. Ironman builds a time machine to save the world. Hulk travels to New York City in 2012 and convinces the Ancient One to give him the Time Stone. Ironman steals the Infinity Stones back from Thanos and uses them to disintegrate Thanos and his army, at the cost of his life.
Other Applications

- **Writing Assistance Tools**
  - recommend formal language
  - recommend structural changes
- Recommend *polite emails*
- **Story Generation**
  - plot, ending, sentiment, topic, persona
- **Content Generation** (websites, descriptions etc)
Overview

Controlled Generation Schema

draft COLING '20

draft ACL '20
Storytelling '19
ACL '18

NAACL '19
EMNLP '18

draft ACL '20
draft ICML '20

WiNLP '19

Style

Content

Structure

Ethical Considerations
Goal

- **Controlled Generation Schema** connects prior work
  - Schema contains 5 modules
  - Identify any architecture as belonging to one of these modules
  - Schema can be used with any algorithmic paradigm

- **Collate knowledge** about different techniques
  - Insights into the advantages of techniques
  - Pave way for new architectures
  - Provide easy access to comparison
Controlled Generation Schema

1. External Input \((h_0)\) → Generator
Controlled Generation Schema

1. External Input ($h_0$) → Generator
2. Generator → Sequential Input ($x_t$)
Controlled Generation Schema

1. External Input ($h_0$)
2. Sequential Input ($x_t$)
3. Generator ($G$)
Controlled Generation Schema

1. External Input ($h_0$)
2. Sequential Input ($x_t$)
3. Generator ($G$)
4. Output ($o_t$)

$\hat{x}_t$
**Controlled Generation Schema**

1. **External Input** ($h_0$)
2. **Sequential Input** ($x_t$)
3. **Generator** ($G$)
4. **Output** ($o_t$)
5. **Training Objective** ($\mathcal{L}$)

\[
\hat{x}_t \rightarrow \mathcal{L} \rightarrow y_t
\]
External Input

- $h_e = \text{input sentence rep}$
- $s = \text{control attribute rep}$
- Arithmetic or Linear Transform
  - $h_0 = [h_e; s]$
  - $h_0 = h_e + s$
  - $h_0 = \tanh(W_e h_e + W_s s + b)$
- Decompose
External Input

- $h_e = \text{input sentence rep}$
- $s = \text{control attribute rep}$
- Arithmetic or Linear Transform
  - $h_0 = [h_e; s]$
  - $h_0 = h_e + s$
  - $h_0 = \tanh(W_e h_e + W_s s + b)$
- Decompose
Analysis

- **Arithmetic or Linear Transform**
  - Concatenating makes the model big
  - Adding loses information
  - Linear Transform might be better than above two

- **Decompose**
  - Provides *interpretable* representations
  - Input should contain signal of control attribute
  - Supervision on decomposed space
Sequential Input

- Arithmetic Changes
  - \( \tilde{X}_t = [x_t; s] \)
  - \( \tilde{X}_t = x_t + s \)
- Changes the input to the generation itself and not the context
- Not shown promising results so far
Controlled Generation Schema

1. External Input \((h_0)\)
2. Sequential Input \((x_t)\)
3. Generator \((G)\)
4. Output \((o_t)\)
5. Training Objective \((\mathcal{L})\)

\[ \hat{x}_t \rightarrow \mathcal{L} \rightarrow y_t \]
Generator Operations

**Controlled Generator Operations**

- $c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t + \tanh(W_d d_t)$
  - $d_t$ = dialogue act representation, change made to LSTM cell
  - Add dialogue act information in the generation process

- $\tilde{h}_t = \tanh(W_h x_t + r_t \odot U_h h_{t-1} + s_t \odot Y g + q_t \odot (1^T_L Z E_t^{new})^T)$
  - $s_t$ = goal select gate; $q_t$ = item select gate, GRU cell
  - recipe generation task
Generator Operations

- **Controlled Generator Operations**
  - \( c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t + \tanh(W_d d_t) \)
  - \( d_t = \) dialogue act representation, change made to LSTM cell
  - Add *dialogue act* information in the generation process

