Generating Fake But Realistic Headlines Using Deep Neural Networks

Ashish Dandekar, Remmy A. M. Zen, Stepháne Bressan

July 18, 2019

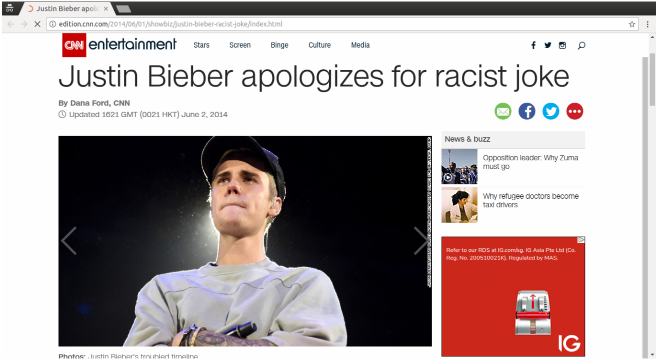
NUS National University of Singapore

Motivation

Can you tell which of these are true?

- First case Chikungunya virus found in Florida: state health care system draws attention
- *The Fault in Our Stars* star Chris Hemsworth and Elsa Pataky reveal
- Samsung sues newspaper over Facebook experiment on users with new profile feature
- Justin Bieber apologizes for racist joke on twitter for gay fans

Motivation



Credit: *http://edition.cnn.com/2014/06/01/showbiz/justin-bieber-racist-joke/index.html*

Motivation

- Social media - a primary source of more than 50% user[1]
- Need to filter out *fake* news
- Use of machine learning models[2]

[1] http://www.digitalnewsreport.org/survey/2016/overview-key-findings-2016/

[2] https://www.theguardian.com/technology/2016/dec/15/facebook-flag-fake-newsfact-check

Introduction

Model is as good as the quality of the data!

Need of a method to generate *fake* but *realistic* headlines for both

- To train *spam* filters
- To evaluated effectiveness of filters

Goodness of model

Ability of capture *syntax* and *semantics* of news headlines

Introduction

Headline Semantics

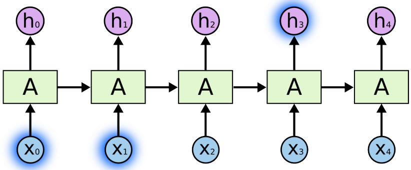
- Ability to learn context of the news headline
- *Contexts*: Topics such as technology, business, entertainment

Headline Syntax

- Ability to learn the sentence structure of the news headline
- *Sentence Structure*: Unconventional grammar to catch attention

NUS National University of Singapore

Recurrent Neural Network (RNN)

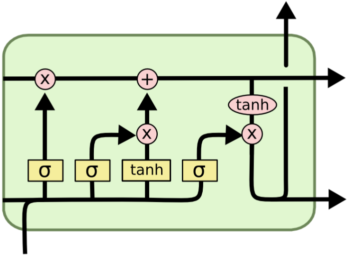


Credit: *http://colah.github.io/posts/2015-08-Understanding-LSTMs/*

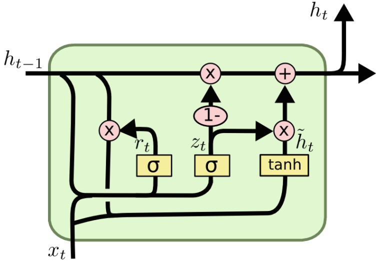
- Deep learning model to capture dependence between previous and current input
- Deep learning model for sequence or temporal data
- *Issue*: Gradient vanishes for the long term dependencies!

RNN variants

Long Short Term Memory (LSTM)

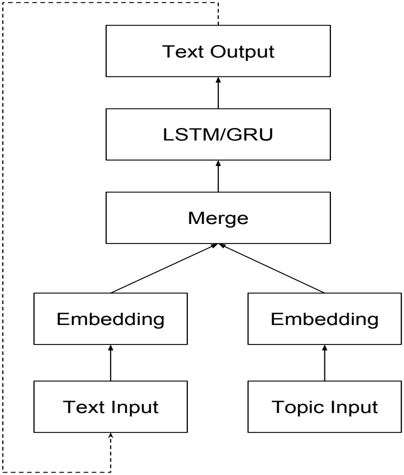


Gated Recurrent Unit (GRU)



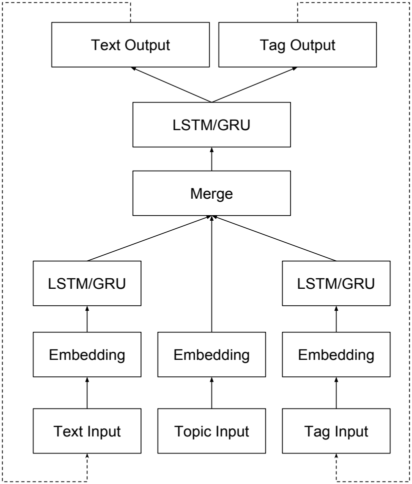
Alleviate vanishing gradient problem by *remembering the history*

