Controllable Text Generation
Controlling Style and Content

Shrimai Prabhumoye, 2022
NVIDIA
NLP Applications are on the rise!
NLP Applications are on the rise!

Personal Assistants
NLP Applications are on the rise!

Personal Assistants

Empathy/Compassion bots
NLP Applications are on the rise!

- Personal Assistants
- Empathy/Compassion bots
- Call-center/FAQ bots
NLP Applications are on the rise!

- Personal Assistants
- Empathy/Compassion bots
- Call-center/FAQ bots
- Ubiquitous Computing
NLP Applications are on the rise!

Personal Assistants

Empathy/Compassion bots

Call-center/FAQ bots

Ubiquitous Computing

Other Applications
NLP Applications are on the rise!

- Style of responses
- Persona of the bot
- Topic of responses

Personal Assistants

Empathy/Compassion bots

Call-center/FAQ bots

Ubiquitous Computing

Other Applications
NLP Applications are on the rise!

- Style of responses
- Persona of the bot
- Topic of responses

Personal Assistants

- Emotions (style)
- Sensitivity of content

Empathy/Compassion bots

Call-center/FAQ bots

Ubiquitous Computing

Other Applications
NLP Applications are on the rise!

- Style of responses
- Persona of the bot
- Topic of responses

Personal Assistants

- Emotions (style)
- Sensitivity of content

Empathy/Compassion bots

- Content needs to be accurate!
- Structured/Unstructured data

Call-center/FAQ bots

Ubiquitous Computing

Other Applications

• Content needs to be accurate!
• Structured/Unstructured data

Emotions (style)
• Sensitivity of content

Personal Assistants

Ubiquitous Computing
NLP Applications are on the rise!

- Style of responses
- Persona of the bot
- Topic of responses

Personal Assistants

- Emotions (style)
- Sensitivity of content

Empathy/Compassion bots

- Content needs to be accurate!
- Structured/Unstructured data

Call-center/FAQ bots

- Accurate information
- Structured data

Ubiquitous Computing

- Content needs to be accurate!
- Structured/Unstructured data

Other Applications
NLP Applications are on the rise!

- Style of responses
- Persona of the bot
- Topic of responses

Personal Assistants

- Accurate information
- Structured data

Ubiquitous Computing

- Emotions (style)
- Sensitivity of content

Empathy/Compassion bots

- Content needs to be accurate!
- Structured/Unstructured data

Call-center/FAQ bots

- Plot of stories (structure)
- Sequence of topics in stories (structure)
- Politeness of emails (style)

Other Applications
Overview
Overview

Style
ACL 2018
ACL 2020

Content
EMNLP 2018
NAACL 2019
NAACL 2021

LMs
ACL Findings 2022
What is Style Transfer?

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

- Style is *culturally diverse*

Usage of “please” to closest friends.
Challenges in style transfer

- Style is *culturally diverse*

- Style is *subtle*

**Indirect:** By the way, do you know when is the deadline?

**1st person plural:** Let us find a good name for this paper.
Challenges in style transfer

- Style is *culturally diverse*
- Style is *subtle*
- *Disentangle* style and content
Challenges in style transfer

- Style is *culturally diverse*
- Style is *subtle*
- *Disentangle* style and content
- *Lack* of parallel data
Goals

☐ Transfer the Style

☐ Preserve the meaning

☐ Representation that reduces stylistic properties

☐ Interpretable Representation
Back-Translation (BST)

(Prabhumoye et al, ACL 2018)

https://github.com/shrimai/Style-Transfer-Through-Back-Translation
Back-Translation (BST)

I thank you, Rep. Visclosky

Machine Translation from English to French, not trained on style specific corpus

(Prabhumoye et al, ACL 2018)
Back-Translation (BST)

I thank you, Rep. Visclosky

MT $e \rightarrow f$

encoder  decoder

Je vous remercie, Rep. Visclosky

MT $f \rightarrow e$

encoder

Only encode the French sentence

I thank you, Rep. Visclosky.

MT $e \rightarrow f$

Je vous remercie, Rep. Visclosky

MT $f \rightarrow e$

Latent representation grounded in meaning and with reduced stylistic properties

(Prabhumoye et al, ACL 2018)

https://github.com/shrimai/Style-Transfer-Through-Back-Translation
Back-Translation (BST)

I thank you, Rep. Visclosky

I thank you, senator Visclosky

I'm praying for you sir.

