Writing Horror Text Using Generative Adversarial Networks with Memory

Monmouth College®
Department of Mathematics and Computer Science

Emily Sheetz

Supervised by Professor Mayfield, Professor Tucker, Professor Utterback

Abstract

The generative adversarial network (GAN) framework is used to train neural networks to generate output. Generating text is a complicated task because meaningful outputs require long-term structure, which can be achieved using a type of recurrent neural network (RNN). By combining a recurrent network architecture with the GAN framework, sentences of horror text were generated on the character level. Based on the results of a survey that put the network to the Turing test, the network's output is not as convincing as text from human authors, but some samples could pass for human.

Generative Adversarial Networks

Generative adversarial networks (GANs) use two competing networks to generate output. The training process for a GAN is depicted in Figure 1 [1]. The generator (G) is fed a random input vector which it uses to generate a sample. The discriminator (D) is given samples and has to classify whether the sample is from the training dataset or G. D's guess is back propagated through the networks, allowing G to learn to generate better samples and D to learn to classify the real and fake samples. Training stops when G generates such good samples that D is randomly guessing the source.

Memory in GANs

To generate long-term structure in the output text, the GAN needs some form of memory. Recurrent neural networks (RNNs) provide memory by feeding outputs back through the network, as can be seen in Figure 2 [5]. This means that each character the network outputs is conditioned on previously generated characters. Gated recurrent units (GRUs) are a particular type of RNN that have shown success when generating text [2]. This project uses GRUs within the generator and discriminator of the GAN framework to output sentences of horror text on the character level. Characters are represented as probability distributions over all possible characters. The network was trained on horror text from Project Gutenberg [3].

Computer or Human?

Try to determine whether each text sample below was generated by a computer or human.

"But after all, the attic was not the most terrible part of the house."
"wasuccuccuccuccuccuccuccuccul"
"in fe- doot was"
"The were we was"
"So long!"
"ok werl we loow were ow we loow were"
"ow"
"I sirat a marte"
"and fis you pain and firin in"
"But I allers left sum hid whar the gal would find it."
"I wis no prifin in"
"we were appay."
"The day."

The Shunned House by Howard Phillips Lovecraft; GAN; GAN; GAN; *Dracula's Guest* by Bram Stoker; GAN; GAN; GAN; GAN; "Sandy's Ghost" in *Twenty-Five Ghost Stories* by W. Bob Holland; GAN; GAN; GAN

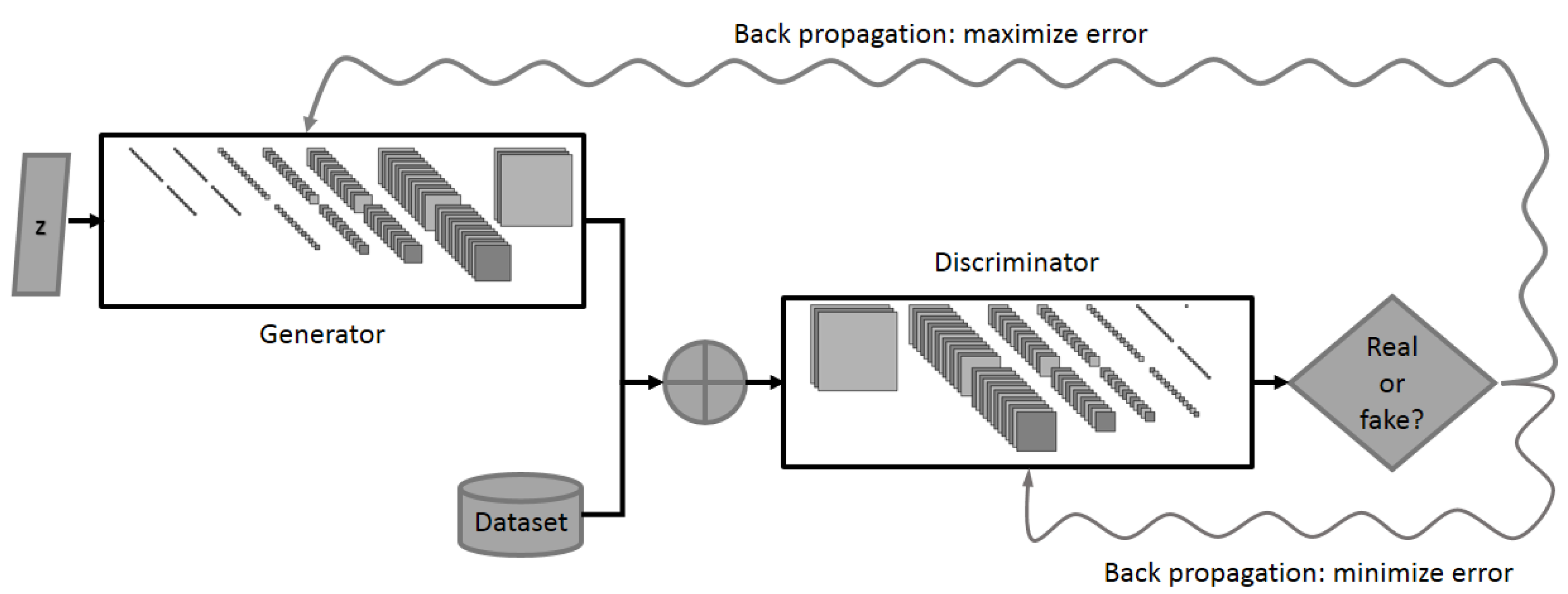

Figure 1: Visualization of the generative adversarial network (GAN) training process. The competing generator and discriminator train to generate text samples comparable to those in the training dataset [1].

