Style Transfer Through Back-Translation

Authors: Shrimai Prabhumoye, Yulia Tsvetkov, Ruslan Salakhutdinov, Alan W Black

What is Style Transfer

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

- 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.”

Goal of the article

To create a representation that is devoid of style but holds the meaning of the input sentence.

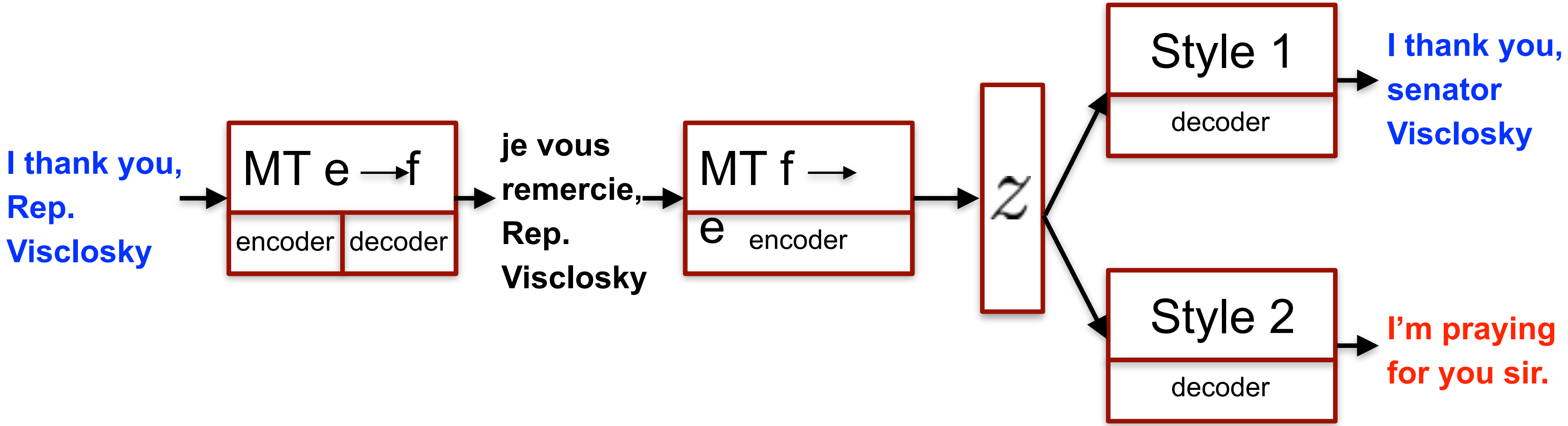
Challenges discussed in the article

- 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.”
- Hard to detect style

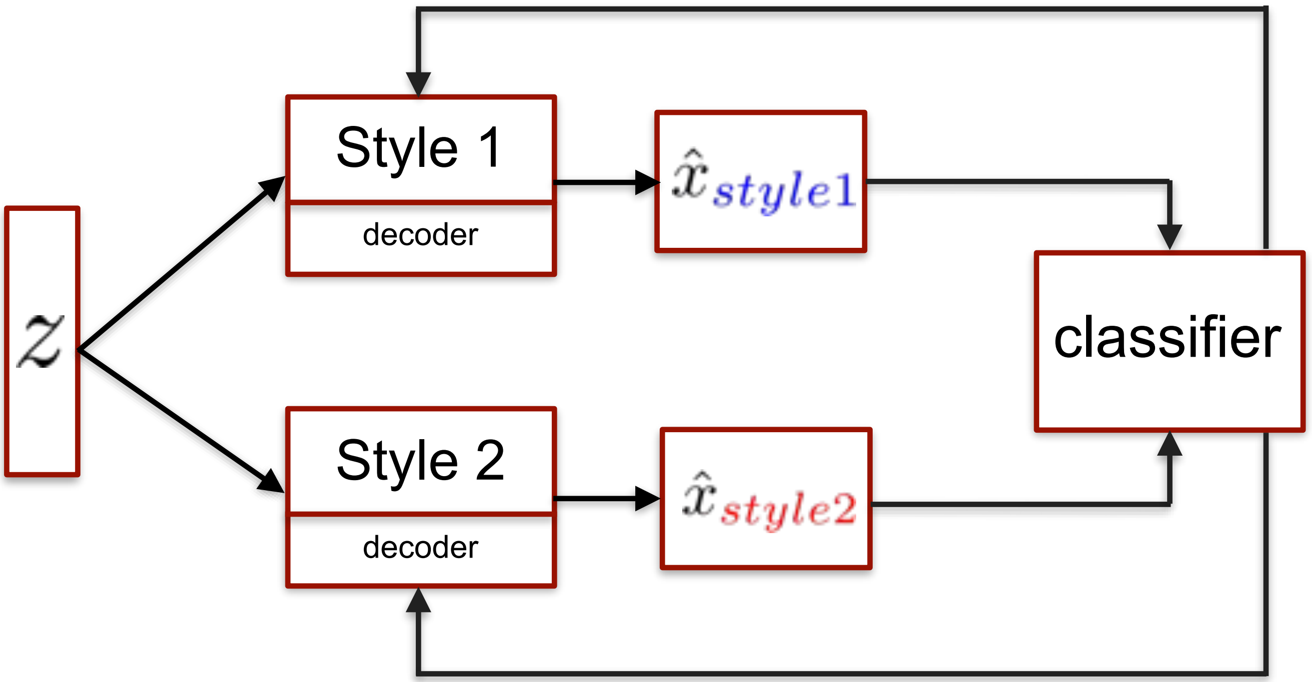
Solution

- Back-Translation
  - Translating an English sentence to a pivot language and then back to English.
- Reduces the stylistic properties
- Helps in grounding meaning