MT e → f
encoder | decoder

Je vous remercie, Rep. Visclosky

MT f → e
encoder

Style 1
decoder

Style 2
decoder

(Prabhumoye et al, ACL 2018)

https://github.com/shrimai/Style-Transfer-Through-Back-Translation
Training of Back-Translation
Training of Back-Translation
Training of Back-Translation
Training of Back-Translation

\[
\min_{\theta_{\text{gen}}} \mathcal{L}_{\text{gen}} = \mathcal{L}_{\text{recon}} + \lambda_c \mathcal{L}_{\text{class}}
\]
Goals achieved by Back-Translation

- Transfer the Style
- Preserve the meaning
- Representation that reduces stylistic properties
- Interpretable Representation
Goals achieved by Back-Translation

- Transfer the Style
  feedback in the form of weighted classifier loss
- Preserve the meaning
- Representation that reduces stylistic properties
- Interpretable Representation
Goals achieved by Back-Translation

- ✔ Transfer the Style
  feedback in the form of weighted classifier loss

- ✗ Preserve the meaning
  styles are surface level, $z$ is in latent space and LSTM

- ☐ Representation that reduces stylistic properties

- ☐ Interpretable Representation
Goals achieved by Back-Translation

- ✔ Transfer the Style
  feedback in the form of weighted classifier loss

- ❌ Preserve the meaning
  styles are surface level, z is in latent space and LSTM

- ✔ Representation that reduces stylistic properties
  translating to another language

- ☐ Interpretable Representation
Goals achieved by Back-Translation

- ✓ Transfer the Style
  feedback in the form of weighted classifier loss
- ✗ Preserve the meaning
  styles are surface level, z is in latent space and LSTM
- ✓ Representation that reduces stylistic properties
  translating to another language
- ✗ Interpretable Representation
  z is in latent space
Tag and Generate

(Madaan et al, ACL 2020)

Tag and Generate

Tagger

Add TAG words at positions where the target phrases can appear OR replace source style phrases with TAG

(Madaan et al, ACL 2020)

Tag and Generate

Add TAG words at positions where the target phrases can appear OR replace source style phrases with TAG

[TAG] send me the data

This is representation z.


(Madaan et al, ACL 2020)
Tag and Generate

Send me the data → Tagger → [TAG] send me the data → Generator

Add TAG words at positions where the target phrases can appear OR replace source style phrases with TAG

GENERATE context appropriate phrases in target style

(Madaan et al, ACL 2020)

Tag and Generate

Send me the data → Tagger → [TAG] send me the data → Generator → Could you please send me the data

Add TAG words at positions where the target phrases can appear OR replace source style phrases with TAG

GENERATE context appropriate phrases in target style

(Madaan et al, ACL 2020)

Tag and Generate

Tagger and Generator are transformers

Add TAG words at positions where the target phrases can appear OR replace source style phrases with TAG

Could you please send me the data

GENERATE context appropriate phrases in target style

Send me the data

(TAG) send me the data

(Madaan et al, ACL 2020)

Goals achieved by Tag and Generate

☐ Transfer the Style

☐ Preserve the meaning

☐ Representation that reduces stylistic properties

☐ Interpretable Representation
Goals achieved by Tag and Generate

- Transfer the Style
  
  [TAG] send me the data -> Generator -> Could you please send me the data

- Preserve the meaning

- Representation that reduces stylistic properties

- Interpretable Representation
Goals achieved by Tag and Generate

- **Transfer the Style**
  - [TAG] send me the data
  - Generator
  - Could you please send me the data

- **Preserve the meaning**
  - [TAG] send me the data

- **Representation that reduces stylistic properties**

- **Interpretable Representation**
Goals achieved by Tag and Generate

- ✔ Transfer the Style
  - [TAG] send me the data → Generator → Could you please send me the data

- ✔ Preserve the meaning
  - [TAG] send me the data

- ✔ Representation that reduces stylistic properties
  - Send me the data → Tagger → [TAG] send me the data

- □ Interpretable Representation
Goals achieved by Tag and Generate

- **Transfer the Style**
  - [TAG] send me the data → Generator → Could you please send me the data

- **Preserve the meaning**
  - [TAG] send me the data

- **Representation that reduces stylistic properties**
  - Send me the data → Tagger → [TAG] send me the data

- **Interpretable Representation**
  - [TAG] send me the data
Politeness Results

- **CAE (Shen et al, NeurIPS 2017)**
- **DRG (Li et al, NAACL 2018)**
- **BST (Prabhumoye et al, ACL 2018)**
- **Tag and Generate**

Accuracy: 99.62, 90.25, 89.5, 60.75, 6.94, 11.83, 10.73, 11.83, 2.55

BLEU-s: 70.44, 41.09, 36.26, 25.71, 10.73, 18.07, 18.07, 18.99, 9.19

METEOR: 90.25, 41.09, 36.26, 25.71, 10.73, 18.07, 18.07, 18.99, 9.19

Results on four transfer tasks

Gender

Accuracy  BLEU-s  METEOR  Rouge-L
0  65.21  54.4  36.29  22.5
22.5  52.76  29.73  22.9  9.25
45  22.57  14.72  22.84  14.27
67.5  37.42  52.76  22.9  22.9
90  74.59  55.55  53.3  54.4

