NYU TANDON SCHOOL OF ENGINEERING

Automatic Generation of Context-specific Fake Reviews

Zhanghao Chen[1], Quanyan Zhu[2]

1. Student, NYU Shanghai 2. Project Advisor, NYU Tandon School of Engineering

Abstract

In this project, we develop a mechanism for automatic generation of context-specific fake reviews. We use an encoder-decoder architecture to generate coherent context-specific reviews. This establishes a better understanding of the possible future attacks towards online review systems and helps us be better prepared for defending against such attacks.

Introduction

- Fake reviews are a major threat to online review systems by spreading misinformation.
- Malicious crowdsourcing forums are currently the major sources of fake reviews, but are limited by the cost of hiring and managing human labors.
- With natural language generation techniques, large-scale and low-cost fake review generation is made possible, but existing methods lack the ability to automatically generate context-specific reviews, although some semi-automatic approach[1] exists.
- We utilize the encoder-decoder architecture for fully automatic generation of context-specific reviews on the Yelp restaurant review dataset.

Methodology

- We adopt the encoder-decoder architecture, originally developed for machine translation by Cho et al.[2]. The encoder captures the context information of reviews (rating star, categories of reviewed restaurant) and the decoder decodes the context information to generate fake reviews.
- Both the encoder and decoder are two-layer stacked GRU (gated recurrent units) recurrent neural networks, but the encoder works in a bidirectional way.

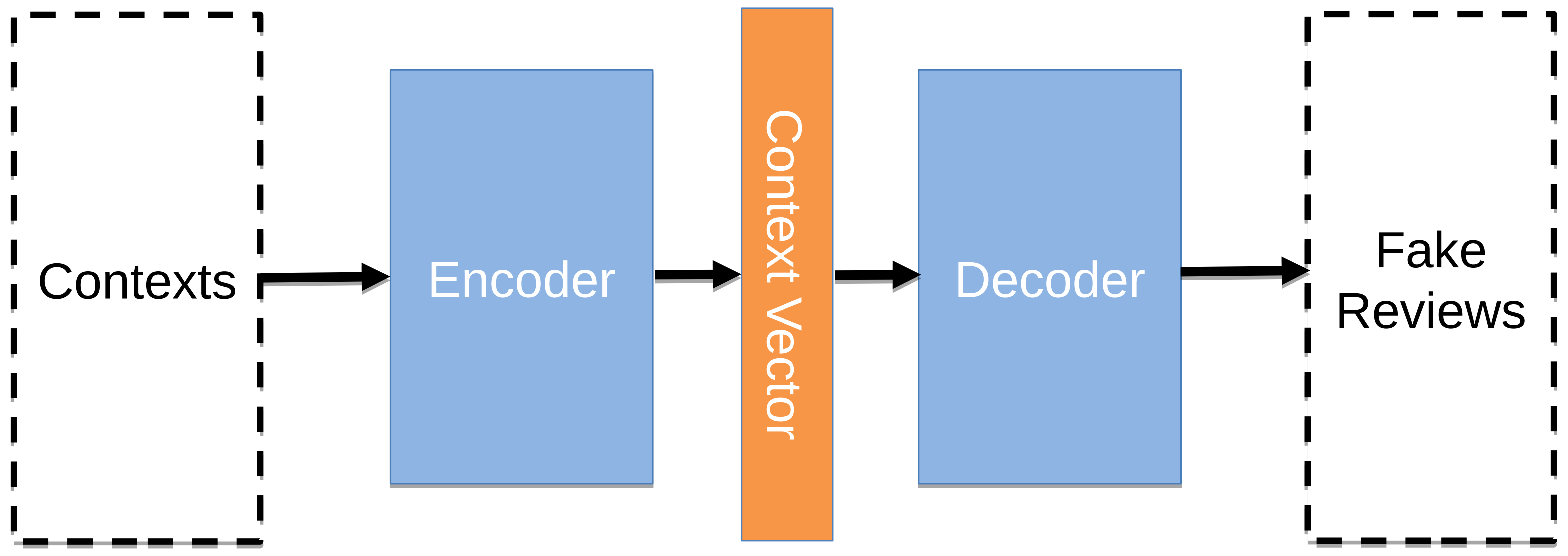

Figure 1: Context-specific review generation model.

