

The **Best** Words

Leon Zhou

lzhou95@gmail.com

LinkedIn: [zhouleon](#)

GitHub: [tetrahydrofuran](#)



**“I know words. I have the best words.”
- Donald J. Trump**

The Data



via precompiled
database



officially released
remarks / transcripts

Generation Methods

Current State	Next Word (p)
---------------	---------------

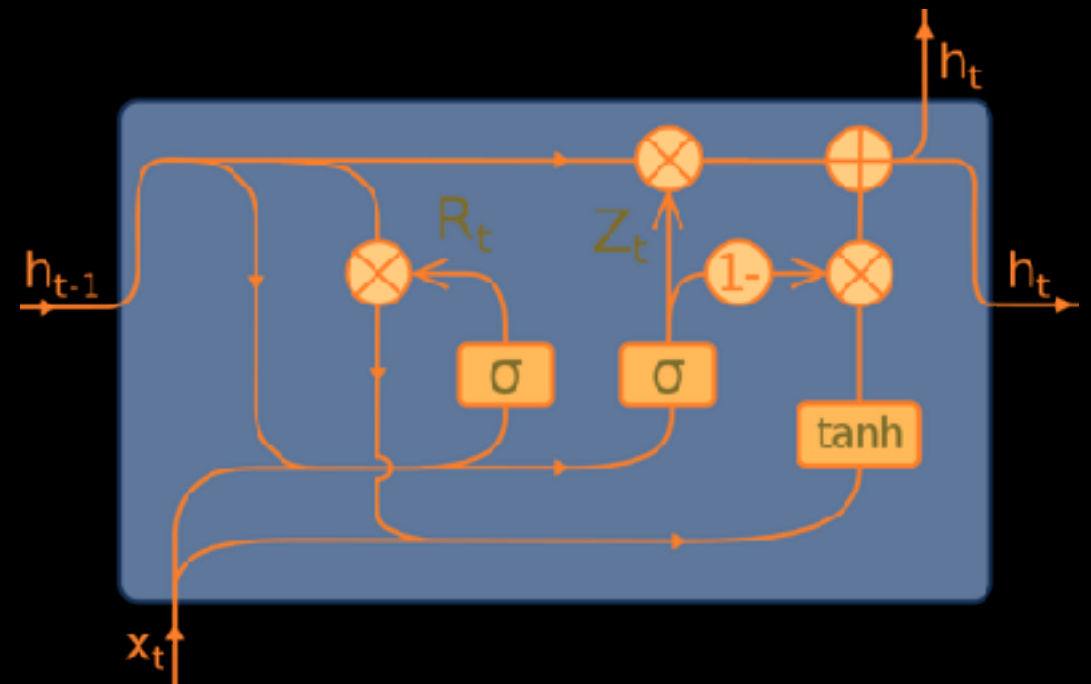
...cut taxes	
----------------	--

	bigly (0.70)
--	--------------

	. (0.23)
--	----------

	soon (0.07)
--	-------------

Markov
Chain



Recurrent
Neural Network

Sample Generated Text

Obama. We'll have a lot of people that do we — okay? I'll tell you they were a little bit of it.

Remarks “Hidden-ELU” Unidirectional, Temperature=1.0

China has agreed to buy massive amounts of the world — and stop what a massive American deal.

Tweets “Flanked-ELU” Bidirectional, Temperature=0.2

FBI are wearing Revlon eyeliner today.

Tweets Markov Chain

@realDonaldTrump or @fakeDonaldTrump?

I want them all to get together and I want people that can look at the farms. So we're going to soon now for a big day.

I want them all to get together and I want them — because everybody wants this solved. But a lot of bad things have happened.

@realDonaldTrump or @fakeDonaldTrump?

I want them all to get together and I want people that can look at the farms. So we're going to soon now for a big day.

Remarks “Flanked-ELU” Bidirectional, Temperature=0.8

I want them all to get together and I want them — because everybody wants this solved. But a lot of bad things have happened.

