What is Theano

jamworld

2016-06-27 Mon 16:15

What is Theano?

Shortest way

Theano is an python library for tensor computation.

Goal

Easy to write and high performance.

- Easy to write and understand Means that we can write down a symbolic **acyclic** graph model for the computation. The computation graph is easy for the human beings to understand.

- High performance Theano will compile the given computation graph into the high performance C code or CUDA code.

Some details about Theano

Graph Structure

Two kinds of nodes

Variable Nodes Represent data, usually tensor. (Input of many Op, Output of at most **one** Op → SSA for compiler optimization)

Apply Nodes Represent the application of mathematical operations.

In theano we organize the graph as

arcs variable

nodes operators

Variable Type System

In order to keep the data consistent for the low level implementation there is an type system for the theano library.

TensorType Type for every n-dimension array

CudaNdarrayType Type for n-dimension cuda array

GpuArrayType More general gpu type

Sparse Sparse Matrix

- **broadcastable pattern** Indicate which dimension will be guaranteed to have dimension 1, thus we can broadcast on it. Also used for the memory layout.

Clone

By cloning the graph we can build new graph from existing graph.

Operator

In order to implement the back propagation method. We need the reverse operator (R-Op)

- R-Operator (Symbolic Differentiation) With chain rule, we can make differentiation in a recursion form.

  grad method Every apply nodes should have the grad method. Grad method here is an method that can implement the automatic differentiation. By automatic differentiation I mean differentiation via the chain rule. What we do here is just break every formula into small piece, the we the only thing we need to do is just implement the differentiation for every single pieces, which is quite simple, than by calling the grad method one by one we can get the differentiation for the whole complex formula.

  scan Since the whole graph is acyclic, hence we need some special operator for the loop. (Otherwise we will got cycle inside our graph). Scan will take care of the communication between the inner and outer computation graph. **IMPORTANT**, scan is responsible to manage the bookkeeping between the different iterations.

R operator R operator will implement the forward operation. By calling the R operation method on each apply node's Op.

The gradient of scan is implemented as another scan operation, which iterates over reversed sequences, computing the same gradient as if the loop had been unrolled, implementing what is known as **back-propagation through time**.Similarly, the R operator is also a scan operation that goes through the loop in the same order as the original scan.

NOTE

- Theano is quite similar to the Spark. In the sense of the computation graph to the high performance executable code. The reason we need something like theano and spark is that it is too difficult for us humans to understand the complex machine code for the complex computation. So as what we have done the mathematics we what more abstraction. So the direction we towards is just more and more high level abstraction. What we want is just high recursion. We what something that can auto-execute the instruction once we describe the base situation and the induction step.

References

Technique-Report-For-Theano.pdf