Pruning Neural Networks Using Cooperative Game Theory

Extended Abstract

Mauricio Diaz-Ortiz Jr
Radboud University
Netherlands
mauricio.diaz-ortizjr@donders.ru.nl

Benjamin Kempinski
Radboud University
Netherlands
benjamin.kempinski@donders.ru.nl

Daphne Cornelisse
New York University
USA
cornelisse.daphne@nyu.edu

Yoram Bachrach
Google DeepMind
United Kingdome
yorambac@google.com

Tal Kachman
Radboud University
Netherlands
talkachman@cerebrnita.com

ABSTRACT

We introduce Game Theoretic Assisted Pruning (GTAP), a method that utlizes power indices from cooperative game theory to efficiently prune deep neural networks without compromising their predictive performance. GTAP identifies and removes less impactful neurons based on their contribution to the network's performance, streamlining the model's size and computational load. Our empirical evaluations show that GTAP outperforms traditional pruning techniques, achieving a better balance between model compactness and accuracy across multiple types of neural networks.

ACM Reference Format:
Mauricio Diaz-Ortiz Jr, Benjamin Kempinski, Daphne Cornelisse, Yoram Bachrach, and Tal Kachman. 2024. Pruning Neural Networks Using Cooperative Game Theory: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

Many of the successes in machine learning in the last two decades are attributed to deep neural networks (DNNs). However, this may require a large number of parameters, resulting in a high compute and resource cost. A network's performance is attained through the collective computation of all neurons, making cooprative game theory well-suited to reduce a networks complexity.

The most direct approach is to reduce the size of the network, in terms of either the number of neurons or the number of connections between them, referred to as *pruning* [2]. Several pruning methodologies exist such as stochastic regularization [12, 16] and co-adaptation [11]. In some cases, there may even exist subnetworks whose prediction accuracy *exceeds* that of the original full network, sometimes referred to as "winning lottery tickets" [10]. However, providing methods for finding such winning tickets is a significant algorithmic challenge [6, 18].

Rather than using heuristics, we seek high performing subnetworks (winning tickets) by viewing the neurons of a network as agents playing a cooperative game, where the neurons work together to maximize network performance [17, 22]. This enables us to use game theoretic tools that measure each participant's contribution to the collective goal [1, 3, 4, 7, 20], with the added benefit of interpretability. These cooperative game principles are model agnostic and thus can be reformulated in other learning problems.

Our Contribution: We propose Game Theory Assisted Pruning (GTAP), a method for pruning neural networks based on cooperative game theory. We define a cooperative game, which views the individual neurons of the trained network as agents working in teams, aiming to produce a highly accurate predictive model.

Under this view, the value of every subset (coalition) of neurons is the quality of the prediction using a network that uses solely these neurons (with all other activations masked out). We then estimate the relative impact of each neuron on the performance of the entire network using solution concepts from cooperative game theory. Given the relative impact estimate of each neuron we construct the sub-network by retaining only the high impact neurons, or adding neurons gradually in decreasing order of impact.

To determine the relative impact of neurons, we use power indices [5] — existing game theoretic solutions designed to estimate the impact that individual members of a team have on the overall team performance, such as the Shapley value [20] and the Banzhaf index [1]. We also propose a parameterized version of the Banzhaf index, called β_d, where d is a parameter reflecting the predicted proportion of neurons required to get a good prediction. To select the parameter d, we apply an *uncertainty estimation process*, akin to the Dropout procedure [21] commonly used to reduce model overfitting in machine learning. Our uncertainty estimation procedure considers randomly eliminating neurons in the trained model, and attempts to characterize the network size where we transition from being relatively certain about making a good prediction to being uncertain about our ability to have a high performing model.

We empirically evaluate our framework by pruning several prominent neural network architectures. For image classification we consider the convolutional neural network LeNet5 [15] on MNIST [15], and for for natural language processing tasks, we consider a feedforward model on news topic classification [8] and emotion classification in social media texts [19]. We also consider the issue of scaling up to large neural networks, reporting results for the AlexNet [13] architecture on Tiny ImageNet [14]. We show that our game theory pruning methods can outperform existing pruning baselines.

For full algorithmic details, results and related work, we refer the reader to the full paper version [9].

