PLACETO: LEARNING GENERALIZABLE DEVICE PLACEMENT ALGORITHMS FOR DISTRIBUTED MACHINE LEARNING

Ravichandra Addanki, Shaileshh Bojja Venkatakrishnan, Shreyan Gupta, Hongzi Mao, Mohammad Alizadeh

Presented by: Obodoekwe Nnaemeka
Problem

Distributed training (GPU and CPU)

Human experts?

Reinforcement learning?
Problem

Sometimes tolerable. Solutions do not generalize

The optimization is done for a single graph. Single computational graph vs Class of computational graph
Placeto

**Efficiency**

Sequence of iterative placement improvements

**Generalizability**

NN architecture that uses graph embedding to encode the computation of graph structure in the placement policy.
Learning method

- Markov Decision Process
POLICY NETWORK ARCHITECTURE
GRAPH EMBEDDING
Training Details

Colocation

Simulator
Experimentation

Deep learning models (Inception-V3, NASNet, NMT)

Synthetic data (cifar10, ptb, nmt)

Single GPU, Scotch, Human Expert, RNN based approach.
Result

- Performance
- Generalizability
## PLACETO VS RNN

<table>
<thead>
<tr>
<th>Model</th>
<th>Placement runtime (sec)</th>
<th>Training time (# placements sampled)</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CPU only</td>
<td>Single GPU</td>
<td>#GPUs</td>
</tr>
<tr>
<td>Inception-V3</td>
<td>12.54</td>
<td>1.56</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>1.15</td>
</tr>
<tr>
<td>NMT</td>
<td>33.5</td>
<td>OOM</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>OOM</td>
</tr>
<tr>
<td>NASNet</td>
<td>37.5</td>
<td>1.28</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>0.84</td>
</tr>
</tbody>
</table>
GENERALIZABILITY
GENERALIZABILITY
Place deep dive

Node traversal order

Alternative architectures

Simple aggregator

Simple partitioner
Critic

+ First attempt to generalize device placement using a graph embedding network

+ Really Impressive performance

- Only optimizes placement decisions

- It shows generalization to unseen graphs, but they are generated artificially by architecture search for a single learning task and dataset.

How does the framework handle failure. Evaluation protocol needs to be more explicit.