

TrumpTwitter Bot

Using an RNN to generate Tweets

Ben Pierce

SCS3546 (Deep Learning - University of Toronto)

Semester 1, 2020

Purpose & Data Source

- Tweet-like text generation using an RNN
- Trained on a large volume of Donald Trump's tweets (data source: <http://www.trumptwitterarchive.com/archive>)
- With a good enough neural network, topics could be provided to the neural net, and opinions or sentiments could be mined and studied without having to interact with real individuals.



Model Types

- Word-based vs Character-based tokens?
 - Word-based tokens mean accuracy of the model is going to be much lower because there are many possible word predictions; however, we don't run the risk of the model predicting words that don't exist.
 - Character-based tokens are going to have a higher accuracy because there are only 26 character predictions (plus symbols); however, we do run the risk of getting nonsense words.

- Rolling sequences

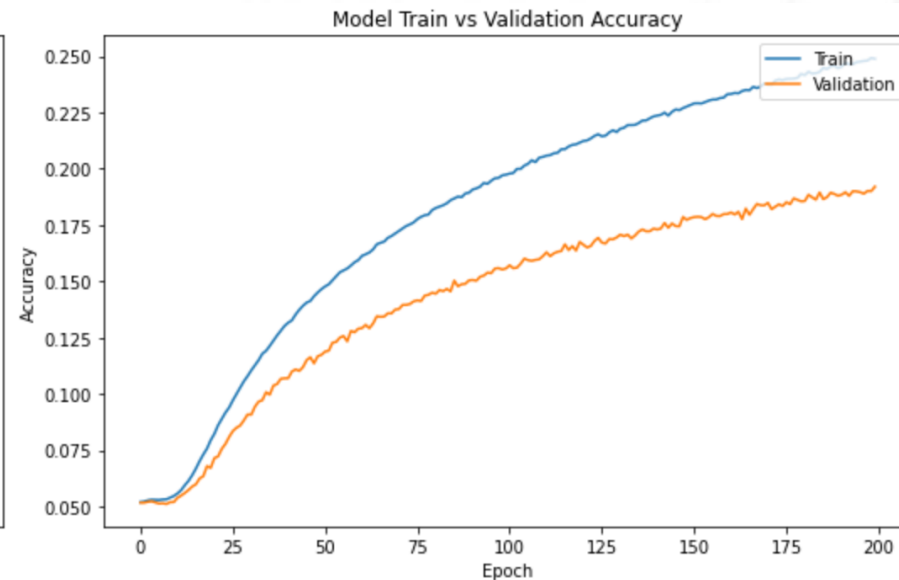
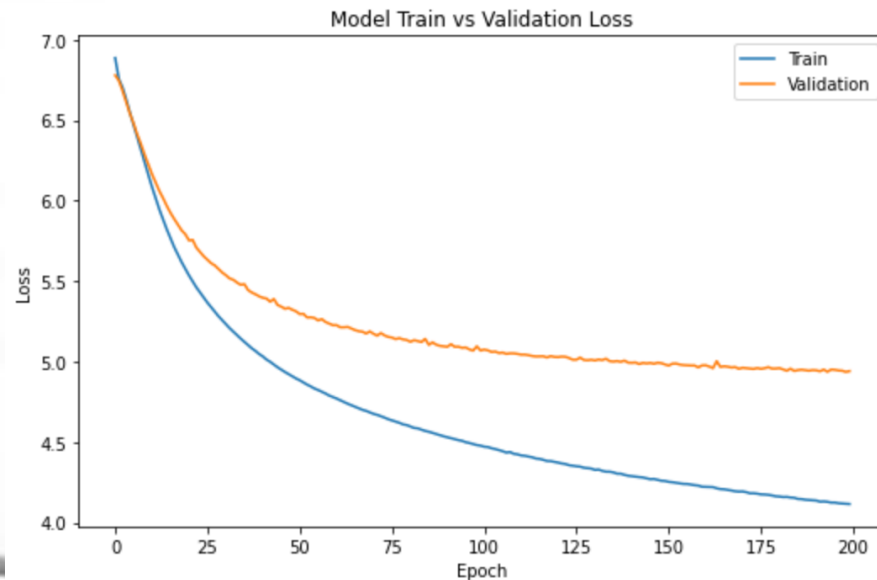
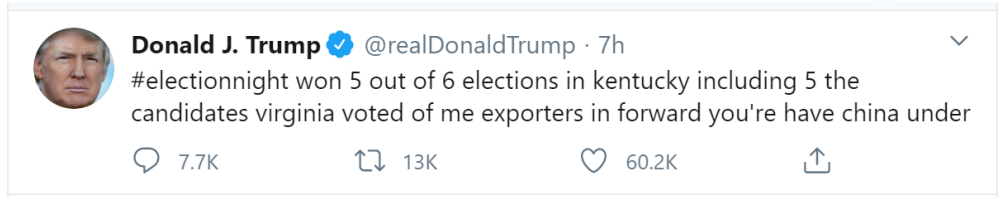
Sequence1	Sequence2	Sequence3	Sequence4	Sequence5	Label
Proud	to	welcome	our	great	Cabinet
to	welcome	our	great	Cabinet	this
welcome	our	great	Cabinet	this	afternoon
our	great	Cabinet	this	afternoon	for
great	Cabinet	this	afternoon	for	our
Cabinet	this	afternoon	for	our	first
this	afternoon	for	our	first	meeting