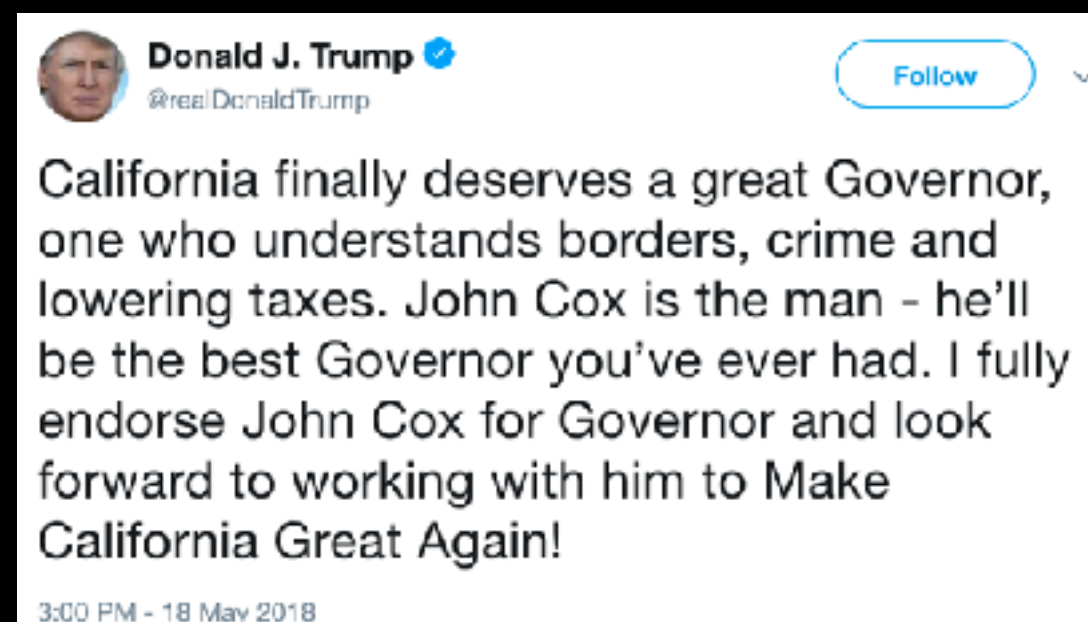
“Remarks by President Trump Before Marine One Departure”

@realDonaldTrump or
@fakeDonaldTrump?

California finally deserves a great
Government to Make America Great
Again! #Trump2016

@realDonaldTrump or
@fakeDonaldTrump?

California finally deserves a great
Government to Make America Great
Again! #Trump2016



Tweet "Hidden-ELU" Bidirectional, Temperature=0.8

Thank You



Leon Zhou
lzhou95@gmail.com
LinkedIn: zhouleon
GitHub: tetrahydrofuran

Appendix

Recurrent Model Architecture

6 variants (3 base + uni/bi-directional)

```
def rnn_basic(X, y, bidirectional=False):
    seq_input, x_ohe, output = __init_shared_layers(X, y)

    x = x_ohe(seq_input)
    gru = GRU(units=256, dropout=0.2, recurrent_dropout=0.2)
    if bidirectional:
        x = Bidirectional(gru)(x)
    else:
        x = gru(x)

    x = output(x)
```

```
def rnn_flanked_elu(X, y, bidirectional=False):
    seq_input, x_ohe, output = __init_shared_layers(X, y)

    x = x_ohe(seq_input)
    x = Dense(units=256, activation='elu')(x)
    gru = GRU(units=256, dropout=0.2, recurrent_dropout=0.2)
    if bidirectional:
        x = Bidirectional(gru)(x)
    else:
        x = gru(x)

    x = Dense(units=512, activation='elu')(x)
    x = output(x)
```

