

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?

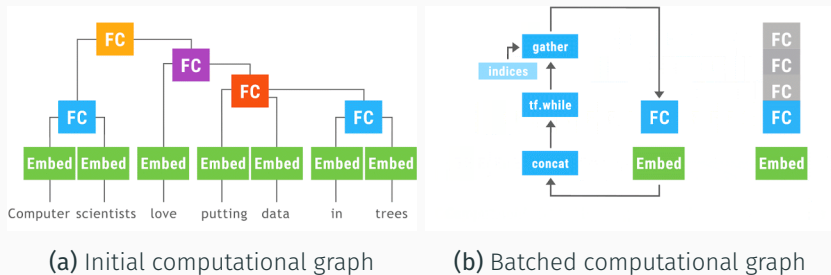


Figure 1: Dynamic batching for a single parse tree [1]

- 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+

- 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)

Workplan

Start date	End date	Days	Task
21 Nov	23 Nov	3	Pre-reading
26 Nov	29 Nov	4	Rebuild experiment from <i>Looks et al. (2017)</i> in PyTorch
4 Dec	7 Dec	4	Rewrite TorchFold for PyTorch 0.4+ and rerun experiment
10 Dec	14 Dec	5	Investigate implementations in other frameworks
17 Dec	20 Dec	4	Gather results and write report

References i



Announcing tensorflow fold: Deep learning with dynamic computation graphs.

<https://ai.googleblog.com/2017/02/announcing-tensorflow-fold-deep.html>, Feb 2017.



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

<https://research.googleblog.com/2017/10/eager-execution-imperative-define-by.html>, Oct 2017.



Moshe Looks, Marcello Herreshoff, DeLesley Hutchins, and Peter Norvig.

Deep learning with dynamic computation graphs.

arXiv preprint arXiv:1702.02181, 2017.



torchfold.

<https://github.com/nearai/torchfold>, Sep 2017.



Kai Sheng Tai, Richard Socher, and Christopher D Manning.

Improved semantic representations from tree-structured long short-term memory networks.

arXiv preprint arXiv:1503.00075, 2015.