Asynchronous training of deep learning models with MCMC algorithms

Paul Gay

GreenAI U.P.P.A.

03-28-2022

The computer bounce effect

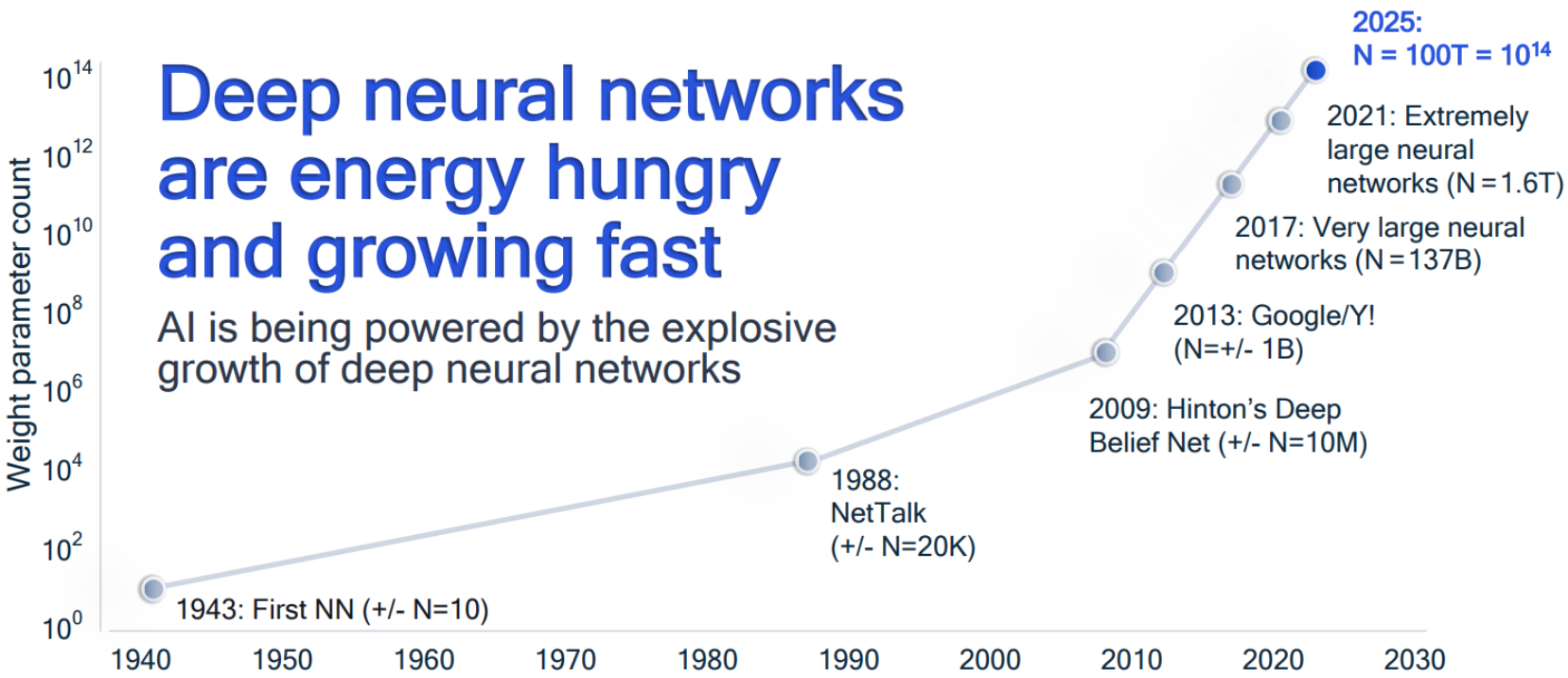

Figure 2: Exponential growth of Deep Learning models [Fournarakis, 2021]

Team GreenAI UPPA

- Team recently created on
  - AI powered Environmental applications
    - AgroEcology, Computer vision for Ecology, social computing and crisis management, Education
    - AI on the edge
  - Environmental Impact of AI
    - Measure of impact : **Coca4ai**
    - New light algorithms (on the edge)


Sébastien Loustau


Paul Gay


Fatou Kiné Sow


Matthieu François


Nicolas Tirel


Simon Lebeaud

Coca4ai Motivation

We plan to build multi-scale models to measure carbone footprint and evaluate the compromise "environmental impact"/"algorithm accuracy" for deep learning algorithms at the scale of a (small) data center

Coca4ai Motivation

We plan to build multi-scale models to measure carbone footprint and evaluate the compromise "environmental impact"/"algorithm accuracy" for deep learning algorithms at the scale of a (small) data center

- **Coca4ai:** one year project funded by RFSI.

The situation

Many joules wasted in data scientists practices [Khan et al., 2019]

- Job crashing
- Brute force optimization to earn a few percents
- Hidden knobs and bad use of the GPUs

The situation

Many joules wasted in data scientists practices [Khan et al., 2019]

- Job crashing
- Brute force optimization to earn a few percents
- Hidden knobs and bad use of the GPUs

A few emerging tools for personal use [Lannelongue et al., 2021, Anthony et al., 2020]

```python
from codecarbon import EmissionsTracker

tracker = EmissionsTracker()
tracker.start()
# GPU Intensive code goes here
tracker.stop()
```

Figure 4: Based on RAPL and Nvidia-Smi

Our cluster

- 12 GPUs equipped nodes dedicated to AI
- Slurm based job management
- over a year :
  - 99 users spread over 5 laboratories : LTCI, LISN, SAMOVAR, L2S, SATIE
  - 20K jobs
- Software tools : Zabbix (network), Idrac (gross power estimation), Graphana

Objectives

- Data collection : Statistics per job and per user
  - Power Consumption (omegawatt + RAPL and Nvidia)
  - Type of models, activity
- Identifying behaviors
- Insights for small and large scale datacenters (Jean-Zay)

Spread Good practices

- Avoid GPU bottle necks
- Normalise your loss to avoid cuda runtime error
- Normalise your layers also
- Most of the time spent by building auxiliary code (evaluation metrics, data formatting, rect or square inference)
  - Training is slow : multiple days
  - You obtain most of the clues with small experiments and unitary tests.

Engagement

- Provide feedbacks to endusers
- Check how receptive are people (which are under pressure to publish)
- A playground to create dialogue (about rebounce effect).

In summary

- 12 machines equiped with GPU
- Consumption of jobs and users
- PowerMeter and Software Measurement tools
- Sensibilise end users (here, data scientists)

References I

Thanks for your attention
Questions ?

[Anthony et al., 2020] Anthony, L., Kanding, B., and Selvan, R. (2020).
Carbontracker: Tracking and predicting the carbon footprint of training deep learning models.
page arXiv preprint https://arxiv.org/abs/2007.03051.

[Fournarakis, 2021] Fournarakis, M. (2021).
A practical guide to neural network quantization.

References II

[Khan et al., 2019] Khan, K., Scepanovic, S., Niemi, T., Nurminen, J., Von Alfthan, S., and Lehto, O. (2019).
Analyzing the power consumption behavior of a large scale data center.
SICS Software-Intensive Cyber-Physical Systems, 34:61–70.

[Lannelongue et al., 2021] Lannelongue, L., Grealey, J., and Inouye, M. (2021).
Green algorithms: Quantifying the carbon emissions of computation.
Advanced science.