Accuracy  BLEU-s  METEOR  Rouge-L
0  77.71  84.44  67.79  87.74
22.5  68.44  51.8  25.69  41.02
45  87.74  69.79  32.09  7.79
67.5  51.8  61.06  19.95  30.02
90  77.51  55.9  7.79  9.61

CAE (Shen et al, NeurIPS 2017)  BST (Prabhumoye et al, ACL 2018)  DRG (Li et al, NAACL 2018)  Tag and Generate

Political Slant

Accuracy  BLEU-s  METEOR  Rouge-L
0  70.99  36.26  47.14  86.6
22.5  86.69  36.69  88.8  55.9
45  70.99  61.06  36.69  88.8
67.5  70.99  61.06  36.69  88.8
90  70.99  61.06  36.69  88.8

Humorous-Romantic-Factual

Accuracy  BLEU-s  METEOR  Rouge-L
0  89.66  93.17  93.17  89.66
22.5  51.01  32.45  2.09  30.02
45  64.32  31.79  9.61  43.57
67.5  32.45  31.79  9.61  43.57
90  79.51  64.32  31.79  9.61

Yelp Sentiment

Accuracy  BLEU-s  METEOR  Rouge-L
0  72.1  36.69  47.14  86.6
22.5  36.69  32.09  36.26  86.6
45  32.09  27.7  32.09  86.6
67.5  36.26  27.7  32.09  86.6
90  70.99  36.26  47.14  86.6

Humorous-Romantic-Factual

Accuracy  BLEU-s  METEOR  Rouge-L
0  89.66  93.17  93.17  89.66
22.5  51.01  32.45  2.09  30.02
45  64.32  31.79  9.61  43.57
67.5  32.45  31.79  9.61  43.57
90  79.51  64.32  31.79  9.61
1. Tag and Generate performs competitively on style transfer accuracy.

2. Tag and Generate performs much better on meaning preservation metric!
Limitations of automated metrics

- BLEU, METEOR, Rouge-L rely on n-gram overlap
- Classifiers can be fooled by polarizing keywords
- Naive Baseline on sentiment transfer task
Limitedations of automated metrics

- BLEU, METEOR, Rouge-L rely on n-gram overlap
- Classifiers can be fooled by polarizing keywords
- Naive Baseline on sentiment transfer task

Good drinks, and good company.  
Positive to Negative  
Good drinks, and good company.  
But overall it sucked.

Salsa is not hot or good.  
Negative to Positive  
Salsa is not hot or good.  
But overall it was perfect.
Limitations of automated metrics

- BLEU, METEOR, Rouge-L rely on n-gram overlap
- Classifiers can be fooled by polarizing keywords
- Naive Baseline on sentiment transfer task

Good drinks, and good company.  
Salsa is not hot or good.

Positive to Negative
Good drinks, and good company.  
But overall it sucked.

Negative to Positive
Salsa is not hot or good.  
But overall it was perfect.

- 91.3% style transfer accuracy
- 61.44 BLEU score
### Examples

<table>
<thead>
<tr>
<th>Input</th>
<th>Output</th>
<th>Politeness Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>What happened to my personal station?</td>
<td>Could you please let me know what happened to my personal station?</td>
<td><strong>could</strong> and <strong>would</strong> are forms which are polite</td>
</tr>
<tr>
<td>Yes, go ahead and remove it.</td>
<td>Yes, we can go ahead and remove it.</td>
<td>First person plural</td>
</tr>
<tr>
<td>Not yet - I’ll try this weekend</td>
<td>Sorry, not yet - I’ll try to make sure this weekend.</td>
<td>Apologizing</td>
</tr>
</tbody>
</table>
Overview

Style
ACL 2018
ACL 2020

Content
EMNLP 2018
NAACL 2019
NAACL 2021

LMs
ACL Findings 2022
Overview

Style
ACL 2018
ACL 2020

Content
EMNLP 2018
NAACL 2019
NAACL 2021

LMs
ACL Findings 2022
Document Grounded Generation
Document Grounded Generation

Context
Document Grounded Generation

Context + Document
Document Grounded Generation

Context + Document = Updated Text
Document Grounded Generation

- generate text given a context and a document
- generated text should *coherently fit* the context and contain new information from the document
Wikipedia Update Generation

(Prabhumoye et al, NAACL 2019)

https://github.com/shrimai/Towards-Content-Transfer-through-Grounded-Text-Generation
Wikipedia Update Generation

(Prabhumoye et al, NAACL 2019)

https://github.com/shrimai/Towards-Content-Transfer-through-Grounded-Text-Generation
Wikipedia Update Generation

Prabhumoye et al, NAACL 2019

https://github.com/shrimai/Towards-Content-Transfer-through-Grounded-Text-Generation
Monkey selfie copyright dispute

From Wikipedia, the free encyclopedia

The monkey selfie copyright dispute is a series of disputes about the copyright status of selfies taken by Celebes crested macaques using equipment belonging to the British nature photographer David Slater. The disputes involve Wikimedia Commons and the blog *Techdirt*, which have hosted the images following their publication in newspapers in July 2011 over Slater’s objections that he holds the copyright, and People for the Ethical Treatment of Animals (PETA), who have argued that the macaque should be assigned the copyright.