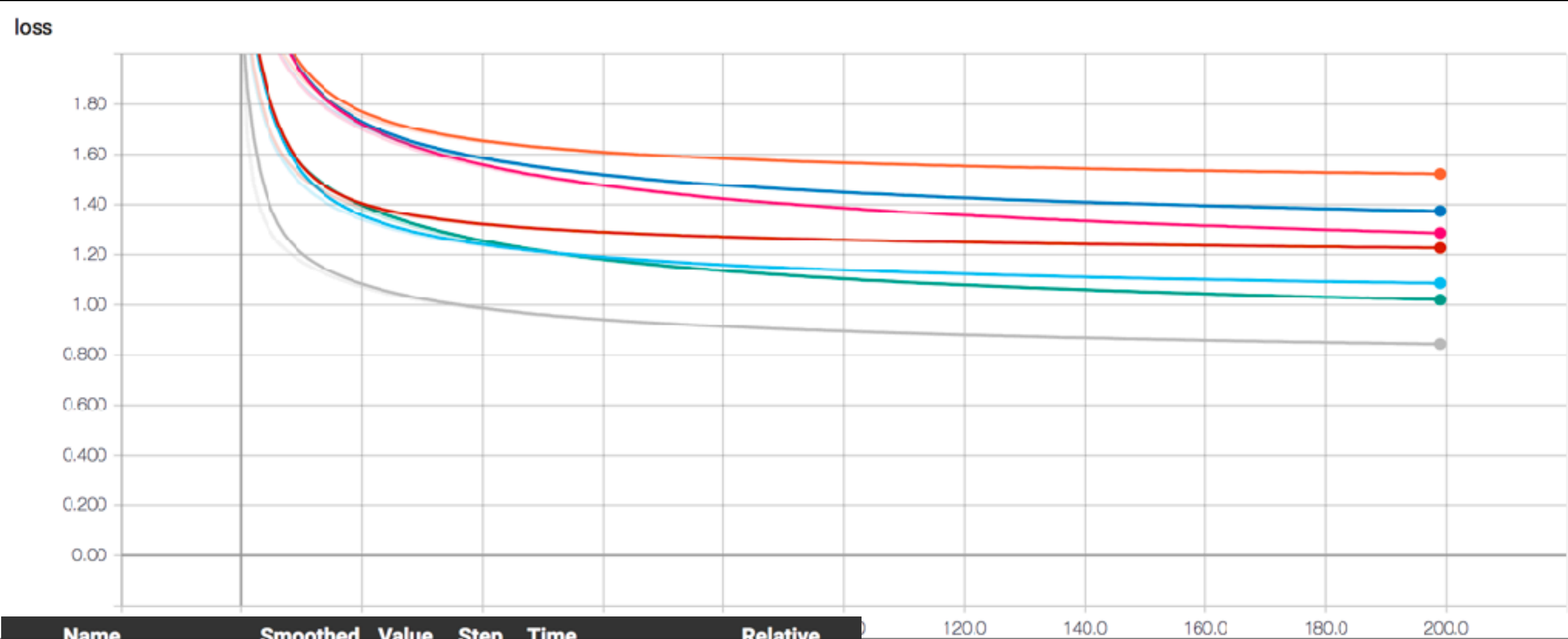
```
def rnn_hidden_elu(X, y, bidirectional=False):
    seq_input, x_ohe, output = __init_shared_layers(X, y)

    # Network architecture
    x = x_ohe(seq_input)
    gru = GRU(units=256, dropout=0.2, recurrent_dropout=0.2)
    if bidirectional:
        x = Bidirectional(gru)(x)
    else:
        x = gru(x)

    x = Dense(units=512, activation='elu')(x)
    x = output(x)
```

Adam Optimizer, 0.01 LR
Categorical Cross-Entropy Loss
200 Epochs
4096 Batch Size
40-Character Window Length