- \( \tilde{h}_t = \tanh(W_h x_t + r_t \odot U_h h_{t-1} + s_t \odot Yg + q_t \odot (1^T_L ZE_t^{new})^T) \)
  - \( s_t = \) goal select gate; \( q_t = \) item select gate, GRU cell
  - *recipe generation* task
Controlled Generation Schema

1. External Input ($h_0$)
2. Sequential Input ($x_t$)
3. Generator ($G$)
4. Output ($o_t$)
5. Training Objective ($L$)
● **Attention**
  ● most effective - especially self and cross
  ● mostly control attribute tokens have been added to source sequence for attention
  ● under explored for controlling attributes but has a lot of potential

● **External Feedback**
● **Attention**
  - most effective - especially self-cross
  - mostly control attribute tokens have been added to source sequence for attention
  - under explored for controlling attributes but has a lot of potential

● **External Feedback**

---

Figure 2: **Global attentional model** – at each time step $t$, the model infers a *variable-length* alignment weight vector $a_t$ based on the current target state $h_t$ and all source states $\bar{h}_s$. A global context vector $c_t$ is then computed as the weighted average, according to $a_t$, over all the source states.
Output

- **Attention**
  - most effective - especially self and cross
  - mostly control attribute tokens have been added to source sequence for attention
  - under explored for controlling attributes but has a lot of potential

- **External Feedback**
Controlled Generation Schema

1. External Input \( (h_0) \)
2. Sequential Input \( (x_t) \)
3. Generator \( (G) \)
4. Output \( (o_t) \)
5. Training Objective \( (L) \)

\[ \hat{x}_t \rightarrow \text{Training Objective (L)} \rightarrow y_t \]
Training Objective

- **General Loss**
  - Cross Entropy Loss
  - Unlikelihood Loss
  - Decoding Strategies
  - Used with any generation task

- **Classifier Loss**
  - design multiple classifier for any control attributes
Proposed Work

- **Empirical Evaluation** of the various technique described on 3 tasks
  - Style transfer
  - Content grounded generation
  - Persona grounded dialogue
- Gain insight into which techniques work better for what type of tasks
Overview

Style

Controlled Generation Schema

Draft COLING ’20

Draft ACL ’20

Storytelling ’19

ACL ’18

Content

NAACL ’19
EMNLP ’18

draft ACL ’20
draft ICML ’20

Structure

Ethical Considerations

BAD

OR

GOOD

WiNLP ’19
What is Style Transfer

- Rephrasing the text to contain specific stylistic properties without changing the intent or affect within the context.

“Shut up! the video is starting!”

“Please be quiet, the video will begin shortly.”
Challenges

● No Parallel Data!

“The movie was very long.”

“I entered the theatre in the bloom of youth and emerged with a family of field mice living in my long, white mustache.”

● Disentangle content from style

● Style is subtle
Our Solution

- Back-Translation
  - Translating an English sentence to a pivot language and then back to English.
- Reduces stylistic properties
- Helps in grounding meaning
- Creates a representation independent of the generative model
- Representation is agnostic to the style task
Architecture

MT $e \rightarrow f$

encoder decoder
I thank you, Rep. Visclosky.

Architecture

MT $e \rightarrow f$

encoder decoder

je vous remercie, Rep. Visclosky
I thank you, Rep. Visclosky
I thank you, Rep. Visclosky

MT $e \rightarrow f$

encoder | decoder

je vous remercie, Rep. Visclosky

MT $f \rightarrow e$

encoder

$\mathcal{Z}$

Style 1

decoder

I thank you, senator Visclosky

Style 2

decoder

I'm praying for you sir.
Train Pipeline

\[ z \]

Style 1
\[
\begin{array}{c}
\text{decoder} \\
\end{array}
\]
\[ \hat{x}_{\text{style1}} \]

Style 2
\[
\begin{array}{c}
\text{decoder} \\
\end{array}
\]
\[ \hat{x}_{\text{style2}} \]
Experimental Settings