Contextual Architecture



- Ghosh et al. [1] used this architecture with LSTM for prediction of next word in a sentence depending upon the provided context

Syntacto-Contextual Architecture



- Syntactic label is given as input
- Learns both dependencies between text inputs and corresponding syntactic labels

Dataset

News Aggregator Dataset[3]

- From 11,242 hostnames
- From 10 March 2014 to 10 August 2014
- 45,000 headlines from from each of four categories: Entertainment, Business, Technology and Medicine

[3] https://archive.ics.uci.edu/ml/datasets/News+Aggregator

Evaluation Metrics

Perplexity

- Quantifies predictive power of a language model
- Defined as

$$Perplexity = 2^{\frac{1}{N}\sum_i \log P^i_{w_t}}$$

- Smaller the better

Syntactic Coherence

- Quantifies the extent to which the generated text adheres to the desired topic
- Defined as classification accuracy over a pre-trained classifier
- Larger the better

NUS National University of Singapore

Evaluation Metrics

Quality

- Quantifies effectiveness in correspondence to a reference
- Defined as the BLEU score considering all news headlines in the corresponding topic as the *gold standard*
- Larger the better

Novelty

- Quantifies *novelty* of generated news headline
- Defined as the number of unique longest sentences between generated news headline and dataset
- Larger the better

Evaluation Metrics

Grammatical Correctness

- Quantifies the extent to which generated headline adheres to the grammar of a certain language
- Defined as the percentage of grammatically correct sentences as classified by a grammar and spell checker tool[4]
- Larger the better

N-gram repetition

- Smaller the better

[4] https://languagetool.org

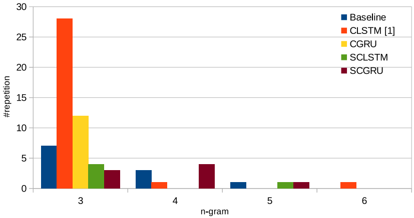
Quantitative Evaluation

	Baseline	CLSTM[1]	CGRU	SCLSTM	SCGRU
Perplexity	108.383	119.10	**92.22**	146.93	175.83
Topical Coherence (%)	-	84.25	77.25	**94.75**	87.50
Quality (BLEU)	0.613	0.637	**0.655**	0.633	0.625
Novelty	21.605	24.67	25.21	**26.57**	25.65
Grammatical Correctness (%)	28.25	49.75	50.75	**75.25**	69.00
n-gram Repetitions	11	30	12	**5**	8

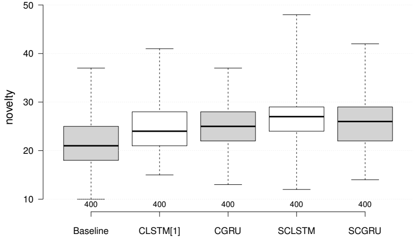
- Baseline: A simple RNN without topic
- CLSTM: Contextual architecture with LSTM
- CGRU: Contextual architecture with GRU
- SCLSTM: Syntacto-contextual architecture with LSTM
- SCGRU: Syntacto-contextual architecture with GRU

NUS National University of Singapore

Qualitative Evaluation



- CLSTM tends to generate headlines with *long* repetitions
- For instance: lorillard *inc nyse wmt wal mart stores* inc nyse wmt wal mart stores



- SCLSTM tends to generate novel headlines on an average

Qualitative Evaluation

Interepreting BLEU score

justin bieber apologizes for racist joke in new york city to take on city (BLEU score of 0.92)

- justin bieber apologizes for racist joke
- uber temporarily cuts fares in new york city to take on city cabs

Higher BLEU score does not necessarily imply good novelty!

Qualitative Evaluation

What do people think?

- Compared the quality of headlines generated by contextual and Syntacto-Contextual architecture
- 66% people agreed that Syntacto-Contextual architecture generates more realistic headlines

Conclusion

- Adapted and extended the model proposed by Ghosh et. al. [1] towards automatic generation of news headlines
- Performed thorough quantitative and qualitative evaluation
- Comparatively showed that the proposed method is better at generating fake but realistic news headlines

References I

Shalini Ghosh, Oriol Vinyals, Brian Strope, Scott Roy, Tom Dean, and Larry Heck.
Contextual lstm (clstm) models for large scale nlp tasks.
arXiv preprint arXiv:1602.06291, 2016.