Slater has argued that he has a valid copyright claim, as he engineered the situation that resulted in the picture by travelling to Indonesia, befriending a group of wild macaques, and setting up his camera equipment in such a way that a “selfie” picture might come about. The *Wikimedia Foundation* in 2014 refused to remove the pictures from its Wikimedia Commons image library based on the understanding that copyright is held by the creator, that a non-human creator (not being a legal person) cannot hold copyright, and that the images are thus in the public domain.

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.[13]

https://github.com/shrimai/Towards-Content-Transfer-through-Grounded-Text-Generation

(Prabhumoye et al, NAACL 2019)
Wikipedia Update Generation

Monkey selfie copyright dispute

The monkey selfie copyright dispute is a series of disputes about the copyright status of selfies taken by Celebes crested macaques using equipment belonging to the British nature photographer David Slater. The disputes involve Wikimedia Commons and the blog Techdirt, which have hosted the images following their publication in newspapers in July 2011 over Slater’s objections that he holds the copyright, and People for the Ethical Treatment of Animals (PETA), who have argued that the macaque should be assigned the copyright.

Slater has argued that he has a valid copyright claim, as he engineered the situation that resulted in the pictures by travelling to Indonesia, befriending a group of wild macaques, and setting up his camera equipment in such a way that a “selfie” picture might come about. The Wikimedia Foundation’s 2014 refusal to remove the pictures from its Wikimedia Commons image library was based on the understanding that copyright is held by the creator, that a non-human creator (not being a legal person) cannot hold copyright, and that the images are thus in the public domain.

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.[13]

Generate text that is:
1. **coherent** to the Wikipedia context
2. contains **new events** from the article

Dataset Size of 636k!

https://github.com/shrimai/Towards-Content-Transfer-through-Grounded-Text-Generation

(Prabhumoye et al, NAACL 2019)
Document Grounded Dialogue

(Zhou, Prabhumoye & Black, EMNLP 2018)

https://github.com/festvox/datasets-CMU_DoG
Alice: Yes this is a Christmas classic …
Bob: Very well said! Some of the critics were …
Alice: Yes I found the first half very entertaining.
Bob: I feel you! So hard to believe it was released in 1990!
Alice: Yes this is a Christmas classic …
Bob: Very well said! Some of the critics were …
Alice: Yes I found the first half very entertaining.
Bob: I feel you! So hard to believe it was released in 1990!
Alice: Yes this is a Christmas classic …
Bob: Very well said! Some of the critics were …
Alice: Yes I found the first half very entertaining.
Bob: I feel you! So hard to believe it was released in 1990!

Alice: I like the quote "sadistic festival of adult-bashing". Between's Kevin's parents forgetting him and the burglars this movie doesn't make adult looks too bright.
Alice: Yes this is a Christmas classic …
Bob: Very well said! Some of the critics were …
Alice: Yes I found the first half very entertaining.
Bob: I feel you! So hard to believe it was released in 1990!

Alice: I like the quote "sadistic festival of adult-bashing". Between’s Kevin's parents forgetting him and the burglars this movie doesn't make adult looks too bright.
Models

- Pre-trained encoder-decoder model
- provide a stronger baseline for the task
Meanwhile, newspapers across the country pick up the story in solidarity with the post and times. The court rules 6-3 in the newspapers' favor, vindicating Graham's decision. Nixon demands that the post should be barred from the white house.

One year later, a security guard discovers a break-in progress at Watergate complex.

It was honorable for the other papers to publish in solidarity with the post and times. That had to prove that the people want to know.
Meanwhile, newspapers across the country pick up the story in solidarity with the post and times. The court rules 6-3 in the newspapers' favor, vindicating graham's decision. Nixon demands that the post should be barred from the white house.

One year later, a security guard discovers a break-in progress at watergate complex.

It was honorable for the other papers to publish in solidarity with the post and times. That had to prove that the people want to know.
Models

- Pre-trained encoder-decoder model
- provide a stronger baseline for the task

- Context Driven Representation (CoDR)
- build a representation of the content in document that is *not present* in the context

- Document Headed Attention (DoHA)
- specific *focus on document* information

(Prabhumoye et al, NAACL 2021)

https://github.com/shrimai/Focused-Attention-Improves-Document-Grounded-Generation
Meanwhile, newspapers across the country pick up the story in solidarity with the post and times. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. Nixon demands that the post should be barred from the white house.

One year later, a security guard discovers a break-in progress at watergate complex.

It was honorable for the other papers to publish in solidarity with the post and times. That had to prove that the people want to know.
Meanwhile, newspapers across the country pick up the story in solidarity with the post and times. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. Nixon demands that the post should be barred from the white house. 

