CS 224N: TensorFlow Tutorial

Lecture and Live Demo

Nishith Khandwala & Barak Oshri

31 January, 2017
Intro to Deep Learning Frameworks

- Scales machine learning code
- Computes gradients!
- Standardizes machine learning applications for sharing
- Zoo of Deep Learning frameworks available with different advantages, paradigms, levels of abstraction, programming languages, etc
- Interface with GPUs for parallel processing

In some ways, rightfully gives Deep Learning its name as a separate *practice*
What is TensorFlow?

- Open source software library for numerical computation using data flow graphs
- Originally developed by Google Brain Team to conduct machine learning research
- "Tensorflow is an interface for expressing machine learning algorithms, and an implementation for executing such algorithms"
Programming model

Big idea: express a numeric computation as a graph.

- Graph nodes are **operations** which have any number of inputs and outputs
- Graph edges are **tensors** which flow between nodes
$h = \text{ReLU}(Wx + b)$
Variables are stateful nodes which output their current value. State is retained across multiple executions of a graph. (mostly parameters)
Programming model

\[ h = \text{ReLU}(Wx + b) \]

**Placeholders** are nodes whose value is fed in at execution time

(inputs, labels, ...)
Programming model

$h = ReLU(Wx + b)$

**Mathematical operations:**

- **MatMul:** Multiply two matrix values.
- **Add:** Add elementwise (with broadcasting).
- **ReLU:** Activate with elementwise rectified linear function.
In code,

1. Create weights, including initialization
   \( W \sim \text{Uniform}(-1, 1) \); \( b = 0 \)

2. Create input placeholder \( x \)
   \( m \times 784 \) input matrix

3. Build flow graph

   \[
   h = \text{ReLU}(Wx + b)
   \]

```python
import tensorflow as tf

b = tf.Variable(tf.zeros((100,)))
W = tf.Variable(tf.random_uniform((784, 100), -1, 1))

x = tf.placeholder(tf.float32, (100, 784))

h = tf.nn.relu(tf.matmul(x, W) + b)
```
But where is the graph?

New nodes are automatically built into the underlying graph!

```
tf.get_default_graph().get_operations():
```

- zeros/shape
- zeros/Const
- zeros
- Variable
- Variable/Assign
- Variable/read
- random_uniform/shape
- random_uniform/min
- random_uniform/max
- random_uniform/RandomUniform
- random_uniform/sub
- random_uniform/mul
- random_uniform
- Variable_1
- Variable_1/Assign
- Variable_1/read
- Placeholder
- MatMul
- add
- **Relu == h**

h refers to an op!
How do we run it?

So far we have defined a graph.

We can deploy this graph with a session: a binding to a particular execution context (e.g. CPU, GPU)
Getting output

```python
import numpy as np
import tensorflow as tf

b = tf.Variable(tf.zeros((100,)))
W = tf.Variable(tf.random_uniform((784, 100), -1, 1))

x = tf.placeholder(tf.float32, (100, 784))
h = tf.nn.relu(tf.matmul(x, W) + b)

csess = tf.Session()
csess.run(tf.initialize_all_variables())
csess.run(h, {x: np.random.random(100, 784)})
```

**Fetches:** List of graph nodes. Return the outputs of these nodes.

**Feeds:** Dictionary mapping from graph nodes to concrete values. Specifies the value of each graph node given in the dictionary.
So what have we covered so far?

We first built a **graph** using **variables** and **placeholders**

We then deployed the graph onto a **session**, which is the **execution environment**

Next we will see how to **train** the **model**
How do we define the loss?

