PYTORCH

An Overview for Text Analytics

DR. SRIJITH RAJAMOHAN

OUTLINE

What is PyTorch?

Comparison to TensorFlow

Basics of PyTorch with CIFAR classifier

Porting to GPU

Using PyTorch for Text Prediction

WHAT IS PYTORCH?

- Evolved from the popular Torch package
- Developed by industry stalwarts like Facebook
- Fastest growing Deep Learning framework
- Designed to be similar to the PyData stack

COMPARISON TO TENSORFLOW

TensorFlow

- Static graphs make it faster and scales well
- However, steeper learning curve
- Can be harder to debug
- Relatively rich functionality

PyTorch

- Easier to debug
- Eager execution can make PyTorch slower
- Functionality and third party libraries for PyTorch are still being developed

CLASSIFICATION ON THE CIFAR DATA

- See https://srijithr.gitlab.io/post/cs4984/ and section Neural Network for Image Classification on the CIFAR10 dataset using PyTorch'
- CIFAR10 has 60,000 32x32 images in 10 classes, 6000 images per class
- 50,000 training images and 10,000 test images

WORD EMBEDDINGS FOR TEXT PREDICTION

- We will do text prediction using Word Embeddings
- Requires much less data and faster than using RNN/LSTM/GRU
- See https://srijithr.gitlab.io/post/word2vec/ for examples on using the Continuous Bag of Words and Skipgram models

WORD EMBEDDINGS FOR TEXT PREDICTION

- Numerical representations of text
- Use n-dimensional vectors
- Words with similar meanings have word vectors that have low distance from each other
- Trained using a Continuous Bag of Words or Skipgram algorithm

CONTINUOUS BAG OF WORDS

Picture courtesy of Wikipedia