

Faculty of Computers and Information Science  
Cairo University

# Style-Transfer Text Paraphrasing

## An Introductory Presentation

Ahmed Hani Ibrahim

April 28, 2018



## Introduction

- Task Definition
- Motivation
- Difficulties and Challenges

## Recent Existing Solutions

- Sequence-to-sequence Variational Autoencoder Model
- Stacked Residual LSTM Sequence-to-sequence Model
- Deep Reinforcement Learning Sequence-to-sequence Model

## Evaluation Metrics

- BLEU, METEOR, WER Evaluation Metrics

## References



- ▶ **Paraphrasing** is core problem in Natural Language Processing that refers to texts that convey the same meaning but **with different expressions**
- ▶ We can consider it as a **transformation** for a given text with keeping the semantic it
- ▶ **Style-Transfer Paraphrasing** preserves the writer's style of writing while generating the paraphrase
- ▶ In other words, Style-Transfer Paraphrasing is the regular Text Paraphrasing **conditioned** on the writing style



### Paraphrase Examples

- ▶ How far is Earth from Sun
- ▶ What is the distance between Sun and Earth
- ▶ How many miles is it from Earth to Sun
- ▶ Distance between Earth and Sun

### Style-Transfer Examples (Shakespeare Poems)

- ▶ JULIA: What shall by these things were a secret fool,  
That still shall see me with the best and force?
- ▶ DUKE SOLINUS: Merchant of Syracuse, plead no more, I  
am not partial to infringe our laws,
- ▶ SCENE III: An ante-chamber. The COUNT's palace.
- ▶ SCENE I: Venice. A street.



- ▶ Paraphrases has numerous applications such as **Information Extraction, Question Answering, Semantic Search** and **Dialogue-based Systems**
- ▶ It can be used as a part of **Plagiarism Detection** for author copyrights ownership and **Text Similarity**
- ▶ **Data Augmentation** for several Natural Language Processing tasks such as **Sentiment Analysis** and **Author Identification and Recognition**
- ▶ Exploring current Machine Learning techniques and their capabilities in text generation and understanding



Due to the complexity of natural language, automatically generating accurate and diverse paraphrases for a given sentence is still very challenging

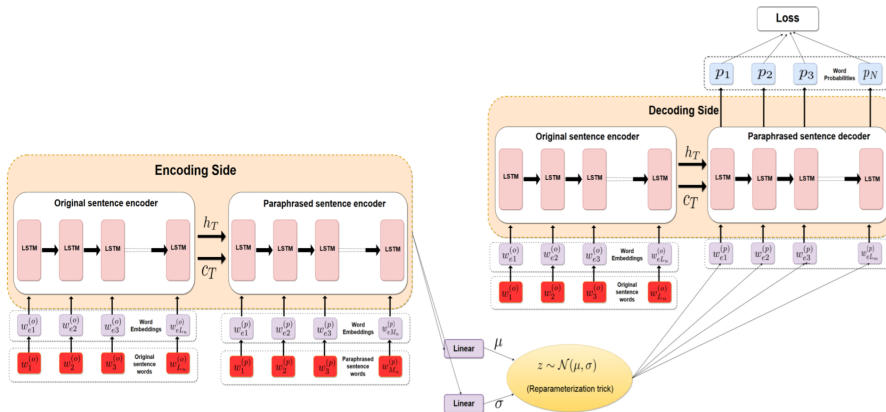
- ▶ Keeping the structure and the language's grammar while generating texts is kind of hard problem
- ▶ Imitating the writing style while keeping the semantic representation of the paraphrased text is unexplored area as far as our knowledge
- ▶ Unsupervised Learning problems are always considered to be challenging problems, specifically in text
- ▶ Semantic level representation of sentences and words are still under exploration and researching

# Recent Existing Solutions

## Sequence-to-sequence Variational Autoencoder Model



### A Deep Generative Framework for Paraphrase Generation. (Gupta et al. 2017)



# Recent Existing Solutions

Sequence-to-sequence Variational Autoencoder Model Cont.



## Encoder Side

- ▶ Two of LSTMs encoders
- ▶ the first one converts the original sentence into vector representation (Skip-thought)
- ▶ The second one encoded the paraphrased sentence as well
- ▶ The two vector representations are fed into Feedforward Network to estimate the VAE's mean and variance

## Sampling

- ▶ Use the estimated parameters to produce a sample from a distribution that is parameterized by the estimated mean and variance





### Decoder Side

- ▶ The VAE's output side uses an LSTM decoder which takes as input the latent representation and the vector representation of the original sentence
- ▶ Both latent representation and the original sentence representation are used to reconstruct the paraphrased sentence
- ▶ In the testing phase, we only concern on the decoder side and ignore the encoding side. We take the decoding side and feed it with the input sentence that we want to get its paraphrased version

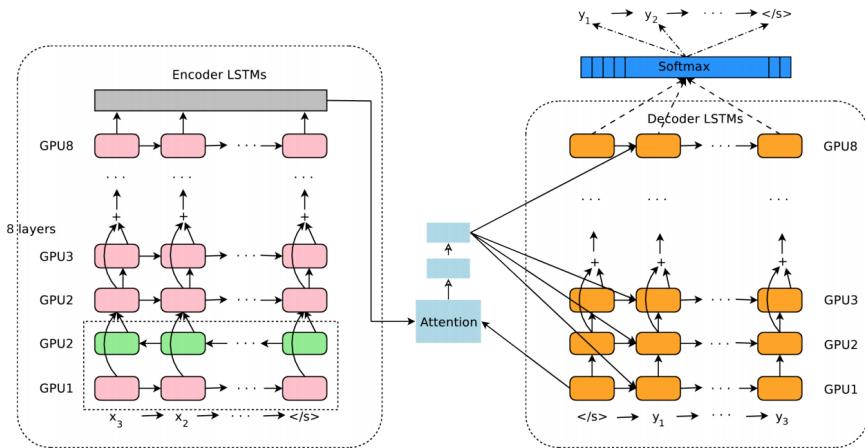


### Model Trivia

- ▶ Simple **Kullback-leibler Divergence** cost function is used to estimate the model loss during the training phase
- ▶ Model is trained and evaluated on MSCOCO and Quora datasets
- ▶ Considered to be the current **state-of-the-art** model in Text Paraphrasing task



