

Deep Learning

Vazgen Mikayelyan

YSU, Krisp

September 24, 2020

1 Stochastic Gradient Descent

2 Introduction to Tensorflow

Stochastic Gradient Descent

Let L be a loss function that we know:

Stochastic Gradient Descent

Let L be a loss function that we know:

$$L(w) = \frac{1}{n} \sum_{i=1}^n (f_w(x_i) - y_i)^2,$$

$$L(w) = \frac{1}{n} \sum_{i=1}^n (-y_i \log f_w(x_i) - (1 - y_i) \log (1 - f_w(x_i))),$$

$$L(w) = \frac{1}{n} \sum_{i=1}^n \left(-y_i^T \log f_w(x_i) \right).$$

Stochastic Gradient Descent

Let L be a loss function that we know:

$$L(w) = \frac{1}{n} \sum_{i=1}^n (f_w(x_i) - y_i)^2,$$

$$L(w) = \frac{1}{n} \sum_{i=1}^n (-y_i \log f_w(x_i) - (1 - y_i) \log (1 - f_w(x_i))),$$

$$L(w) = \frac{1}{n} \sum_{i=1}^n \left(-y_i^T \log f_w(x_i) \right).$$

Do you see problems in finding minimum of these functions using GD?

Note that in each case we can represent the loss function by the following form:

$$L(w) = \frac{1}{n} \sum_{i=1}^n L_i(w).$$

Stochastic Gradient Descent

Note that in each case we can represent the loss function by the following form:

$$L(w) = \frac{1}{n} \sum_{i=1}^n L_i(w).$$

SGD algorithm is the following:

- Choose an initial vector of parameters w and learning rate α .

Stochastic Gradient Descent

Note that in each case we can represent the loss function by the following form:

$$L(w) = \frac{1}{n} \sum_{i=1}^n L_i(w).$$

SGD algorithm is the following:

- Choose an initial vector of parameters w and learning rate α .
- Repeat until an approximate minimum is obtained.

Stochastic Gradient Descent

Note that in each case we can represent the loss function by the following form:

$$L(w) = \frac{1}{n} \sum_{i=1}^n L_i(w).$$

SGD algorithm is the following:

- Choose an initial vector of parameters w and learning rate α .
- Repeat until an approximate minimum is obtained.
 - Randomly shuffle examples in the training set.

Stochastic Gradient Descent

Note that in each case we can represent the loss function by the following form:

$$L(w) = \frac{1}{n} \sum_{i=1}^n L_i(w).$$

SGD algorithm is the following:

- Choose an initial vector of parameters w and learning rate α .
- Repeat until an approximate minimum is obtained.
 - Randomly shuffle examples in the training set.
 - For $i = 1, 2, \dots, n$, do $w \leftarrow w - \alpha \nabla L_i(w)$.

Stochastic Gradient Descent

Note that in each case we can represent the loss function by the following form:

$$L(w) = \frac{1}{n} \sum_{i=1}^n L_i(w).$$

SGD algorithm is the following:

- Choose an initial vector of parameters w and learning rate α .
- Repeat until an approximate minimum is obtained.
 - Randomly shuffle examples in the training set.
 - For $i = 1, 2, \dots, n$, do $w \leftarrow w - \alpha \nabla L_i(w)$.

Do you see problems in this optimization method?

Mini-Batch Gradient Descent

MBGD algorithm is the following:

- Choose an initial vector of parameters w , learning rate α and batch size B .

Mini-Batch Gradient Descent

MBGD algorithm is the following:

- Choose an initial vector of parameters w , learning rate α and batch size B .
- Repeat until an approximate minimum is obtained.

Mini-Batch Gradient Descent

MBGD algorithm is the following:

- Choose an initial vector of parameters w , learning rate α and batch size B .
- Repeat until an approximate minimum is obtained.
 - Randomly shuffle examples in the training set.

Mini-Batch Gradient Descent

MBGD algorithm is the following:

- Choose an initial vector of parameters w , learning rate α and batch size B .
- Repeat until an approximate minimum is obtained.
 - Randomly shuffle examples in the training set.
 - For $i = 1, 2, \dots, \lceil \frac{n}{B} \rceil$, do

$$w \leftarrow w - \alpha \nabla \frac{1}{B} \sum_{k=(i-1) \cdot B + 1}^{i \cdot B} L_k(w).$$

Outline

1 Stochastic Gradient Descent

2 Introduction to Tensorflow

Difference Between



CPU



GPU



TPU

CPU vs GPU vs TPU

CPU vs GPU vs TPU

- Central Processing Unit is the electronic circuitry, which work as a brain of the computer that perform the basic arithmetic, logical, control and input/output operations specified by the instructions of a computer program.