Proposed Architecture



Train Pipeline



Experimental Settings

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

$$\min_{\theta_{gen}} \mathcal{L}_{gen} = \mathcal{L}_{recon} + \lambda_c \mathcal{L}_{class}$$

Baseline

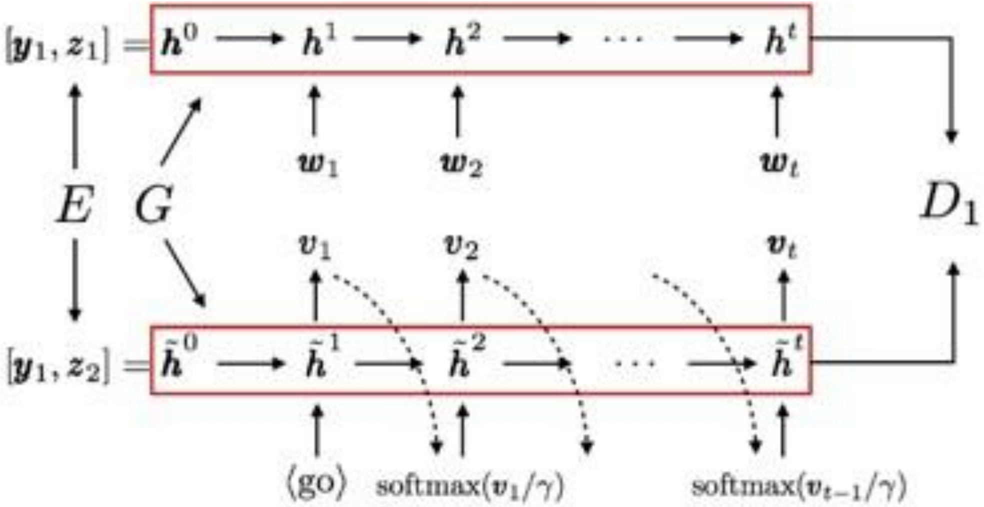


Figure 2: Cross-aligning between x_1 and transferred x_2. For x_1, G is teacher-forced by its words $w_1 w_2 \cdots w_t$. For transfered x_2, G is self-fed by previous output logits. The sequence of hidden states $h^0, \cdots, h^t$ and $\tilde{h}^0, \cdots, \tilde{h}^t$ are passed to discriminator D_1 to be aligned. Note that our first variant aligned auto-encoder is a special case of this, where only h^0 and $\tilde{h}^0$, i.e. z_1 and z_2, are aligned.