- Encoder-Decoders follow sequence-to-sequence framework (Sutskever et al., 2014; Bahdanau et al., 2015)

$$\min_{\theta_{\text{gen}}} \mathcal{L}_{\text{gen}} = \mathcal{L}_{\text{recon}} + \lambda_{c} \mathcal{L}_{\text{class}}$$
Style Transfer Accuracy

<table>
<thead>
<tr>
<th></th>
<th>Baseline (Shen et al, 2017)</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>60.4</td>
<td>57.04</td>
</tr>
<tr>
<td>Political Slant</td>
<td>75.82</td>
<td>88.01</td>
</tr>
<tr>
<td>Sentiment</td>
<td>80.43</td>
<td>87.22</td>
</tr>
</tbody>
</table>
Preservation of Meaning

Which transferred sentence maintains the same semantic intent of the source sentence while changing the political position.

Which transferred sentence is semantically equivalent to the source sentence with an opposite sentiment.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Political Slant</th>
<th>Sentiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline (Shen et al, 2017)</td>
<td>Ours</td>
<td>No Preference</td>
</tr>
<tr>
<td>15.23</td>
<td>43.41</td>
<td>39.55</td>
</tr>
<tr>
<td>23.18</td>
<td>41.36</td>
<td>45.9</td>
</tr>
<tr>
<td>35.91</td>
<td>40.91</td>
<td></td>
</tr>
</tbody>
</table>

Would you like to discuss any of these points further?
<table>
<thead>
<tr>
<th>Category</th>
<th>Baseline (Shen et al, 2017)</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>2.42</td>
<td>2.81</td>
</tr>
<tr>
<td>Political Slant</td>
<td>2.79</td>
<td>2.87</td>
</tr>
<tr>
<td>Sentiment</td>
<td>2.7</td>
<td>3.09</td>
</tr>
<tr>
<td>Overall</td>
<td>2.7</td>
<td>2.91</td>
</tr>
<tr>
<td>Overall short</td>
<td>3.05</td>
<td>3.11</td>
</tr>
<tr>
<td>Overall Long</td>
<td>2.18</td>
<td>2.62</td>
</tr>
</tbody>
</table>
Proposed Work

- **Style Representations**
  - Cross domain tasks
    - Example: Generate stories in a particular persona
  - Low Resource Setting
  - Preserve Privacy
    - Distribute representation and not data

- **Understanding Style**
  - Formulate Transformations
  - Better Style Transfer models
Style Representations

- **Average**

\[ S_{y_i} = \sum_{x_j \in y_i} BERT(x_j) \]

- **Task**: classifier to predict if two sentences belong to the same style or not (BERT-model)
  - Subtract *style representation* and predict (BERT-Style)
  - Subtract *random representation* and predict (BERT-Random)
Style Representations

- PASTEL dataset (Kang, 2019)
- Gender, Age and Education

<table>
<thead>
<tr>
<th>Model</th>
<th>Gender</th>
<th>Age</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>BERT-model</td>
<td>56.10</td>
<td>58.42</td>
<td>58.94</td>
</tr>
<tr>
<td>BERT-Style</td>
<td>54.49</td>
<td>56.61</td>
<td>50.96</td>
</tr>
<tr>
<td>BERT-Random</td>
<td>54.98</td>
<td>58.10</td>
<td>58.94</td>
</tr>
</tbody>
</table>
Style Representation

- PASTEL dataset
- Design techniques to extract style
- Evaluate style representation
  - Style Transfer Task
  - Cross Domain
    - Gender with Yelp and PASTEL
Understanding Style

- *Ablation* studies on classifier
- *Lexical Understanding*
  - N-gram features
  - distribution of function and content words
- *Structural Understanding*
  - Parts of Speech N-grams
  - Parse Tree Features
Understanding Style

- PASTEL dataset
  - Gender: 3 classes, Age: 8 classes, Education: 9 classes
- **POS-model**: SVM based on POS N-grams
- **BERT-model**: numbers taken from (Kang, 2019)

<table>
<thead>
<tr>
<th>Task</th>
<th>BERT-model</th>
<th>POS-model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>73.0</td>
<td>71.7</td>
</tr>
<tr>
<td>Age</td>
<td>46.3</td>
<td>40.5</td>
</tr>
<tr>
<td>Education</td>
<td>42.5</td>
<td>38.0</td>
</tr>
</tbody>
</table>
Content Transfer

- After *graduate form* Columbia University, Obama worked in Chicago.
- After *graduating from* Columbia University, Obama worked in Chicago.
- After graduating from *Carnegie Mellon University*, Obama worked in Chicago.
- After graduating from *Columbia University*, Obama worked in Chicago.

