Investigating scalability of recurrent network using dynamic batching in PyTorch

Devin Taylor
November 27, 2018

Computer Laboratory, University of Cambridge, United Kingdom
What is dynamic batching?

(a) Initial computational graph

(b) Batched computational graph

**Figure 1**: Dynamic batching for a single parse tree [1]
TensorFlow Fold

- Simplified API for adding dynamic batching to TensorFlow
- Last commit on 31 October 2017 (but not deprecated) - TensorFlow Eager prioritised [2]
- Poor evaluation when compared to TensorFlow
  - Very little insight into whether dynamic batching is actually useful
  - Only evaluated on binary trees
  - All trees had the same shape and size
  - “Best case scenario”
  - Inference timing results excluded time to construct static computation graph
Dynamic batching in PyTorch

- Want to evaluate if concept of dynamic batching is more efficient
- PyTorch dynamic computation graphs support direct batching of variable inputs
- Can test on real data
- Reconduct experiments from *Looks et al. (2017)* for PyTorch [3]
- Implementation already exists - TorchFold [4]
  - Last commit on 7 July 2018
  - No support for PyTorch 0.4+
Evaluation

• Sentiment classification with TreeLSTM network [5]
  • Direct batching
  • Dynamic batching

• Measure inference time for variable batch sizes

• Compare to results obtained using TensorFlow Fold

• Investigate implementing in additional frameworks for further comparisons
  • TensorFlow Eager
  • Knet (Julia)
<table>
<thead>
<tr>
<th>Start date</th>
<th>End date</th>
<th>Days</th>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>21 Nov</td>
<td>23 Nov</td>
<td>3</td>
<td>Pre-reading</td>
</tr>
<tr>
<td>26 Nov</td>
<td>29 Nov</td>
<td>4</td>
<td>Rebuild experiment from <em>Looks et al. (2017)</em> in PyTorch</td>
</tr>
<tr>
<td>4 Dec</td>
<td>7 Dec</td>
<td>4</td>
<td>Rewrite TorchFold for PyTorch 0.4+ and rerun experiment</td>
</tr>
<tr>
<td>10 Dec</td>
<td>14 Dec</td>
<td>5</td>
<td>Investigate implementations in other frameworks</td>
</tr>
<tr>
<td>17 Dec</td>
<td>20 Dec</td>
<td>4</td>
<td>Gather results and write report</td>
</tr>
</tbody>
</table>
Announcing tensorflow fold: Deep learning with dynamic computation graphs.

Eager execution: An imperative, define-by-run interface to tensorflow.

Moshe Looks, Marcello Herreshoff, DeLesley Hutchins, and Peter Norvig.
Deep learning with dynamic computation graphs.
torchfold.
