Deep Learning DD2424 - Assignment 4

Anton Stagge

April 2019

1 Changes made to make a Trump Bot

First and foremost there was a lot of changed to be made in how to read the data. This was quite easily done in python however. I simply saved all the tweets in an array and the added a end-of-tweet character Ω at the end of every string. I made a guess that Trump never would have tweeted the character Ω. During training, I would then get *seq_length* number of characters from each tweet. If there were fewer than *seq_length* characters left in the tweet I would simply just return the remaining characters. This was to ensure that the end-of-tweet character was included in training for each an every tweet. When a new tweet is introduced, I reset the $h0$ vector to the zero vector, as this marked the beginning of a new tweet.

When synthesizing new tweets, I made sure that the maximum length was 140 characters and that if it ever predicted the next character to be Ω, I terminated the process. $h0$ was initialized as the zero vector but $x0$ can be specified, that way you can choose which character the generate tweets should start with, leading to some interesting results.

2 Result

I did some minor experiments with the value of *seq_length*. I started by having *seq_length* = 15, then I tried it with *seq_length* = 140 meaning the whole tweet, but I did not see any improvement, finally I tried *seq_length* = 5, which according to me was the best model. I base this on the fact that this model very fast managed to spit out text like https://t.co/RANDOM and &, which are two very common strings in trumps tweets. Below you can see some of the loss graphs and synthesized tweets from the models. I regret that I did not specify what epoch the model was in during training when recording the synthesized text, since the iteration count is not a fair progress indication between two models with different *seq_length*s.

2.1 Model with *seq_length* = 15

2.1.1 Loss graphs

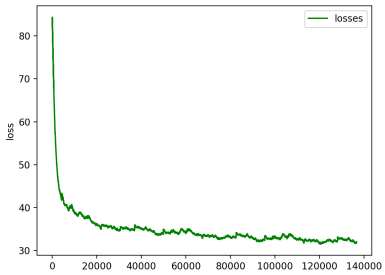

Figure 1: The graph of the loss from the first 3 epochs of training with *seq_length* = 15.

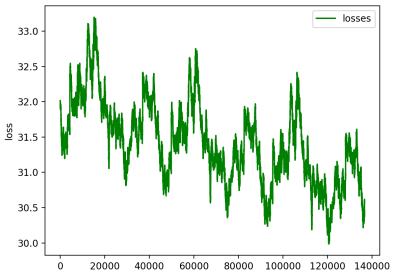

Figure 2: The graph of the loss from the second 3 epochs (number 4-6) of training with *seq_length* = 15.

2.1.2 Example synthesized tweets during training

Synthesized text at iteration: 0 with smooth loss: 83.918753
ó.? ,$2_{b|rE"”mFEM/éE0J[wI]Tf/gnb1NfMUG@'c=}$

Synthesized text at iteration: 500 with smooth loss: 71.968721
, lhenho rHsgeiltvaisu,glosd e eoao atbat sbkorxmopiatbtavretotyvud

Synthesized text at iteration: 1000 with smooth loss: 61.889268
earstghd ehl tho. j bN Br Sashhltgts w negs tgyg ertd itiledht nieghwinh doesataorrredrtI“ so2 hg ite td ts ttephg ts arbhkks thtdiiriwa

Synthesized text at iteration: 2000 with smooth loss: 50.114993
rirac wy to

Synthesized text at iteration: 5000 with smooth loss: 41.853774
aratEave giertreris cor @o s. anee femauporreeS.Saordig P ole cusaricpa!

Synthesized text at iteration: 50000 with smooth loss: 34.811514
uf Dax is oor sett!

Synthesized text at iteration: 100000 with smooth loss: 33.193011
ond. Tews gevangited crove id friys sexs Dell 9apemjor!

Synthesized text at iteration: 197500 with smooth loss: 32.655971
. Of) omsing lorutrermoroctalle. **Trump** Drooffal man Nouseormerranf. . . ust forth tid ale wryerlyone and pagl; cow Coes goufps://t.co/500S6qVE #

2.1.3 Example synthesized tweets with chosen x0 (after 6 epochs)

Starting with F:
FBI lamigets tomillal the doanisnark a toar ps timiving Mid the 41 fivered theevion stop the Un_Ams you bet. Weulla: dive scory event Hary.