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

Pruning Neurons Based on Power Indices. In pruning neural networks to enhance efficiency, we explore various methods centered around power index calculations. The simplest, non-iterative approach involves a one-time estimation of power indices to select the top r most powerful neurons for retention, known as Top-n Pruning. On the other hand, Iterated Pruning and Iterated Building represent more complex, iterative methods that adjust the network over several cycles. Iterated Pruning progressively removes the weakest neurons and recalculates power indices, aiming to reduce the network to r neurons by excluding the least powerful ones in each iteration. Conversely, Iterated Building starts with an empty network, gradually adding neurons based on their power indices, recalculating and including the most influential neurons in each step until the network reaches the desired size. While iterative methods promise greater precision by constantly refining neuron selection, they demand significantly more computational resources compared to the simpler single-estimation Top-n Pruning approach.

1 EMPIRICAL EVALUATION

We empirically evaluate the performance of GTAP and contrast it to multiple baselines [2, 10], by pruning both feedforward neural networks and convolutional neural networks. We carry our evaluation on two image classification datasets, MNIST [15] and Tiny ImageNet [14], and on two natural language processing datasets, one for topic classification [8] and one for social media text emotion classification [19]. For the full range of results we refer the reader to the full paper [9]. The neural architectures we prune are feedforward and convolutional neural networks. Networks were trained from random weight initialization, so as to perform a clean slate retraining of the networks to verify the winning ticket's success.

Experiments: We apply our GTAP method for neural network pruning using power indices (Shapley, Banzhaf, Biased Banzhaf) and compare it against traditional pruning baselines [10]. We conducted experiments on modified LeNet-300-100 and LeNet5 architectures, focusing on selective pruning of layers while fully discussing the implications of uncertainty bands for pruning decisions. This approach aims to optimize pruning efficiency by carefully selecting neurons based on their calculated power indices. For full discussion of the uncertinity bands and their implications see the full paper [9].

1.1 Game Theoretic Pruning

We examine the ability of GTAP to prune neural networks while retaining high accuracy. We show "compression curves", where the x-axis is the target size for the pruned network, and the y-axis is the accuracy of the pruned model. Better pruning methods have curves that are higher for a wide range of pruned network sizes.

1.1.1 LeNet5. Using the bias parameter d, that maximizes the uncertainty, we applied Top-n and Iterated Pruning for LeNet5 with MNIST, using d-biased and plain Banzhaf index and Shapely values. Figure 1 compares the performance under these indices.

Figure 1 indicates a significant improvement in the performance of our GTAP method over the baselines (for all ranges of pruned network size) showing that at least for these architectures and datasets, game theory can enable strong pruning methods. [1] We note that Top-n Pruning exhibits higher improvement over the baseline, indicating that a one-shot selection of the highest power indices does not fully capture the importance of neuron interactions.

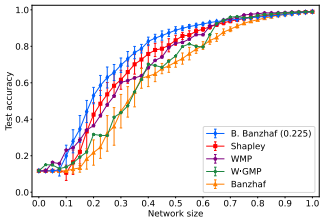

Figure 1: Comparing GTAPto baselines on LeNet5.

1.1.2 NLP Tasks. We show that GTAP also achieves good results in natural language processing.

We examine two text classification tasks: topic classification [8], identifying the news topic of a text, and emotion classification [19], pinpointing the specific emotion conveyed by the text. A simple neural network model using a binary term frequency vector for text representation and consisting of three layers with 256 neurons each is employed for classification. After training, we apply the GTAP pruning method to this model and assess its performance against weight-based pruning baselines. We include one example here, with the additional results in the full paper.

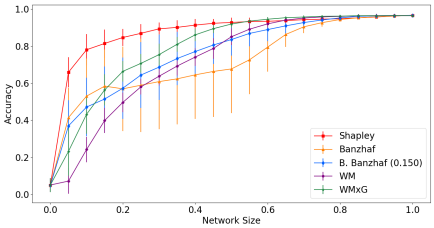

Figure 2: NLP tasks: Comparison of GTAP performance to baselines of Weight Magnitude Pruning (WMP) and Weight Gradient Magnitude Pruning ($W \cdot GMP$).