Neural Machine Translation

- WMT 15 data
  - News, Europarl and Common Crawl
  - ~5M parallel English - French sentences

Model	BLEU
English - French	32.52
French - English	31.11

Style Tasks

Task	Labels	Corpus
Gender	Male, Female	Yelp
Political Slant	Republican, Democratic	Facebook Comments
Sentiment Modification	Negative, Positive	Yelp

Evaluation

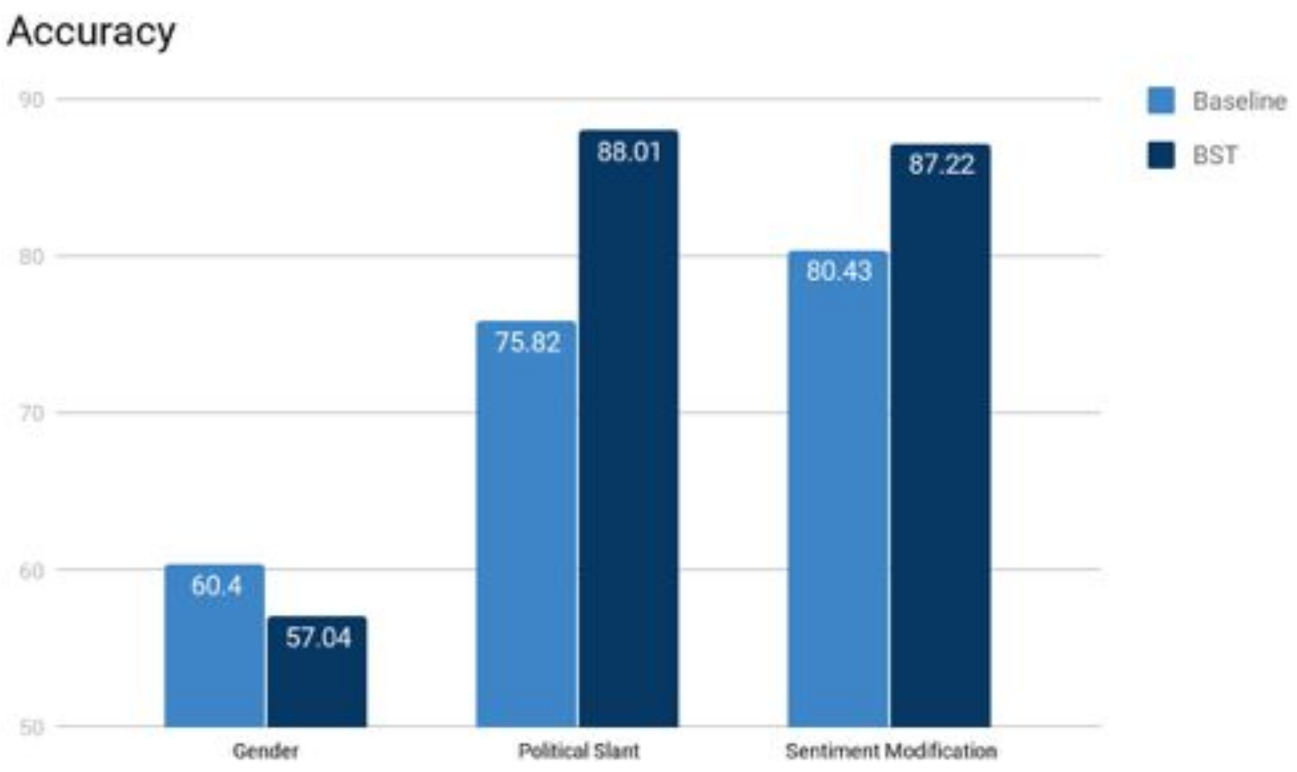
- Style Transfer Accuracy
- Meaning Preservation
- Fluency

Style Transfer Accuracy

- Generated sentences are evaluated using a pre-trained style classifier
- Transferred the style of test sentences and tested the classification accuracy of the generated sentences for the desired label.

Classifier Model	Accuracy
Gender	82%
Political Slant	92%
Sentiment Modification	93.23%