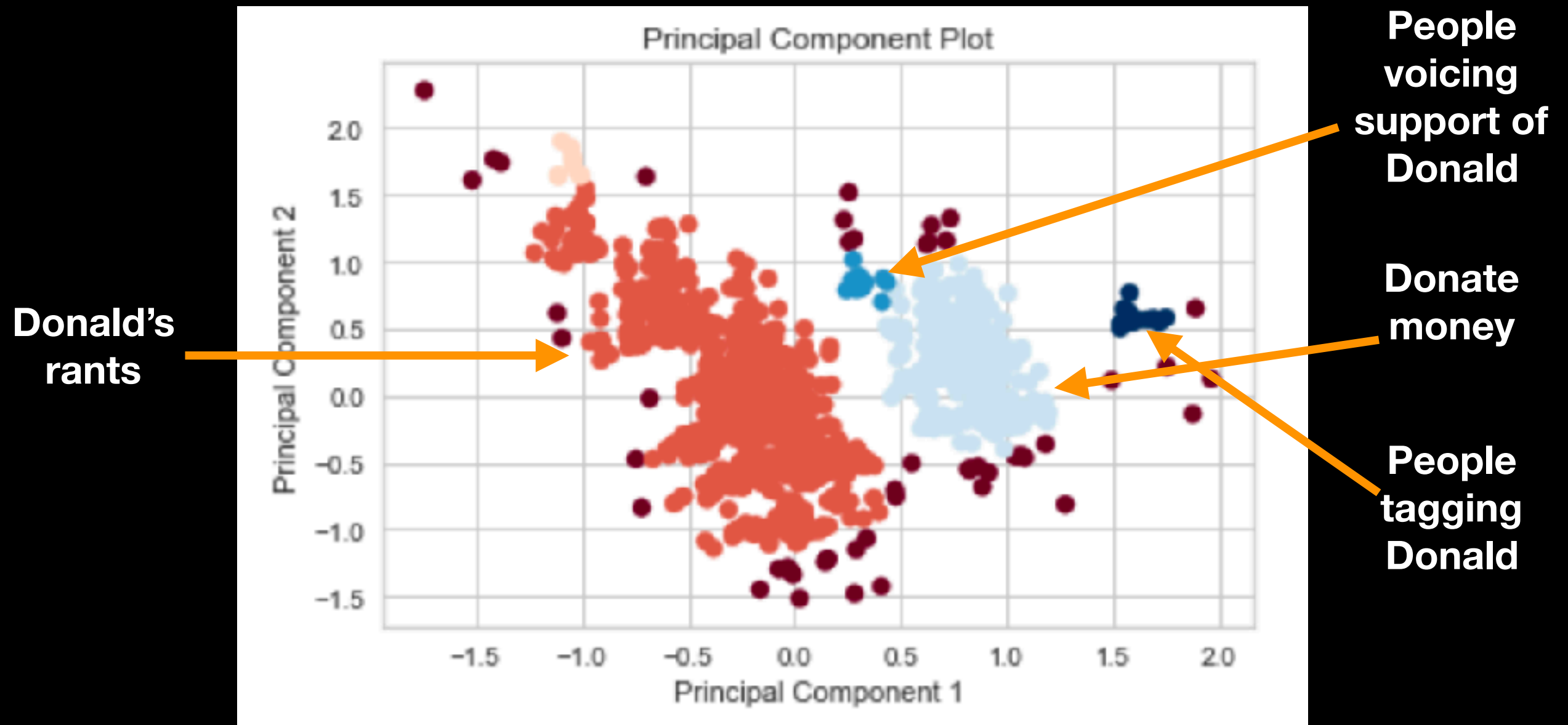
Model Performance



	Name	Smoothed	Value	Step	Time	Relative
●	remarks_basic_uni	1.227	1.226	199.0	Tue May 29, 12:04:36	5h 43m 32s
●	remarks_flanked_bi	0.8441	0.8434	199.0	Wed May 30, 18:04:59	3h 21m 55s
●	remarks_hidden_uni	1.087	1.086	199.0	Tue May 29, 17:52:14	5h 45m 51s
●	tweets_basic_uni	1.522	1.522	199.0	Tue May 29, 02:24:30	3h 51m 31s
●	tweets_flanked_uni	1.287	1.286	199.0	Wed May 30, 12:10:45	2h 1m 6s
●	tweets_hidden_bi	1.020	1.019	199.0	Wed May 30, 14:34:44	2h 23m 11s
●	tweets_hidden_uni	1.374	1.373	199.0	Tue May 29, 06:19:06	3h 53m 24s

