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
Application

I have a talk to present today!
Application

I have a talk to present today!

Best of luck!
Application

I have a talk to present today!

Best of luck!

Style
Application

I have a talk to present today!

May the force be with you!
Application
Application

Hey, have you watched Avengers Endgame
Application

Hey, have you watched Avengers Endgame

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

Yeah, I loved the movie!
oh great! Can you tell me the story!
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.
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.
oh great! Can you tell me the story!

Yeah, 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

Style

Content

Ethical Considerations

ACL ’20
Storytelling ’19
ACL ’18

NAACL ’19
EMNLP ’18

WiNLP ’19
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.”
Applications

**Anonymization**: To preserve anonymity of users online, for personal security concerns (Jardine, 2016), or to reduce stereotype threat (Spencer, 1999).

**Demographically-balanced training data** for downstream applications.
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
Architecture

MT $e \rightarrow f$

encoder  decoder
I thank you, Rep. Visclosky

Architecture

MT e → f
encoder decoder

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

Architecture

MT e $\rightarrow$ f

encoder decoder

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

Architecture

MT e → f
encoder decoder

je vous remercie, Rep. Visclosky

MT f → e
encoder
I thank you, Rep. Visclosky

je vous remercie, Rep. Visclosky

Architecture
I thank you, Rep. Visclosky

Architecture
I thank you, Rep. Visclosky.

je vous remercie, Rep. Visclosky.

Architecture
I thank you, Rep. Visclosky.

Architectural Diagram:

- **MT e → f**: English to French translation.
- **MT f → e**: French to English translation.

- **Style 1**: Decoder
- **Style 2**: Decoder

Architecture diagram illustrating the flow of translation and style options.
I thank you, Rep. Visclosky

MT e → f
encoder decoder

je vous remercie, Rep. Visclosky

MT f → e
encoder

Architecture

Style 1
decoder

Style 2
decoder

I thank you, senator Visclosky

I’m praying for you sir.
Train Pipeline
Train Pipeline

\[ Z \]

Style 1
decoder
\[ \hat{x}_{style1} \]

Style 2
decoder
\[ \hat{x}_{style2} \]
Train Pipeline

- Style 1
  - Decoder
  - $\hat{x}_{style1}$
- Style 2
  - Decoder
  - $\hat{x}_{style2}$
- Classifier
Train Pipeline

Style 1

Style 2

decoder

decoder

\(z\)

\(\hat{x}_{style1}\)

\(\hat{x}_{style2}\)

classifier
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

- Gender: Baseline (Shen et al, 2017) 60.4, Ours 57.04
- Political Slant: Baseline (Shen et al, 2017) 75.82, Ours 88.01
- Sentiment: Baseline (Shen et al, 2017) 80.43, Ours 87.22
Preservation of Meaning

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

- Gender
  - Baseline (Shen et al, 2017): 15.23
  - Ours: 43.41
  - No Preference: 41.36

- Political Slant
  - Baseline (Shen et al, 2017): 14.55
  - Ours: 39.55
  - No Preference: 45.9

- Sentiment
  - Baseline (Shen et al, 2017): 35.91
  - Ours: 23.18
  - No Preference: 40.91

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

- Gender
  - Baseline (Shen et al, 2017): 14.55
  - Ours: 39.55
  - No Preference: 45.9

- Political Slant
  - Baseline (Shen et al, 2017): 35.91
  - Ours: 23.18
  - No Preference: 40.91

- Sentiment
  - Baseline (Shen et al, 2017): 40.91
  - Ours: 45.9
  - No Preference: 39.55
Al assistance deals with form (grammar, style, etc.)

Our goal is to control for content
What is our task?
What is our task?

Monkey selfie copyright dispute

From Wikipedia, the free encyclopedia
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]
What is our task?

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]
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
Primary Contribution

curated text (context)
Primary Contribution

+ curated text (context)
Primary Contribution

curated text (context) + document
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
Applications
Applications

Software Documentation
Applications

Software Documentation

Wikipedia Articles

Inbox

Legal Summarization

Document

Legal Articles
Overview

Style

Content

Ethical Considerations

ACL ’20
Storytelling ’19
ACL ’18

NAACL ’19
EMNLP ’18

WiNLP ’19
Overview

Style

- ACL ’20
- Storytelling ’19
- ACL ’18

Content

- NAACL ’19
- EMNLP ’18

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
Ethical Frameworks

- Understanding ethics and the techniques designed to make systems ethical should be well-grounded in the relevant literature outside AI.
  - Generalization principle
    
    [An ethical decision-maker] must be rational in believing that the reasons for action are consistent with the assumption that everyone with the same reasons will take the same action.
  - Utilitarian principle
    
    An action is ethical only if it is not irrational for the agent to believe that no other action results in greater expected utility.
Thank You!

- Alan W Black (co-advisor)
- Ruslan Salakhutdinov (co-advisor)
- Yulia Tsvetkov (committee member)
- Jason Weson (committee member)
- Michel Galley (collaborator)
- Chris Quirk (collaborator)