Starting with h:
https://t.co/c8R

Starting with T:
The Gollions beingriet by and becaugs to Trump a the 2/tight woul 2reat peoploxeainney to to waking baconitt? A Hy7cush the lorto rorg. beche

Starting with &:
&as atrump; civpent so very tidnt has shis marth Coluttiagonicaugh Trump Lick comous col Wasin dest Breetratent on needer we luete have yeth

2.2 Model with $seq\_length = 5$

2.2.1 Loss graphs

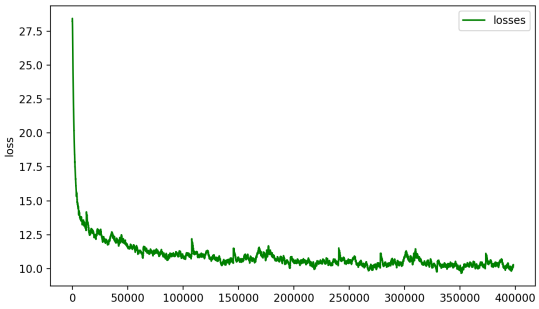

Figure 3: The graph of the loss from the first 3 epochs of training with $seq\_length = 5$.

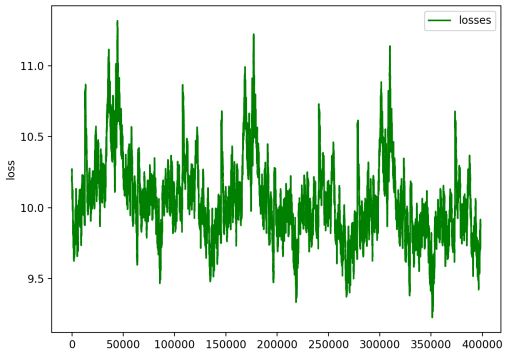

Figure 4: The graph of the loss from the second 3 epochs (number 4-6) of training with $seq\_length = 5$.

2.2.2 Example synthesized tweets during training

Synthesized text at iteration: 0 with smooth loss: 28.249919
HMK‘DB8p-,7Hv’y DQ8’]—ápaxwY 30óiKZ5P dr Y K V] Teeeto Y rv)

Synthesized text at iteration: 1000 with smooth loss: 22.464318
yp gvuc ewa y ti ooelu hiey v teehieyeo0tut lyayTtniWcs ltTn ean nsedt’t ttreir r

Synthesized text at iteration: 2000 with smooth loss: 18.525713
ntode:dblc ne nehede mi Ds o w eae nthete owEn nohaae e as nt Cd iazos aaahind s g wean p i rnlrwdtr aw hId wha Ueithioowg c ! O.noww ane

Synthesized text at iteration: 3000 with smooth loss: 16.562826
t @ip tonG tog th'otanho 'v defw bon co l ise d th uo. Eras C Bad tm nn P as trini oartneon oEx yag ve he p. tht .t ok et oUlray b In !in

Synthesized text at iteration: 5000 with smooth loss: 14.717798
ctbl'ayrisy, Con 1opSt HhaulSSSrpeete wust orse worerSzcatngons herperyy Rosis Nouy SarigoNl srer. ersK IeSdec in bearers un t hd ariallbrese

Synthesized text at iteration: 37000 with smooth loss: 12.429031
nton thesofustedtppif pard mave foren ws on. fs theas ithornenoud in afdings Snat fanis Melss @re https://Nens wlocusof stest. T. HRer Lor th

Synthesized text at iteration: 157000 with smooth loss: 10.674533
nDifperal bataldants of Crime of fild! Cho offich stritsa As o Stang he https://t.co/GScOR

Synthesized text at iteration: 384000 with smooth loss: 10.316723
to for winn a neail and in al dilstrad qullion Ut the mast ara letitary Mast. Uy amp; ming than mayd fork is "PI JAMyX2SB roLW

Synthesized text at iteration: 572000 with smooth loss: 10.331208
al The Deed their toom made of fume 7 loway the compelforianal, of lladeay seoraldaluatol Mbigting this Trump a me's great at thanl @Mfick ex

Synthesized text at iteration: 778000 with smooth loss: 9.925909
nif guker lave you not with Jums!o stin or wils is trump Wougatt Unitincer (fancens mand arisent to Canident the Kars. in the is nollaying of

2.2.3 Example synthesized tweets with chosen x0 (after 6 epochs)

Starting with F:
Fake Trodever, on there ploak douss will up resellge. Cho wast echoTle our the UNIxCTITH damast a wenflica gushy of Spuscy Re! Gut jush fort

Starting with h:
https://t.co/6LP5z99V

Starting with T:
The U.S. Det!

Starting with &:
& gussident on Rusial #Asting coun tha upn noal yeak one MACT-Metcen but it ray and bembees - &amp https://t.co/wy

Starting with M:
MAGA Car fust on mave!).dBR