One year later, a security guard discovers a break-in progress at watergate complex.

It was honorable for the other papers to publish in solidarity with the post and times. That had to prove that the people want to know.
Meanwhile, newspapers across the country pick up the story in solidarity with the post and times. The court rules 6-3 in the newspapers' favor, vindicating Graham's decision. Nixon demands that the post should be barred from the white house. ... One year later, a security guard discovers a break-in progress at Watergate complex.

It was honorable for the other papers to publish in solidarity with the post and times. That had to prove that the people want to know.
Meanwhile, newspapers across the country pick up the story in solidarity with the post and times. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. Nixon demands that the post should be barred from the white house.

One year later, a security guard discovers a break-in progress at watergate complex.

It was honorable for the other papers to publish in solidarity with the post and times. That had to prove that the people want to know.
Meanwhile, newspapers across the country pick up the story in solidarity with the post and times. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. Nixon demands that the post should be barred from the white house. One year later, a security guard discovers a break-in progress at watergate complex.

It was honorable for the other papers to publish in solidarity with the post and times. That had to prove that the people want to know.
Meanwhile, newspapers across the country pick up the story in solidarity with the post and times. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. Nixon demands that the post should be barred from the white house.

One year later, a security guard discovers a break-in progress at watergate complex.

It was honorable for the other papers to publish in solidarity with the post and times. That had to prove that the people want to know.
Meanwhile, newspapers across the country pick up the story in solidarity with the post and times. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. Nixon demands that the post should be barred from the white house.

One year later, a security guard discovers a break-in progress at watergate complex.

It was honorable for the other papers to publish in solidarity with the post and times. That had to prove that the people want to know.
Meanwhile, newspapers across the country pick up the story in solidarity with the post and times. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. Nixon demands that the post should be barred from the white house.

One year later, a security guard discovers a break-in progress at watergate complex.

It was honorable for the other papers to publish in solidarity with the post and times. That had to prove that the people want to know.
Meanwhile, newspapers across the country pick up the story in solidarity with the post and times. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. Nixon demands that the post should be barred from the white house.

One year later, a security guard discovers a break-in progress at watergate complex.

It was honorable for the other papers to publish in solidarity with the post and times. That had to prove that the people want to know.
Meanwhile, newspapers across the country pick up the story in solidarity with the post and times. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. Nixon demands that the post should be barred from the white house.

One year later, a security guard discovers a break-in progress at watergate complex.
Wikipedia Update Generation

**Content Transfer** (Prabhumoye et al, NAACL 2019)  
**BART**  
**CoDR**  
**DoHA**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Content Transfer</th>
<th>BART</th>
<th>CoDR</th>
<th>DoHA</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLEU-4</td>
<td>1.23</td>
<td>9.2</td>
<td>11.31</td>
<td>10.86</td>
</tr>
<tr>
<td>METEOR</td>
<td>6.21</td>
<td>12.9</td>
<td>14.38</td>
<td>14.28</td>
</tr>
<tr>
<td>Rouge-L</td>
<td>10.08</td>
<td>22.39</td>
<td>23.48</td>
<td>23.49</td>
</tr>
</tbody>
</table>
Wikipedia Update Generation

- CoDR and DoHA outperform BART and Content Transfer

<table>
<thead>
<tr>
<th></th>
<th>BLEU-4</th>
<th>METEOR</th>
<th>Rouge-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content Transfer</td>
<td>1.23</td>
<td>6.21</td>
<td>10.08</td>
</tr>
<tr>
<td>BART</td>
<td>9.2</td>
<td>12.9</td>
<td>22.39</td>
</tr>
<tr>
<td>CoDR</td>
<td>11.31</td>
<td>14.38</td>
<td>23.48</td>
</tr>
<tr>
<td>DoHA</td>
<td>10.86</td>
<td>14.28</td>
<td>23.49</td>
</tr>
</tbody>
</table>
CMU_DoG

Low-Res (Zhao et al, ICLR 2020)  BART  CoDR  DoHA

BLEU-4: 1.2  16.91  20.68  20.9

METEOR: 12.59  14.47  14.55

Rouge-L: 19.3  20.41  20.44
CMU_DoG

- 19.7 BLEU-4 improvement over Low-Res

Low-Res (Zhao et al, ICLR 2020)  | BART  | CoDR  | DoHA
---|---|---|---
BLEU-4: 1.2 | 16.91 | 20.68 | 20.9
METEOR: 12.59 | 14.47 | 14.55
Rouge-L: 19.3 | 20.41 | 20.44
**CMU_DoG**

- 19.7 BLEU-4 improvement over Low-Res
- 24% BLEU-4 improvement over BART

<table>
<thead>
<tr>
<th>Metric</th>
<th>Low-Res (Zhao et al, ICLR 2020)</th>
<th>BART</th>
<th>CoDR</th>
<th>DoHA</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLEU-4</td>
<td>1.2</td>
<td>16.91</td>
<td>20.68</td>
<td>20.9</td>
</tr>
<tr>
<td>METEOR</td>
<td>12.59</td>
<td>14.47</td>
<td>14.55</td>
<td></td>
</tr>
<tr>
<td>Rouge-L</td>
<td>19.3</td>
<td>20.41</td>
<td>20.44</td>
<td></td>
</tr>
</tbody>
</table>
Relevance Human Evaluation

Pick the option which contains information from the document and fits the context coherently.
Pick the option which contains information from the document and fits the context coherently.

