Delete, Retrieve, Generate: A Simple Approach to Sentiment and Style Transfer

Juncen Li\textsuperscript{1}, Robin Jia\textsuperscript{2}, He He\textsuperscript{2}, and Percy Liang\textsuperscript{2}

\textsuperscript{1}Tencent \hspace{2cm} \textsuperscript{2}Stanford University
Text Attribute Transfer

Original Sentence: “The gumbo was bland.”
Original Attribute: negative sentiment
Target Attribute: positive sentiment
New Sentence: “The gumbo was tasty.”
No parallel data
<table>
<thead>
<tr>
<th>English</th>
<th>French</th>
</tr>
</thead>
<tbody>
<tr>
<td>The blue house is old.</td>
<td>La maison bleue est vieille.</td>
</tr>
<tr>
<td>The music was loud.</td>
<td>La musique était forte</td>
</tr>
<tr>
<td>The boat left.</td>
<td>Le bateau est parti</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Negative

The gumbo was bad

Very rude staff

Poorly lit

...

Positive

The beignets were tasty

I like their jambalaya

Very affordable

...
Delete, Retrieve, Generate

- **Delete**
  - I *hated* the gumbo

- **Retrieve**
  - *love it*

- **Generate**
  - I *love the gumbo*
Outline

• Prior work with adversarial methods
• Simple baselines
• Simple neural methods
Outline

- Prior work with adversarial methods
- Simple baselines
- Simple neural methods
Basic auto-encoder

Encoder $\rightarrow$ Decoder

The gumbo was bland. $\rightarrow$ The gumbo was bland.

Target = negative

Shen et al. (2017); Fu et al. (2018)
Basic auto-encoder

Encoder \[\Rightarrow\] \[\Rightarrow\] Decoder

Target = positive

The beignets were tasty. The beignets were tasty.

Shen et al. (2017); Fu et al. (2018)
Basic auto-encoder

Encoder ➔ Decoder

Target = positive

The gumbo was bland. ➔ The gumbo was tasty.

Shen et al. (2017); Fu et al. (2018)
Basic auto-encoder

Encoder \[\rightarrow\] Decoder

Target = positive

The gumbo was bland. \[\rightarrow\] The gumbo was bland.

Can copy input and ignore target attribute

Shen et al. (2017); Fu et al. (2018)
The gumbo was bland.  The gumbo was tasty.

Make discriminator unable to predict attribute

Shen et al. (2017); Fu et al. (2018)
Error Cases

Input: “Think twice -- this place is a dump.”
Output: “Think twice -- this place is a dump.”

No Attribute Transfer
Error Cases

Input: “The queen bed was horrible!”
Output: “The seafood part was wonderful!”
Error Cases

Poor grammar

Input: “Simply, there are far superior places to go for sushi.”

Output: “Simply, there are far of vegan to go for sushi.”
A balancing act

Attribute Transfer

Content Preservation

Grammaticality
Outline

- Prior work with adversarial methods
- Simple baselines
- Simple neural methods
Pick two out of three

Attribute Transfer

Content Preservation

Grammaticalilty
Content + Grammar

Content Preservation

Grammaticalitivity

Just return the original sentence...
Attribute + Grammar

- Any sentence in the target corpus works!
- **Retrieve** one that has similar content as input
The beignets were tasty
Great prices!
The gumbo was delicious
My wife loved the po’boy
...

The gumbo was bland
The gumbo was bland

The beignets were tasty
Great prices!
**The gumbo was delicious**
My wife loved the po’boy

...
I hated the shrimp

The beignets were tasty
Great prices!
The gumbo was delicious
My wife loved the po’boy

...
Content + Attribute

Attribute Transfer

Content Preservation
Content + Attribute

My wife *hated* the shrimp

- **Delete** markers of the source attribute
- **Replace** them with markers of the target attribute
Attribute Markers

Negative
hated
very disappointed
won’t be back
...

Positive
great place for
well worth
delicious
...