Use **placeholder** for labels

Build loss node using labels and **prediction**

```python
prediction = tf.nn.softmax(...)  # Output of neural network
label = tf.placeholder(tf.float32, [100, 10])

cross_entropy = -tf.reduce_sum(label * tf.log(prediction), axis=1)
```
How do we compute Gradients?

\[
\text{train\_step} = \text{tf.train.GradientDescentOptimizer(0.5).minimize(cross\_entropy)}
\]

- `tf.train.GradientDescentOptimizer` is an **Optimizer** object
- `tf.train.GradientDescentOptimizer(lr).minimize(cross\_entropy)` adds optimization **operation** to computation graph

  TensorFlow graph **nodes** have **attached gradient operations**

  Gradient with respect to **parameters** computed with **backpropagation**

  ...automatically
Creating the train_step op

prediction = tf.nn.softmax(...)  
label = tf.placeholder(tf.float32, [None, 10])

cross_entropy = tf.reduce_mean(-tf.reduce_sum(label * tf.log(prediction), reduction_indices=[1]))

train_step = tf.train.GradientDescentOptimizer(0.5).minimize(cross_entropy)
Training the Model

1. Create Session
2. Build training schedule
3. Run `train_step`

```python
sess = tf.Session()
sess.run(tf.initialize_all_variables())

for i in range(1000):
    batch_x, batch_label = data.next_batch()
    sess.run(train_step, feed_dict={x: batch_x, label: batch_label})
```
Variable sharing
Variable sharing: naive way

```
variables_dict = {
    "weights": tf.Variable(tf.random_normal([782, 100]),
                            name="weights")
    "biases": tf.Variable(tf.zeros([100]), name="biases")
}
```

Not good for encapsulation!
What’s in a Name?

`tf.variable_scope()` provides simple name-spacing to avoid clashes

`tf.get_variable()` creates/accesses variables from within a variable scope

```
with tf.variable_scope("foo"):
  v = tf.get_variable("v", shape=[1])  # v.name == "foo/v:0"

with tf.variable_scope("foo", reuse=True):
  v1 = tf.get_variable("v")  # Shared variable found!

with tf.variable_scope("foo", reuse=False):
  v1 = tf.get_variable("v")  # CRASH foo/v:0 already exists!
```
In Summary:

1. Build a graph
   a. Feedforward / Prediction
   b. Optimization (gradients and train_step operation)

2. Initialize a session

3. Train with session.run(train_step, feed_dict)
Acknowledgments

Jon Gauthier, Natural Language Processing Group, Symbolic Systems

Bharath Ramsundar, PhD Student, Drug Discovery Research

Chip Huyen, Undergraduate, teaching CS20SI: TensorFlow for Deep Learning Research!
Visual Dialog

Abhishek Das, Satwik Kottur, Khushi Gupta, Avi Singh, Deshraj Yadav, José M.F. Moura, Devi Parikh, Dhruv Batra

Presented by: Alan Luo
Introduction  
Natural Language Processing + Computer Vision

- Aiding visually impaired users in understanding their surroundings or social media content
- Interacting with an AI assistant
Related Work  Image/Video Captioning

Image Captioning

1. Input Image
2. Convolutional Feature Extraction
3. RNN with attention over the image
4. Word by word generation

Video Summary

Input video:

Our output: A cat is playing with a toy.
Humans: A ferret and cat fighting with each other. / A cat and a ferret are playing. / A kitten is playing with a ferret. / A kitten and a ferret are playfully wrestling.

a man is throwing a frisbee in a park

a man riding a wave on top of a surfboard
Related Work  Visual-Semantic Alignments

Datasets

Regions
- the white bowls on the table
- the woman
teaching the little
girl to cook
- the liquid in the
bowl
- the utensil on the
food
- the pink shirt on
the woman
- the pink shirt on
the little girl
- little girl with a
pink too on next to
Question Answers

Attributes
- bowl is white
- pan is silver
- kitchen utensils is
kitchen
- kitchen utensils is
behind
- utensils is behind
- pizza is unripe
- hand is white
- shirt is pink
- girl is young
- spoon is large
- spoon is silver

Relationships
- top
- lady WEARING
shirt
- woman helping
girl
- sauce ON bread
- tin foil under
bread
- paper hanging on
wall
- lady helping girl
- child standing at
table
- woman IN kitchen

What color is the girl's shirt?
Where was the photo taken?
What are the people doing?
How many adults in the kitchen?
What color is the sauce being put onto the food?
Related Work  Visual Q&A

