Critical Survey of Bidirectional Generative Adversarial Networks for Neural Machine Translation

Mica Haney
University of North Texas
micahaney@my.unt.edu

Abstract

This paper is a critical survey of (Zhang et al., 2018). An overview of the idea, architecture, and results will be given as well as a discussion of the areas where this paper was unclear or lacking information. Overall while the paper is excellent and the discussed idea provides notable results, there are areas of the paper that are lacking. Fortunately, the rest of this well-done paper far outweighs the issues.

1 Introduction

The idea put forth in (Zhang et al., 2018) is a novel model architecture that is a Bidirectional Generative Adversarial Network (BGAN) for Neural Machine Translation (NMT). The proposed architecture has a generator that is a source-to-target translator, while the discriminator is a target-to-source translator. Each translator is used as a generator for one language, and its pair checks the translation and generates in the opposite direction. The idea behind the proposed architecture is to reduce or eliminate some of the issues that come with training a GAN for machine translation tasks. Mainly it looks to target the inadequate training problem.

The inadequate training problem, also known as the exposure bias problem, as described in (Zhang et al., 2018) notes that when training a discriminator for a GAN for NLP purposes, often few samples of correct-incorrect groups of strings are used to train, and often there is a low ratio of correct to incorrect. This is done to bring training times and computational costs low enough for it to be practical to train, as a sufficiently sized dataset could be impossible to practically train. This reduces the dataset to an insufficient size to train on and introduces a bias towards the correct translation in the discriminator that results in the discriminator model quickly overfitting while the generator is still training. Overall the result is a model that underperforms after training has been completed.

By using a translator in the reverse direction as the discriminator, the architecture in (Zhang et al., 2018) reduces the training instability of GANs for translation by providing the ability to generate a much larger sample size with a better ratio of correct to incorrect translations. The proposed architecture also leverages the generators' language models to act as a discriminator, which in theory is a stronger discriminator than the usual neural networks.

2 Architecture

The section describing the architecture has multiple unclear points that make understanding this portion of the paper difficult.

The section 3.1 Training Objective goes over much of the math behind the training policies for the various parts of the GANs. The first and largest confusion comes from the concept of flipping the generator and discriminator. It is stated that this flip forms the second GAN, meaning that the formulas should be flipped with the second discriminator having the formulas from the first generator. However, this is not the case as found in the description of the math for the second GAN. Instead the formulas remain unchanged except for the notation of the new locations. Nowhere is it described how this is done. It is likely that each module of the GANs have the formulas for both the generator and discriminators, however this comes from a logical conclusion rather than the paper. Given that this is the paper where this architecture is first proposed, this is an unfortunate oversight.

Another issue is that the discussion of the architecture only covers the construction of the GANs and the math behind the NMT system. No supporting structures are discussed. It is stated elsewhere that there are embedding layers attached

to the system, but they are never mentioned in the discussion of the model architecture where it should be described above all other sections.

3 Data

For the German-English translation task, the IWSLT2014 evaluation tasks corpus was used. For the Chinese-English translation task, a number of LDC datasets were used. Notably in this section, the data preprocessing for the LDC datasets was discussed (word limits, vocabulary limits, etc.) but not for the IWSLT2014. Nor was it mentioned that data preprocessing was deemed unnecessary. Given that the IWSLT2014 corpus contains TED and TEDx talks, preprocessing should have been necessary and the exclusion of such makes it difficult to implement the model and train with the data necessary to replicate these results.

4 Experiments

In the description of the experimental setup, there is a point of confusion as well as a point of ambiguity.

First, the point of confusion is in the first paragraph of section 4.1.3 Training Details where the setting of the parameters is discussed. Unusually, the only parameters mentioned appears to be gradient parameters, and that is only implied. It requires looking up the referenced paper (Glorot and Bengio, 2010) to understand that this is the discussed parameter. This point is realized when structure rows and columns are briefly discussed, when nowhere else in the paper is any structure with rows and columns mentioned. It should also be noted that the lack of any other parameters, either for the model or for the training, make the replication of these results exceedingly difficult.

As for the point of ambiguity, it is mentioned that the German-English translation model has its hidden states and embedding sizes set to values that allow for the model to be easily compared to previous work. However the Chinese-English model does not have the same values for its hidden states and embeddings, nor is it stated that the chosen values allow for comparison to other models. It is possible, as all but one of the models that the Chinese-English model was compared to were not also in the German-English comparison, but it is not stated or discussed. It begs the questions, are the values different? If so, should that be rectified or did the German-English model need to have its values set to match that of other models?

5 Results

Experiments were conducted on both German-English and Chinese-English translation tasks. Both models were compared with implementations of other translation models using BLEU scores. For both tasks, the proposed BGAN-NMT model outperformed all of the other models that the architecture was compared with. It was reported that for the German-English task, the smallest improvement seen was by 1.14 points, while the Chinese-English task saw a minimum score improvement of 1.03.

The conclusions put forth are that these notable score improvements, combined with an ablation experiment, showcase that the model does in fact gain a great deal of training stability and robustness against the exposure bias problem. On top of these benefits is an increase in the quality of translations generated by the system.

6 Related Work

Unusually the related work section has been relegated to the end of (Zhang et al., 2018). Normally this section is placed at the beginning of the paper to provide readers with supporting information to help them better understand the paper. This is likely because this section is densely packed with references to papers with a short mention of what they did. This gives the impression that the authors did not want readers looking at the related work, or that they found the work that forms the basis of their own research unimportant.

7 Conclusion

Overall this paper proposed a novel idea that produced results supporting its validity. The explanations given in the paper were largely well-done with only a few sections giving cause for confusion. Sufficient mathematical background for the model was provided, something that not all papers do unfortunately. Some information was omitted that leaves questions about how the model was trained, but overall the vast majority of necessary information was presented. On the whole this paper presents well and is worth reading.

References

Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *AISTATS*.

Zhirui Zhang, Shujie Liu, Mu Li, Ming Zhou, and Enhong Chen. 2018. Bidirectional Generative Adversarial Networks for Neural Machine Translation. In *Proceedings of the 22nd Conference on Computational Natural Language Learning*.