- AI assistance deals with *form* (grammar, style, etc.)
- Our goal is to control for *content*
What is our task?

Background [edit]

On 4 July 2011 several publications including the Daily Mail,[8][10] The Telegraph, and The Guardian[11] picked up the story and published the pictures along with articles that quoted Slater as describing the photographs as self-portraits taken by the monkeys: "Monkey steals camera to snap himself" (The Telegraph),[12] "a camera on a tripod" triggered by the monkeys (The Guardian),[13] and a camera started by a monkey "Fascinated by her reflection in the lens".[10] The articles also contained Slater quotes such as "He must have taken hundreds of pictures by the time I got my camera back." The following day, Amateur Photographer reported that Slater gave them further explanation as to how the photographs were created, downplaying the way newspaper articles had described them; Slater said reports that a monkey ran off with his camera and "began taking self-portraits" were incorrect and that the portrait was shot when his camera had been mounted on a tripod, with the primates playing around with a remote cable release as he fended off other monkeys.[14]
On 4 July 2011 several publications including the Daily Mail,[8] The Telegraph, and The Guardian[11] picked up the story and published the pictures along with articles that quoted Slater as describing the photographs as self-portraits taken by the monkeys: "Monkey steals camera to snap himself" (The Telegraph),[12] "a camera on a tripod" triggered by the monkeys (The Guardian),[13] and a camera started by a monkey "Fascinated by her reflection in the lens".[10] The articles also contained Slater quotes such as "He must have taken hundreds of pictures by the time I got my camera back." The following day, Amateur Photographer reported that Slater gave them further explanation as to how the photographs were created, downplaying the way newspaper articles had described them;

Slater said reports that a monkey ran off with his camera and "began taking self-portraits" were incorrect and that the portrait was shot when his camera had been mounted on a tripod, with the primates playing around with a remote cable release as he fended off other monkeys.[14]
What is our task?

On 4 July 2011 several publications including the *Daily Mail*, *The Telegraph*, and *The Guardian* picked up the story and published the pictures along with articles that quoted Slater as describing the photographs as self-portraits taken by the monkeys: “Monkey steals camera to snap himself” (*The Telegraph*), “a camera on a tripod” triggered by the monkeys (*The Guardian*), and a camera started by a monkey “Fascinated by her reflection in the lens.” The articles also contained Slater quotes such as “He must have taken hundreds of pictures by the time I got my camera back.”

The following day, *Amateur Photographer* reported that Slater gave them further explanation as to how the photographs were created, downplaying the way newspaper articles had described them; Slater said reports that a monkey ran off with his camera and “began taking self-portraits” were incorrect and that the portrait was shot when his camera had been mounted on a tripod, with the primates playing around with a remote cable release as he fended off other monkeys.
Primary Contribution

curated text (context)
Primary Contribution

curated text (context) + document
Primary Contribution

curated text (context) + document = updated text
Primary Contribution

- design a task to perform content transfer from an unstructured source of information
- release dataset
Data Creation Process
Data Creation Process

Context

Update
Data Creation Process

Context

Update

HTML News Article

Plain Text of News Article

Common Crawl
### Data Creation Process

**Total Data Size: 636K**

<table>
<thead>
<tr>
<th>News Article</th>
<th>Wikipedia Context</th>
<th>Update</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Models

Generative Models

• Context Agnostic Generative Model (CAG) — Baseline
• Context Informed Generative Model (CIG)
• Context Responsive Generative Model (CRG)

All models have global attention

Extractive Models

• SumBasic
• Context Informed SumBasic
• Oracle

All models are simplistic to infer if context helps in generation
Context Agnostic Model (CAG) - Baseline