- The whole network is optimized with stochastic gradient decent to minimize the negative log likelihood loss between the real reviews and the generated reviews given the same context information.

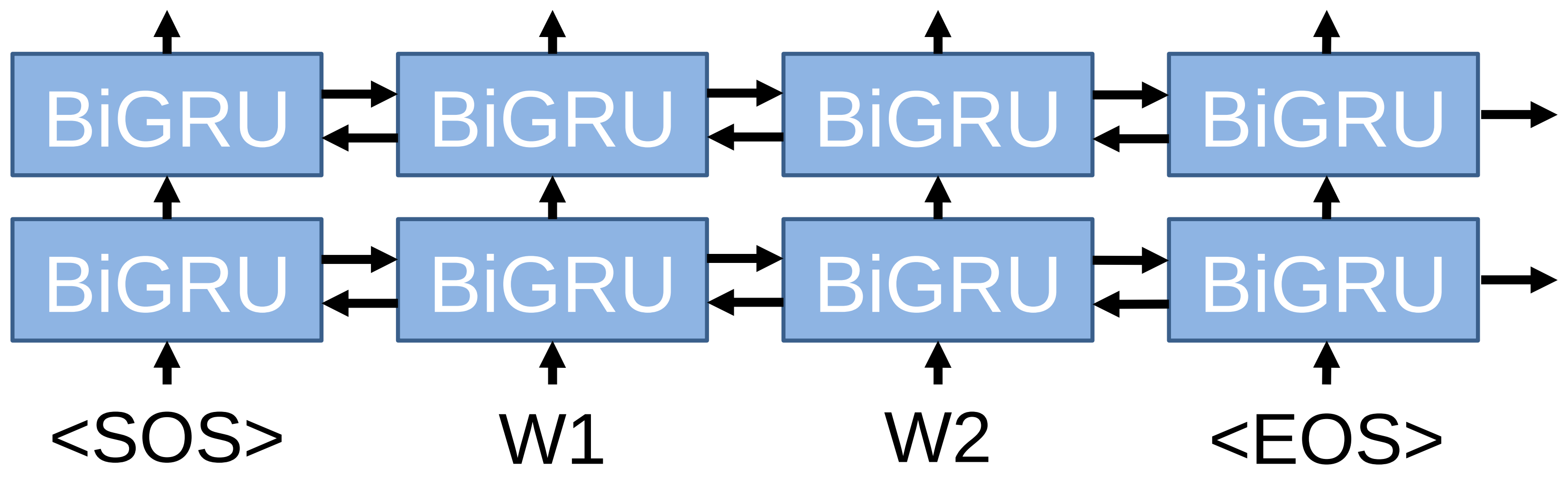

Figure 2: Two-layer stacked bidirectional GRU RNN..

Results

Context	Review (which are fake?)
1.0 restaurants american (traditional) seafood	crabmeat was slimy and they no longer have firegrilled flavor , and my crabdaddy was missing the king crab
	this particular location is filthy , disgusting and horrible . if you don't mind these things you'll have a great time .

Context	Review
3.0 nightlife bars restaurants american (new)	great bartenders . decent food . my burger was so salty i couldn't eat it . fries are good . cheap drinks .
	i love the club , and the music . on a dive bar .
3.0 restaurants vietnamese	cook like friendly and truly from vietnam . food is good and best time to go is weekdays .
	its pretty good pho in a nice quiet location , charming .
4.0 chinese restaurants	last two visits service and food were spot on . when it's good it's the best chinese food in the university area .
	!!! first time to this , but i'm a chicken it is very quick and . . my . . .

Challenges and Future Work

- The most challenging thing is to debug the network due to its sensitivity to parameters. Efficient implementation is also required due to the heavy computation and large amount of data.
- For future work, one direction is to incorporate the generative adversarial (GAN) architecture into our model. Another direction is to develop a detecting mechanism against this type of attack.

Works Cited

[1] Yao, Yuanshun, et al. "Automated crowdturfing attacks and defenses in online review systems." Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2017.

[2] Cho, Kyunghyun, et al. "Learning phrase representations using RNN encoder-decoder for statistical machine translation." arXiv preprint arXiv:1406.1078 (2014)..

Acknowledgement

The authors thank NYU Tandon School of Engineering's Office of Undergraduate Academics for generous funding of the project.