MSc Project Progress Report: Machine Learning Software Frameworks

Abhijit Singh (s1788323)

July 3, 2018

1 Goal of the project

This project aims to compare some of the popular Machine Learning frameworks. This is being done by implementing a recent language modeling paper, which introduces a new model called "Mixture of Softmaxes" (MOS). The metrics which are being used to compare the different frameworks are:

- Speed of processing on single GPU
- Scalability to multiple GPUs
- Hardware utilization
- Ease of use of the framework
- Perplexity of the neural language model

As this is partially a group project, each member is using one framework. I am using Keras with a Tensorflow backend. Towards the end of the project, all of us will share our results so that we can compare all frameworks.

2 Methods

We begin by implementing two simple baseline models, a simple single layer LSTM model and a slightly more complex two layer LSTM with dropout applied between the Input to LSTM layer connections, between the two LSTM layers, and LSTM to Output layer connections.

The dataset being used is the Penn Tree Bank (PTB) data. We try to push the limits of the simple baseline models, and see how well they perform on the PTB data (in terms of perplexity). Once this is completed, we implement the MOS model. While doing so, we also subjectively document how difficult it is to write new functions in the framework, as this MOS is a novel concept and hence is not available as a pre-defined function in any framework yet. Finally, we compare the performance of the different frameworks based on the criteria mentioned above.

3 Work accomplished

Until now, I have implemented the baseline models and run several experiments with them. The primary aim of these experiments was to see how sensitive they were to different hyper-parameters of the model, and what effect did having an effective learning rate schedule make to them. Some key results are mentioned below (these are averages of several runs) :

Table 1: Perplexity scores of some key models

Model	Perplexity
1 layer LSTM	141.17
2 layer LSTM (w/o dropout)	158.38
2 layer LSTM (with dropout)	134.15
1 layer LSTM with learning rate schedule	139.07
2 layer LSTM with learning rate schedule	129.02
2 layer LSTM with best hyper-parameters	113.07

4 Work remaining

Currently I am building the MOS model. As soon as that is done, I will begin evaluating that model. Them I will proceed to analyzing the performance of the framework, and write the dissertation.