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>How many pickles are on the plate?</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>What is the shape of the plate?</td>
<td>circle</td>
</tr>
<tr>
<td></td>
<td>round</td>
</tr>
<tr>
<td></td>
<td>round</td>
</tr>
<tr>
<td>What does the sign say?</td>
<td>stop</td>
</tr>
<tr>
<td></td>
<td>stop</td>
</tr>
<tr>
<td></td>
<td>stop</td>
</tr>
<tr>
<td></td>
<td>yield</td>
</tr>
<tr>
<td>What shape is this sign?</td>
<td>octagon</td>
</tr>
<tr>
<td></td>
<td>octagon</td>
</tr>
<tr>
<td></td>
<td>octagon</td>
</tr>
<tr>
<td></td>
<td>diamond</td>
</tr>
<tr>
<td></td>
<td>octagon</td>
</tr>
<tr>
<td></td>
<td>round</td>
</tr>
</tbody>
</table>
Contributions

1. Propose a new AI task: Visual Dialog

2. Develop a novel two-person chat data-collection protocol and introduce a new dataset

3. Introduce a family of neural encoder-decoder models for Visual Dialog
Technical Details With Late Fusion Encoder

Image I
Do you think the woman is with him?

Question $Q_t$

The man is riding his bicycle on the sidewalk. Is the man wearing a helmet? No he does not have a helmet on. ... Are there any people nearby? Yes there's a woman walking behind him.

$t$ rounds of history
(concatenated)

Decoder

No I don't think they are together

Answer $A_t$
**Dataset** VisDial

**Qualitative**

Caption: A statue depicting a bear breaking into a car.
Person A (1): how big is statue
Person B (1): about size of real full grown bear
Person A (2): so is car full size then as well
Person B (2): yes replica of car
Person A (3): is statue all 1 color
Person B (3): no brown and black
Person A (4): what color is car
Person B (4): dark red
Person A (5): where is this, do you think
Person B (5): in wooded area someplace
Person A (6): do you see any people in image
Person B (6): yes 1 man
Person A (7): how old is man
Person B (7): 35-40
Person A (8): what is man doing
Person B (8): sitting in car behind replica
Person A (9): do you see any signs
Person B (9): yes, on car door warning sign
Person A (10): what else can you tell me about this image
Person B (10): there are many trees in background

**Quantitative**

- [Graph showing percentage of questions and answers vs. number of words]
- [Graph showing percentage coverage vs. unique answers (x10000)]
Results

Qualitative Results

- **Visual Chatbot**
  - Hi, I’m a visual chatbot, capable of answering a sequence of questions about images. Please upload an image and fire away!

  ![Image: A group of bikers parked in a parking lot under large lights]