News Article

Encoder Vector

LSTM
LSTM
LSTM

\[ x_1 \]
\[ x_2 \]
\[ x_3 \]

\[ y_1 \]
\[ y_2 \]
\[ y_3 \]

Update
Context Informed Model (CIG)

\[ x_1 \rightarrow LSTM \rightarrow x_i \rightarrow LSTM \rightarrow x_{i+1} \rightarrow LSTM \rightarrow x_n \rightarrow LSTM \rightarrow \text{Encoder Vector} \rightarrow LSTM \rightarrow y_1 \rightarrow LSTM \rightarrow y_2 \rightarrow LSTM \rightarrow y_3 \rightarrow \text{Update} \]

News Article + Wiki Context
Context Receptive Model (CRG)

**News Article**
- $x_1$
- $x_2$
- $x_n$

**Wiki Context**
- $s_1$
- $s_2$
- $s_n$

**Encoder Vector**
- LSTM
- LSTM
- LSTM

**Update**
- $y_1$
- $y_2$
- $y_3$
## Automated Evaluation

<table>
<thead>
<tr>
<th>Model</th>
<th>ROUGE-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>SumBasic</td>
<td>5.6 (5.6-5.7)</td>
</tr>
<tr>
<td>Context Informed SumBasic</td>
<td>7.0 (7.0-7.1)</td>
</tr>
<tr>
<td>Context Agnostic Generative Model</td>
<td>9.1 (9.0-9.2)</td>
</tr>
<tr>
<td>Context Informed Generative Model</td>
<td>16.0 (15.9-16.1)</td>
</tr>
<tr>
<td>Context Receptive Generative Model</td>
<td>14.7 (14.6-14.8)</td>
</tr>
<tr>
<td>Oracle</td>
<td>28.8 (28.7-29.0)</td>
</tr>
</tbody>
</table>

* METEOR and BLEU numbers are consistent with ROUGE-L
Relative Human Evaluation

Which system output is closest in meaning to the reference update?

- CAG: 15.8
- CIG: 30.8
- Neither: 53.3

Which system output is more accurate relative to the background information given in the snippet of the article?

- CAG: 30.8
- CIG: 39.2
- Neither: 53.3

Which system output is more accurate relative to the background information given in the snippet of the article?
Absolute Quality Evaluation

<table>
<thead>
<tr>
<th>Category</th>
<th>CAG</th>
<th>CIG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammar</td>
<td>2.6</td>
<td>4.3</td>
</tr>
<tr>
<td>Non-redundancy</td>
<td>1.8</td>
<td>3.9</td>
</tr>
<tr>
<td>Referential Quality</td>
<td>2.7</td>
<td>3.6</td>
</tr>
<tr>
<td>Focus</td>
<td>2.6</td>
<td>3.5</td>
</tr>
<tr>
<td>Structure and Coherence</td>
<td>2.4</td>
<td>3.2</td>
</tr>
</tbody>
</table>
Proposed Work

- **Methodology**
  - Extending existing architectures for new tasks
  - Attention is under explored

- **Evaluation**
  - No automatic metric to check *content fidelity*
  - Information generated exists in document
Methodology

- Attention based technique to incorporate unstructured data for grounded generation

\[ p_t = \sigma(W_o o_t) \]

\[ \tilde{p}_t = p_t W_1 a^c_t + (1 - p_t) W_2 a^s_t \]

\[ \tilde{o}_t = \tanh(W_p [\tilde{p}_t; o_t]) \]

decide whether to focus on Wiki context or News article to generate current token
Methodology

- Attention based technique to incorporate unstructured data for grounded generation

\[ p_t = \sigma(W_od_t) \]
\[ \tilde{p}_t = p_t W_1 a_t^c + (1 - p_t) W_2 a_t^s \]
\[ \tilde{o}_t = \tanh(W_p[\tilde{p}_t; o_t]) \]

\( a_t^c \) = attention weights of Wiki context
\( a_t^s \) = attention weights of News article
Methodology

