The Carbon footprint of Machine Learning Models

Aniruddha Das
Computer Science and Engineering Department
Haldia Institute of Technology
Haldia,India
das.aniruddha53@gmail.com

Avisikta Modak
Dietetics and Community nutrition management
Women's College Calcutta
Kolkata ,India
modak.avisikta@gmail.com

***Abstract*—Machine Learning models are growing increasingly powerful in their abilities, whether that might be in processing natural language, tackling the intricacies of computer vision or any other number of exciting application that are emerging . But the environmental impact of machine learning models is increasingly receiving attentions. Here ,the works to focus on the carbon footprint of language models, as these models grow larger and larger, do their corresponding carbon footprints, especially when it comes to creating and training complex models. Here we will take a look at some concrete example of carbon emissions from machine learning models, will present tools that can be used to estimate the carbon footprint of a machine learning models. Finally present ideas for how to reduce the carbon footprint.**

Keywords—machine learning models; carbon footprint;

I. INTRODUCTION

The environmental impact of machine learning models are increasingly receiving attention, mostly from academia. Modern AI models consume a massive amount of energy and these energy requirements are growing at a breathtaking rate. In deep learning era, the computational resources needed to produce a best in class.AI model has on average doubled every 3.4 months[21] and has a meaningful carbon footprint today, and if industry trends continue it will soon became much worse.

II. THE ENVIRNMENTAL IMPACT OF MACHINE LEARNING

All software and the apps that run on our phones to the data science pipelines that run in the cloud – consume electricity and as long as not all our electricity is generated by renewable energy source, electricity consumption will have a carbon footprint. This is why machine learning models can have a carbon footprint."Carbon footprint" refer to the amount of CO2e emission, where "e" is short for "equivalents". Since other gas such as methane, nitrous oxide or even water vapor also have a warming effect, a standardized measure for describing how much warming a given amount of gas will have is often provided in CO2e for simplification purpose. However

electricity consumption from data centers reported in 2015 and 2016 ranged from 3-5% of global electricity consumption [2][3].More recently, some claim data centers account for 1% of global electricity use[4][5].If we take a look at the energy consumption from machine learning at the organization level, Google says that 15% of the company's total energy consumption went towards machine learning related computing across research, development and production[21].NVIDIA has estimated that 80 – 90% of machine learning workload is inference processing[7].Similarly, Amazon Web Services have stated that 90% of the machine learning demand in the cloud is for inferences[8].

Let now take a look at some concrete example of the carbon footprint of machine learning models.

TABLE 1 : Energy consumption of 7 large deep learning models , adopted from [6] and [11]

Model	Energy Consumption, MWh	CO2e emissions,tons
Evolved Transformer	7.5	3.2
T5	85.7	46.7
Meena	232	96.4
Gshard 600B	24.1	4.8
Switch Transformer	179	72.2
GPT-3	1,287	552.1
PaLM	3,181	271

Looking at few example of the carbon footprint of running inference with language model, Facebook has estimated that the carbon footprint of their Transformer-based universal language model for text translation is dominated by the inference phase, using much higher inferences(65%) as compared to training(35%)[20].The average carbon footprint for ML training tasks at Facebook is 1.8 times larger than that of Meena used in modern conversational agent and 0.3 times of GPT-3's carbon footprint.

III. ESTIMATE THE CARBON FOOTPRINT OF A MACHINE LEARNING

Formulae for computing carbon footprint

Carbon footprint = E*C (1)

Where E is the number of electricity units consumed during some computational procedure. This can be quantified as kilowatt hours(kWh). C is the amount of CO_2e emitted from producing one of said unit of electricity.

IV . OVERVIEW OF ENERGY AND CO2 EQUIVALENT EMMISIONS FOR ML TRAINING [6]

We estimate energy and carbon footprint using the following terms:

1) Metric tons are the common CO_2e unit measure, abbreviated tCO_2e, representing1000kg

*2)*Mega watt hours measure energy;1 Mwh equals 1 millions W of electricity used continuously for 1h.One terawatt hour equals 1 million Mwh.