Epochs

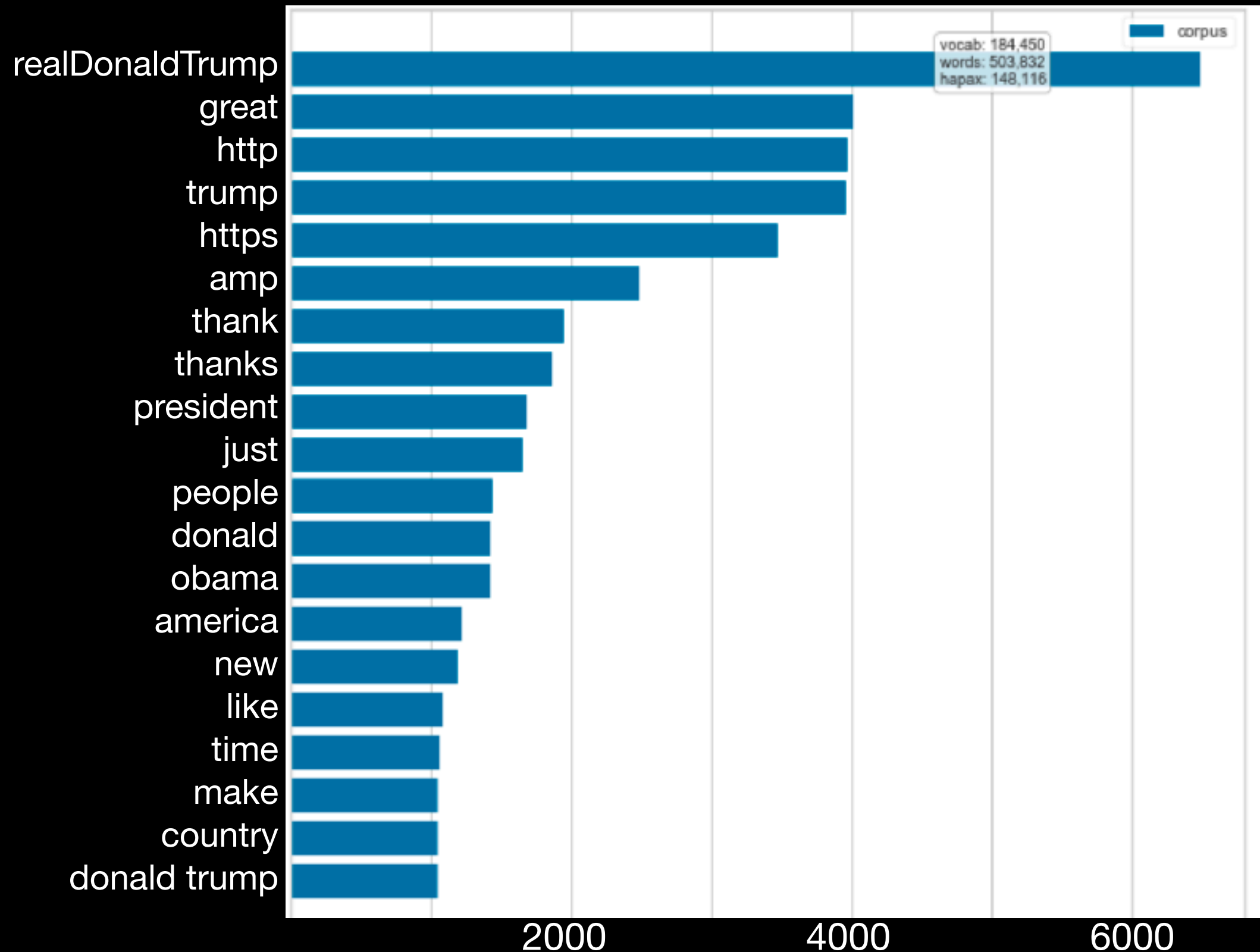
PCA and DBSCAN



eps = 0.15
min_samples = 10

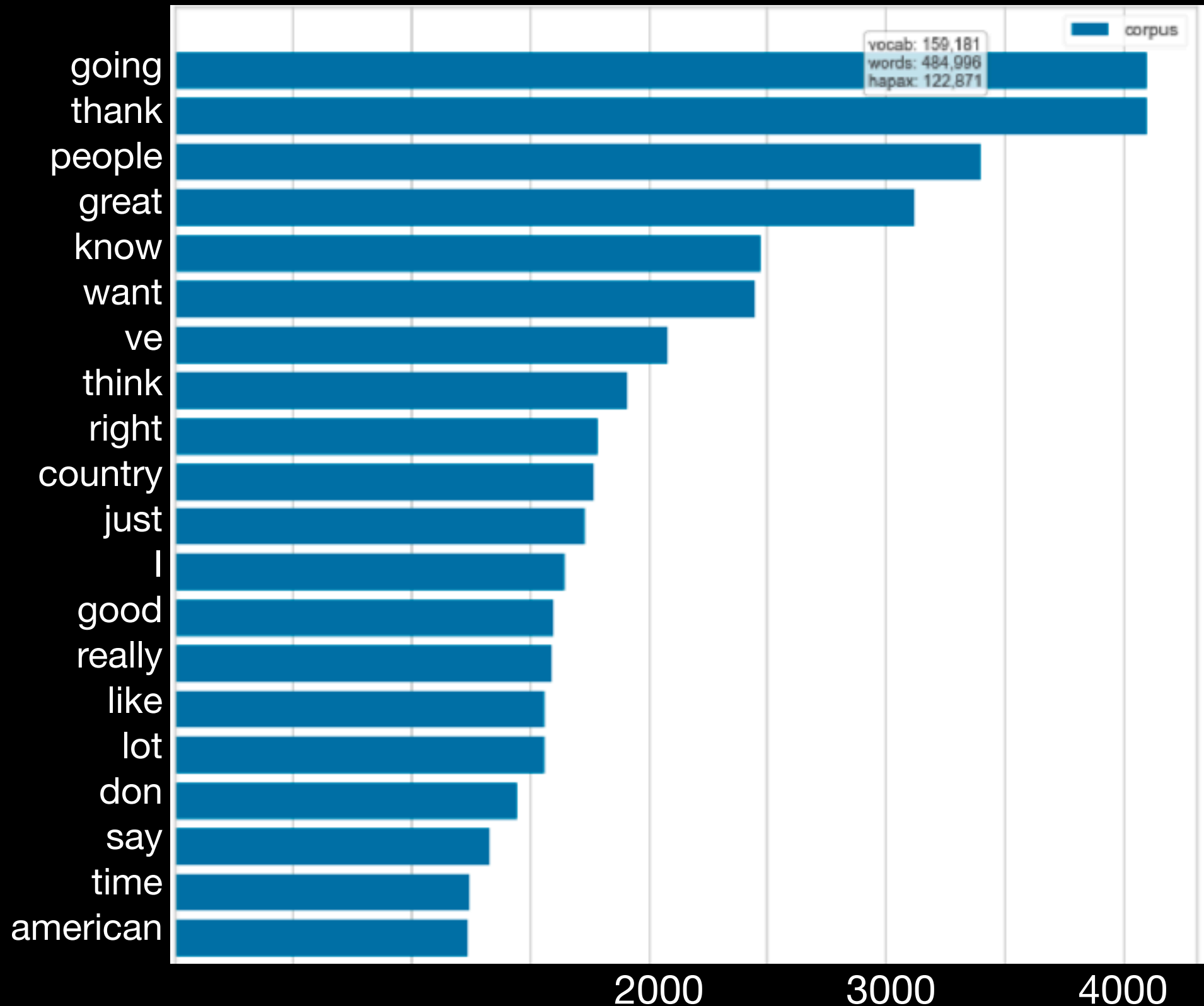
Top Token Frequency

Tweets



Top Token Frequency

Remarks



REGULAR EXPRESSION

65 matches, 156164 steps (~46ms)

```
:/ ([PRESIDENT TRUMP|THE PRESIDENT]): (.*) ([A-Z']+) ?[A-Z']+?:|Q ?|END)
```

TEST STRING

SWITCH TO UNIT TESTS ▶

and it's just an endless honor to be a part of it, and even in any a small way. So thank you, sir. **THE PRESIDENT:** Thank you, Stephen. That's great. A great job you do, too. **GASPAR:** Good afternoon, Mr. President. It's an honor to be here. I'm Kristin Gaspar representing the largest county here today, San Diego County. I have 3.5 million constituents that I'm responsible for their public safety. If you look around this room, your tiny but mighty team, this is what Governor Brown classifies as low-life politicians. Well, here we are. You've heard about the problems. You've heard about the statistics. And I could have thought of a million things to say to you. I have a stack of 3,000 emails in my office. These are the emails that have come in – thank yous, people supporting what we're doing. And I have a tiny little stack of less than 50 where people are very upset with what we're doing in San Diego County. **THE PRESIDENT:** How is the wall going? How is the wall? (Laughter.) **GASPAR:** It's going. It's going. **THE PRESIDENT:** We're getting it built, right? **GASPAR:** It is being built. **THE PRESIDENT:** They wanted it so badly – San Diego. They wanted it so badly. And I said, you know, if we build it, we will lose a big constituency, because there won't be anybody saying, "We want the wall." But we had to build it. So I know they're very happy about it. **GASPAR:** And I'd like to share with you a story, because sometimes humanizing the issue is really important. And a family reached out to me, and I brought with me one single photo on that plane, since the stack of 3,000 emails is a little difficult to carry. But that photo was the last photo taken