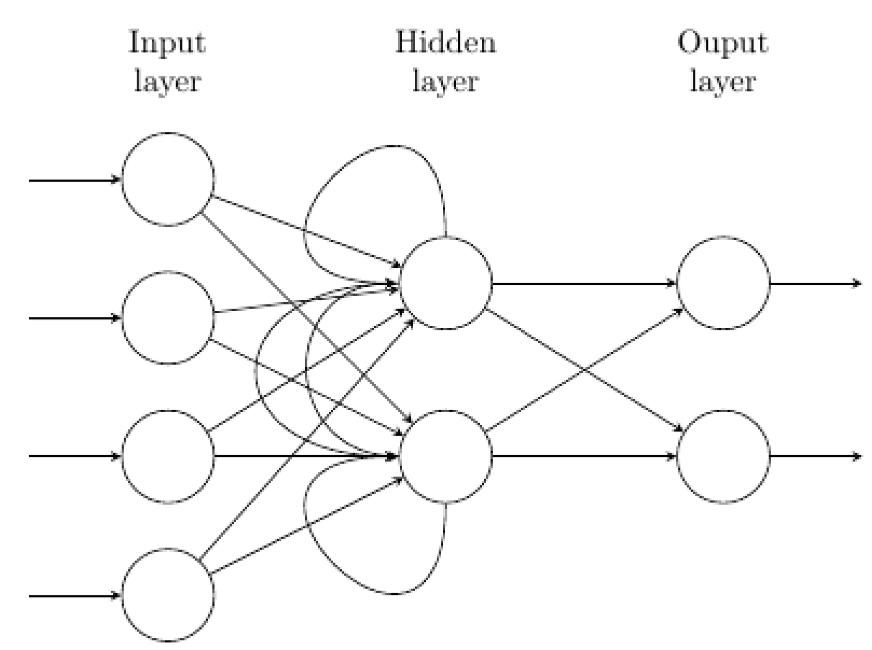

Figure 2: Visualization of recurrent neural networks (RNNs). RNNs can be used for memory within the GAN framework by feeding previous outputs back through the network [5].

Turing Test Survey

To qualitatively evaluate the text generated by the GAN, a survey was sent out to students and faculty. The survey contained 10 text samples from human authors and 10 text samples from the GAN. Participants were asked to determine whether each sample was generated by a computer or human. The results of the survey are shown in Table 1. Some human text samples were confused for computer generated samples. The GAN's samples rarely passed for human samples; most of the time participants correctly determined that the samples were generated by a computer. This Turing test shows that the GAN does not write horror text as well as human authors.

Text Source	Incorrect Responses
GAN	28.5%
Human Authors	32.8%

Table 1: Results of Turing test survey. The table shows how often the GAN's text samples passed for human text and, for comparison, how often text written by human authors was mistaken for computer generated text.

Conclusions

Combining GANs with RNNs—specifically GRUs—for memory allows the neural network to learn long-term structure and generate words and phrases. However, generating complete sentences is a more challenging task. A qualitative comparison of text from computers and human authors shows that the GAN requires more training to generate better text. Though the GAN does not generate quality text, it learned how to spell words by character, showing that the network learned about the structure of language.

References

[1] Heinrich, Greg. "Photo Editing with Generative Adversarial Networks (Part 1)." *NVIDIA, Parallel Forall*, devblogs.nvidia.com/parallelforall/photo-editing-generative-adversarial-networks-1/. 2017. Accessed Aug. 2017.

[2] Press, Ofir, et al. "Language Generation with Recurrent Generative Adversarial Networks without Pre-training." *arXiv preprint arXiv:1706.01399*. 2017.

[3] Project Gutenberg. "Free ebooks – Project Gutenberg." *Project Gutenberg*, www.gutenberg.org/. 2017. Accessed Oct. 2017.

[4] TensorFlow. "Getting Started With TensorFlow." *TensorFlow*, www.tensorflow.org/get_started/get_started. 2017. Accessed Oct. 2017.

[5] TeX StackExchange. "Drawing an Unfolded Recurrent Neural Network." *TeX StackExchange*, tex.stackexchange.com/questions/364413/drawing-an-unfolded-recurrent-neural-network. 2017. Accessed Apr. 2018.