2 DISCUSSION AND CONCLUSIONS

We explored the impact of game theortic concepts in neural network pruning. We proposed GTAP, which utilizes cooperative game solution concepts such as the Shapley value and the Banzhaf index. We demonstrated GTAP's effectiveness across various datasets, including MNIST and Tiny ImageNet for image classification, and topic and emotion classification for NLP, showing its ability to significantly reduce neural network size and computational demands while maintaining robust predictive performance. Our findings reveal that GTAP not only surpasses traditional pruning benchmarks like Weight Magnitude Pruning and Weight Gradient Magnitude Pruning but also offers key benefits including elimination of the need for retraining, high parallelizability, and model agnosticism, making it applicable to a wide array of machine learning models beyond those reliant on weight magnitudes.

[1] Similar trends hold for LeNet-300-100.

REFERENCES

[1] J. F. Banzhaf III. Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review*, 19(2):317–344, 1965 1964.

[2] D. Blalock, J. J. Gonzalez Ortiz, J. Frankle, and J. Guttag. What is the state of neural network pruning? *Proceedings of machine learning and systems*, 2:129–146, 2020.

[3] R. Branzei, D. Dimitrov, and S. Tijs. *Models in cooperative game theory*, volume 556. Springer Science & Business Media, 2008.

[4] G. Chalkiadakis, E. Elkind, and M. Wooldridge. Computational aspects of cooperative game theory. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 5(6):1–168, 2011.

[5] G. Chalkiadakis, E. Elkind, and M. Wooldridge. Computational aspects of cooperative game theory. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 5(6):11–23, 2011.

[6] T. Chen, J. Frankle, S. Chang, S. Liu, Y. Zhang, Z. Wang, and M. Carbin. The lottery ticket hypothesis for pre-trained bert networks. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 15834–15846. Curran Associates, Inc., 2020.

[7] D. Cornelisse, T. Rood, Y. Bachrach, M. Malinowski, and T. Kachman. Neural payoff machines: Predicting fair and stable payoff allocations among team members. *Advances in Neural Information Processing Systems*, 35:25491–25503, 2022.

[8] R. Costa. Twitter financial news topic dataset, 2022. Available in https://huggingface.co/datasets/zeroshot/twitter-financial-news-topic.

[9] M. Diaz-Ortiz Jr, B. Kempinski, D. Cornelisse, Y. Bachrach, and T. Kachman. Using cooperative game theory to prune neural networks. *arXiv preprint arXiv:2311.10468*, 2023.

[10] J. Frankle and M. Carbin. The lottery ticket hypothesis: Training pruned neural networks. *CoRR*, abs/1803.03635, 2019.

[11] S. Han, J. Pool, J. Tran, and W. Dally. Learning both weights and connections for efficient neural network. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

[12] B. Hassibi, D. Stork, and G. Wolff. Optimal brain surgeon and general network pruning. In *IEEE International Conference on Neural Networks*, pages 293–299 vol.1, 1993.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[14] Y. Le and X. Yang. Tiny imagenet visual recognition challenge. *CS 231N*, 7(7):3, 2015.

[15] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[16] Y. LeCun, J. S. Denker, and S. A. Solla. Optimal brain damage. In *Advances in neural information processing systems*, pages 598–605, 1990.

[17] F. Leon. Optimizing neural network topology using shapley value. In *2014 18th International Conference on System Theory, Control and Computing (ICSTCC)*, pages 862–867. IEEE, 2014.

[18] E. Malach, G. Yehudai, S. Shalev-Schwartz, and O. Shamir. Proving the lottery ticket hypothesis: Pruning is all you need. In H. D. III and A. Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 6682–6691. PMLR, 13–18 Jul 2020.

[19] E. Saravia, H.-C. T. Liu, Y.-H. Huang, J. Wu, and Y.-S. Chen. CARER: Contextualized affect representations for emotion recognition. In *EMNLP*, 2018.

[20] L. S. Shapley. A value for n-person games, contributions to the theory of games, 2, 307–317, 1953.

[21] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[22] J. Stier, G. Gianini, M. Granitzer, and K. Ziegler. Analysing neural network topologies: a game theoretic approach. *Procedia Computer Science*, 126:234–243, 2018.