Fake News: Generative Adversarial Networks and Data Creation

Joe Strinka, *Member, IEEE*
RIT Electrical Engineering
jrs5885@rit.edu

***Abstract*— Generative adversarial networks (GANs) are a class of artificial intelligence algorithms used in unsupervised machine learning, implemented by a system of two neural networks contesting with each other in a zero-sum game framework. Introduced in 2014, GANs are a current hot topic in machine learning but the main problems remain of GANs having are extremely sensitive hyper parameters and uninformative loss functions. Deep generative models based on Generative Adversarial Networks (GANs) have demonstrated impressive sample quality but in order to work they require a careful choice of architecture, parameter initialization, and selection of hyper-parameters. Another problem with current GANs is the loss functions (both the generator's and discriminator's) are not informative which means that while the generated samples may start to closely resemble the true data, this behavior can't be indexed to a trend of the losses in general. This means that we can't just run a hyper-parameter optimizer such as skopt using the losses and must instead iteratively tune them manually. The fragility of GANs is due to a dimensional mismatch between the distribution of the model and the true data. As an advanced machine learning technique, an easy to understand description is critical. A GAN is a model made up of two entities: a generator and a discriminator. The discriminator's parameters are optimized to maximize the probability of correctly distinguishing real from fake data (from the generator) while the generator's goal is to maximize the probability of the discriminator failing to classify its fake samples as fake. It is found in this paper that GANs are not well suited to data creation for sets that have data constraints because if there is insufficient data to train the discriminator the generator will tend to copy the input data which results in no new data.**

***Index Terms*—Adversarial learning, Artificial Intelligence, Data Creation, Deep Learning, Generative Adversarial Networks, Generative Models, Knowledge Engineering,**

I. Introduction

Generative Adversarial Networks (GANs) are a type of generative model introduced in 2014. [1] The basic principle of GANs is that the two parts, the generator and the discriminator, are locked in a zero-sum game where the models strive to make their loss functions converge. In the zero sum architecture, the generator and the discriminator learn simultaneously which means that as the generator is learning which data will be classified as fake the discriminator is adjusting the parameters as to what classifies fake data. In general the generator tries to develop parameters that characterize the real data and then generate new data based on those parameters, while the discriminator works to classify the output of the generator as either real or fake data. Since there are only two possible outcomes for the discriminator therefore it follows that the discriminator will be a binary classifier.

GANs are very useful in areas such including image generation and classification, speech and language processing, and image resolution enhancement. [12] It is now possible to use GANs to create photorealistic images of faces, animals, and settings as well as take a low resolution image and make it into a higher resolution image. There are many applications for this technology such as facial recognition and generation of new concepts anywhere that requires a visualization of a specified area.

This paper examines the current state of GANs as well as bringing up a previously unexplored topic, using GANs to generate heartrate data, in order to counteract data availability constraints in the medical field. The reasoning behind the creation of this paper are explored first, followed by a formal definition of GANs. We then examine the advantages and disadvantages of GANs as compared to alternatives such as Hidden Markov Models. After the advantages and disadvantages are established, we look at what GANs are currently being used for and why generating data with low amounts of data is different. Lastly, we will reflect on future work to be done on the subject.

II. Paper Justification

A. Current Generative Approaches

In recent years, the amount of available computing power has skyrocketed making things that were previously impossible possible for the average computer. The whole field of artificial intelligence has also gained traction in real world problem scenarios such as detecting cancer and autonomous locomotion making more researchers and businesses prone to experiment with the newest and greatest forms of artificial intelligence. Generative approaches such as the ones found in this paper would not have been possible thirty years ago due to the lack of available computing power and electrical hardware.

GANs use the concept of deep neural networks to overcome parameter training [4], which is computationally expensive, in order to solve highly complex problems such as image classification and speech recognition.

Generative models are essential to the current landscape of the artificial intelligence community. The generation of data is particularly interesting for this paper because we will be examining the generation of new data for use in the medical

field where data may not be readily available. The generative methods involve forming a hypothesis based on learned parameters from a given dataset and then constructing an artificial piece of data that fits those determined criteria. This generation of data signifies that, at some level, the model understands what makes valid data. There is a possibility that these machines will eventually be able to construct data that is better than what humans could construct given the same starting data. For example, when looking at a heart rate signal a human would typically try to copy the shape and frequency of the heart rate signal while a machine could look at many more variables in order to make a convincing signal.

