WICCAN: (deep) learning directly from the future

Amanda Coston, Alan Mishler

Abstract

Deep learning methods are extremely popular but suffer from a number of limitations, including computational and conceptual complexity, fragility to input variation, and poor generalizability. Perhaps most worryingly, they rely on data from the past to train models and generate predictions about the future, raising questions about the validity of their output.

The dark arts offer potential solutions to a number of these issues. Deep learning currently relies on a handful of alchemical techniques, but it has yet to take advantage of the full array of available magical methods. Here, we propose a novel type of deep learning: Weakly Independent Concurrent Convolutional Adversarial Networks (WICCAN). WICCAN eschews the reliance on the past that characterizes other techniques. It is model-free, data-free, and space-time-free, and can predict unseen or unrealized labels, outcomes, and events. WICCAN performs comparably to the state of the dark art, with the advantage that it runs in O($\emptyset$) time and does not require access to deceased prophets at runtime.

1 Introduction

1.1 Background

Deep learning methods are extremely popular and perform at state-of-the-art in a wide range of prediction and labeling tasks. However, they suffer from a number of well known limitations. First, they typically require very large amounts of training data, which may be expensive or impractical to acquire and annotate. Second, they can be fragile; for example, a change in a few pixels can cause a convolutional neural net to fail to correctly label an image. Finally, they can be extremely computationally expensive to train and store.

Deep learning suffers from two additional limitations that have been underappreciated in the literature. First, like other machine learning methods, deep learning overwhelmingly relies on data from the past to train models and generate predictions about the future. This approach requires that the data-generating mechanism remain stable over time, an unrealistically strong assumption since a lot of things happened in the past which might not happen again. Furthermore, something new might happen.

Second and relatedly, deep learning models only generalize to the type of data that they learned from. For example, a convolutional neural net trained on images of animals is unlikely to perform well when asked to predict the location and timing of future earthquakes.

Human prophecy has a nearly complementary set of strengths and limitations. Like deep learning, prophecy also aims to uncover the unseen and predict the future. Unlike deep learning, however, prophecy is not fragile and does not require training data; in fact, it has proven remarkably robust to the presence of data. This means that training time is not constrained by the size of the data and that it generalizes equally way to nearly anything. Finally, prophecy takes the sensible approach of peering directly into the future, rather than the somewhat backward approach of looking into the past in order to make guesses about the future.

Although deep learning and the dark arts are often invoked together, there has been surprisingly little discourse between researchers in the two fields. Deep learning has traditionally utilized only a few basic alchemical techniques, leaving untouched a wide array of other methods in the dark arts.

1.2 WICCAN: A new type of neural net

Here, we propose a new method that builds on existing areas of overlap while drawing on additional strengths of the two approaches. The method is called Weakly Independent Concurrent Convolutional Adversarial Networks (WICCAN). WICCAN predicts unseen and/or as-yet-unrealized labels, events, and facts, while eschewing the reliance on data that characterizes other methods. WICCAN is therefore immune to overfitting and generalizes perfectly to any type of task, even when the future does not resemble the past. Notably, WICCAN does not require any pretraining or parameter tuning; the choice of a familiar is all that is required at both train and test time.

We empirically evaluate our method on standard benchmarks as well as new tasks that we believe to be more representative of real-world problems. Our method performs comparably to the state of the dark art, with the advantage that it runs in O($\emptyset$) time and does not require access to deceased prophets at runtime.

The remainder of this paper is organized as follows: In section 2, we summarize relevant work in deep learning and the dark arts. Section 3 details the procedure, to the extent allowable by the Ardanes. Section 4 situates this work within the recently explosive literature on fairness. Section 5 presents our empirical results, which show that WICCAN outperforms existing methods on both benchmarks and unseen real world prediction tasks. Finally, section 6 uses our method to propose future work. Sections 3 and 6 are written in the grimoire tradition and may be incomprehensible to mortals.[1]

2 Related Work

Both machine learning and the dark arts have witnessed significant progress in the past centuries, and this paper builds on the formidable literature of both disciplines. Machines are now outperforming humans on vision tasks and games like Jeopardy and Go [1] [2]. Horcruxes are now able to achieve immortality, and wielders of white magic have demonstrated that a patronus can defeat Dementors [3] [4]. The literature in both fields is too vast to properly review; in this section we discuss a select few relevant works.