## Neural Paraphrase Generation with Stacked Residual LSTM Networks (Prakash et al. 2016)





### Encoder LSTMs

- ▶ Residual Stacked LSTMs - They used only four layers
- ▶ They used one-hot encoding for words representation
- ▶ The input sentence is represented as a Skip-thought vector after applying the stacked LSTMs processes, which is fed into the decoder side

### Decoder LSTMs

- ▶ Residual Stacked LSTMs - They used only four layers
- ▶ The input representation is fed to the decoder as the LSTMs initial state
- ▶ Softmax layer to get the predicted word for each time step

# Recent Existing Solutions

Stacked Residual LSTM Sequence-to-sequence Model Cont.



## Model Trivia

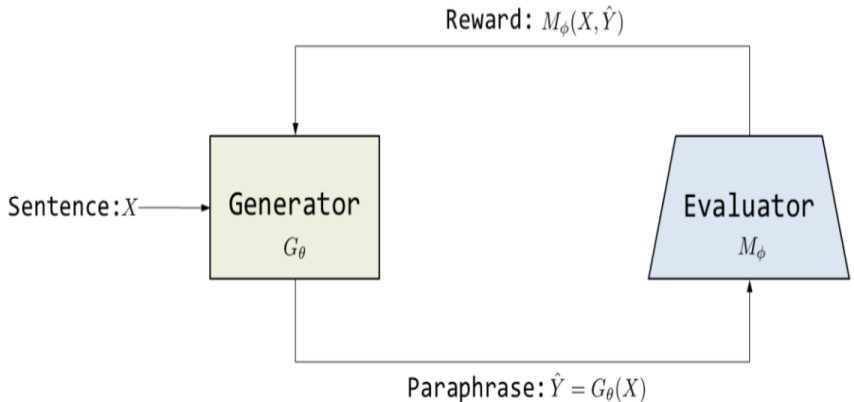
- ▶ The model is inspired from Google's Neural Machine Translation (NMT) Model. But they only used four stacked LSTMs instead of eight
- ▶ Beam Search algorithm was used to rank the best paraphrases candidates
- ▶ Model is trained and evaluated on MSCOCO and WikiAnswers and PPDB datasets
- ▶ Considered to be the first attempt in Text Paraphrasing task using Deep Learning model

# Recent Existing Solutions

Deep Reinforcement Learning Sequence-to-sequence Model



## Paraphrase Generation with Deep Reinforcement Learning (Zichao et al. 2018)





### Generator

- ▶ A simple RNN sequence-to-sequence model
- ▶ The reinforcement part is that every paraphrased word is considered to be an action. The target is to maximize the profit and (the cross entropy in this situation) while generating the words

### Evaluator

- ▶ Trained as a Supervised Learning model
- ▶ The model is trained using set of positive and negative examples (paraphrase pairs) as cross-entropy loss
- ▶ A well trained Evaluator forces the Generator to generate highly accurate paraphrased text
- ▶ This technique called Reinforced by Matching with Supervised Learning (RbM-SL)

# Recent Existing Solutions

Deep Reinforcement Learning Sequence-to-sequence Model Cont.



## Model Trivia

- ▶ The model is considered to be the first attempt on using Reinforcement Learning in Text Paraphrasing problem
- ▶ Outperforms some of sequence-to-sequence models, such as **Vanilla Seq2Seq**, **Seq2Seq with Attention** and **Seq2Seq with Copy**
- ▶ Model is trained and evaluated on Quora dataset





- ▶ Regularly, **BLEU**, **METEOR** and **WER** are the evaluation metrics that are used in any Text Generation task
- ▶ In text paraphrasing, we will focus on **METEOR** metric as it matches between the input and output including word synonyms and hypernyms

## METEOR vs. BLEU

- ▶ **METEOR** word matches between input and output semantic equivalent
- ▶ **METEOR** relies more on words ordering instead of **BLEU's** n-grams approach
- ▶ **METEOR** has better correlation with human judgments, specially in short sentence level
- ▶ **BLEU** works better in large sentences and paragraphs

## References

- A Deep Generative Framework for Paraphrase Generation - <https://arxiv.org/abs/1709.05074>
- Neural Paraphrase Generation with Stacked Residual LSTM Networks - <https://arxiv.org/abs/1610.03098>
- Paraphrase Generation with Deep Reinforcement Learning - <https://arxiv.org/abs/1711.00279>
- Sequence to Sequence Learning with Neural Networks - <https://arxiv.org/abs/1409.3215>
- Generating Sentences from a Continuous Space - <https://arxiv.org/abs/1511.06349>
- Variational Recurrent Auto-Encoders - <https://arxiv.org/abs/1412.6581>
- METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments - <https://www.cs.cmu.edu/~alavie/papers/BanerjeeLavie2005-final.pdf>
- BLEU: A Method for Automatic Evaluation of Machine Translation <http://aclweb.org/anthology/P/P02/P02-1040.pdf>



Thank you!