- Attention based technique to incorporate unstructured data for grounded generation

\[ p_t = \sigma(W_o o_t) \]

\[ \tilde{p}_t = p_t W_1 a_t^c + (1 - p_t) W_2 a_t^s \]

\[ \tilde{o}_t = \tanh(W_p[\tilde{p}_t; o_t]) \]
Evaluation

- Information Extraction System (IE)
  - OpenIE, Comet, RAKE

- \( i_r = IE(\text{reference}), i_g = IE(\text{generation}), \)
  \( i_s = IE(\text{context}), i_d = IE(\text{document}) \)

- **Close to reference:** \( \text{Sim}(i_r, i_g) \)

- **Coherent to context:** \( \text{Sim}(i_g, i_s) \)

- **Information from source:** \( \text{Sim}(i_g, i_d) \)

- \( \text{Sim}() \) is Jaccard similarity, BLEU, BERT space similarity
Evaluation

- Information Extraction System (IE)
  - OpenIE, Comet, RAKE
- $i_r = IE(\text{reference}), i_g = IE(\text{generation})$
  
  $i_s = IE(\text{context}), i_d = IE(\text{document})$

- Close to reference: $\text{Sim}(i_r, i_g)$
- Coherent to context: $\text{Sim}(i_g, i_s)$
- Information from source: $\text{Sim}(i_g, i_d)$

- $\text{Sim}()$ is Jaccard similarity, BLEU, BERT space similarity
Evaluation

- Information Extraction System (IE)
  - OpenIE, Comet, RAKE
- $i_r = IE(\text{reference}), i_g = IE(\text{generation})$, $i_s = IE(\text{context}), i_d = IE(\text{document})$
- Close to reference: $\text{Sim}(i_r, i_g)$
- Coherent to context: $\text{Sim}(i_g, i_s)$
- Information from source: $\text{Sim}(i_g, i_d)$
- $\text{Sim}()$ is Jaccard similarity, BLEU, BERT space similarity
Ironman builds a time machine to save the world.

Ironman steals the Infinity Stones back from Thanos and uses them to disintegrate Thanos and his army, at the cost of his life.

Hulk travels to New York City in 2012 and convinces the Ancient One to give him the Time Stone.

Thor decapitates Thanos.

Five years later, AntMan escapes from the quantum realm.

Ironman builds a time machine to save the world.
Ironman builds a time machine to save the world.

Ironman steals the Infinity Stones back from Thanos and uses them to disintegrate Thanos and his army, at the cost of his life.

Hulk travels to New York City in 2012 and convinces the Ancient One to give him the Time Stone.

Thor decapitates Thanos.

Five years later, AntMan escapes from the quantum realm.

Sentence Ordering Task

After

Thor decapitates Thanos.

Five years later, AntMan escapes from the quantum realm.

Ironman builds a time machine to save the world.

Hulk travels to New York City in 2012 and convinces the Ancient One to give him the Time Stone.

Ironman steals the Infinity Stones back from Thanos and uses them to disintegrate Thanos and his army, at the cost of his life.
Problem Framing

- Prior Work
  - Sequence Prediction Task
  - Hierarchical models to learn document structure

- Our Approach
  - Constraint Solving Problem
Methodology

- For a document with $n$ sentences ($\{s_1 \ldots s_n\}$)
  
  \[ |\mathcal{C}| = \binom{n}{2} \] constraints

- Predicted constraints of the form $s_1 < s_2$
- 4 sentences in a document then 6 constraints
  - $\{s_1 < s_2, s_1 < s_3, s_1 < s_4, s_2 < s_3, s_2 < s_4, s_3 < s_4\}$

- Topological sort to find an order given $\mathcal{C}$
  - Graph: $s_1 \rightarrow s_2$ if $s_1 < s_2$
Constraint Learning

- **BERT based Representation (B-TSort)**
  - Next Sentence Prediction
  - MLP(BERT($s_1$[SEP]$s_2$))

- **LSTM based Representation (L-TSort)**
  - $h_1 = LSTM(s_1); h_2 = LSTM(s_2)$
  - MLP([h$_1$; h$_2$])
Baselines

- **Attention Order Network (AON)**
  - LSTM: sentence representation
  - Transformer: document representation
  - LSTM decoder: generate order