Compare Frequency
My wife hated the shrimp
My wife _____ the shrimp

- loved
- tasty
- polite
...
My wife ______ the shrimp

- loved
- tasty
- polite
...
My wife _____ the shrimp

- loved
- tasty
- polite
- ...
My wife _____ the shrimp

- loved
- tasty
- polite
- ...
My wife _____ the shrimp

Retrievec attribute markers from similar contexts

The beignets were tasty
Great prices!
The gumbo was delicious
My wife loved the po’boy

...
My wife _____ the shrimp

Retrieve attribute markers from similar contexts

The beignets were tasty
Great prices!
The gumbo was delicious
My wife loved the po’boy

...
Experiments

- Average over 3 datasets
  - Sentiment for Yelp reviews (Shen et al., 2017)
  - Sentiment for Amazon reviews (He and McAuley, 2016; Fu et al., 2018)
  - Factual to Romantic/Humorous style for image captions (Gan et al., 2017)
Experiments

• Human Evaluation
  • Likert scale from 1-5 for
    • Attribute transfer
    • Content preservation
    • Grammaticality
  • Overall success: get ≥ 4 on each category
## Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Attribute</th>
<th>Content</th>
<th>Grammar</th>
<th>Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>StyleEmbedding (Fu et al., 2018)</td>
<td></td>
<td></td>
<td></td>
<td>12%</td>
</tr>
<tr>
<td>MultiDecoder (Fu et al., 2018)</td>
<td></td>
<td></td>
<td></td>
<td>11%</td>
</tr>
<tr>
<td>CrossAligned (Shen et al., 2017)</td>
<td></td>
<td></td>
<td></td>
<td>12%</td>
</tr>
</tbody>
</table>
## Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Attribute</th>
<th>Content</th>
<th>Grammar</th>
<th>Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>StyleEmbedding (Fu et al., 2018)</td>
<td></td>
<td></td>
<td></td>
<td>12%</td>
</tr>
<tr>
<td>MultiDecoder (Fu et al., 2018)</td>
<td></td>
<td></td>
<td></td>
<td>11%</td>
</tr>
<tr>
<td>CrossAligned (Shen et al., 2017)</td>
<td></td>
<td></td>
<td></td>
<td>12%</td>
</tr>
<tr>
<td>Retrieval Baseline</td>
<td></td>
<td></td>
<td></td>
<td>23%</td>
</tr>
<tr>
<td>Template Baseline</td>
<td></td>
<td></td>
<td></td>
<td>24%</td>
</tr>
</tbody>
</table>
## Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Attribute</th>
<th>Content</th>
<th>Grammar</th>
<th>Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>StyleEmbedding (Fu et al., 2018)</td>
<td>2.6</td>
<td>3.2</td>
<td>3.3</td>
<td>12%</td>
</tr>
<tr>
<td>MultiDecoder (Fu et al., 2018)</td>
<td>3.0</td>
<td>2.8</td>
<td>3.1</td>
<td>11%</td>
</tr>
<tr>
<td>CrossAligned (Shen et al., 2017)</td>
<td>3.2</td>
<td>2.4</td>
<td>3.3</td>
<td>12%</td>
</tr>
<tr>
<td>Retrieval Baseline</td>
<td><strong>3.7</strong></td>
<td><strong>2.7</strong></td>
<td><strong>4.1</strong></td>
<td><strong>23%</strong></td>
</tr>
<tr>
<td>Template Baseline</td>
<td></td>
<td></td>
<td></td>
<td><strong>24%</strong></td>
</tr>
</tbody>
</table>
## Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Attribute</th>
<th>Content</th>
<th>Grammar</th>
<th>Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>StyleEmbedding (Fu et al., 2018)</td>
<td>2.6</td>
<td>3.2</td>
<td>3.3</td>
<td>12%</td>
</tr>
<tr>
<td>MultiDecoder (Fu et al., 2018)</td>
<td>3.0</td>
<td>2.8</td>
<td>3.1</td>
<td>11%</td>
</tr>
<tr>
<td>CrossAligned (Shen et al., 2017)</td>
<td>3.2</td>
<td>2.4</td>
<td>3.3</td>
<td>12%</td>
</tr>
<tr>
<td>Retrieval Baseline</td>
<td>3.7</td>
<td>2.7</td>
<td>4.1</td>
<td>23%</td>
</tr>
<tr>
<td>Template Baseline</td>
<td>3.5</td>
<td>3.9</td>
<td>3.2</td>
<td>24%</td>
</tr>
</tbody>
</table>
## Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Attribute</th>
<th>Content</th>
<th>Grammar</th>
<th>Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>StyleEmbedding (Fu et al., 2018)</td>
<td>2.6</td>
<td>3.2</td>
<td>3.3</td>
<td>12%</td>
</tr>
<tr>
<td>MultiDecoder (Fu et al., 2018)</td>
<td>3.0</td>
<td>2.8</td>
<td>3.1</td>
<td>11%</td>
</tr>
<tr>
<td>CrossAligned (Shen et al., 2017)</td>
<td>3.2</td>
<td>2.4</td>
<td>3.3</td>
<td>12%</td>
</tr>
<tr>
<td>Retrieval Baseline</td>
<td>3.7</td>
<td>2.7</td>
<td>4.1</td>
<td>23%</td>
</tr>
<tr>
<td>Template Baseline</td>
<td>3.5</td>
<td>3.9</td>
<td>3.2</td>
<td>24%</td>
</tr>
<tr>
<td>Human</td>
<td>4.1</td>
<td>4.1</td>
<td>4.4</td>
<td>58%</td>
</tr>
</tbody>
</table>
Outline

• Prior work with adversarial methods
• Simple baselines
• Simple neural methods
Content separation revisited

Adversarial Discriminator

Encoder

Decoder

Target = positive

The gumbo was bland. The gumbo was tasty.