- vs Padded sequences

Sequence1	Sequence2	Sequence3	Sequence4	Sequence5	Label
0	0	0	0	Proud	to
0	0	0	Proud	to	welcome
0	0	Proud	to	welcome	our
0	Proud	to	welcome	our	great
Proud	to	welcome	our	great	Cabinet
0	0	0	0	to	welcome
0	0	0	to	welcome	our

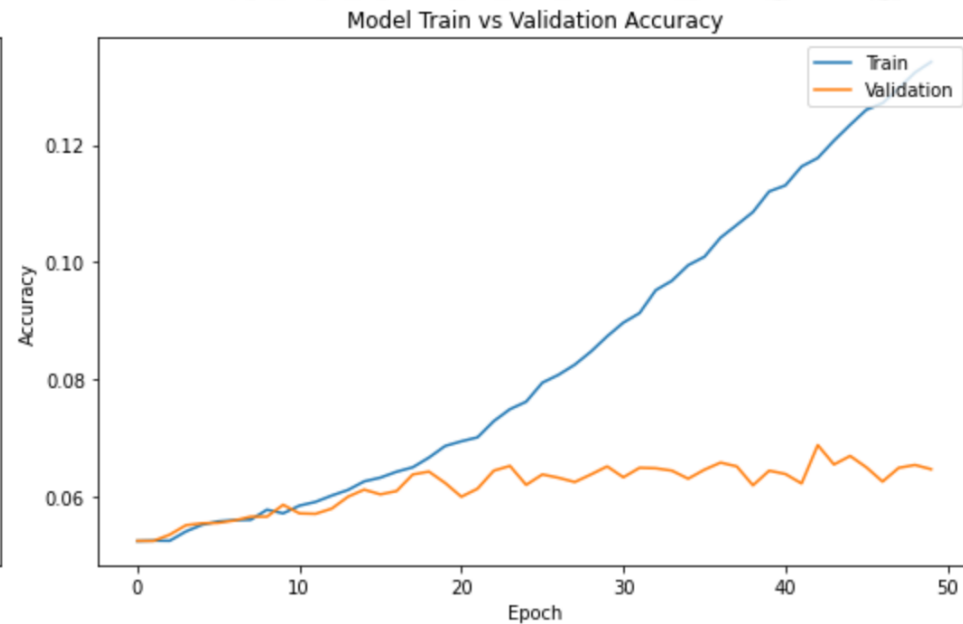
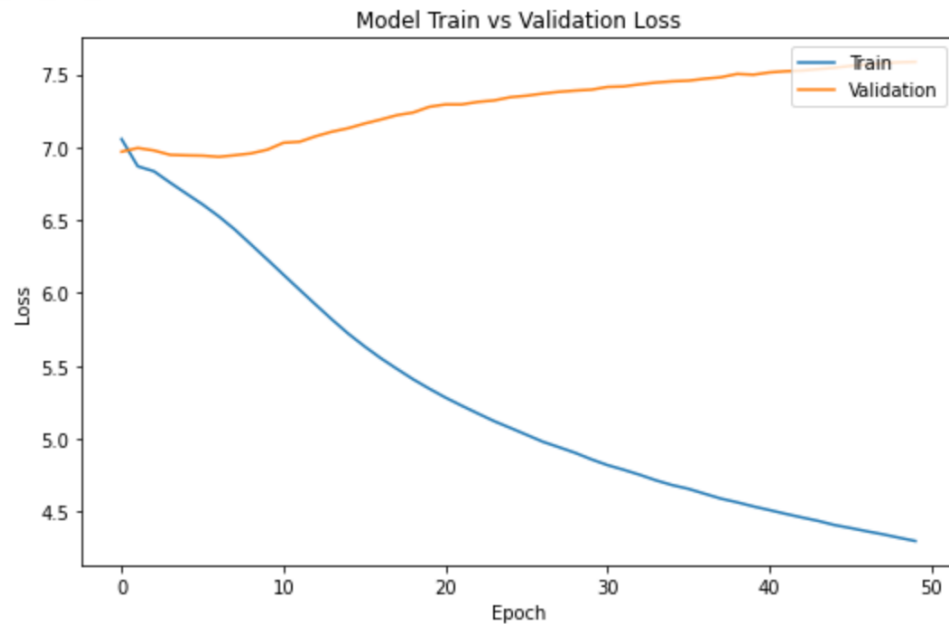
Word-Based / Padded Sequences

- Padded sequences outperformed Rolling sequences.
- Sequence length of 7 worked best for padded sequences.
- As expected, only about 20% accuracy but still, decent learning curves.
- Interestingly, was able to generate Farsi tweets.