CPU vs GPU vs TPU

- Central Processing Unit is the electronic circuitry, which work as a brain of the computer that perform the basic arithmetic, logical, control and input/output operations specified by the instructions of a computer program.
- The Graphics Processing Unit is a specialized electronic circuit designed to render 2D and 3D graphics together with a CPU. GPU also known as Graphics Card in the Gammer's culture. Now GPU are being harnessed more broadly to accelerate computational workloads in areas such as financial modeling, cutting-edge scientific research, deep learning, analytics and oil and gas exploration etc.

CPU vs GPU vs TPU

- Central Processing Unit is the electronic circuitry, which work as a brain of the computer that perform the basic arithmetic, logical, control and input/output operations specified by the instructions of a computer program.
- The Graphics Processing Unit is a specialized electronic circuit designed to render 2D and 3D graphics together with a CPU. GPU also known as Graphics Card in the Gammer's culture. Now GPU are being harnessed more broadly to accelerate computational workloads in areas such as financial modeling, cutting-edge scientific research, deep learning, analytics and oil and gas exploration etc.
- Tensor Processing Unit is a custom-built integrated circuit developed specifically for machine learning and tailored for TensorFlow, Google's open-source machine learning framework. TPU's have been powering Google data centers since 2015.

Why Tensorflow?



Interest Over Time

<input checked="" type="radio"/> pytorch Search term	<input checked="" type="radio"/> tensorflow Search term	<input type="radio"/> keras Search term	<input type="radio"/> caffe Search term	+
---	--	--	--	---

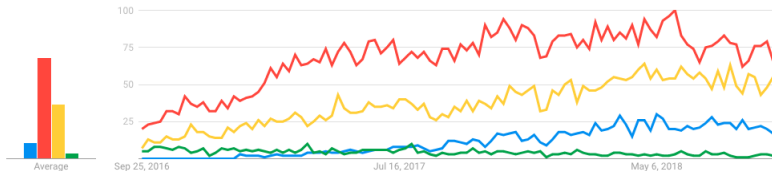
Worldwide ▾

9/20/16 - 9/21/18 ▾

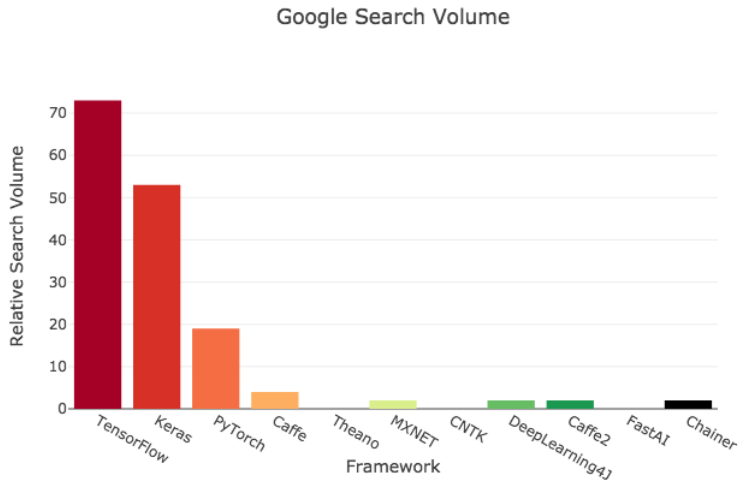
Machine Learning & Arti... ▾

Web Search ▾

Interest over time ?

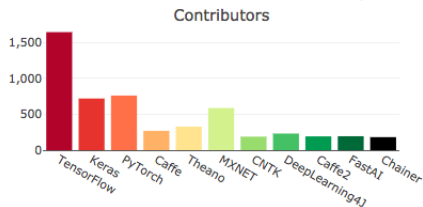
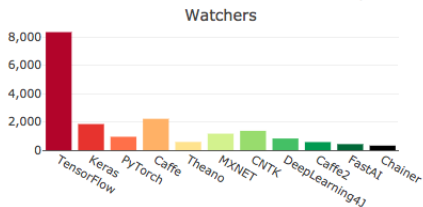
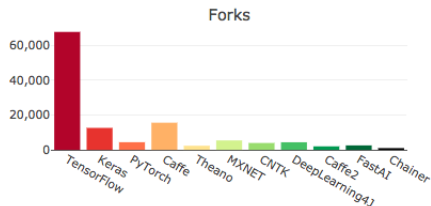
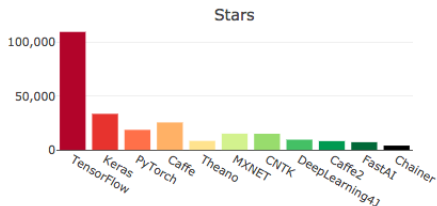