- No Preference is chosen most number of times.
Pick the option which contains information from the document and fits the context coherently.

- No Preference is chosen most number of times.
- Both CoDR and DoHA outperform BART.
Wikipedia Update Generation
### Wikipedia Update Generation

<table>
<thead>
<tr>
<th>Error Class</th>
<th>%</th>
<th>Reference</th>
<th>Generation</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linguistic Variation: Generation is appropriate but a linguistic variation of the reference.</td>
<td>43</td>
<td>December 12 - The Smiths play Brixton Academy, their last ever gig before their dissolution.</td>
<td>December 12 - The Smiths perform their final show, at Brixton Academy in London.</td>
<td>41</td>
</tr>
</tbody>
</table>

Generation is a paraphrase of the Reference
Still gets a Rouge-L score of 41!
Inadequate automated metrics
<table>
<thead>
<tr>
<th>Error Class</th>
<th>%</th>
<th>Reference</th>
<th>Generation</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Linguistic Variation:</strong> Generation is appropriate but a linguistic variation of the reference.</td>
<td>43</td>
<td>December 12 - The Smiths play Brixton Academy, their last ever gig before their dissolution.</td>
<td>December 12 - The Smiths perform their final show, at Brixton Academy in London.</td>
<td>41</td>
</tr>
<tr>
<td><strong>Partial Hallucination:</strong> Generation is either missing or hallucinates some information</td>
<td>23</td>
<td>America Online and Prodigy (online service) offered access to the World Wide Web system for the first time this year, releasing browsers that made it easily accessible to the general public.</td>
<td>The World Wide Web was first introduced on January 17, 1995 on Prodigy.</td>
<td>17</td>
</tr>
</tbody>
</table>

The generation has captured some information of the reference but missed some.
<table>
<thead>
<tr>
<th>Error Class</th>
<th>%</th>
<th>Reference</th>
<th>Generation</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linguistic Variation:</td>
<td>43</td>
<td>December 12 - The Smiths play Brixton Academy, their last ever gig before their dissolution.</td>
<td>December 12 - The Smiths perform their final show, at Brixton Academy in London.</td>
<td>41</td>
</tr>
<tr>
<td>Partial Hallucination:</td>
<td>23</td>
<td>America Online and Prodigy (online service) offered access to the World Wide Web system for the first time this year, releasing browsers that made it easily accessible to the general public.</td>
<td>The World Wide Web was first introduced on January 17, 1995 on Prodigy.</td>
<td>17</td>
</tr>
</tbody>
</table>

Wikipedia pages that are in the form of lists
1340s, Timeline of DC Comics (1950s)
50% times the generation is grounded in the document and close to reference
### Wikipedia Update Generation

<table>
<thead>
<tr>
<th>Error Class</th>
<th>%</th>
<th>Reference</th>
<th>Generation</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Linguistic Variation:</strong> Generation is appropriate but a linguistic variation of the reference.</td>
<td>43</td>
<td>December 12 - The Smiths play Brixton Academy, their last ever gig before their dissolution.</td>
<td>December 12 - The Smiths perform their final show, at Brixton Academy in London.</td>
<td>41</td>
</tr>
<tr>
<td><strong>Partial Hallucination:</strong> Generation is either missing or hallucinates some information</td>
<td>23</td>
<td>America Online and Prodigy (online service) offered access to the World Wide Web system for the first time this year, releasing browsers that made it easily accessible to the general public.</td>
<td>The World Wide Web was first introduced on January 17, 1995 on Prodigy.</td>
<td>17</td>
</tr>
<tr>
<td><strong>Incoherent Reference:</strong> The reference does not coherently follow the context</td>
<td>22</td>
<td>“The Naked Ape” by Desmond Morris, is published.</td>
<td>Zoologist Desmond Morris publishes “The Naked Ape”.</td>
<td>26</td>
</tr>
<tr>
<td><strong>Incorrect:</strong> The generation is either not appropriate or is not grounded (completely hallucinates the information).</td>
<td>7</td>
<td>The year 2000 is sometimes abbreviated as “Y2K” (the “Y” stands for “year”, and the “K” stands for “kilo-” which means “thousand”)</td>
<td>The Y2K conspiracy theory claimed that a secret nuclear attack by the United States on 2 January 2000 was planned to begin World War 2.</td>
<td>9</td>
</tr>
</tbody>
</table>