3) Power usage effectiveness(PUE) is the industry standard metric of data center efficiency, defined as the ration between total energy use divided by the energy directly consumed by a datacenter's computing equipment. The average industry data centre PUE in 2020 was 1.58, while cloud provides had PUE of ~1.1[5]

The energy consumption of the servers performing a training task is proportional to the number of processors used and the duration of the training run

MWh = hours to train * number of processors * average power per processor (2)

We include all server components in "processors" (including local memory, network links and so on), additionally data center consumes energy to power and cool hardware (for example, voltage transformations losses and cooling equipment),which is capturing by the PUE. Thus, the final formula for energy consumption :

MWh = (hours to train * number of processors * average power per processor) * PUE (3)

We can then turn energy into carbon by multiplying it with the carbon intensity of the energy supply

tCO_2e = MWh * tCOe per Mwh (4)

V. HOW TO REDUCE THE CARBON FOOTPRINT OF MACHINE LEARNING MODELS

1) Select efficient ML Model

Selecting computationally efficient ML model architectures while advancing ML quality, such as sparse model verses dense models , can reduce computation.

2) Don'nt train a model from scratch

The majority of researchers don't train from scratch, training a heavy model from scratch can burn up a lot of CPU/GPU power. Hence using a pre trained model can reduce the carbon footprint significantly as less electricity is necessary. Azure offers various sophisticated pre-trained models for vision, speech, language & search. These are already trained on massive dataset using many compute hours training the model, the end-user only has to transfer the knowledge of these models to their own dataset

3) Use of federated learning(FL)

FL methods can be advantageous in the area of carbon footprint. For example, a study conducted by experts at the University of Cambridge that FL has an advantage because of the cooling requirements of data centers. Though GPU and TPUs are becoming more efficient in terms of processing powers given per unit of energy spent, the requirement for a powerful and energy consuming cooling system persists. So the FL can always benefited from hardware developments.

4) Train models where the energy is cleaner

You can choose which region to run your computational procedure. Research has shown that emission can be reduced by up to 30X just by running experiments in regions powered by more renewable energy source [21]. Table below shows how many kg CO_2e are emitted by using 100 hours of compute on an A100 GPU on the Google Cloud platform in various regions. It is clear that the carbon intensity of electricity varies greatly between regions

TABLE 2 : kg CO2 eq by region emitted by 100 hours of compute on A100 PCle 40/80 GB GPU in a Google Cloud data center[19]

Region	Kg CO2eq
Asia- east1	14
Asia-east2	17.5
Asia-northeast1	13
Asia-northeast2	13
Asia-south1	23
Asia-southeast1	10.5
Australia-south east1	20
Europe – north 1	5.25
Europe – west 1	6.75

Europe-west 2	15.5
Europe-west 3	15.25
Europe-west 4	14.25
Europe-west 6	0.5
Northamerica-northeast1	0.75
Southamerica-east1	5
US-central 1	14.25
US-east1	9.25
US-east4	9.25
US –west1	7.5

5) *Train models when the energy is cleaner*

The carbon intensity of electricity can vary from day to day and even from hour to hour as illustrated by Fig1. Which show the average carbon footprint in g/kWh for each hour of the day and each day of the week respectively in eastern Denmark in the period January1 2022 – October 7 , 2022. This data shows that the electricity is cleaner around noon than at night.

We can reduce the carbon footprint of our work by scheduling heavy workloads to those periods when the energy is cleaner. If you are not in an extreme hurry to train a new model, a simple idea is to start your model training when the carbon intensity of electricity in your cloud region is below a certain threshold, and put the training on hold when the carbon intensity is above some threshold.