Style Transfer Accuracy



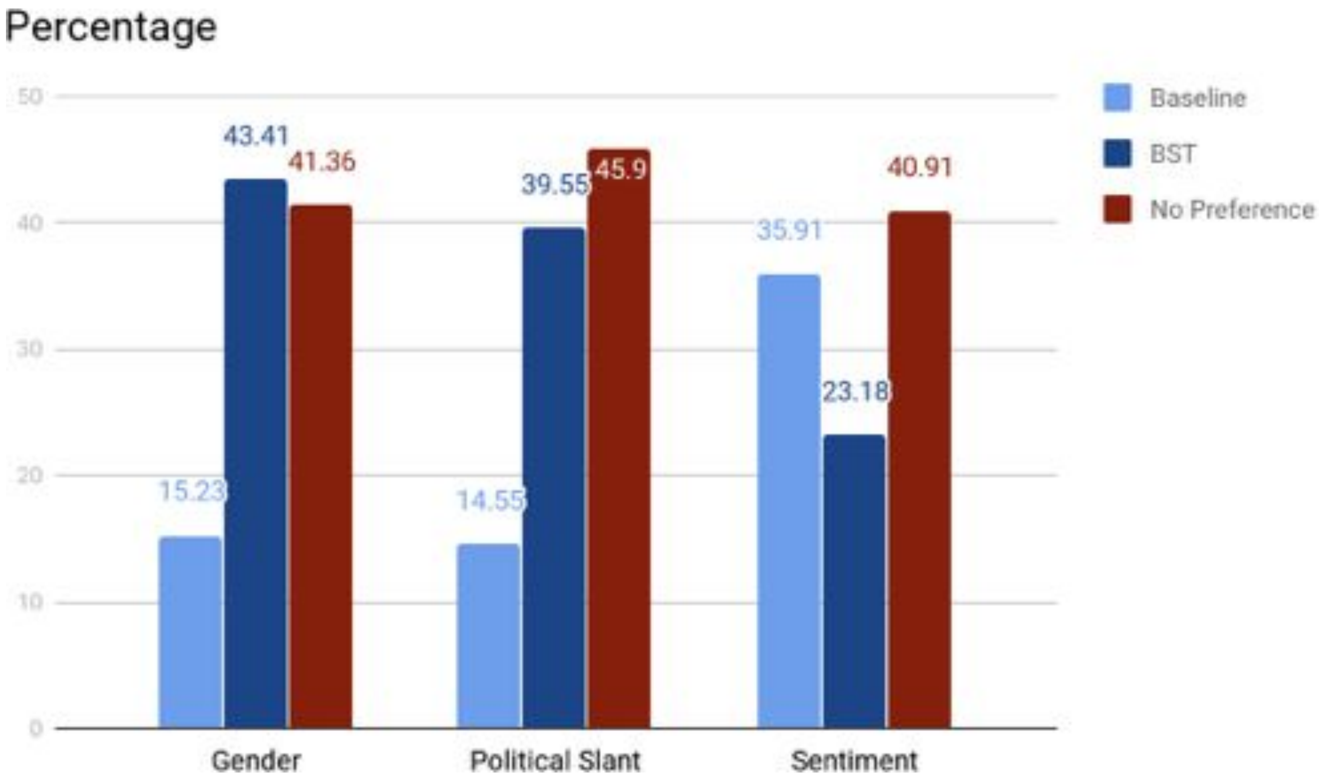
Preservation of Meaning

- Human Annotation: A/B Testing
- The annotators were given instructions.
- Annotators were presented with the original sentence.

Instructions

- “Which transferred sentence maintains the same sentiment of the source sentence in the same semantic context (i.e. you can ignore if food items are changed)”
- “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”

