Parallelized Training of Deep NN

Comparison of Current Concepts and Frameworks

Sebastian Jäger, Hans-Peter Zorn, Stefan Igel, Christian Zirpins

Rennes, Dec 10, 2018
Motivation

- Need to scale the training of neural networks horizontally
- Kubernetes based technology stack
- Scalability of concepts and frameworks
Distributed Training Methods

Data Parallelism
Data Parallelism

Centralized Parameter Server

TensorFlow: https://www.tensorflow.org
Data Parallelism

Decentralized Parameter Server
Experimental Setup

Environment

- Google Kubernetes Engine
- CPU: 2.6 GHz
- Ubuntu 16.04
- TensorFlow 1.8.0
- MXNet 1.3.0
## Experimental Setup

### Networks

<table>
<thead>
<tr>
<th>Convolutional NN</th>
<th>Recurrent NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>LeNet-5</td>
<td>LSTM</td>
</tr>
<tr>
<td>5 layer</td>
<td>2 layer</td>
</tr>
<tr>
<td>10 classes</td>
<td>200 units</td>
</tr>
<tr>
<td>Fashion MNIST</td>
<td>Penn Tree Bank</td>
</tr>
<tr>
<td>28x28 gray-scale</td>
<td>1.000.000 words</td>
</tr>
</tbody>
</table>
Experimental Setup

Metrics

\[ \text{throughput}_n = \frac{\text{no. examples} \times \text{epochs} \times \text{no. workers}}{\text{training time}_n} \]

\[ \text{speedup}_n = \frac{\text{throughput}_n}{\text{throughput}_1} \]
Results

Convolutional Neural Network
Results

Convolutional Neural Network
Results

Recurrent Neural Network
Summarizing the Experiments

Decentralized Parameter Server ...
 › more robust regarding increasing communication effort
 › scales better for small NN

For bigger/ more complex NN ...
 › no significant difference between concepts
Conclusion

MXNet ...

› for small NN better scalability and throughput
› for bigger NN higher throughput
› less and less complicated code
› easier to scale up training
Thank you

Sebastian Jäger
@se_jaeger

inovex GmbH
Ludwig-Erhard-Allee 6
76131 Karlsruhe

sebastian.jaeger@inovex.de