Fig . 1. Avg carbon intensity of electricity in eastern Denmark per hour of the day

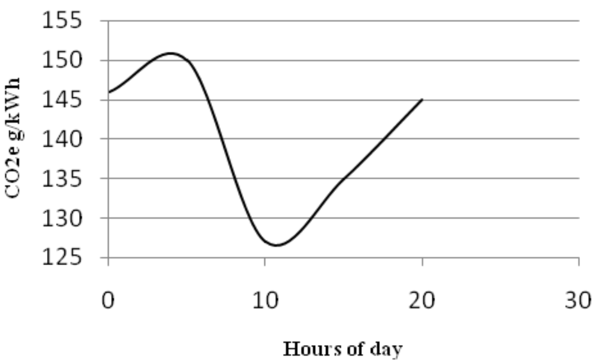
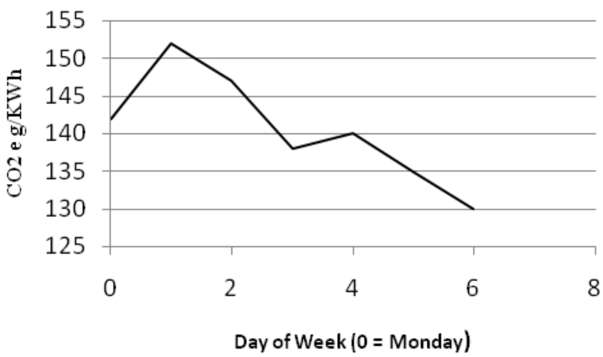


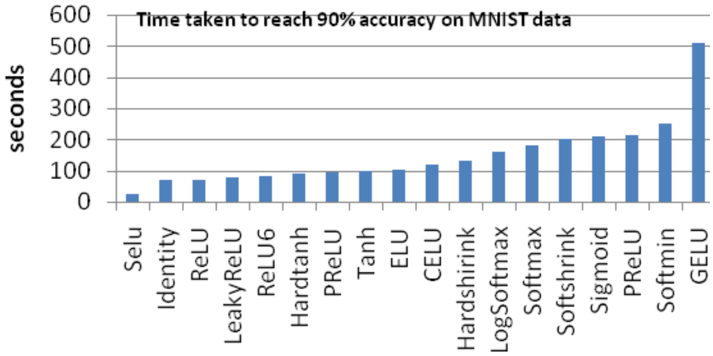
Fig. 2. Avg carbon intensity of electricity in eastern Denmark per day of the week



6) *Efficient activation function*

The selection of an activation function can greatly influence the time that your model takes to train. As seen in Fig. below, Dercynski demonstrated that the time it took to train an image classification model on the MNIST dataset to 90% accuracy varied from a few seconds to more than 500 seconds. Aside from demonstrating that the choice of activation function influences training time. Dercynski also found that

- Activation function choice appears to have more effect in situations where inference is performed over smaller sets at a time
- Applications should be analyzed and tuned on the target hardware if one is to avoid particularly costly activation function



VI. CONCLUSION

ML workloads have rapidly grown in importance, raising legitimate concerns about their energy use. This paper determine the environmental impact of ML. This article is an attempt to address these issues and is written for practitioners and researchers alike who do hands on machine learning. The article also shows few ideas for how we might reduce the carbon footprint of machine learning models. Another perspective is that some consider the carbon footprint to be erased entirely if a cloud provider matches 100% of its energy consumptions with renewable energy as Google and Facebook have done.

Acknowledgment

I had a great deal of help from others along the way for an earlier study[16] that eventually led to this version of the article. Kasper Groes , Albin Ludvigsen shared information about GPT. They did a great deal of work to measure the performance and power of GPUs and TPUs in Google data centers. Kimberley Mok helped with the data and policies for energy and CO2e at Google. Arindam Giri provided helpful suggestion on how to write better technical content.

References

[1] N.C Thompson,K. Greenewald,K.Lee, and G.F.Manso,"Deep Learning's diminishing returns: The cost of improvement is becoming unsustainable," EIEEE Spectr.,vol.58,no.10,pp.50-55,2021,doi:10.1109/MSPEC.2021.9563954.

[2] E.Strubell,A.Ganesh,and A.McCallum,"Energy and policy consideration for deep learning in NLP," in Proc.Annu.Meeting Assoc.Comput. Linguistics,2019,pp.13,693-13,696.

