Technology Review: An Introduction of TensorFlow, its implementation and comparison with Pytorch

Jatin Saxena

University of Illinois Urban Champaign

CS 410 Text Information Systems

November 6, 2021

Introduction: Machine learning is a complex task and implementing machine learning models is more complex without the use of any framework. Google TensorFlow framework implementing machine learning models is far less daunting and difficult than it used to be. It eases the process of acquiring data, training models, serving predictions, and refining future results.

TensorFlow allows developers to create dataflow graphs—structures that describe how data moves through a graph, or a series of processing nodes. Each node in the graph represents a mathematical operation, and each connection or edge between nodes is a multidimensional data array, or tensor.
Nodes and tensors in TensorFlow are Python objects, and TensorFlow applications are themselves Python applications.

The actual math operations, however, are not performed in Python. The libraries of transformations that are available through TensorFlow are written as high-performance C++ binaries. Python just directs traffic between the pieces, and provides high-level programming abstractions to hook them together.

There are multiple spaces where Google TensorFlow is getting used but we will mostly focus on the usage of TensorFlow in the text processing space.

Implementation of TensorFlow:

TensorFlow is getting used extensively in text based applications such as language detection, sentimental analysis, fraud detection and threat detection.

Below are few of the specific use cases of TensorFlow in applications:

- **Text summarization**: Google found out that for shorter texts, summarization can be learned with a technique called sequence-to-sequence learning. This can be used to summarize the text and can be used in generating headlines from new articles.

To get an idea of what the model produces, you can take a look at some examples below provided on Google blog. The first column shows the first sentence of a news article which is the model input, and the second column shows what headline the model has written.

Input: Article 1st sentence	Model-written headline
metro-goldwyn-mayer reported a third-quarter net loss of dlrs 16 million due mainly to the effect of accounting rules adopted this year	mgm reports 16 million net loss on higher revenue
starting from July 1, the island province of Hainan in southern china will implement strict market access control on all incoming livestock and animal products to prevent the possible spread of epidemic diseases	hainan to curb spread of diseases
australian wine exports hit a record 52.1 million liters worth 260 million dollars (143 million us) in september, the government statistics office reported on monday	australian wine exports hit record high in september

Text summarization problem has many useful applications. It can create titles and short summaries for user generated content on the website. If someone wants to read a lot of articles and doesn't have time to do that, a virtual assistant can summarize the main points from these articles for you.

It can be implemented to generate minutes of meetings from meeting recordings which will be very helpful for the person who doesn't attend the meeting and doesn't want to listen to the lengthy recording.

Another use case of text summarization is Smart Reply implemented in google mail. Smart Reply suggests up to three responses based on the emails you get. For those emails that only need a quick response, it can take care of the thinking and save precious time spent typing. And for those emails that require a bit more thought, it gives you a jump start so you can respond right away.

TF Ranking: Ranking is a process of keeping most relevant information on the top. This is applicable in multiple systems like search engines, recommender engines. It became a challenge to keep the most relevant information on the top when the data is huge. TensorFlow provided a library to apply learning to rank at scale.

Twitter implemented TensorFlow in their platform to “Ranked Timeline” which shows the most relevant Tweets at the top of the timeline, ensuring users never miss their best Tweets.

Comparison between TensorFlow vs Pytorch:

Pytorch: PyTorch is an open source machine learning library that specializes in tensor computations, automatic differentiation, and GPU acceleration. It was developed by Facebook and it is a strong competitor of TensorFlow. We will not deep dive in PyTorch as it is out of scope of this document.

We will go through some of the high level differences between Pytorch and TensorFlow.

- **Graph Definition -** Both frameworks operate on tensors and view any model as a directed acyclic graph (DAG), but they differ drastically on how you can define them. In TensorFlow you can define graph statically before a model can run. In PyTorch things are way more imperative and dynamic: you can define, change and execute nodes as you go, no special session interfaces or placeholders. For this reason, PyTorch is often preferred in research since it is more suitable to create custom models, and being dynamic might make it easier to interact with the models' internals.

- **Model Deployment -** TensorFlow has the TensorFlow Serving, which is a built-in model deployment tool used to deploy machine learning models as well as gRPC servers. PyTorch introduced TorchServe, which is a model

deployment tool. This tool provides the basic set of features, such as metrics, an API endpoint specification, a model archiver tool, and so on.

- **Visualization -** PyTorch and TensorFlow support visualization tools, which facilitates debugging and allows the user to visualize the results quickly and have a broad view of the model's training process. TensorFlow has Tensorboard, which offers a suite of apps that allow the user to comprehend the deep learning model through five different visualizations: (1) graphs; (2) audio; (3) images; (4) distributions and histograms; (5) scalars. PyTorch does not necessarily have a dedicated visualization tool, but it does have Visdom, a minimalistic visualization tool. Visdom can be used with Numpy or PyTorch. It provides limited basic features, but it is also reasonably easy to use, flexible and supports PyTorch tensors.

- **Debugging -** Pytorch uses Python's standard debuggers - e.g. PyCharm debugger and pdb. To debug tensorflow code, user must learn the library's debugger - tfdbg. In this sense, debugging tensorflow code is little complex as compared to Pytorch.

Reference:

https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html

https://pytorch.org/

https://www.tensorflow.org/about

https://ai.googleblog.com/2016/08/text-summarization-with-tensorflow.html

https://ai.googleblog.com/2015/11/computer-respond-to-this-email.html

https://ai.googleblog.com/2018/12/tf-ranking-scalable-tensorflow-library.html