Word-Based / Rolling Sequences

- Sequence length of 4 worked best for rolling sequences.
- Model performed very poorly:
 - Low validation accuracy
 - Overfit almost immediately



Character-Based / Padded Sequences

- Validation loss & accuracy was very good (~85% accuracy) with a sequence length of 25.
- Created some nonsense words when predictions went wrong *“getting ready to meet thede orely in madoaooos”*
- *Showed good promise for further experimentation.*



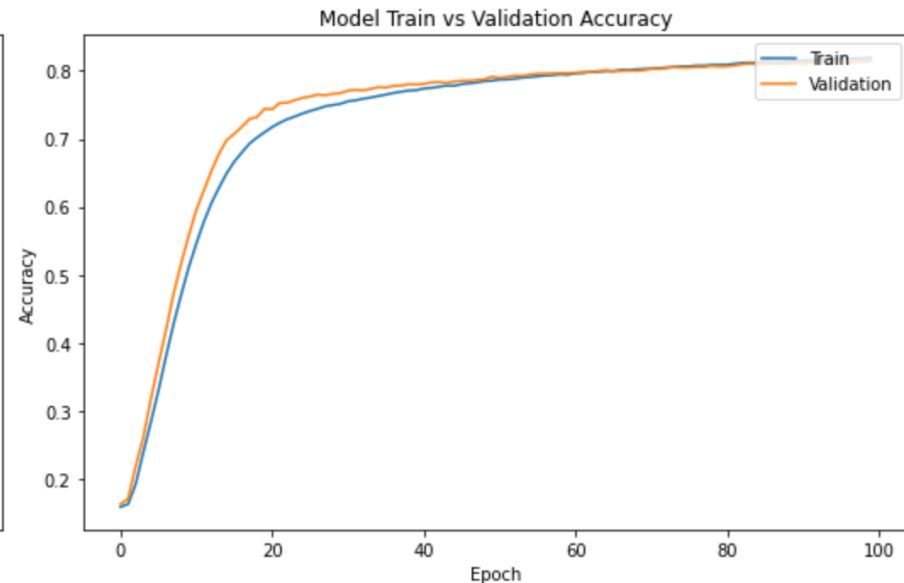
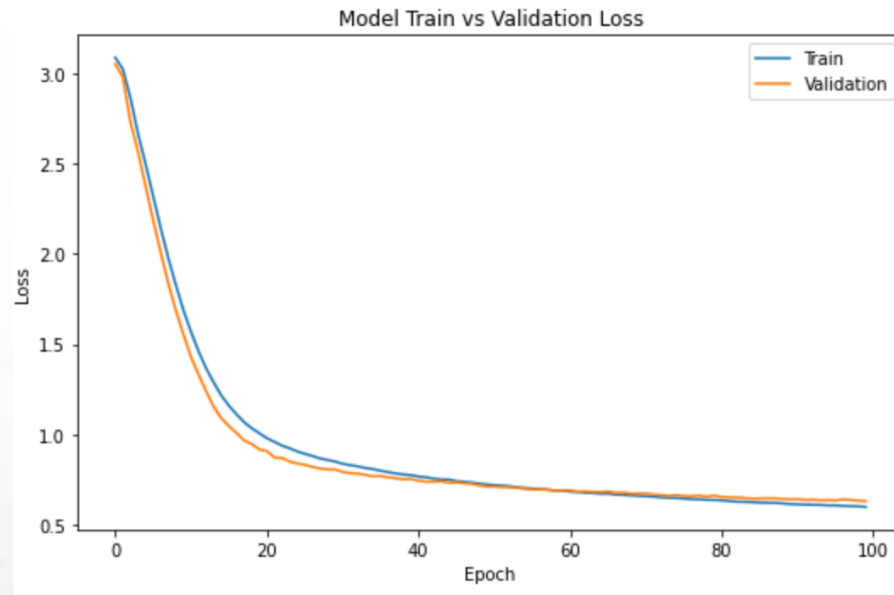
Donald J. Trump @realDonaldTrump · 5h

new polls say most americans and the world america will get it done! winh
siat r gitn nuper t he lost states that he easily would have won