Adversarial machine learning is a combination of game theoretic machine learning and competition. [4] The success of these types of machines is proven in many forms such as AlphaGo [8] which used this approach as one of its intermediate layers when learning how to beat human masters of the board game Go.

B. Potential Applications

Generative Adversarial Networks are an exciting area because there are so many possible uses. Some of the applicable uses are language processing, image generation, and image resolution enhancement.

One of the uses of generative adversarial networks is language processing. In order to make significant progress in the field of natural language processing a few issues the original GAN architecture must be examined and altered. The common GAN framework has difficulties generating discrete values, and requires a complete generative sequence for the generator to look at. SeqGAN [10] is a sequence generation framework that models the data generator as a stochastic policy which allows it to bypass the generator differentiation problem. In the paper, Rajeswar et al. [10] introduce a baseline that addresses the problem of generating discrete outputs without using gradient exploration which cannot be used with the GAN architecture. SeqGAN was very successful because they were able to generate sentence from context free information. Included below in Fig 1. is a sample of sentences that SeqGAN has created.

Level	Method	1-billion-word
Word	LSTM	An opposition was growing in China . This is undergoing operation a year . It has his everyone on a blame . Everyone shares that Miller seems converted President as Democrat . Which is actually the best of his children . Who has The eventual policy and weak ?
	CNN	Companies I upheld , respectively patented saga and Ambac. Independence Unit have any will MRI in these Lights It is a wrap for the annually of Morocco The town has Registration matched with unk and the citizens
Character	CNN	To holl is now my Hubby , The gry timers was faller After they work is jith a But in a linter a revent

Figure 1: Results of SeqGAN

Seeing understandable sentences generated completely within the generative adversarial network are encouraging for the field of language processing. Since GANs are a relatively new technology, the innovations of the model are still being discovered.

The second use of GANs to be examined is image generation. One of the state of the art image generators is Boundary Equilibrium Generative Adversarial Networks (BeGAN). BeGAN used a new equilibrium forcing method that balances the generator and discriminator. [11] Using this method the results in Fig. 2 were generated.

Figure 2: BeGAN Generated Images

The images in Fig 2. are completely computer generated and photorealistic human beings which is an incredible development in the field of Generative Adversarial Networks.

Another use for GANs is image resolution enhancement. [12] Resolution enhancement can be used in a multitude of different scenarios such as criminal identification and social media. There have been some fantastic improvements in the world of image resolution enhancement such as found in Fig. 3 below:

Figure 3: ENHANCE! Image Resolution Enhancement

Using the CSI style presentation of ENHANCE!, it is seen that GANs could have a long lasting effect on the law enforcement industry by identifying criminal from low quality images. [12]

III. What are Generative Adversarial Networks?

A. Formal Introduction

The idea of a generative adversarial network comes from a theory called Nash Equilibrium. Nash Equilibrium is a stable state of a system involving the interaction of different participants, in which no participant can gain by a unilateral change of strategy if the strategies of the others remain unchanged.

In order to apply GANs to real world problems, it is first required to understand the math behind them. The main parts of understating generative adversarial networks are training the discriminator to detect fake data and training the

generator to learn how to create better fake data using an iterative process.

The discriminator is a binary classifier which works to minimize the loss function of itself given the data from the generator. Due to the nature of the loss function this is often done by minimizing the cross entropy of the system [1]. The loss function is cited below:

$$Obj^{D}(\theta_D\theta_G) = -\frac{1}{2}E_{x\sim pdata(x)}(\log D(x)) - \frac{1}{2}E_{x\sim pdata(x)}\big(\log(1 - D\big(g(z)\big)\big) \quad (1)$$

where x is a sample set of the real data $p_{data}(x)$, z is sampled from the prior distribution $p_z(z)$, and E() represents the expectation model. [1] Given the generator the goal is to minimize the loss function of the discriminator which results in the equation given below:

$$D^*_G(x) = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)} \quad (2)$$

Using (2), the discriminator basically guesses the ratio of the probabilities and updates accordingly. This ratio of the probabilities is what separates GANs from Markov chain models. [1] Since the system is under Nash equilibrium, it follows that the generator's cost function is the negative of the discriminator's cost function which forces the GAN to constantly struggle with itself. Using this logic, the optimization of the entire system can be formulated as a minimax problem [1]

$$\min_G \max_D \{f(D,G) = nE_{x\sim p_{\text{data}}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1 - D(G(z)))]\}. \quad (3)$$

