Improved Neural Text Attribute Transfer with Non-parallel Data

Igor Melnyk, Cicero Nogueira dos Santos, Kahini Wadhawan, Inkit Padhi, Abhishek Kumar

IBM Research AI
T. J. Watson Research Center
Yorktown Heights, NY

igor.melnyk@ibm.com, cicerons@us.ibm.com, kahini.wadhawan@ibm.com, inkit.padhi@ibm.com, abhishk@us.ibm.com

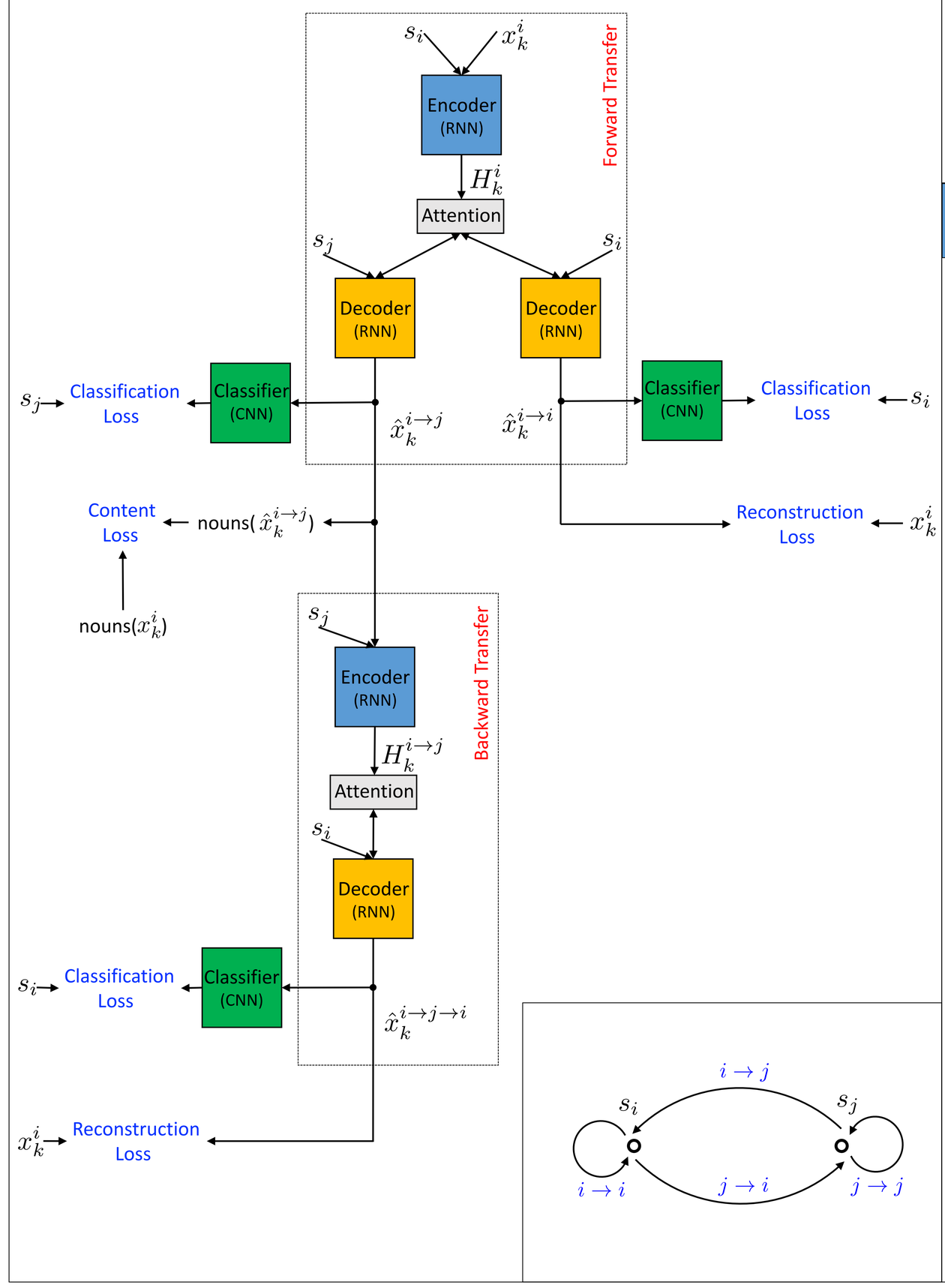
Introduction

- Text attribute transfer with non-parallel data
  - Change certain attributes of text
  - Preserve content
  - No access to parallel data
  - Example:

Input		Output
attribute1 (Language = English) attribute2 (Sentiment = Positive)	➡	attribute1 (Language = French) attribute2 (Sentiment = Negative)

- Examples of other attribute transfers:
  - Professional medical text → Colloquial medical text
  - Scientific paper → Patent
  - Short sentence → Long sentence
  - Past tense → Present tense