About 85% times, the generation is either completely or partially grounded if the reference is grounded.
## Wikipedia Update Generation

<table>
<thead>
<tr>
<th>Error Class</th>
<th>%</th>
<th>Reference</th>
<th>Generation</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Linguistic Variation:</strong> Generation is appropriate but a linguistic variation of the reference.</td>
<td>43</td>
<td>December 12 - The Smiths play Brixton Academy, their last ever gig before their dissolution.</td>
<td>December 12 - The Smiths perform their final show, at Brixton Academy in London.</td>
<td>41</td>
</tr>
<tr>
<td><strong>Partial Hallucination:</strong> Generation is either missing or hallucinates some information</td>
<td>23</td>
<td>America Online and Prodigy (online service) offered access to the World Wide Web system for the first time this year, releasing browsers that made it easily accessible to the general public.</td>
<td>The World Wide Web was first introduced on January 17, 1995 on Prodigy.</td>
<td>17</td>
</tr>
<tr>
<td><strong>Incoherent Reference:</strong> The reference does not coherently follow the context</td>
<td>22</td>
<td>“The Naked Ape” by Desmond Morris, is published.</td>
<td>Zoologist Desmond Morris publishes “The Naked Ape”.</td>
<td>26</td>
</tr>
<tr>
<td><strong>Incorrect:</strong> The generation is either not appropriate or is not grounded (completely hallucinates the information).</td>
<td>7</td>
<td>The year 2000 is sometimes abbreviated as “Y2K” (the “Y” stands for “year”, and the “K” stands for “kilo-“ which means “thousand”)</td>
<td>The Y2K conspiracy theory claimed that a secret nuclear attack by the United States on 2 January 2000 was planned to begin World War 2.</td>
<td>9</td>
</tr>
<tr>
<td><strong>Reference is not grounded</strong></td>
<td>5</td>
<td>This was achieved under dead calm conditions as an additional safety measure, whereas the Wrights flew in a 25 mph+ wind to achieve enough airspeed on their early attempts.</td>
<td>This was verified by a video crew present at the test flight.</td>
<td>14</td>
</tr>
</tbody>
</table>
## Wikipedia Update Generation

<table>
<thead>
<tr>
<th>Error Class</th>
<th>%</th>
<th>Reference</th>
<th>Generation</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Linguistic Variation:</strong> Generation is appropriate but a linguistic variation of the reference.</td>
<td>43</td>
<td>December 12 - The Smiths play Brixton Academy, their last ever gig before their dissolution.</td>
<td>December 12 - The Smiths perform their final show, at Brixton Academy in London.</td>
<td>41</td>
</tr>
<tr>
<td><strong>Partial Hallucination:</strong> Generation is either missing or hallucinates some information</td>
<td>23</td>
<td>America Online and Prodigy (online service) offered access to the World Wide Web system for the first time this year, releasing browsers that made it easily accessible to the general public.</td>
<td>The World Wide Web was first introduced on January 17, 1995 on Prodigy.</td>
<td>17</td>
</tr>
<tr>
<td><strong>Incoherent Reference:</strong> The reference does not coherently follow the context</td>
<td>22</td>
<td>“The Naked Ape” by Desmond Morris, is published.</td>
<td>Zoologist Desmond Morris publishes “The Naked Ape”.</td>
<td>26</td>
</tr>
<tr>
<td><strong>Incorrect:</strong> The generation is either not appropriate or is not grounded (completely hallucinates the information).</td>
<td>7</td>
<td>The year 2000 is sometimes abbreviated as “Y2K” (the “Y” stands for “year”, and the “K” stands for “kilo-” which means “thousand”)</td>
<td>The Y2K conspiracy theory claimed that a secret nuclear attack by the United States on 2 January 2000 was planned to begin World War 2.</td>
<td>9</td>
</tr>
<tr>
<td><strong>Reference is not grounded</strong></td>
<td>5</td>
<td>This was achieved under dead calm conditions as an additional safety measure, whereas the Wrights flew in a 25 mph+ wind to achieve enough airspeed on their early attempts.</td>
<td>This was verified by a video crew present at the test flight.</td>
<td>14</td>
</tr>
</tbody>
</table>
# Observations on CMU_DoG

<table>
<thead>
<tr>
<th>Error Class</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference and generation are grounded</td>
<td>35</td>
</tr>
<tr>
<td>Reference is grounded but generation is not</td>
<td>3</td>
</tr>
<tr>
<td>Reference is not grounded but generation is</td>
<td>5</td>
</tr>
<tr>
<td>Reference and generation are not grounded</td>
<td>43</td>
</tr>
</tbody>
</table>
# Observations on CMU_DoG

<table>
<thead>
<tr>
<th>Error Class</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference and generation are grounded</td>
<td>35</td>
</tr>
<tr>
<td>Reference is grounded but generation is not</td>
<td>3</td>
</tr>
<tr>
<td>Reference is not grounded but generation is</td>
<td>5</td>
</tr>
<tr>
<td>Reference and generation are not grounded</td>
<td>43</td>
</tr>
</tbody>
</table>

1. References are not grounded if they are follow up *questions, opinions or experiences*.

2. *Context dictates* grounding

3. *Conversational subtleties not captured* by the automated metrics

4. 90% times the generation is grounded if the reference is grounded and generation is not grounded if the reference is not grounded

5. 86.7% generated responses are *appropriate to the context!*
Reference vs CoDR/DoHA

Pick the option that is most appropriate to the given context.