B. Understandable Introduction

The formal intro to any machine intelligence concept can be confusing for machine learning hobbyists and beginning machine learning students. Generative Adversarial Networks are fairly simple to understand and implement. There are GANs that are implemented in less than 50 lines of code. [3] The math basis for the GANs is called a zero-sum framework which basically means that the loss functions are inversely proportional. So as the discriminator performs better, the generator is performing worse. Then the generator would update until the generator is performing better than the discriminator. The main goal of the generator is to trick the discriminator and it follows that the samples that do trick a properly trained discriminator will be realistic. The general flow of a GAN is outlined in Fig 4:

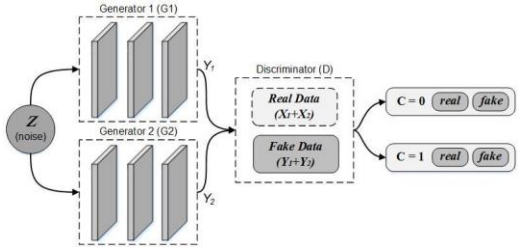

Figure 4: Generative Adversarial Network Framework

IV. Advantages of GANs

Generative adversarial networks have some distinct advantages over traditional generative methods such as the Hidden Markov Model. Two of the main advantages of GANs are that the generated data is easily interpreted by humans, and that GANs can be trained solely with backpropagation.

A. Generates data that can be interpreted by humans

One of the most obvious advantages of GANs are that the data that is generated by this framework is easily interpreted by the human mind. For example, BeGAN generates images that look like human faces, SeqGAN generates sentences that are intelligible to humans, and the entire goal of Enhance is to make images easier for humans to understand.

B. Trained Solely By Backpropagation

Another advantage of the GAN structure is that it is trained solely using backpropagation. Backpropagation is a simple term in the machine learning community that means that given an artificial neural network and an error function, backpropagation calculates the gradient of the error function with respect to the neural network's weights. That is a complicated way to explain that the weights in the neural network structure are updated iteratively based on the inputs and the resulting error function.

V. Disadvantages of GANs

While GANs have numerous advantages as compared with other frameworks, they also have some disadvantages that are associated with their neural network based structure.

A. Cannot understand inner layer data of Neural Nets

One of the disadvantages of a neural network structure is that it acts as a logical black box where the internal weights are not meaningful. This makes troubleshooting GANs particularly frustrating.

B. Needs a lot of input data

The most notable disadvantage of generative adversarial networks in the scope of this paper is that, as with all deep learning techniques, GANs require a lot of input data in order to produce meaningful results. If there is not enough training data given to the discriminator, then the discriminator will not have enough parameters to properly classify data that is generated from the actual input data. This results in the generator simply copying the input data and therefore does not actually result in any new variation of the data.

VI. Training GANs with Limited Data.

In this paper, the amount of data that is needed to train the discriminator is examined since there are some sets of data that are considered private information such as medical data. If the discriminator does not have enough data to be properly trained, then the generator will be able to completely copy the input data and the discriminator will not have developed enough parameters to properly classify the generated samples. [2]. Current Generative Adversarial Networks require a lot of data, like most other deep learning models. There are numerous

approaches being considered in order to address the data constraint problem such as DeliGAN. [13]

VII. Current Approaches to Limited Data

DeliGan addresses the data constraint problem by reparametrizing the latent generative space as a mixture model and learn the mixture model's parameters along with those of GAN.[13] What this means is that in DeliGAN both the GAN parameters, the data characteristics, and the internal weights are learned. This allows for much more control at the expense of the time that it takes to manually tune these extra hyper parameters. Using this simple tweak to the basic structure of generative adversarial networks, results can be seen which enables diverse results despite limited data. DeliGAN has been tested on the MNIST dataset and increased the class variation using a subsample of MNIST as training data. The future of generative adversarial networks is constantly changing as the framework is still very new in the machine learning community. One can expect many more innovations in the realm of generative adversarial networks in a relatively short amount of time.