In a seminal vision paper, researchers from the University of Toronto showed that deep convolutional neural networks (DCNN's) vastly outperformed other methods on ImageNet classification [5]. The architecture consists of five convolutional layers as well as several fully connected and max pooling layers. The authors proposed methods to reduce the risk of overfitting and to speed up training. While similar in spirit to DCNN's, our method requires only an arbitrary number of layers, trains in sub-constant time, and has no risk of overfitting. ImageNet led to an explosion of followup work, with deep learning researchers investing considerable effort on designing architectures and appropriate activation functions and on developing methods to reduce training time. We believe our method will considerably advance the field since it requires no specification of an architecture and runs in near-trivial time, which we denote as "time-free."

Our method also builds upon the recent work of Nostradamus, whose method of foresight outperformed nearly all political pundits in predicting the 2016 election of Donald Trump [6]. Notably, Donald Trump had never before been elected, so many existing methods were unable to predict this "black swan" event. However, our method significantly improves upon Nostradmus's prophecies, which are characteristically vague albeit never incorrect. Our method inherits his perfect accuracy while specifying the details of the event in question, which aids in the interpretability of the model.

Previous work in the prestigious SIGBOVIK conference considered whether parapsychology can be used to influence people's minds, ultimately and unfortunately finding that the author was "supernaturally unpersuasive" [7]. Our method also concerns the inner workings of mortal minds, with an emphasis on reading versus influencing them, which we find to be a more tractable problem.

[1] If you have the misfortune of being mortal, we recommend reading these sections upside down, and we offer our condolences in the likely event that you are turned into a bat.

(a) My friend's dog

(b) My dog

Figure 1: The optimal familiars for EEVEE

Our method is also relevant to future work in machine learning, which will use more layers and more activation functions to accomplish more things. Perhaps the most relevant of these future methods is EEVEE, Empirical Evidence Variational Expert Encodings [8]. EEVEE uses a familiar-stacking architecture and finds that Tiggy and Eevee, who are displayed in Figure 1 are optimal familiars and are incidentally optimally cute.

As far as the authors are aware, we are the first to propose a paradigm that leverages the benefits of both machine learning and sorcery.

3 Methods

WICCAN consists of an input layer, an arbitrary number of medial layers, and an output layer. The network is trained using SGD, as described in Algorithm 1. Unlike previous methods, training does not require data, which must be necessarily collected in the past relative to training time and which therefore bears only a tenuous relationship to the post-training future.

WICCAN is also distinct from other neural nets in several other respects:

- The input layer can accommodate any size or type of input, as long as it can be framed as a statistical estimation task or as a space-time-based query, such as "will this paper be published in a prestigious proceedings?" or "will a locust swarm devastate the coastal croplands?"
- Unlike other neural nets, WICCAN does not distinguish between hidden and non-hidden layers; all layers in WICCAN, including the output layer, are hidden from mortal sight.
- The only hyperparameter which needs to be selected is the choice of familiar, e.g. cat, owl, snake, or undergraduate research assistant. Other hyperparameters such as the learning rate, learning rate schedule, mini-batch size, paranormality, momentum, convolutional stride, regularization coefficients, broom length, and the number and structure of the medial layers are not explicitly specified and do not need to be separately learned. All parameters, hyper-parameters, pseudoparameters, supernatural parameters, and non-parameters are learned simultaneously in a single burst of power.
- Early stopping: It is extremely dangerous to stop the training spell early.

WICCAN is trained until training is complete, at which point it can be used to predict any unseen or as-yet-unrealized labels, events, or facts.

3.0.1 Complexity

WICCAN retains the full complexity and mystery of the natural world in which it operates. Since no data is used in training, the training time is O(1), although the associated constant is unknown and unknowable. Runtime is O($\emptyset$): answers to queries are instantly available.

Algorithm 1: SGD: Spell-based General Divination

for *number of incantations* k **do**

update the network by intoning the following:

$$\delta_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[\log D\left(x^{(i)}\right) + \log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right) \right] \tag{1}$$

end

Note that k is not set prior to training but is revealed over the course of training.

Table 1: MNIST empirical evaluation

Method	Test accuracy
WICCAN	Slightly above perfect
Support vector machines	0.994
DCNN	0.997
EEVEE	1.00
k-Nearest Neighbors	0.994
Asking my cat	0.996

4 Fairness

In most machine learning contexts, the test set is used as a proxy for the future, yet machine learning researchers regard it as "unfair" to use data from the test set to train the network. We find this stance strange, since the future is ultimately the object of interest. In fact, we regard it as substantially more unfair to use data from the past to make claims about a possibly entirely different future.

We also think pejorative references to deep learning as "alchemy" are unfair. Both alchemy and deep learning have seen widespread successful application [9, 10], despite having their share of misses [11, 12].

