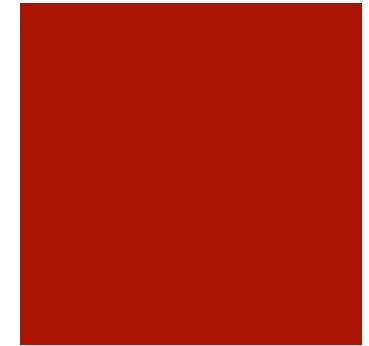




# Tweet like Trump

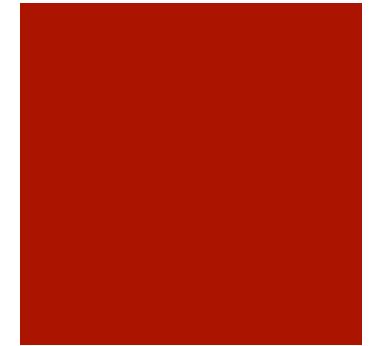
An exploration of deep learning and natural language processing into the mind of D. Trump

Presentation by Ed Chin  
November 11, 2016



# What did I do with

- 26,204 tweets from Donald Trump from 2009-2016
- 10 campaign speeches in 2016
- And many sleepless hours agonizing over the results of the presidential election
- I built a character-based text generation model using a recurrent neural network (LSTMs variant) to simulate **Trump's tweets**

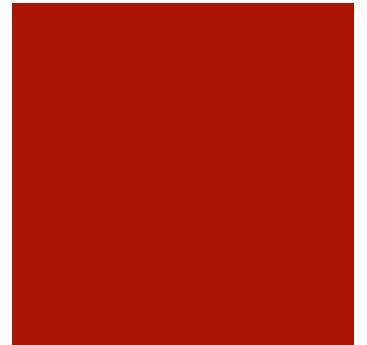


# Why Donald Trump?

- Because he is a polarized figure who trolls the twitter-verse like nobody's business
- Because he exuded so much negative sentiment in his daily rhetoric it would be a complete waste of time performing a sentiment analysis on him
- Because you are going to hear from him for at least four more years...might as well get used to it

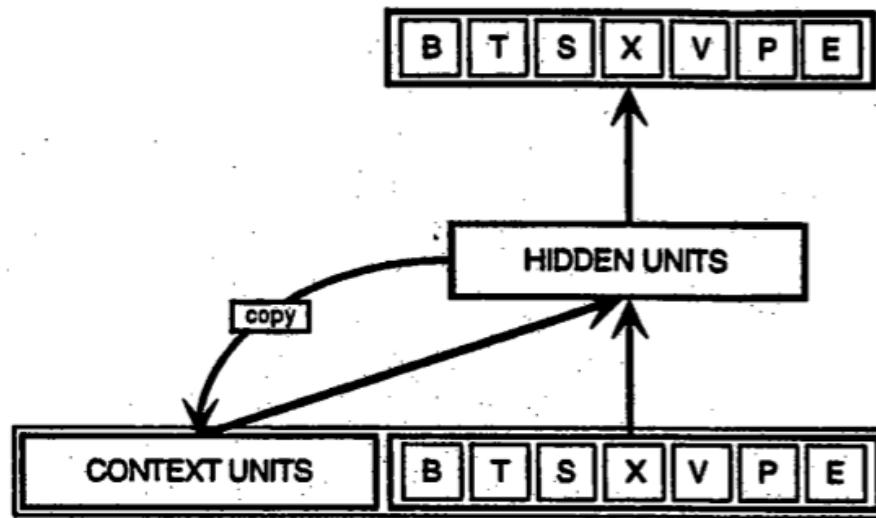
# Ok, let's get the boring technical stuff out of the way

- Going Deep...



# Why Recurrent Neural Network?

- RNN takes as their input not just the current input it sees, but also one step back in time
- It combines the present and the recent past to respond to new data, much as we do in life



- With its ability to have “memory”, RNN dramatically improves performance over CNN (Word2Vec) or Markov Chain Generator

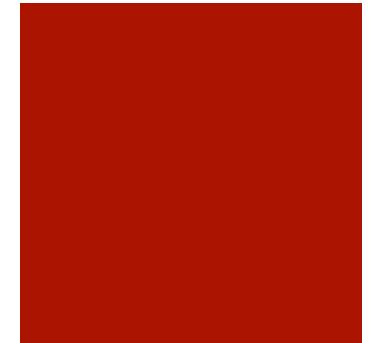
# What is Long Short Term Memory Units (LSTMs)?

