Recurrent Neural Network
The Deepest of All Deep Learning

Slides by Chen Liang
Deep Learning

With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart.

AI Is Transforming Google Search. The Rest of the Web Is Next

IN A HUGE BREAKTHROUGH, GOOGLE’S AI BEATS A TOP PLAYER AT THE GAME OF GO

Drive.ai to test 'deep learning' autonomous cars on California roads

Baidu’s Deep-Learning System Rivals People at Speech Recognition
Beta: Measuring search interest in *topics* is a beta feature which quickly provides accurate measurements of overall search interest. To measure search interest for a specific *query*, select the "search term" option.
Deep learning works like the human brain?

Demystify Deep Learning
Deep Learning: Building Blocks

- Linear transformations
  \[ h_{k+1} = Wh_k \]

- Non-linear activation functions
  \[ h_{k+2} = f(h_{k+1}) \]
Deep Learning: Deep Composition

- A deep representation is a composition of many functions

\[ x \xrightarrow{w_1} h_1 \xrightarrow{w_2} \ldots \xrightarrow{w_n} h_n \xrightarrow{w_{n+1}} y \]

- Its gradient can be backpropagated by the chain rule

\[
\frac{\partial h_1}{\partial x} \leftarrow \frac{\partial h_2}{\partial h_1} \leftarrow \ldots \leftarrow \frac{\partial y}{\partial h_n} \leftarrow \frac{\partial}{\partial y} \\
\downarrow \quad \downarrow \quad \downarrow \\
\frac{\partial h_1}{\partial w_1} \quad \ldots \quad \frac{\partial h_n}{\partial w_n} \quad \frac{\partial y}{\partial w_{n+1}}
\]
Deep Learning: Gradient Descent

- A loss function $l(y)$ measures goodness of output $y$, e.g.
  - Mean-squared error $l(y) = \|y^* - y\|^2$
  - Log likelihood $l(y) = \log P [y^* | x]$

- The loss is appended to the forward computation

  $x \xrightarrow{w_1} h_1 \xrightarrow{w_2} \ldots \xrightarrow{w_n} h_n \xrightarrow{w_{n+1}} y \xrightarrow{} l(y)$

- Gradient of loss is appended to the backward computation

\[
\frac{\partial l(y)}{\partial y} \xleftarrow{\frac{\partial y}{\partial h_n}} \xrightarrow{\frac{\partial h_n}{\partial w_n}} \ldots \xrightarrow{\frac{\partial h_1}{\partial w_1}} \frac{\partial l(y)}{\partial h_1} \xrightarrow{\frac{\partial h_1}{\partial x}} \frac{\partial l(y)}{\partial x}
\]
Deep Learning: Weight Sharing

Recurrent neural network shares weights between time-steps

Convolutional neural network shares weights between local regions
Recurrent Neural Network

Deepest of Deep learning?

- Can be infinitely deep

Equation

\[ h_t^l = \tanh W^l \begin{pmatrix} h_t^{l-1} \\ h_{t-1}^l \end{pmatrix} \]

RNN, LSTM illustrations from Christopher Olah's blog
BPTT: Backpropagation Through Time
RNN is Turing complete, but...

Short term dependency

Long term dependency

Exploding/vanishing gradient
LSTM: Long-Short Term Memory
LSTM: Long-Short Term Memory

Let the gradient flow to earlier steps
RNN: A General Framework

- one to one: Image Caption Generation
- one to many: Sentiment Analysis
- many to one: Image Recognition
- many to many: Machine Translation, Speech recognition, Language Modeling
Char-RNN

How it works?

Vocabulary:

[“h”, “e”, “l”, “o”]

Training sequence:

“hello”
Char-RNN

Linux

Latex

Wikipedia

Music

…

Check out the blog:

The Unreasonable Effectiveness of RNN
What does the Neuron do?

The sole importance of the crossing of the Berezina lies in the fact that it plainly and indubitably proved the fallacy of all the plans for cutting off the enemy's retreat and the soundness of the only possible line of action---the one Kutuzov and the general mass of the army demanded---namely, simply to follow the enemy up. The French crowd fled at a continually increasing speed towards all its energy was directed to reaching its goal. It fled like a wounded animal and it was impossible to block its path. This was shown not so much by the arrangements it made for crossing as by what took place at the bridges. When the bridges broke down, unarmed soldiers, people from Moscow and women with children who were with the French transport, all carried on by inertia---pressed forward into boats and into the ice-covered water and did not surrender.

"You mean to imply that I have nothing to eat out of... on the contrary, I can supply you with everything even if you want to give dinner parties," warmly replied Chichagov, who tried by every word he spoke to prove his own rectitude and therefore imagined Kutuzov to be animated by the same desire.