[3] J.Koomey and E,Masanet,"Does not compute:Avoiding pitfalls assessing the Internet's energy and carbon impacts," Joul,vol.5,no. 7,pp.1625-1628,2021,doi: 10.1016/j.joule.2021.05.007.

[4] E.Masanet,A.Shehabinuoa,L.Smith,and J.Koomey,"Recalibrating global datacenter energy-use estimates,"Science,vol.367,no 6481,2020,doi:10.1126/science.aba3758.

[5] D.Patterson et al.,"Carbon emissions and large neural network training." 2021,arxiv:2104.10350

[6] N.C Thonpson, K. Greenewald, K. Lee , and G.F.Manso,"The computational limits of deep learning." 2020, arxiv: 2007.05558.

[7] C.Freitag,M. Berners-Lee,K.Widdicks,B.Knowles,G.S.Blair,and A.Friday,"The real climate and transformative impact of ICT: A critique of estimates,trends ,and regulations," Patterns,vol.2,no.9,p.100,340,2021,doi:10.1016/j.patter.20121.100340.

[8] D.R.So, C.Liang and Q.V.Le, "The evolved transformer," in Proc.Int.Conf.Mach.Learn,pp.6010-6022.

[9] D.R.So, W.Marike, H.Liu, Z. Dai, N.Shazeer, and Q.V.Le,"Primer: Searching for efficient transformers for language modeling," in Proc.Conf.Neural Inf.Process.Syst.,2021.

[10] N. Jouppi et al., " Ten lessons from three generation shaped Google's TPUv4i," in Proc.Int.Symp.Comput.Arch.,2021,pp.1-14.

[11] T.B. Brown et al ., "Language models are few-shot learners," In Proc.Conf.Neural Inf.Process.Syst.,pp.1877-1901,2020.

[12] N.Du et al., "Glam: Efficient scaling of language models with mixture-of experts," 2021,arxiv:2112.06905.

[13] B. Evans,"Amazons shocker: CEO Jassy says cloud less than 5% of all IT spending," Cloudwars,2021. https://cloudwars.co/amazon/anazon-shocker-ceo-jassy-cloud-less-than -5-percent-it-spending/

[14] S.Schechner,"Amazon and other tech giants race to buy up renewable energy." The Wall Street Journal,Jun.23,2021.[Online]. Available:https://www.wsj.com/article/amazon-and-other-tech-giants-race-to-buy-up-renewable-energy-11624438894

[15] S. Li et al.,"Searching for fast model families on datacenter accelerators," in Proc. Conf. Comput. Vis. Pattern Recognit., 2021.

[16] D.Patterson et al., "Energy analysis of smartphones and machine learning's role," submitted for publication

[17] P. Henderson, J.Hu, J.Romoff,E.Brunskill,D.Jurafsky, and J.Pineau, "Towards the systematic reporting of the energy and carbon footprints of machine learning," J.March.Learn.Res.,pp.8085-8095,2020.

[18] R.Schwartz,J.Dodge,N.A.Smith,and O.Etzioni,"Green AI," Commun. ACM,vol.63,no.12,pp.54-63,2020,doi:10.1145/3381831.

[19] A.Lacoste,A.Luccioni,V.Schmidt,and T.Dandres,"Quantifying the carbon emissions of machine learning,"2019 ,arxiv:1910.09700

[20] E.M.Bender, T.Gebru,A.Mcmillan-Majore,and S.Shmitchell,"On the dangers of stochastic parrots: Can language models be too big?" in Proc.ACM Conf.Fairness,Accountability,Transparency,2021, pp.610-623,doi:10.1145/3442188.3445922.

[21] D.Patterson,J. Gonzalez, U. Hilzle, Quoc Le, Chen Liang , D. Royhchild, D.R So, M.Texier and Jeff Dean ,"The Carbon Footprint of Maching Learning Training Wiil Plateau, Then Shirnk" doi: 10.1109/MC.2022.3148714