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

IrEne-viz can help
➢ Identify specific bottlenecks in a model in order to increase energy efficiency
➢ Tailor models for a specific use case e.g., battery-powered mobile devices

Energy Information
Name: root.encoder.&.attention
Predicted energy: 189.326 mJ
Instance Type: robertaattention

Memory (bytes): 189.750
CPU: 5.862
Flops: 2472.739
GPU: 94.000
GPU memory: 19.846
GPU energy mean: 125.260 mJ
Level: 3
Level type: module
Level % Energy: 127.72
Model % Energy: 3.64

Identify Bottlenecks

<table>
<thead>
<tr>
<th>Node Name</th>
<th>Pred. Energy (mJ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>transformer</td>
<td>2015.527</td>
</tr>
<tr>
<td>transformer3</td>
<td>307.248</td>
</tr>
<tr>
<td>transformer0</td>
<td>307.247</td>
</tr>
<tr>
<td>transformer1</td>
<td>307.247</td>
</tr>
<tr>
<td>transformer2</td>
<td>307.247</td>
</tr>
<tr>
<td>transformer3</td>
<td>307.247</td>
</tr>
</tbody>
</table>

Check out the Demo!
Or, come find us
stonybrooknlp.github.io/irene/demo

References