Kutuzov, shrugging his shoulders, replied with his subtle penetrating smile: "I meant merely to say what I said."

A large portion of cells are not easily interpretable. Here is a typical example:

```
static int _dequeue_signal(struct sigpending *pending, sigset_t *mask, siginfo_t *info)
{
    int sig = next_signal(pending, mask);
    if (sig) {
        if ((current->notifier) (SIGSETNOTIFY(sigmask(current->notifier_mask, sig))) ||
            (current->notifier)(current->notifier_data))
            clear_thread_flag(TIF_SIGPENDING);
    }

    collect_signal(sig, pending, info);

    return sig;
}
```

A large portion of cells are not easily interpretable. Here is a typical example:
Seq2seq: sequence-to-sequence learning
Sequence is even longer now => Attention

Again, let the gradient flow to earlier steps

Diagram derived from Fig. 3 of Bahdanau, et al. 2014
More Seq2seq fun: Chatbots
More Seq2seq fun: Programmers
Summary

- Deep Learning is like Lego Blocks
  => Compositionality and Backpropagation
- Creative ways to combine the blocks
  => New applications
- When you have problem
  => Get new blocks (LSTM, attention) to let gradient flow

Food for thought: is our brain just a bunch of Lego blocks?
TensorFlow

TensorFlow™ is an open source software library for numerical computation using data flow graphs.
import tensorflow as tf
import numpy as np

# Create 100 phony x, y data points in NumPy, y = x * 0.1 + 0.3
x_data = np.random.rand(100).astype(np.float32)
y_data = x_data * 0.1 + 0.3

# Try to find values for W and b that compute y_data = W * x_data + b
# (We know that W should be 0.1 and b 0.3, but TensorFlow will
# figure that out for us.)
W = tf.Variable(tf.random_uniform([1], -1.0, 1.0))
b = tf.Variable(tf.zeros([1]))
y = W * x_data + b

# Minimize the mean squared errors.
loss = tf.reduce_mean(tf.square(y - y_data))
optimizer = tf.train.GradientDescentOptimizer(0.5)
train = optimizer.minimize(loss)
TensorFlow: Computation Graph

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

# Create 100 phony x, y data points in NumPy, y = x * 0.1 + 0.3
x_data = np.random.rand(100).astype(np.float32)
y_data = x_data * 0.1 + 0.3

# Try to find values for W and b that compute y_data = W * x_data + b
# (We know that W should be 0.1 and b 0.3, but TensorFlow will
# figure that out for us.)
W = tf.Variable(tf.random_uniform([1], -1.0, 1.0))
b = tf.Variable(tf.zeros([1]))
y = W * x_data + b

# Minimize the mean squared errors.
loss = tf.reduce_mean(tf.square(y - y_data))
optimizer = tf.train.GradientDescentOptimizer(0.5)
train = optimizer.minimize(loss)
```

Synthesize some noisy data from a linear model
```python
import tensorflow as tf
import numpy as np

# Create 100 phony x, y data points in NumPy, y = x * 0.1 + 0.3
x_data = np.random.rand(100).astype(np.float32)
y_data = x_data * 0.1 + 0.3

# Try to find values for W and b that compute y_data = W * x_data + b
# (We know that W should be 0.1 and b 0.3, but TensorFlow will
# figure that out for us.)
W = tf.Variable(tf.random_uniform([1], -1.0, 1.0))
b = tf.Variable(tf.zeros([1]))
y = W * x_data + b