3.1K

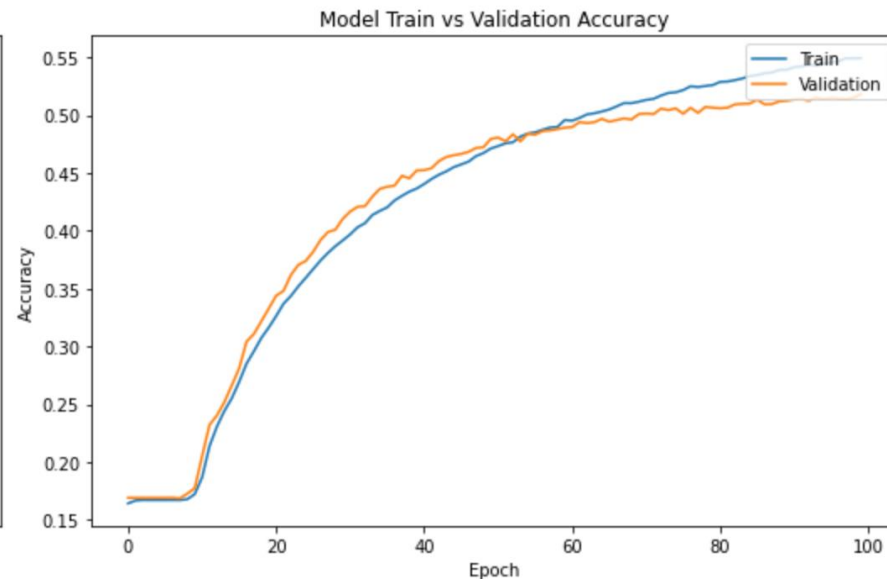
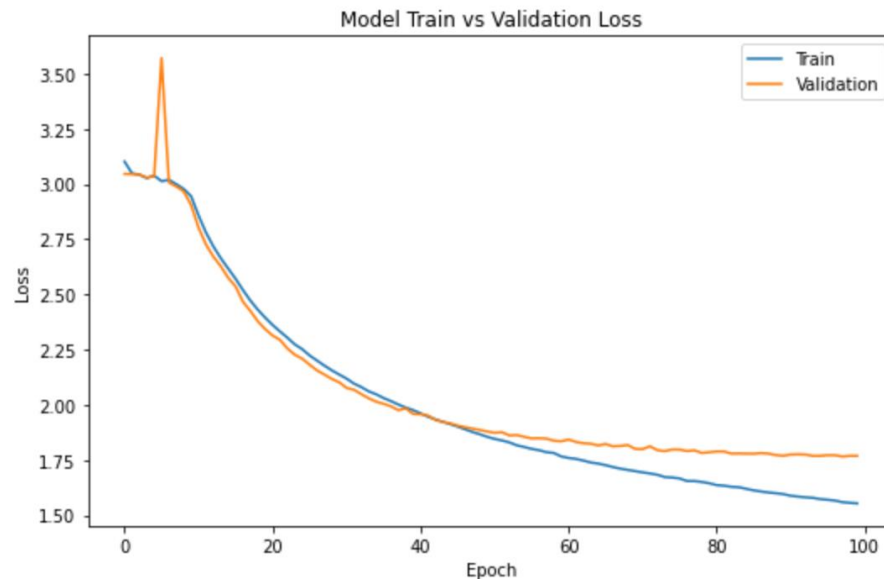
10.2K

41.2K



Character-Based / Rolling Sequences

- Validation loss & accuracy was mediocre (~50% accuracy) with a sequence length of 50.
- Created more nonsense words when predictions went wrong *“the democrat prosicule saadeing the presid”*
- *Would choose padded sequences over rolling sequences*



Summary

- Padded sequences greatly outperforms rolling sequences.
- *Word-based vs Character-based models both had good training metrics.*
- *Areas for improvement:*
 - *Additional data cleanup of messy Twitter text*
 - *Train on much more data*
 - *Train in a reserved, cloud-based, runtime rather than on Colab.*
- *Would choose word-based / padded sequences for future research.*

Thank You

- Questions & Answers



Donald J. Trump  @realDonaldTrump · 5h

it has been a scam to don't more and our impeaching possibility was no in the nothing new participated for not president and formal and conflicting

 3.1K

 10.2K

 41.2K



Donald J. Trump  @realDonaldTrump · 5h

breaking news: the president of they party couldn't that fact great the a hoax because that are a this far on mississippi people the will a thing

 3.1K

 10.2K

 41.2K



Donald J. Trump  @realDonaldTrump · 5h

وزیر امور خارجه ایران می گوید ایران خواستار مذاکره با ایالات متحده است اما می خواهد که تحریم ها برداشته شود @oann@foxnews نه، مرسی! to the wonderful

 3.1K

 10.2K

 41.2K



Donald J. Trump  @realDonaldTrump · 7h

the decision not to prosecute the the their a see bernie for it path in nancy that is senate for who us chris to federal trump want credibility

 7.7K

 13K

 60.2K