- LSTMs is a variation of RNN that is surprisingly effective in combating the vanishing gradient problem
- LSTM helps preserves the error that can be back-propagated through time and layers
- Text contains **recurrent themes at varying intervals**, something that LSTMs are good at capturing

# How did I train the model?

- Single character classification problem with 52 classes (unique characters) and as such is defined as optimizing log loss, as opposed to classification accuracy
- 2,237,658 characters in corpus
- Two layers LSTM with 256 hidden nodes in first layer and 512 nodes in the second layer, with 0.2 dropout after each layer (~1.9mm parameters)
- Model was optimized after ~**35** epochs at 1.0306 loss rate over a 2 day period

# Techniques and Tools



gensim

 python™  
Natural Language Analyses  
with NLTK



<HTML>  
HTML



{CSS}  
CSS



(JS)  
JavaScript



and the awesome

 **jQuery**

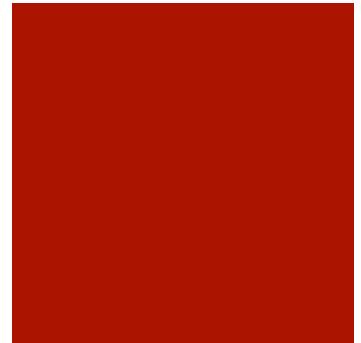
Now comes the fun stuff...  
Boom-shaka-laka

- App is live - <http://52.55.95.182:8080>



# Step 1: Which Trump do you want?

- Increasing temperature, aka “Idiocy Factor”, will give more diversity but at cost of more mistakes



**Idiocy Factor: 0.1**

Robotic Troll   Angry Drunk



Prime your own Trump's Tweet: (Optional)

Seed Text:

Trumpify

**Topic: Random Topic**

[Random](#)   [Losers](#)   [Media](#)   [Obama](#)   [Hillary](#)   [Dumbest](#)   [Great](#)   [Real](#)

**Idiocy Factor: 1.5**

Robotic Troll   Angry Drunk



Prime your own Trump's Tweet: (Optional)

Seed Text:

Trumpify

**Topic: Random Topic**

[Random](#)   [Losers](#)   [Media](#)   [Obama](#)   [Hillary](#)   [Dumbest](#)   [Great](#)   [Real](#)

# Step 2: Choose a Topic

- The key words were chosen based on a topic modeling exercise using LDA

**Topic: Media Disdain**

Random   Losers   **Media**   Obama   Hillary   Dumbest   Great   Real



**Topic: Dumb, Terrible or the Worst**

Random   Losers   Media   Obama   Hillary   **Dumbest**   Great   Real

## Step 3: Click “Trumpify” to get live tweets

- Adjust the sentence length using the slider
- Wait ~5 seconds for the model to generate results



Is @realdonaldtrump debating crooked  
@hillaryclinton and this country is a great guy.  
The press was making a good president of the  
country in the world. I have to speak the people  
of new york, ive got to walk. And they dont want  
to ask anybody to be working at the polls. You  
know,



RETWEETS LIKES  
**15,075** **37,794**

8:10 PM - 10 Nov 2016

# Sample Outputs from Model

- Republicans must stop relying on losers like @karlrove and jeb bush and what do it was a total disaster. not going to be a ticket of the polls. I mean, I mean, we'll have a country. I dont know
- Is @realdonaldtrump debating crooked @hillaryclinton said they want a great thing in the world trade deals. I got the money and they have to do it. So were going to be able to do well with the world. I was a tremendous deal with mexico and they dont want to do that. I dont know what
- Jon stewart is a joke, not very bright and totally overrated in charlotte of the world to call for a record the real work not to see each other for a support. The republicans have scandal country that i had a long time i was a no time for the history of the president obama is a total disaster

# Reflecting on Model Results

- Very impressive considering the model had to learn English completely from scratch and know where to put commas and spaces
- The trained LSTM model shows glimmer of insights
  - It frequently supports its own argument
  - It rambles just like Trump
- Model does a good job walking the fine line between memorization and creativity

# Future Work

- Remove punctuations from source text, and therefore reduce model's vocabulary for training
- Train model on padded sentence rather than random sequences of work
- Add more layers and neurons and tune the dropout percentage.
- Train over 100 epochs
- Apply regularization to combat overfitting

# Contact Info

- Ed Chin
- Email: [echin6@gmail.com](mailto:echin6@gmail.com)
- Linkedin:
  - <https://www.linkedin.com/in/edwin-chin-62392b1>
- Blog: [echin6.github.io](http://echin6.github.io)
- Repo: [echin6/metis\\_projects](https://github.com/echin6/metis_projects)