# Minimize the mean squared errors.
loss = tf.reduce_mean(tf.square(y - y_data))
optimizer = tf.train.GradientDescentOptimizer(0.5)
train = optimizer.minimize(loss)
```

TensorFlow: Computation Graph
import tensorflow as tf
import numpy as np

# Create 100 phony x, y data points in NumPy, y = x * 0.1 + 0.3
x_data = np.random.rand(100).astype(np.float32)
y_data = x_data * 0.1 + 0.3

# Try to find values for W and b that compute y_data = W * x_data + b
# (We know that W should be 0.1 and b 0.3, but TensorFlow will
# figure that out for us.)
W = tf.Variable(tf.random_uniform([1], -1.0, 1.0))
b = tf.Variable(tf.zeros([1]))
y = W * x_data + b

# Minimize the mean squared errors.
loss = tf.reduce_mean(tf.square(y - y_data))
optimizer = tf.train.GradientDescentOptimizer(0.5)
train = optimizer.minimize(loss)
# Before starting, initialize the variables. We will 'run' this first.
init = tf.initialize_all_variables()

# Launch the graph.
sess = tf.Session()
sess.run(init)

# Fit the line.
for step in range(201):
    sess.run(train)
    if step % 20 == 0:
        print(step, sess.run(W), sess.run(b))

# Learns best fit is W [0.1], b: [0.3]
# Before starting, initialize the variables. We will 'run' this first.
init = tf.initialize_all_variables()

# Launch the graph.
sess = tf.Session()
sess.run(init)

# Fit the line.
for step in range(201):
    sess.run(train)
    if step % 20 == 0:
        print(step, sess.run(W), sess.run(b))

# Learns best fit is W [0.1], b: [0.3]
Tensorboard Demo

TensorBoard

- Fit to screen
- Download PNG

Run
- train

Session runs
- 0

Upload
- Choose File

Color
- Structure
- Device
- Color: same substructure
- Gray: unique substructure

Graph
- Expandable
- Namespace
- OpNode
- Unconnected series
- Connected series
- Constant
- Summary
- Dataflow edge
- Control dependency edge
- Reference edge

Diagram:
- cross_entropy
- accuracy
- layer2
- dropout
- layer1
- input
- train
Tensorboard Demo

Graph showing the accuracy over steps.

- Horizontal Axis: STEP
- Runs: test, train

Metrics available:
- cross entropy
- dropout_keep_probability
- max
- mean
Tensorboard Demo
Now the part that everybody hates...
Homework

Part 1: Backpropagation and gradient check

- NumPy

Part 2: Char-RNN

- Undergrad/Grad Descent
  - Gradient descent => graduate descent
  - Systematic search of hyperparameters

- Do something fun with it!
TRUMP:
See you here in thy land with a bloody servant.

BENVOLIO:
Be peace?

BARNISAR:
We will not speak, sworn the husband's death.

BRUTUS:
What is the people?

MENENIUS:
How with his hands, what haste the point than was full of the war;
And I will make the people and all:
And I will not have shown the firmy may land,
Where with some men speak it that you have heard the rest.

AUFIDIUS:
Horse, look to the gods shall her marriage on their lawful and so so die all.

KATHARINA:
What news is this that we shall frest the people and warrant
I am a state to see thy lips his face.

LEONTES:
I will not be bring him to him to him,
Which is think us the way
To be more be in this thing I had in the earth,
And I can say 'tis not for the duke of Plantagenet, but you have the power,
To make a man all the dearly of the country's law
And thus I thank you, say how is he that doth he was a truth
Than the county dead is come.

CLARENCE:
I will not show you the dishonourable.
我鼓起很大的勇气，才决定写这一封痴情的情书，想让你知道我有多么的爱你，我不知道你是否心有所属，但是，我不会放弃的，我早已认定你就是我此生的唯一，不管你是否接受我，我对你的爱，永远不变！

我不知道我是从什么时候开始喜欢上你的，只知道，在很久以前，你的身影总是会出现在我的脑海中，开始我并不知道那是爱，但，这是我给你的时间话，会说了算缘的答应，我的继续卖水果了，不然这个月工资会扣得只剩下几块钱，我得努力赚钱让你过得好，为了我们的未来，我会认真卖水果，要知道，我是在为你而努力！证明我对你的爱!

每一次激动的眼泪都因为你的情深意重，体贴入微。“你累了吗？不要再熬夜了！”面对心疼我的你，我只能放下手上未完的文字，揣着一怀甜蜜的满足，欣然入梦。安静的梦中依然如故，满脸的泪花径自地流，耳边响起你的呢喃：到哪里找那么好的人？不管天长地久，也要牵你的手入梦。

我真的不想这样，我不想在想念你的时候哭泣。可是，我做不到，不让眼泪流下，为什么我控制不了它？不让它从我的眼眶中流出，朋友说想念是一件很痛苦的事情，今天的我才真正的体会到，原来想真的让人很痛苦，也让眼泪绵绵不绝，只为想念你而流。。。

Sincerely,

Chatbot1022
Sampled text is:
who here loves to eat lemons? work up the to the fire a chat we comp cookin
PART
U156 ...
U115 have you want to some cont to never to becorder
i am to shoot of you wanna be that bear all the book hou U110
U102?
what work to man ..... lol
i have a more to here U116 .. how are you goin and for can play whil and a chat
the pic?
hi U35
hi U35
not to be U12?
PART
hey U15
lol U31
PART
lol U35
Hi U15
PART
lol U116 hun with a long me
PART
lol U31
hi U33
any guys wanna chat U17 ... lol
ok i wanna chat the botter to the start of you me

JOIN
. ACTION wants to U19?
any guys U19

Second sample. “Any guys wanna chat U17 ... lol”
Generated samples:

- Seed: 2500, temperature: 0.7, start_text: “Harry”:

```
Harry had gone angry look out of them when it was pretending up at winky, pointed on a wear of a differ red now. He and Dumbledore had their easy to the end of the night but the pale of his leg dead happe with him the chair better as he had finished by its rather closer hands he was socker than he had been speaking very down in the room on.