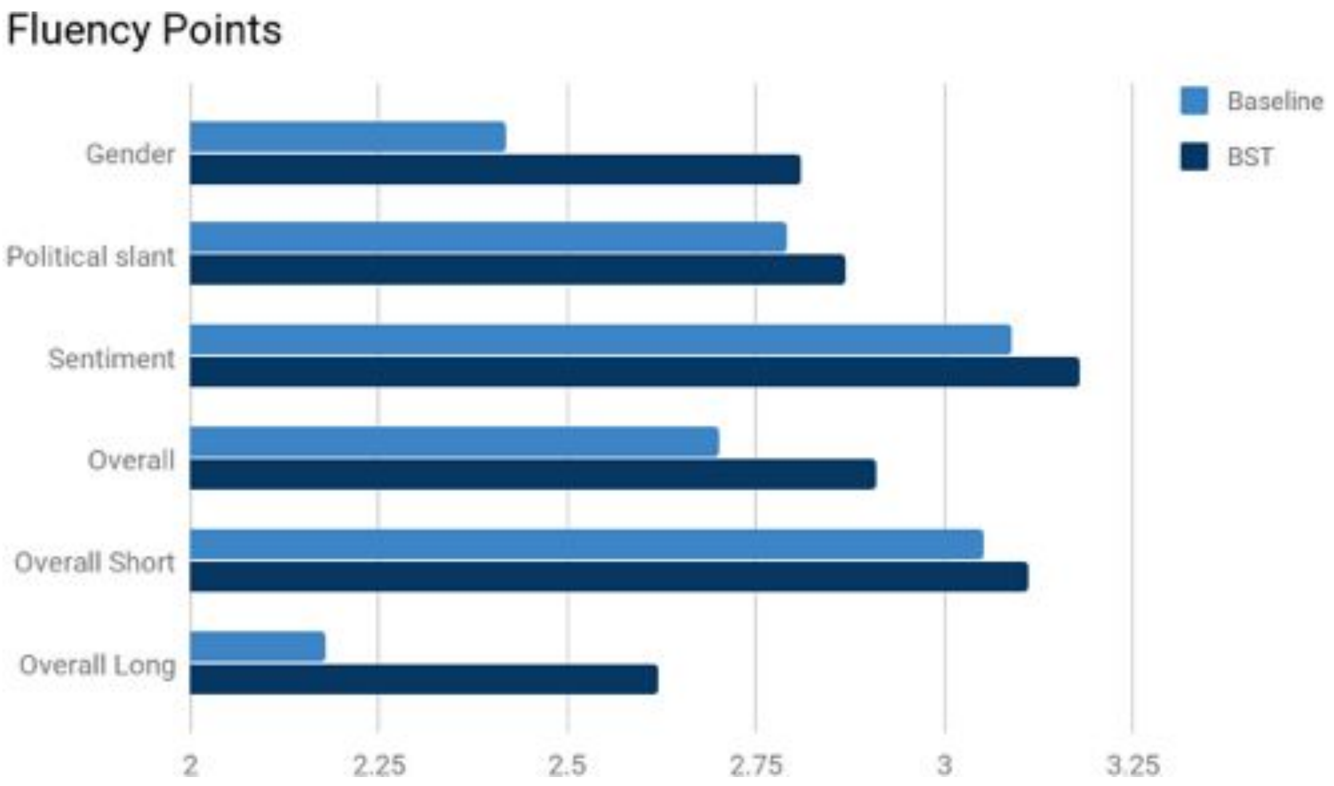
Preservation of Meaning



Fluency

- Human annotators were asked to annotate the generated sentences for fluency on a scale of 1-4.
- 1: Unreadable
- 4: Perfect

Fluency


Fluency Points
Gender
Political slant
Sentiment
Overall
Overall Short
Overall Long
2
2.25
2.5
2.75
3
3.25
Baseline
BST

Discussion

- The loss function of the generators includes two competing terms, one to improve meaning preservation and the other to improve the style transfer accuracy.
- Sentiment modification task in the article is not well-suited for evaluating style transfer
- The style-transfer accuracy for gender is lower for BST model but the preservation of meaning is much better for the BST model, compared to CAE model and to “No preference” option.

Gender Examples

- Male -- Female

  my wife ordered country fried steak and eggs.

  My husband ordered the chicken salad and the fries.

- Female -- Male

  Save yourselves the huge headaches,

  You are going to be disappointed.

Political Slant Examples

- Republican -- Democratic

  I will continue praying for you and the decisions made by our government!

  I will continue to fight for you and the rest of our democracy!

- Democratic -- Republican

  As a hoosier, I thank you, Rep. Vislosky.

  As a hoosier, I'm praying for you sir.

Sentiment Modification Examples

- Negative -- Positive

  This place is bad!

  This place is amazing!

- Positive -- Negative

  The food is excellent and the service is exceptional!

  The food is horrible and the service is terrible.

Future Directions Mentioned

- Enhance back-translation by pivot through several languages to learn a better grounded latent meaning representation.
- Use multiple target languages with single source language as described in (Johnson et al., 2016) to see whether pivoting via multiple languages captures better semantic representations.

- Deploy the system in a real world conversational agent to analyze the effect on user satisfaction
- Caring for more styles!

References

- Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural net- works. In Proc. NIPS, pages 3104–3112.
- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Ben- gio. 2015. Neural machine translation by jointly learning to align and translate. In Proc. ICLR.
- Eric Jardine. 2016. Tor, what is it good for? political repression and the use of online anonymity-granting technologies. New Media & Society.
- Steven J. Spencer, Claude M. Steele, and Diane M. Quinn. 1999. Stereotype Threat and Women's Math Performance. Journal of Experimental Social Psy- chology, 35:4–28.

- Melvin Johnson, Mike Schuster, Quoc V Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Vie´gas, Martin Wattenberg, Greg Corrado, et al. 2016. Google's multilingual neural machine translation system: enabling zero-shot translation. arXiv preprint arXiv:1611.04558.
- Tianxiao Shen, Tao Lei, Regina Barzilay, and Tommi Jaakkola. 2017. Style transfer from non-parallel text by cross-alignment. In Proc. NIPS.