  **Caption:** A group of bikers parked in a parking lot under large lights

- **moped**
- **on street**
- **cloudy sky looks clear**

Quantitative Results

<table>
<thead>
<tr>
<th>Model</th>
<th>MRR</th>
<th>R@1</th>
<th>R@5</th>
<th>R@10</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answer prior</td>
<td>0.311</td>
<td>19.85</td>
<td>39.14</td>
<td>44.28</td>
<td>31.56</td>
</tr>
<tr>
<td>NN-Q</td>
<td>0.392</td>
<td>30.54</td>
<td>46.99</td>
<td>49.98</td>
<td>30.88</td>
</tr>
<tr>
<td>NN-QI</td>
<td>0.385</td>
<td>29.71</td>
<td>46.57</td>
<td>49.86</td>
<td>30.90</td>
</tr>
<tr>
<td>LF-Q-G</td>
<td>0.403</td>
<td>29.74</td>
<td>50.10</td>
<td>56.32</td>
<td>24.06</td>
</tr>
<tr>
<td>LF-QH-G</td>
<td>0.425</td>
<td>32.49</td>
<td>51.56</td>
<td>57.80</td>
<td>23.11</td>
</tr>
<tr>
<td>LF-QI-G</td>
<td>0.437</td>
<td>34.06</td>
<td>52.50</td>
<td>58.89</td>
<td>22.31</td>
</tr>
<tr>
<td>LF-QIH-G</td>
<td>0.430</td>
<td>33.27</td>
<td>51.96</td>
<td>58.09</td>
<td>23.04</td>
</tr>
<tr>
<td>HRE-QH-G</td>
<td>0.430</td>
<td>32.84</td>
<td>52.36</td>
<td>58.64</td>
<td>22.59</td>
</tr>
<tr>
<td>HRE-QIH-G</td>
<td>0.442</td>
<td>34.37</td>
<td>53.40</td>
<td>59.74</td>
<td>21.75</td>
</tr>
<tr>
<td>HREA-QIH-G</td>
<td>0.442</td>
<td>34.47</td>
<td>53.43</td>
<td>59.73</td>
<td>21.83</td>
</tr>
<tr>
<td>MN-QH-G</td>
<td>0.434</td>
<td>33.12</td>
<td>52.14</td>
<td>58.11</td>
<td>22.14</td>
</tr>
<tr>
<td>MN-QIH-G</td>
<td>0.443</td>
<td>34.62</td>
<td>53.74</td>
<td>60.18</td>
<td>21.69</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>MRR</th>
<th>R@1</th>
<th>R@5</th>
<th>R@10</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>LF-Q-D</td>
<td>0.482</td>
<td>34.29</td>
<td>63.42</td>
<td>74.31</td>
<td>8.87</td>
</tr>
<tr>
<td>LF-QH-D</td>
<td>0.505</td>
<td>36.21</td>
<td>66.56</td>
<td>77.31</td>
<td>7.89</td>
</tr>
<tr>
<td>LF-QI-D</td>
<td>0.502</td>
<td>35.76</td>
<td>66.59</td>
<td>77.61</td>
<td>7.22</td>
</tr>
<tr>
<td>LF-QIH-D</td>
<td>0.511</td>
<td>36.72</td>
<td>67.46</td>
<td>78.30</td>
<td>7.63</td>
</tr>
<tr>
<td>HRE-QH-D</td>
<td>0.489</td>
<td>34.74</td>
<td>64.25</td>
<td>75.40</td>
<td>8.32</td>
</tr>
<tr>
<td>HRE-QIH-D</td>
<td>0.502</td>
<td>36.26</td>
<td>65.67</td>
<td>77.05</td>
<td>7.79</td>
</tr>
<tr>
<td>HREA-QIH-D</td>
<td>0.508</td>
<td>36.76</td>
<td>66.54</td>
<td>77.75</td>
<td>7.59</td>
</tr>
<tr>
<td>MN-QH-D</td>
<td>0.524</td>
<td>36.84</td>
<td>67.78</td>
<td>78.92</td>
<td>7.25</td>
</tr>
<tr>
<td>MN-QIH-D</td>
<td>0.529</td>
<td>37.33</td>
<td>68.47</td>
<td>79.54</td>
<td>7.03</td>
</tr>
<tr>
<td>SANI-Q-D</td>
<td>0.506</td>
<td>36.21</td>
<td>67.08</td>
<td>78.16</td>
<td>7.74</td>
</tr>
<tr>
<td>HisCoAt-QD</td>
<td>0.509</td>
<td>35.54</td>
<td>66.79</td>
<td>77.94</td>
<td>7.68</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>MRR</th>
<th>R@1</th>
<th>R@5</th>
<th>R@10</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human-Q</td>
<td>0.441</td>
<td>23.10</td>
<td>67.37</td>
<td>74.19</td>
<td>4.19</td>
</tr>
<tr>
<td>Human-QIH</td>
<td>0.485</td>
<td>30.31</td>
<td>70.53</td>
<td>73.91</td>
<td>3.91</td>
</tr>
<tr>
<td>Human-QI</td>
<td>0.619</td>
<td>46.12</td>
<td>82.54</td>
<td>82.92</td>
<td>2.92</td>
</tr>
<tr>
<td>Human-QIH</td>
<td>0.635</td>
<td>48.03</td>
<td>83.76</td>
<td>82.83</td>
<td>2.83</td>
</tr>
</tbody>
</table>

**What is this?**
**Where is this?**
**how is the weather?**