VIII. Conclusions and Future Work

Unfortunately, it does not seem that current GAN frameworks are well suited to the task of creating diverse data with a limited amount of training data. While there are some innovations being introduced, there is still a long way to go in order to make the elimination of vast databases a reality. There is a future where private information may be able to be kept private due to the generation capabilities of general adversarial networks, but that reality is idealistic at best given the current landscape and frameworks being implemented. There certainly will be future work in this area in the future.

References

[1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2017). Generative Adversarial Networks. [online] Arxiv.org. Available at: https://arxiv.org/abs/1406.2661 [Accessed 12 Dec. 2017].

[2] Roth, K., Lucchi, A., Nowozin, S. and Hofmann, T. (2017). Stabilizing Training of Generative Adversarial Networks through Regularization. [online] Arxiv.org. Available at: https://arxiv.org/abs/1705.09367 [Accessed 12 Dec. 2017].

[3] Ferreira, P. (2017). Towards data set augmentation with GANs – Towards Data Science. [online] Towards Data Science. Available at: https://towardsdatascience.com/towards-data-set-augmentation-with-gans-9dd64e9628e6 [Accessed 12 Dec. 2017].

[4] Salimans, T. and Goodfellow, I. (2017). Improved Techniques for Training GANs. [online] Available at: http://papers.nips.cc/paper/6125-improved-techniques-for-training-gans.pdf [Accessed 15 Dec. 2017].

[5] Wang, K., Gou, C., Duan, Y., Lin, Y., Zheng, X. and Wang, F. (2017). Generative adversarial networks: introduction and outlook. IEEE/CAA Journal of Automatica Sinica, 4(4), pp.588-598.

[6] He, Z., Liu, H., Wang, Y. and Hu, J. (2017). Generative Adversarial Networks-Based Semi-Supervised Learning for Hyperspectral Image Classification. Remote Sensing, 9(10), p.1042.

[7] He, D. and Chen, W. (2017). A Game-theoretic Machine Learning Approach for Revenue Maximization in Sponsored Search. [online] Semantic Scholars. Available at: https://pdfs.semanticscholar.org/9235/63627a1422455ae5ab91e675ad9b98e78e5a.pdf?_ga=2.239858627.1519667248.1511726438-1127838621.1511726438 [Accessed 15 Dec. 2017].

[8] Silver, D., Huang, A., Maddison, C., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T. and Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587), pp.484-489.

[9] Rajeswar, S., Subramanian, S., Dutil, F., Pal, C. and Courville, A. (2017). Adversarial Generation of Natural Language. [online] Arxiv.org. Available at: https://arxiv.org/abs/1705.10929 [Accessed 15 Dec. 2017].

[10] Yu, L., Zhang, W., Wang, J. and Yu, Y. (2017). SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient. [online] Arxiv.org. Available at: https://arxiv.org/abs/1609.05473 [Accessed 15 Dec. 2017].

[11] Berthelot, D., Schumm, T. and Metz, L. (2017). BEGAN: Boundary Equilibrium Generative Adversarial Networks. [online] Arxiv.org. Available at: https://arxiv.org/abs/1703.10717 [Accessed 15 Dec. 2017].

[12] Litt, G. (2017). ENHANCE!: Upscaling images CSI-style with generative adversarial neural networks. [online] Geoffreylitt.com. Available at: http://geoffreylitt.com/2017/06/04/enhance-upscaling-images-with-generative-adversarial-neural-networks.html [Accessed 15 Dec. 2017].

[13] Gurumurthy, S., Sarvadevabhatla, R. and Radhakrishnan, V. (2017). DeLiGAN : Generative Adversarial Networks for Diverse and Limited Data. [online] Arxiv.org. Available at: https://arxiv.org/abs/1706.02071 [Accessed 15 Dec. 2017].