Google Search Activity

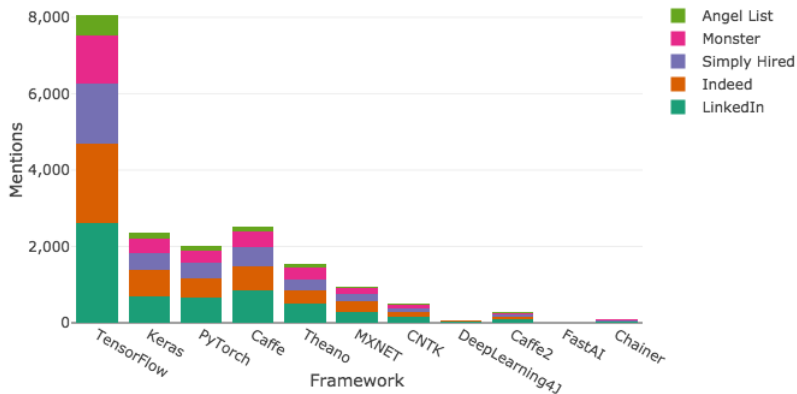


GitHub Activity

GitHub Activity

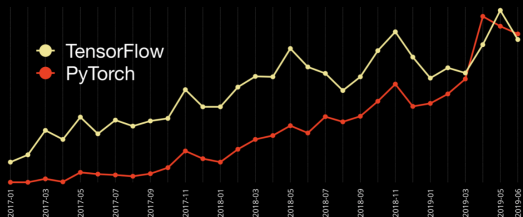


Online Job Listings

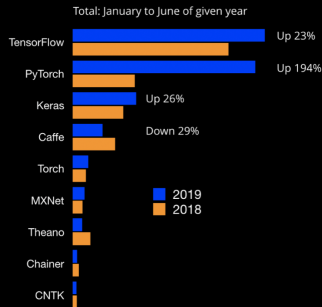


Tensorflow vs PyTorch

Number of papers on arxiv.org that mention a given framework



Source: Data from RISELab ; graphic from gradientflow.com



What can do Tensorflow?

- TensorFlow lets you deploy computation to one or more CPUs or GPUs in a desktop, server, or mobile device in a very simple way. This way the things can be done very fast.

What can do Tensorflow?

- TensorFlow lets you deploy computation to one or more CPUs or GPUs in a desktop, server, or mobile device in a very simple way. This way the things can be done very fast.
- TensorFlow lets you express your computation as a data flow graph.

What can do Tensorflow?

- TensorFlow lets you deploy computation to one or more CPUs or GPUs in a desktop, server, or mobile device in a very simple way. This way the things can be done very fast.
- TensorFlow lets you express your computation as a data flow graph.
- TensorFlow lets you visualize the graph using the in-built tensorboard. You can inspect and debug the graph very easily.

What can do Tensorflow?

- TensorFlow lets you deploy computation to one or more CPUs or GPUs in a desktop, server, or mobile device in a very simple way. This way the things can be done very fast.
- TensorFlow lets you express your computation as a data flow graph.
- TensorFlow lets you visualize the graph using the in-built tensorboard. You can inspect and debug the graph very easily.
- TensorFlow gives the best performance with an ability to iterate quickly, train models faster and run more experiments.

What can do Tensorflow?

- TensorFlow lets you deploy computation to one or more CPUs or GPUs in a desktop, server, or mobile device in a very simple way. This way the things can be done very fast.
- TensorFlow lets you express your computation as a data flow graph.
- TensorFlow lets you visualize the graph using the in-built tensorboard. You can inspect and debug the graph very easily.
- TensorFlow gives the best performance with an ability to iterate quickly, train models faster and run more experiments.
- TensorFlow runs on nearly everything: GPUs and CPUs—including mobile and embedded platforms—and tensor processing units (TPUs), which are specialized hardware to do the tensor math on.

- View functions as computational graphs.

- View functions as computational graphs.
- First build a computational graph, and then use a session to execute operations in the graph.

Basic Code Structure

- View functions as computational graphs.
- First build a computational graph, and then use a session to execute operations in the graph.
- This is the basic approach, there is also a dynamic approach implemented in the recently introduced eager mode.

Computational Graphs

