Improved Wasserstein Generative Adversarial Networks for the Generation of Reviews

Cameron Ibrahim

September 2017

Abstract

I propose we research the feasibility of the improved Wasserstein Generative Adversarial Network for the purposes of text generation, specifically with regards to the generation of human-like reviews on Yelp. This is a yet evolving topic which has seen significant progress in recent months; this would be a chance to contribute to this progress, and take new strides into WGANs and into the field of text generation.

1 Research Context

Generative Adversarial Networks operate by maintaining a generator and a discriminator, sometimes called the "critic"; these separate pieces of the model play off of each other as the generator learns to produce samples while the critic learns how to tell if they're real or not. GANs have been historically powerful models which suffer from severe training problems, and which are rarely, if ever, used for text generation; instead, GANs have almost solely been used for image recognition. However, this quite recently began to change with the introduction of the Wasserstein GAN, which aimed to address the issues with optimizing GANs, render them more stable, and ease the restrictive conditions for balancing the generator and discriminator. This was primarily achieved by adopting and approximation of the Earth Mover's distance as a measure of distance in the model.

This is not to say that the WGANs fixed every problem with GANs; rather, it actually added one of it's own. To force EM distance to conform to the requirements of the model, the initial paper introducing WGANs merely clipped the weights, and offered up the search for a better solution to the reader. This call was later heard and answered, resulting in a paper detailing the improved training of Wasserstein GANs.

This improvement on the Wasserstein GAN had the additional effect of allowing the WGAN to operate on discrete data; the paper where it was introduced even went so far as to perform a basic character by character text generation example, which when compared to the abysmal display of text generation

by the standard GAN, was astounding. It is natural then to wonder if we are not at a point in time where we may viably explore the possibility of generating phrases, sentences, passages by making further modifications to these improved WGANs.

2 Data/Design

The Yelp dataset is an abundant source of text, available in the form of reviews. While additional data may be nice to experiment with, either to further the objective of generating reviews (Amazon, IMDB data?), or to explore message generation (Messenger, Google + data?), it is not entirely necessary. It may be necessary to start small, however, so breaking up larger reviews into smaller, more manageable pieces such as sentences may be necessary.

The first objective in the research would be to evaluate the current WGAN's ability to generate text, to see if it is capable of significant word or phrase generation. From there, we would attempt to expand the current capabilities of the WGAN through the attempted application of existing Natural language processing techniques to the model, perhaps making use of Recurrent Neural networks in the generator stage.

3 Methods

As above, the main focus of this research would be the improved Wasserstein Generative Adversarial Network. It generally functions by learning to generate samples which are difficult to differentiate from the real thing. With the additional possibility of using discrete data, we may feasibly be able to generate reviews which mirror the real item.

Additional tools to consider would be recurrent neural networks, tools prevalent in the field of Natural Language Processing today.

Because the generated data would be simple plaintext, analysis would be difficult, beyond reading it manually. Options for automating analysis would be to build a simple program to check for spelling mistakes, or to make a model to verify sentence quality.

4 Significance

This research topic could either be used as a cautionary tale for the company, or as a business opportunity. The advent of bots which may generate comments difficult to distinguish from those left by humans could make a serious ding in the sites credibility; however, it's also entirely plausible that they may themselves offer a service to generate descriptive reviews.

Additionally, this topic of study may lead to a variety of different possible topics in the future. For example, it could influence a new generation of AI based chat bots. Alternatively, by making use of sentiment analysis, one could

create a model which generates phrases based on learned emotional input, in addition to the text itself.

5 References

https://arxiv.org/pdf/1701.07875.pdf

https://arxiv.org/pdf/1704.00028.pdf

https://web.stanford.edu/class/cs224n/reports/2761133.pdf (This is pretty much this proposal, found it when I was looking for more references :( )