

Technology Review

TensorFlow and NLP

Jay Kim

November 7, 2021

1 Background

Machine Learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. It is a complex problem but implementing machine learning models is far less difficult than it used to be since there are machine learning frameworks. One of the reputable frameworks is Google's TensorFlow, which ease the process of acquiring data, training models, serving predictions, and refining future results.

2 Introduction to TensorFlow

TensorFlow is an end-to-end open-source platform that has a comprehensive, flexible ecosystem of tools, libraries, and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML powered applications.

It is designed to help you learn to build models easily. With a set of APIs that makes it simple to learn and implement machine learning, deep learning, and scientific computing, it provides you with a rich collection of tools for building models. These include data pre-processing, data ingestion, model evaluation, visualization, and serving. But it's not just for building models. You can easily train and deploy your model anywhere with TensorFlow. It's designed to be highly portable, running on a variety of devices and platforms. It can scale from a single CPU to a GPU or cluster of GPUs, all the way up to a multi-node TPU infrastructure. So, it's simple to build models that work on any platform taking advantage of the power of each.

TensorFlow also allows for powerful experimentation. With the flexibility to quickly implement state-of-the-art models, TensorFlow can power your research into new techniques to solve novel problems, such as helping rural farmers spot diseased plants, and predicting extreme weather conditions.

On top of these, TensorFlow is driven by a worldwide community. TensorFlow derives its roadmap from the needs of its users with a vision to provide a machine learning platform that's for everyone.

3 Natural Language Processing using TensorFlow

Natural Language Processing (NLP) is the automatic manipulation of natural language like speech and text and concerns with giving computers the ability to understand text and spoken words in much the same way human beings can.

TensorFlow allows you to do tokenization, sequencing, sentimental analysis, Convolutional Neural Network, Recurrent Neural Network, Long Short-Term Memory, Gated Recurrent Unit, and even training an AI to create poetry.

- Tokenization: A task of separating a text corpus into atomic units like words.
- Sequencing: A task of creating a sequence of numbers from the sentences and using tools to process them to make them ready for teaching a neural network.
- Sentimental Analysis: A task of analyzing the subjective information in an expression, that is, the opinions, emotions, or attitudes towards a topic, person, or entity.
- Convolutional Neural Network: A class of neural network, mostly applied to analyze visual imagery.
- Recurrent Neural Network: A class of neural network that are helpful in modeling sequence data.
- Long Short-Term Memory & Gated Recurrent Unit: Advanced version of the Recurrent Neural Network.

TensorFlow provides for helping with NLP and Python programming language provides a wide range of tools and libraries for tackling specific NLP tasks.

4 Comparison between TensorFlow and PyTorch

TensorFlow and PyTorch are by far the two most powerful major deep learning frameworks. TensorFlow is developed by Google while PyTorch is developed by Facebook. Both frameworks are being used for research and production needs. However, PyTorch is not a simple set of wrappers to support popular language, it was rewritten and tailored to be fast and feel native.

They are very close in performance and in the features that they provide. Overall, there is a difference in the coding syntax between the two and this is what makes a difference when researchers and developers working on a machine learning project.

Additionally, TensorFlow is suitable for developing models for production since it has strong visualization capabilities and several options to use for high-level model development. On the other hand, PyTorch is good for doing a research or the production's non-functional requirements are not very demanding as it allows you to create a custom model or a custom dataset. PyTorch uses object-oriented programming style, therefore, you will most likely create a new class that inherits the given PyTorch libraries and then adapt your own methods.

Both frameworks are powerful tools and have been improving continuously, making it very difficult to choose which one is better. While TensorFlow is considered to be more friendly

with high-level model development, PyTorch has gained a lot of popularity among research-oriented developers, supporting highly customizable implementations.

References

- [1] IBM, Machine Learning, <https://www.ibm.com/cloud/learn/machine-learning>
- [2] TensorFlow, <https://www.tensorflow.org>
- [3] IBM, Natural Language Processing, <https://www.ibm.com/cloud/learn/natural-language-processing>
- [4] Wikipedia, Convolutional Neural Network, https://en.wikipedia.org/wiki/Convolutional_neural_network
- [5] Wikipedia, Recurrent Neural Network, https://en.wikipedia.org/wiki/Recurrent_neural_network
- [6] Towards Data Science, TensorFlow vs. PyTorch, <https://towardsdatascience.com/pytorch-vs-tensorflow-spotting-the-difference-25c75777377b>