- **BERT Attention Order Network (B-AON)**
  - BERT: sentence representation
Results for NIPS abstracts

<table>
<thead>
<tr>
<th>Measure</th>
<th>AON</th>
<th>L-TSort</th>
<th>B-AON</th>
<th>B-TSort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect Match</td>
<td>16.25</td>
<td>12.19</td>
<td>19.9</td>
<td>32.59</td>
</tr>
<tr>
<td>Sentence Acc</td>
<td>50.5</td>
<td>43.08</td>
<td>55.23</td>
<td>61.48</td>
</tr>
<tr>
<td>Kendall Tau</td>
<td>67</td>
<td>64</td>
<td>73</td>
<td>81</td>
</tr>
<tr>
<td>Rouge-S</td>
<td>80.97</td>
<td>80.08</td>
<td>83.65</td>
<td>76.29</td>
</tr>
<tr>
<td>LCS</td>
<td>74.38</td>
<td>71.11</td>
<td>76.29</td>
<td>83.45</td>
</tr>
</tbody>
</table>

*entire sequence was correctly predicted*
Results for NIPS abstracts

- **Perfect Match**: AON 16.25, L-TSort 12.19, B-AON 19.9, B-TSort 32.59, absolute position was correctly predicted
- **Sentence Acc**: AON 50.5, L-TSort 43.08, B-AON 55.23, B-TSort 61.48
- **Kendall Tau**: AON 67, L-TSort 64, B-AON 73, B-TSort 81
- **Rouge-S**: AON 80.97, L-TSort 80.08, B-AON 83.65, B-TSort 87.97
- **LCS**: AON 74.38, L-TSort 71.11, B-AON 76.29, B-TSort 83.45
### Results for NIPS abstracts

<table>
<thead>
<tr>
<th>Metric</th>
<th>AON</th>
<th>L-TSort</th>
<th>B-AON</th>
<th>B-TSort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect Match</td>
<td>16.25</td>
<td>12.19</td>
<td>19.9</td>
<td>32.59</td>
</tr>
<tr>
<td>Sentence Acc</td>
<td>50.5</td>
<td>43.08</td>
<td>55.23</td>
<td>61.48</td>
</tr>
<tr>
<td>Kendall Tau</td>
<td>67</td>
<td>64</td>
<td>73</td>
<td>81</td>
</tr>
<tr>
<td>Rouge-S</td>
<td>80.97</td>
<td>80.08</td>
<td>83.65</td>
<td>76.29</td>
</tr>
<tr>
<td>LCS</td>
<td>74.38</td>
<td>71.11</td>
<td>80.08</td>
<td>76.29</td>
</tr>
</tbody>
</table>

The chart illustrates the number of inversions to reach the correct order for different metrics and algorithms.
Results for NIPS abstracts

- AON
- L-TSort
- B-AON
- B-TSort

Pairs with correct relative order:

<table>
<thead>
<tr>
<th>Measure</th>
<th>AON</th>
<th>L-TSort</th>
<th>B-AON</th>
<th>B-TSort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect Match</td>
<td>16.25</td>
<td>12.19</td>
<td>19.9</td>
<td>32.59</td>
</tr>
<tr>
<td>Sentence Acc</td>
<td>50.5</td>
<td>55.23</td>
<td>64</td>
<td>67</td>
</tr>
<tr>
<td>Kendall Tau</td>
<td>61.48</td>
<td>64</td>
<td>73</td>
<td>81</td>
</tr>
<tr>
<td>Rouge-S</td>
<td>83.65</td>
<td>87.97</td>
<td>83.65</td>
<td>87.97</td>
</tr>
<tr>
<td>LCS</td>
<td>80.08</td>
<td>80.97</td>
<td>74.38</td>
<td>71.11</td>
</tr>
<tr>
<td>LCS</td>
<td>83.45</td>
<td>83.65</td>
<td>76.29</td>
<td>83.45</td>
</tr>
<tr>
<td>Metric</td>
<td>AON</td>
<td>L-TSort</td>
<td>B-AON</td>
<td>B-TSort</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------</td>
<td>---------</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>Perfect Match</td>
<td>16.25</td>
<td>12.19</td>
<td>19.9</td>
<td>32.59</td>
</tr>
<tr>
<td>Sentence Acc</td>
<td>50.5</td>
<td>43.08</td>
<td>55.23</td>
<td>61.48</td>
</tr>
<tr>
<td>Kendall Tau</td>
<td>61.48</td>
<td>64</td>
<td>73</td>
<td>81</td>
</tr>
<tr>
<td>Rouge-S</td>
<td>80.97</td>
<td>80.08</td>
<td>83.65</td>
<td>87.97</td>
</tr>
<tr>
<td>LCS</td>
<td>71.11</td>
<td>76.29</td>
<td>80.97</td>
<td>83.45</td>
</tr>
</tbody>
</table>
Results for NIPS abstracts