5 Empirical Results

We demonstrate the performance of our methods empirically, on a standard machine learning benchmark and on two new prediction tasks which we believe are more representative of real-world tasks. We compare these to current and future state of the art methods.

MNIST is a dataset of handwritten digits where the prediction task is to classify each digit as one of 0-9 [13]. MNIST is used as a standard benchmark for evaluating machine learning algorithms. Table 1 shows that WICCAN outperforms all other methods on MNIST, achieving over 100% accuracy. The authors note that not only was our method able to classify the number correctly but it was also able to reconstruct the users' thoughts as they wrote the digit, a significant improvement upon existing methods. For exposition we include a few excerpts in Table 2.

Motivated by real world challenges, we introduce three new prediction tasks to the literature: predicting 1) Bigfoot sightings, 2) who will dance together at the Yule Ball, and 3) next year's holidays. Understanding the location and movement of Bigfoot is important for maintaining the ecological stability and balance of the forests of the Pacific Northwest. Ecologists, who have asked for years for better prediction of Bigfoot, have extolled our method as telling them *"exactly what they didn't need to know"* and providing *"unvaluable information"* about Bigfoot's whereabouts. Our method was able to predict 14 Bigfoot sightings in the next 1000 years, which compares to 0 predicted by current methods like logistic regression, reputable newspaper articles, and science.

The Yule Ball is a major social event at Hogwarts that can determine the course of one's romantic and social trajectory for the immediate future (on the order of hours). Wizard behavior is notoriously difficult to characterize or predict, but yet our method is able to not only predict the couples who will dance together but also is able to itself dance an enchanting foxtrot. While DCNN's are able to achieve similar results, our method's runtime vastly outperforms DCNN's, which takes 10 lifetimes to run.

Table 2: MNIST mind exposition

Thought	Digit Written
I used to be 8 once	8
6 doesn't really deserve to be a number...kinda creepy	6
Did I meet that guy at the dentist last year?	0
My handwriting is so pretty	4
Why did they tell me to write 5?	6

Table 3: WICCAN prediction for holiday dates in 2020

Holiday	Prediction
Valentine's Day	February 14
April 1st	April 2nd
Halloween	Every day
Someone's Birthday	October 6
Fourth of July	July 4th

Knowledge of the dates of future holidays like Thanksgiving and Valentine's Day is important to most people who want to celebrate with their friends and families. A significant challenge to holiday planning is the uncertainty about when holidays will be. Our method is able to predict some future holidays for 2020, which are demonstrated in Table 3.

6 Conclusion and Future Work

WICCAN will be used in the future to predict future work.

Acknowledgments

We would like to acknowledge the mere mortals who will give us invaluable advice in the distant future: Duncan McElfresh, Anjalie Fields, Kevin Lin, Lily Potter, Channing Huang, Riccardo Fogliato, Benjamin LeRoy, Nic Dalmasso, Kwangho Kim, and Siddharth Ancha. We are also grateful to the immortals who will read this paper and think about it for a little bit, which include Lisa Lee and God. This research is sponsored by the Graduate Research Fellowship from the Ministry of Magic.

References

[1] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, *et al.*, "Mastering the game of go with deep neural networks and tree search," *nature*, vol. 529, no. 7587, p. 484, 2016.

[2] D. A. Ferrucci, "Introduction to "this is watson"," *IBM Journal of Research and Development*, vol. 56, no. 3.4, pp. 1–1, 2012.

[3] J. K. Rowling, *Harry Potter and the deathly hallows*, vol. 7. Bloomsbury Publishing, 2013.

[4] J. K. Rowling, *Harry Potter and the prisoner of Azkaban*. Bloomsbury publishing, 2015.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[6] M. Nostradamus, *Les propheties*. Fayard/Mille et une nuits, 1998.

[7] D. Edelstein, "This grad student studied parapsychology — and you won't believe what he found!," in *SIGBOVIK*, 2018.

[8] L. Lee, "Empirical evidence variational expert encodings: A better method," in *International Conference on Machine Learning*, pp. 100–1105, 2021.

[9] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks," pp. 1–9, 2014.

[10] J. K. Rowling, *Harry Potter and the sorcerer's stone*. Bloomsbury publishing, 2013.

[11] A. Ghorbani, A. Abid, and J. Zou, "Interpretation of Neural Networks is Fragile," no. Lipton 2016, 2017.

[12] "Chinese alchemical elixir poisoning," *Wikipedia*, accessed 03/01/2019.

[13] Y. LeCun, C. Cortes, and C. Burges, "Mnist handwritten digit database," *AT&T Labs [Online]. Available: http://yann. lecun. com/exdb/mnist*, vol. 2, p. 18, 2010.