Make discriminator unable to predict attribute

Shen et al. (2017); Fu et al. (2018)
The gumbo was bland.

Adversarial Discriminator

Encoder

Decoder

Target = negative

The gumbo was bland.
The gumbo was bland.

The gumbo was better.

Target = positive

Delete and Generate
## Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Attribute</th>
<th>Content</th>
<th>Grammar</th>
<th>Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>StyleEmbedding (Fu et al., 2018)</td>
<td>2.6</td>
<td>3.2</td>
<td>3.3</td>
<td>12%</td>
</tr>
<tr>
<td>MultiDecoder (Fu et al., 2018)</td>
<td>3.0</td>
<td>2.8</td>
<td>3.1</td>
<td>11%</td>
</tr>
<tr>
<td>CrossAligned (Shen et al., 2017)</td>
<td>3.2</td>
<td>2.4</td>
<td>3.3</td>
<td>12%</td>
</tr>
<tr>
<td>Retrieval Baseline</td>
<td><strong>3.7</strong></td>
<td>2.7</td>
<td><strong>4.1</strong></td>
<td>23%</td>
</tr>
<tr>
<td>Template Baseline</td>
<td>3.5</td>
<td><strong>3.9</strong></td>
<td>3.2</td>
<td>24%</td>
</tr>
<tr>
<td>Delete and Generate</td>
<td>3.6</td>
<td>3.6</td>
<td>3.4</td>
<td><strong>27%</strong></td>
</tr>
<tr>
<td>Human</td>
<td>4.1</td>
<td>4.1</td>
<td>4.4</td>
<td>58%</td>
</tr>
</tbody>
</table>
Context Cues

- Can retrieved attribute markers help the model?
The gumbo was bland.

The gumbo was bland.
The gumbo was bland.
The beignets were tasty.
Great prices!
The shrimp is delicious.
My wife loved the po' boy...

The gumbo was.

Marker = is delicious

The gumbo was delicious.

The gumbo was bland.
## Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Attribute</th>
<th>Content</th>
<th>Grammar</th>
<th>Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>StyleEmbedding (Fu et al., 2018)</td>
<td>2.6</td>
<td>3.2</td>
<td>3.3</td>
<td>12%</td>
</tr>
<tr>
<td>MultiDecoder (Fu et al., 2018)</td>
<td>3.0</td>
<td>2.8</td>
<td>3.1</td>
<td>11%</td>
</tr>
<tr>
<td>CrossAligned (Shen et al., 2017)</td>
<td>3.2</td>
<td>2.4</td>
<td>3.3</td>
<td>12%</td>
</tr>
<tr>
<td>Retrieval Baseline</td>
<td>3.7</td>
<td>2.7</td>
<td>4.1</td>
<td>23%</td>
</tr>
<tr>
<td>Template Baseline</td>
<td>3.5</td>
<td>3.9</td>
<td>3.2</td>
<td>24%</td>
</tr>
<tr>
<td>Delete and Generate</td>
<td>3.6</td>
<td>3.6</td>
<td>3.4</td>
<td>27%</td>
</tr>
<tr>
<td>Delete, Retrieve, Generate</td>
<td>3.7</td>
<td>3.6</td>
<td>3.7</td>
<td>34%</td>
</tr>
<tr>
<td>Human</td>
<td>4.1</td>
<td>4.1</td>
<td>4.4</td>
<td>58%</td>
</tr>
</tbody>
</table>
Deleting too much...

Input: “Worst customer service I have ever had.”
Output: “Possibly the best chicken I have ever had.”
Deleting too little…

Input: “I am actually afraid to open the remaining jars.”

Output: “I am actually afraid to open the remaining jars this is great.”
Thank you!

I don’t like NLP
Delete

love it
Retrieve

I love NLP
Generate

CodaLab
http://tiny.cc/naacl2018-drg

GitHub
https://github.com/lijuncen/Sentiment-and-Style-Transfer