IrEne-viz: Visualizing Energy Consumption of Transformer Models

Yash Kumar Lal, Qingqing Cao, Harsh Trivedi, Reetu Singh, Aruna Balasubramanian, Niranjan Balasubramanian
Stony Brook University
{ylal, qicao, hjtrivedi, reesingh, arunab, niranjan}@cs.stonybrook.edu

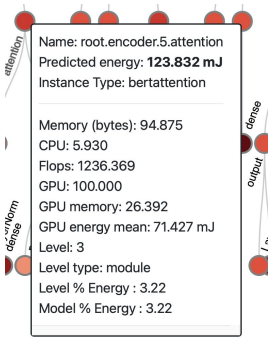
Overview

- Software-based energy measurements are inaccurate (Cao et al, 2020)
- IrEne-viz presents an interactive demonstration of energy consumption of different models and their components
- Easy to use interface for fine-grained resource usage and energy analysis of transformer models
- Extensible to new transformer models

Knowing energy usage can help

- Prioritize R&D of models and save costs
- Target bottlenecks inside a model architecture to increase efficiency
- Tailor models for a specific use case e.g., battery-powered mobile devices

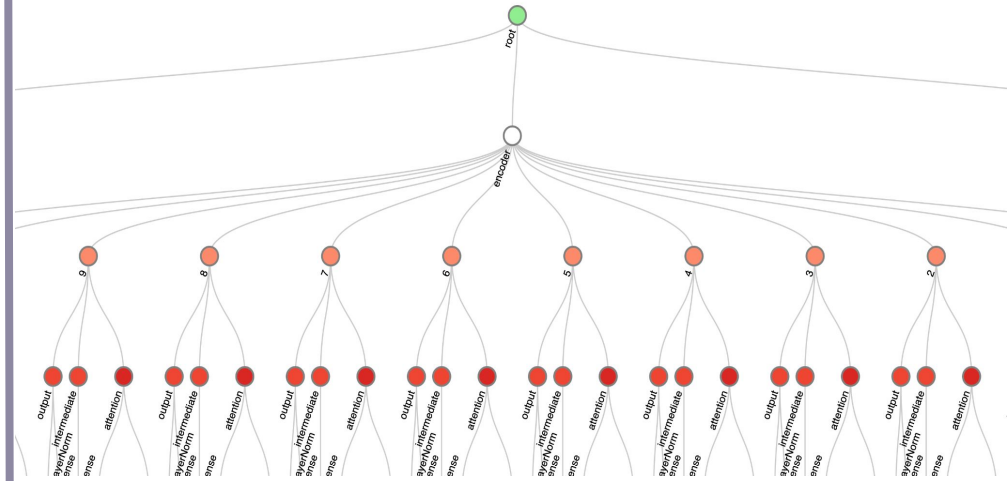
Fine-grained information



Node Energies

Node Name	Pred. Energy (mJ)
encoder	4244.749
encoder.0	323.662
encoder.1	323.662
encoder.2	323.662
encoder.3	323.662
encoder.4	323.662

Interpretable energy of a model



Check out the Demo!

Or, come find us

stonybrooknlp.github.io/irene/demo