A Hierarchical Reinforced Sequence Operation Method for Unsupervised Text Style Transfer

Chen Wu[1], Xuancheng Ren[2], Fuli Luo[2], Xu Sun[2,3]

Presented at ACL-2019

[1]Tsinghua University

[2]MOE Key Laboratory of Computational Linguistics, School of EECS, Peking University

[3]Center for Data Science, Beijing Institute of Big Data Research, Peking University

Outline

Background

Our Approach

Main Experiment

Take-Home Messages

Outline

Background

Our Approach

Main Experiment

Take-Home Messages

Unsupervised Text Style Transfer

☺ I will be going back and enjoying this great place !

→ ☹ I will not be going back and avoid this horrible place !

Training data (Unsupervised)

- $\mathcal{X}_1 = \{x_1^{(1)}, \cdots, x_1^{(n)}\}$ of style s_1
- $\mathcal{X}_2 = \{x_2^{(1)}, \cdots, x_2^{(m)}\}$ of style s_2
- Non-aligned!

Goal

- $p(x_{1\to 2}|x_1)$ that transfers style s_1 into style s_2
- $p(x_{2\to 1}|x_2)$ that transfers style s_2 into style s_1

Prior Work

Disentanglement Approach

Disentangling latent style and content[1,2,3]

Two-step Approach

Neutralization (deletion) + stylization (reconstruction)[4,5]

---

[1] Hu et al. "Toward Controlled Generation of Text". *ICML-17*.

[2] Fu et al. "Style Transfer in Text: Exploration and Evaluation". *AAAI-18*.

[3] John et al. "Disentangled Representation Learning for Non-Parallel Text Style Transfer". *ACL-19*.

[4] Li et al. "Delete, Retrieve, Generate: A Simple Approach to Sentiment and Style Transfer". *NAACL-HLT-18*.

[5] Xu et al. "Unpaired Sentiment-to-Sentiment Translation: A Cycled Reinforcement Learning Approach". *ACL-18*.

Challenges

Poor Content Preservation

Original.	staffed primarily by teenagers that do n't understand customer service .
System 1.	staffed , the best and sterile by flies , how fantastic customer service .
System 2.	staffed established each tech feel when great customer service professional .
System 3.	staffed distance that love customer service .

Lack of Interpretability

Outline

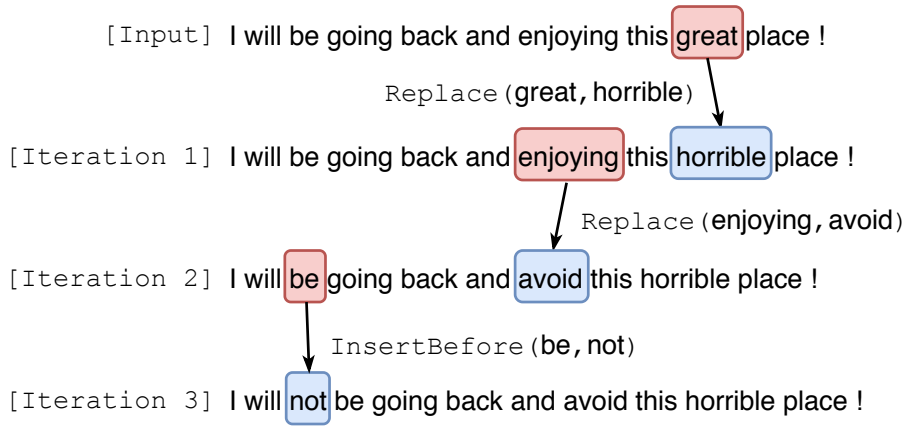
Background

Our Approach

Main Experiment

Take-Home Messages

An Example Case



A Hierarchy of Agents

The Options Framework

An HRL framework proposed by Sutton et al. (1999)

High-Level Agent

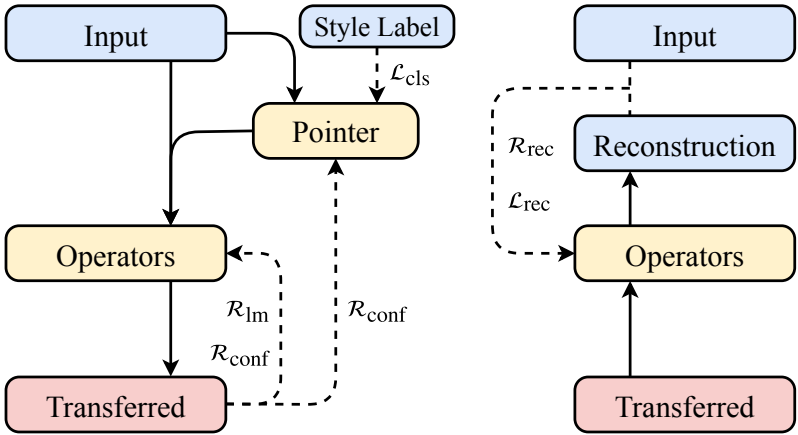
Propose a position to be operated around

Low-Level Agent

Select an operator from the table and generate a word $\hat{w}$ (*optional*)