SUBSTITUTION

Example of complex regex extractions

A Unique Style



Donald J. Trump ✓

@realDonaldTrump

Follow



Sorry losers and haters, but my I.Q. is one of the highest -and you all know it! Please don't feel so stupid or insecure, it's not your fault

6:37 PM - 8 May 2013



Donald J. Trump ✓

@realDonaldTrump

Follow



A TOTAL WITCH HUNT!!!

4:08 AM - 10 Apr 2018



Donald J. Trump ✓

@realDonaldTrump

Follow



....to President of the United States (on my first try). I think that would qualify as not smart, but genius....and a very stable genius at that!

4:30 AM - 6 Jan 2018

Quirks

Tweet-trained models deal with uncertainty with excessive username mentions:

This is too good to be true for Republicans are a total job to community and endorsement is a complete course in the country is a complete to see you can t know it will be a careful course in the country is a complete course in chicago. -- @realdonaldtrump @realdonaldtrump @foxnews @realdonaldtrump @realdonaldtrump @mittromney @celebapprentice @realdonaldtrump @realdonaldtrump @realdonaldtrump @realdonaldtrump @realdonaldtrump @realdonaldtrump @cnn @cnn @cnn @cnn @cnn @cnn @cnn @cnn @jerrysalmond @realdonaldtrump @realdonaldtrump @cnn @cnn @cnn @cnn @cnn @cnn @cnn @cnn @jack_wallown @trumpchicago @realdonaldtrump @celebapprentice @realdonaldtrump @foxnews @realdonaldtrump @realdonaldtrump @macys @realdonaldtrump @celebapprentice @realdonaldtrump @realdonaldtrump @macys @realdonaldtrump @realdonaldtrump @realdonaldtrump @apprenticenbc @realdonaldtrump @realdonaldtrump @trumpchicago @realdonaldtrump @realdonaldtrump @realdonaldtrump @realdonaldtrump @mittromney @jerrysalor @realdonaldtrump @realdonaldtrump @realdonaldtrump @re

Tweet “Hidden-ELU” Unidirectional, Temperature=0.2

Quirks

... or hyperlinks:

Why didn't President Obama do something it was a complete course in the country is a complete course in the country is a complete security and we will continue to care of the country will solve the special country and the country is a complete course in the country is a complete course in chicago - and will be a great country great again! <https://t.co/usyaccclw>
<https://t.co/syukcbycck> <https://t.co/cuygpyccecy> <https://t.co/aqasyvice0>
<https://t.co/qcyccsyccps> <https://t.co/qxckcyccc4> <https://t.co/agec2yhccc>
<https://t.co/lswxxxsccc> <https://t.co/jcsyywwyoxx> <https://t.co/oc3uycocpcc>
<https://t.co/qysygcccol> <https://t.co/cssyacywcc> <https://t.co/ecymsycccccw>
<https://t.co/icycccyccc> <https://t.co/ycbuyccecc> <https://t.co/ysecxacco4ch>
<https://t.co/ec6rcs4crt> <https://t.co/as0cxckyco> <https://t.co/s4xocwpycc>
<https://t.co/xcsc3wyycc> <https://t.co/c7yschcooqy> <https://t.co/s4csmgrcccc>
<https://t.co/c4yxcpcccws> <https://t.co/syskcxxsxxh> <https://t.co/csyp4iiccc>
<https://t.co/aasyscrccc> <https://t.co/s4wawc5ycc> <https://t.co/ycyjpgswys>
<https://t.co/csyj>

Tweet “Hidden-ELU” Unidirectional, Temperature=0.2

Quirks

Remarks-trained models enjoy thanking the audience:

[illegible]

Remarks “Hidden-ELU” Unidirectional, Temperature=0.5

Markov Chain Text Samples

America the Emir of steel.

Democrats seem desperate.

Obama in the Clinton and ISIS like — enjoyed being in the haters and extreme leftist.

Russia thinks that I really cares about my presentation.

Immigration is coming.

Loser newspaper MUST READ.