<table>
<thead>
<tr>
<th></th>
<th>Reference</th>
<th>No Pref</th>
<th>CoDR/DoHA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wikipedia</strong></td>
<td>33.9</td>
<td>28.3</td>
<td>37.8</td>
</tr>
<tr>
<td><strong>Update</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Generation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CMU_DoG</strong></td>
<td>22.8</td>
<td>45.6</td>
<td>31.6</td>
</tr>
</tbody>
</table>

0 25 50 75 100
Reference vs CoDR/DoHA

Pick the option that is most appropriate to the given context.

1. Automated metrics are inadequate
2. Sole reference should not be considered as the only correct response to the context
Overview

Style

Content

LMs

ACL 2018
ACL 2020

EMNLP 2018
NAACL 2019
NAACL 2021

ACL Findings 2022
Overview

Style
ACL 2018
ACL 2020

Content
EMNLP 2018
NAACL 2019
NAACL 2021

LMs
ACL Findings 2022
Controllable Text Generation
Using Pre-trained Language Models

- Language generation is now fluent and grammatical!
- How useful are LMs in steering the output in the desired direction?
- Harness the knowledge from LMs for QA tasks and classification tasks
  - Can it be used for controllable text generation tasks?
Challenges in building Large Language Models

- High compute requirements
  - Made efficient due to NVLink, NVSwitch and Mellanox 200Gbps Infiniband cards

- Model Parameters don’t fit in GPU memory
  - Efficient and scalable 3D parallel system capable of combining data, pipeline and tensor-slicing based parallelism

- Evaluations
  - Benchmark testing to understand bias and toxicity
  - NLP benchmark tests may not be enough
Multi-Stage Dialogue Prompting Framework (MSDP)
Multi-Stage Dialogue Prompting Framework (MSDP)

Inputs

- Topic
- Dialogue History

(Liu et al, ACL Findings 2022)
Multi-Stage Dialogue Prompting Framework (MSDP)

(Liu et al, ACL Findings 2022)
Multi-Stage Dialogue Prompting Framework (MSDP)

(Liu et al, ACL Findings 2022)
Knowledge Generator
Knowledge Generator

Support Repository

Typically Train sets!

Topic

Dialogue History
Knowledge Generator

Exemplars selected based on similarity metrics

(Last Turn 1) Topic 1 ==> Knowledge-1
(Last Turn 2) Topic 2 ==> Knowledge-2
...
(Last Turn k) Topic k ==> Knowledge-k
(Query Last Turn) Query-Topic ==>
Knowledge Generator

Support Repository

Exemplars selected based on similarity metrics

(Last Turn 1) Topic 1 ==> Knowledge-1
(Last Turn 2) Topic 2 ==> Knowledge-2
... 
(Last Turn k) Topic k ==> Knowledge-k

(Query Last Turn) Query-Topic ==> Pre-trained LM

Knowledge
Response Generator
Response Generator
Response Generator

[Topic-1] System: s-turn User: u-turn We know that: Knowledge-1 System replies: Response-1

... [Topic-k] System: s-turn User: u-turn We know that: Knowledge-k System replies: Response-k

[Query Topic-1] System: s-turn User: u-turn We know that: Knowledge-Generated System replies:
Response Generator

Support Repository

Exemplars selected based on similarity metrics

Pre-trained LM

Response
Results

- **PPLM** [Dathathri et al., 2019]
- **FCM w/ DPR (seen)**
- **FCM w/ DPR (Wiki)**
- **FCM w/ MSDP-KG**
- **MSDP**

FCM - Fine-tuned Conversation Model

DPR - Dense Passage Retrieval [Karpukhin et al., 2020] fine-tuned on WoW

Wizard of Wikipedia (Seen Test)

Wizard of Wikipedia (Unseen Test)
Model Size and Response Generation

Wizard of Wikipedia (Seen Test)

Wizard of Wikipedia (Unseen Test)
Summary

• The retriever is limited to the support repository set
  • Can retrieve information that is not relevant to dialogue context
• Model generates information relevant to the dialogue context
• More details in the paper
  • Knowledge generation
  • Human Evaluation
  • Ablation on choosing the type on encoder for semantic similarity
  • Ablation on changing the size of k (number of shots)