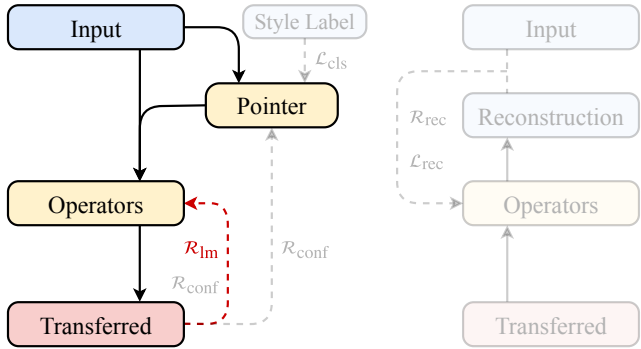
Operator	Operation
IF_{ϕ_1}	**I**nsert a word $\hat{w}$ in the **F**ront
IB_{ϕ_2}	**I**nsert a word $\hat{w}$ **B**ehind
Rep_{ϕ_3}	**Rep**lace the word with $\hat{w}$
DC	**D**elete the **C**urrent word
DF	**D**elete the word in the **F**ront
DB	**D**elete the word **B**ehind
Skip	Do not change anything

Graphical Overview for Training



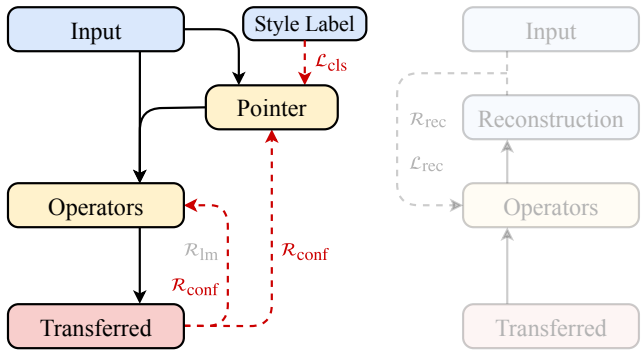
Hierarchical Policy Learning

- Language model reward



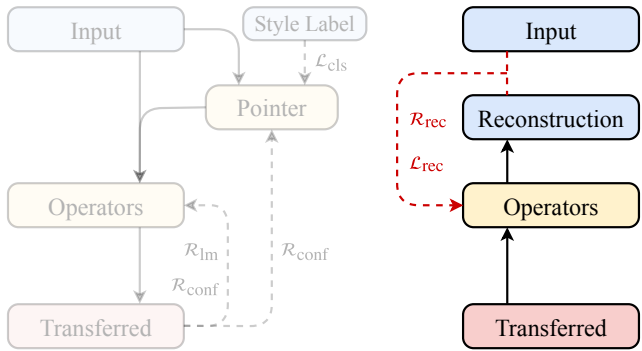
Hierarchical Policy Learning

- Classification confidence reward
- Auxiliary task: style classification



Hierarchical Policy Learning

- Self-supervised reconstruction loss
- Reconstruction reward



Training with Single-Option Trajectory

Motivations for Single-Step Training

- High variance of policy gradients
- Ambiguity in self-supervised reconstruction

Iterative and Dynamic Inference

Basic Ideas

- Enumerate the operators and select the *best* one in each iteration
- Greedy modifications
- Mask-mechanisms
- Until the sentence does not show the original style anymore
  (or beyond maximum iterations)

Outline

Background

Our Approach

Main Experiment

Take-Home Messages

Automatic Evaluation

	Yelp		Amazon	
	Acc	BLEU	Acc	BLEU
CrossAligned	74.7	9.06	75.1	1.90
MultiDecoder	50.6	14.54	69.9	9.07
StyleEmbedding	8.4	21.06	38.2	15.07
TemplateBased	81.2	22.57	64.3	34.79
DeleteOnly	86.0	14.64	47.0	33.00
Del-Ret-Gen	88.6	15.96	51.0	30.09
BackTranslate	94.6	2.46	**76.7**	1.04
UnpairedRL	57.5	18.81	56.3	15.93
UnsuperMT	**97.8**	22.75	72.4	33.95
Human	74.7	-	43.2	-
Point-Then-Operate	91.5	**29.86**	40.2	**41.86**

- Classification accuracy is low for human references
- BLEU of our method outperforms baselines by a large margin

Human Evaluation[6]

	Yelp				Amazon			
	Flu.	Cont.	Sty.	Suc	Flu.	Cont.	Sty.	Suc
TemplateBased	3.47	3.76	3.25	68.0 %	3.46	4.08	2.15	9.0 %
Del-Ret-Gen	3.82	3.73	3.52	70.3 %	4.02	4.31	2.69	21.0 %
UnpairedRL	3.54	3.59	2.90	53.8 %	2.58	2.55	2.44	4.5 %
UnsuperMT	4.26	4.24	**4.03**	**82.5 %**	4.24	4.13	3.05	35.5 %
Point-Then-Operate	**4.39**	**4.56**	3.78	81.5 %	**4.28**	**4.47**	**3.31**	**47.0 %**

- Better overall performance
- Sacrificed style polarity on Yelp

[6]Baselines for human evaluation are selected based on automatic evaluation

Outline

Background

Our Approach

Main Experiment

Take-Home Messages

Take-Home Messages

1. A sequence operation method with hierarchical reinforcement learning (HRL) for unsupervised text style transfer
2. Address two challenges
   - Content preservation
   - Interpretability
3. Provide an iterative and dynamic mask-based inference algorithm that allows for single-option trajectory training

We make our code public.