Text Attribute Transfer



Model Details

- Setup
  - Data contains two non-parallel corpora: $X = X_0 \cup X_1,\ |X| = N$
  - Sentence x_k^i, where $k \in 1, \ldots, N$, attribute s_i, for $i \in \{0, 1\}$
- Model Components
  - Encoder: $E(x_k^i, s_i) = H_k^i$
  - Decoder/Generator: $G(H_k^i, s_j) = \hat{x}_k^{i\to j}$
  - Classifier: $C(\hat{x}_k^{i\to j}) = p_C(s_j|\hat{x}_k^{i\to j})$
- Model Losses

Forward Transfer:
- Reconstruction Loss: $\mathcal{L}_{rec} = \mathbb{E}_{x_k^i \sim X}\left[-\log p_G\left(x_k^i | E(x_k^i, s_i), s_i\right)\right]$
- Content Loss: $\mathcal{L}_{cnt} = \mathbb{E}_{\left(x_k^j = \{\ldots, w_{kr}^j \ldots\} \sim X\right)}\left[-\log p_G\left(x_k^i = \{\ldots, w_{kr}^i \ldots\} | E(x_k^i, s_i), s_j\right)\right]$
  - where w_{kr}^j is a noun in x_k^j ; $w_{kr'}^{i\to j}$ is a noun in $\hat{x}_k^{i\to j}$ and (r, r') is a correspondence pair established by attention mechanism
- Classification Loss: $\mathcal{L}_{class\_td} = \mathbb{E}_{\left(\hat{x}_k^{i\to j} \sim \hat{X}\right)}\left[-\log p_C(s_j|\hat{x}_k^{i\to j})\right]$

Backward Transfer:
- Reconstruction Loss: $\mathcal{L}_{back\_rec} = \mathbb{E}_{x_k^i \sim X}\left[-\log p_G\left(x_k^i | E(\hat{x}_k^{i\to j}, s_j), s_i\right)\right]$
- Classification Loss: $\mathcal{L}_{class\_btd} = \mathbb{E}_{\left(\hat{x}_k^{i\to j} \sim \hat{X}\right)}\left[-\log p_C\left(s_i | G(E(\hat{x}_k^{i\to j}, s_j), s_i)\right)\right]$

- Classification Loss on Original data: $\mathcal{L}_{class\_od} = \mathbb{E}_{x_k^i \sim X}\left[-\log p_C(s_i|x_k^i)\right]$

Results

- Evaluated on single attribute transfer: sentiment transfer (positive ⟷ negative)
- Data
  - Yelp restaurant reviews
    - Positive (179K, 25K, 51K), Negative (268K, 38K, 76K)
  - Amazon customer reviews
    - Positive (265K, 33K, 33K), Negative (265K, 33K, 33K)
- Evaluation
  - Sentiment accuracy
    - Pre-trained classifier accuracy is 97.4% for Yelp and 82.02% for Amazon
  - Perplexity score
    - Pre-trained language model perplexity is 23.5 for Yelp and 25.5 for Amazon
  - Content preservation

	Yelp			Amazon		
	Sentiment	Content	Perplexity	Sentiment	Content	Perplexity
Shen et. al	86.5	38.3	27.0	32. 8	71.6	27.3
Our Method	94.4	77.1	80.1	59.5	77.5	43.7

- Examples of transfer (positive → negative) on Yelp dataset

Original	their food was definitely delicious	love the southwestern burger
Shen et. al	there was so not spectacular	avoid the pizza sucks
Our Method	their food was never disgusting	avoid the grease burger
Original	restaurant is romantic and quiet	the facilities are amazing
Shen et. al	the pizza is like we were disappointed	the drinks are gone
Our Method	restaurant is shame and unprofessional	the facilities are ridiculous

- Examples of transfer (negative → positive) on Yelp dataset

Original	sorry they closed so many stores	these people will try to screw you over
Shen et. al	thanks and also are wonderful	these guys will go to work
Our method	amazing they had so many stores	these people will try to thank you special
Original	i wish i could give them zero stars	seriously , that 's just rude
Shen et. al	i wish i love this place	clean , and delicious ...
Our method	i wish i 'll give them recommended stars	seriously , that 's always friendly

References:

- T. Shen, T. Lei, R. Barzilay, and T. Jaakkola. ***Style transfer from non-parallel text by cross-alignment***. NIPS, 2017
- Z. Hu, Z. Yang, X. Liang, R. Salakhutdinov, and E. P. Xing. ***Towards controllable generation of text***. ICML, 2017
- ***Language Style Transfer from Non-Parallel Text with Arbitrary Styles***, under review in ICLR 2018
- Z. Fu, X. Tan, N. Peng, D. Zhao, R. Yan. ***Style Transfer in Text: Exploration and Evaluation***, arXiv, 2017
- J. Ficler and Y. Goldberg. ***Controlling linguistic style aspects in neural language generation***. arXiv, 2017
- J. Mueller, D. Gifford, and T. Jaakkola. ***Sequence to better sequence: continuous revision of combinatorial structures***, ICML, 2017