"It was drown in a hundred crowd perhaps for you where there realized her magical champion of Hogwarts"

"Yeah, you know." said Harry.

Malfoy and Hermione, the pair of the gab. Madavia Maxime would fit and looked as though he was had on e of the morning. Mulgerin really approached into her too look at Moody was occurring a very early like a wizard with his head had seen its spot, in her sudden helped over the side of everybody wouldn't t hey himself in the gool, which was lay now dead, had gone or save his mood. He unfolded his magi cal eye who had mentioned for this head with the Daily Prophet grig. He remembered Harry watched them any other
```

- Seed: 8000, temperature: 0.56, start_text: “Hermione”:

```
Hermione was trying to keep his face and he took a noteing bird. He pulled off the stairs, he had said a slight with a slight floor and went on a hand from the top of his face was still surprising at him. He was going to fill right in his hand and sprinting at Harry had been so surface as they had been read of his head of the sliver closer than he had been passing Harry and Ron was beaming at the castle and started to his feeling sort of a few point, then he had very seen in his head and disturbed the eyes and started something later as he was the tip, his four and settled out her stood at the cold voice, and the terth at his eyes were standing the ground with a periness of the hair and called with him. He had stood up the back of ground and saw the dark break the back on the Quidditch world cup to the same with them. He was working at Hermione and her musital before the stood of the lake of the deep floor, and then swept off the side of the ground and stared
```

- Seed: 8687, temperature: 0.6, start_text: “Ron”:

```
Ron was still distracted the trouble show and his best head of Gryffindor Tower. They were fifteen on the stands with a great voice.

"Professor Karkaroff and Charlie and they trouble wizard what yeh yeh what the first by the house-elf to someone the places of noticest anyone else, and went on the second name on a slight and pointed the mouns of the floor, while going his going in the fire. A serious at the grining shones of the stand of his voice eyes around them. Harry stared at his eyes a his skewers and said, "Thz ey'll have to come to the castle, who was an anything next to the unforgoones in the fire sarrow a dress, the third task had told then standing on a paunching the ball of a night of his face was sayin g to stay there in the color, but for his hands was now said a great night. The first back at Harry h appily. "I don't know where it were the only observe accorn of parents with a bit of a rush were read ing him. He was turned his own from her window. He
```
Sampled text is:
DOWNEY: at a burrito and service. It's a little onions; that's why I go back to can a bacon St. mine in the front with the food was good, they walk out is realizing my comporental name. I was disappointing. Beef told the scallops, and the sauce and having for it as well, but I will really stake you to realized cardeatle. Being in the back if you do chang to their food - and it is both time me mine, so we could come back as we ordered it eat with the bagels for $7 for the staff was good, relaxed by the subs or "major Lemon Grance Falafel was a nice and opachists.

I received maybe nice place and it's going here. I like and offer somewhat clearly and back there...but so that my own were that we promped in the window sometimes not some comongs up on the area is awful feeling us on the filet mignon. It's your dishes sole know. I get some tofu even at The Ponele's spicy being sizes for it. The food was great - and with my 2nd and sat here because it's an hugive, due to the chaise, or it's th
References

Christopher Colah’s Blog: http://colah.github.io/

Andrej Karpathy’s Blog: http://karpathy.github.io/2015/05/21/rnn-effectiveness/

David Silver’s Talk: http://videolectures.net/rldm2015_silver_reinforcement_learning/

Geoffrey Hinton’s Coursera Talk: https://class.coursera.org/neuralnets-2012-001/lecture
When the Tao on its ald not tract to it in the worl seems and not the Tao.
When the Tao is nowhere is the world is a than then is the world.
The Master does in the Tao, to the is loth seaming of the Tao.
When the Master wan the Tao is the wager.
When you wand to ene ot less the is tral things; thas is why the Tao, the mistons the harken is the fore is nothing. If you on to prose, you and there is nothing seare ar in no forsed.

The Master does he see and not in donent, you are things himself to the people and not in the ardous an hemple. 21
The groes the condernt be a pain of contern. It is his lidern in the Tao will all things and in fored. The Tao is the Tao.
The Master dont to bach fol the Tao is the stouss and goen both is it in the are is nothone is that they dont to the core frear the Tao.
The Master to he doesnt man ould with the Tao seept experse, and complentent thin does an harmine the sill for without the Tao is nothing is bore and feem the people ar