- **AON**
- **L-TSort**
- **B-AON**
- **B-TSort**

<table>
<thead>
<tr>
<th>Metric</th>
<th>AON</th>
<th>L-TSort</th>
<th>B-AON</th>
<th>B-TSort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect Match</td>
<td>16.25</td>
<td>12.19</td>
<td>19.9</td>
<td>32.59</td>
</tr>
<tr>
<td>Sentence Acc</td>
<td>50.5</td>
<td>43.08</td>
<td>55.23</td>
<td>61.48</td>
</tr>
<tr>
<td>Kendall Tau</td>
<td>67</td>
<td>64</td>
<td>73</td>
<td>81</td>
</tr>
<tr>
<td>Rouge-S</td>
<td>80.97</td>
<td>80.08</td>
<td>83.65</td>
<td>87.97</td>
</tr>
<tr>
<td>LCS</td>
<td>74.38</td>
<td>71.11</td>
<td>76.29</td>
<td>83.45</td>
</tr>
</tbody>
</table>
Results for Human Evaluation

B-TSort vs B-AON
- B-TSort: 41
- No Preference: 28
- B-AON: 31

B-TSort vs Gold
- B-TSort: 26
- No Preference: 20
- Gold: 54

B-AON vs Gold
- B-AON: 22
- No Preference: 24
- Gold: 54
Results

- B-TSort performs best in all metrics for SIND captions, NSF abstract, AAN abstract datasets
- Analysis of displaced sentences
- Analysis of documents with more than 10 sentences
- Percentage of mismatch in input and output for AON
Overview

Controlled Generation Schema

Style

draft COLING ’20

draft ACL ’20

Storytelling ’19

ACL ’18

Content

NAACL ’19

EMNLP ’18

Structure

draft ACL ’20

draft ICML ’20

Ethical Considerations

WiNLP ’19
Ethical Considerations

- Swear words, obscenity, bias, hate speech
- Broader Impact of controllable text generation
- Social good and bad applications
  - Generate persuasive tweets to spread awareness about climate change
  - Generate persuasive social media content to keep people away from vaccines
Proposed Work

- **Generating Balanced Datasets**
  - Preserve distributional properties of style transfer
  - Analyze the distribution on sentiment class for the gender transfer and vice versa
  - New way of evaluating style transfer
- Does demographically balanced dataset lead to better sentiment analysis models?
Overview

Controlled Generation Schema

Style

Content

Structure

Ethical Considerations

draft COLING ’20

draft ACL ’20
Storystelling ’19
ACL ’18

NAACL ’19
EMNLP ’18

draft ACL ’20
draft ICML ’20

WiNLP ’19
Timeline

May - Aug 2020
- Internship at Salesforce
  - Empirical evaluation of the controllable text generation techniques
  - Attention based technique for content grounded generation

Sep - Dec 2020
- Techniques to build effective style representation
  - Ablation analysis of classifier to computationally understand style

Jan - Feb 2021
- Ethical considerations, analysis of style transfer models,
  - generate demographically balanced dataset

Mar - Apr 2021
- Apply for job and write thesis
Thank You!

Alan W Black (co-advisor)

Ruslan Salakhutdinov (co-advisor)

Yulia Tsvetkov

Jason Weston