# Assignment 4 - BONUS part- DD2424

Alexandra Hotti

June 13, 2019

## 1 Introduction

This report contains the results from the bonus part of Assignment 4 in the course DD2424, were an RNN was used to synthesize English text consisting of tweets from Donald Trump. Data from the years 2015-2018 was collecccted from the following page, see Donald Trump Tweets. Tweets form previous years were excluded from the dataset due to memory limitations. AdaGrad was used for optimization.

## 2 Implementation Changes

The following changes were done to the implementation in the basic part of the assignment.

- The character § was used to signify the end of each tweet in the data set since it was not included in the dataset of tweets.
- A constraint was added to the function *GenerateSequence* which can be used to synthesize text. The constraint was that a maximum text of length 140 characters could be generated. This limit was used since this number of characters is a limit used by twitter.

  When a user requests a synthesized text of a specif length the synthesizing function makes sure that every tweet that is generated has a maximum length of 140 characters. However, if the user requests a tweet longer than 140 characters and an end of tweet character §is generated after for instance 138 characters. Then the next character generated would be considered the first character in a separate tweet.
- During training hprev was re-initialized to its default value after each training tweet.
- Also, at the end of each training tweet, if the *end-of-tweet* character was encountered at the middle of a training sequence, the training sequence was chopped at the character § such that the input sequence at that point became shorter. However, when moving on to the next training tweet the input sequence length was reset again and hprev was set to zero. This implementation can be found in the function *MiniBatchGD*.

Next, additional changes were implemented in order to improve the performance of the network.

- Several sequence lengths were used for training.
- Decayed learning rate.
- Gradient clipping.

# 3 Dataset Analysis

Many others have previously analyzed Donald Trumps twitter via for instance word counts. Thus, for reference here are some commonly used words in Donald Trumps tweets: people, hillary, trump2016, trump, america, #makeamericagreatagain, preseident, time, obama, tonight, clinton, country, bad, poll, crooked, join, amazing, enjoy, night, tomorrow, great, good, news, vote, love , big, jobs, years, foxnews, american, u.s, barackobama, true, cnn, democrats, border, wall, security, florida, fake, rimes, states, military, united, state and security [2], [1].

# 4 A Vanilla RNN

First a RNN without any of the optimization techniques above was trained.

## 4.1 Training Evolution

Below are synthesized text and smooth loss from the vanilla RNN during 100 000 update steps at every 10 000 update step. As training went on the network seems to learn short, accurately spelled words. Also, common words tweeted by Trump such as or close to: Trump, Fake, Fox and American.

iter = 1, smooth loss = -

□□‼訪□□e.□り⬇間8□t□ナ□☆□
大_:《□をう)□□f揺□V□トnす□□Vの大□統XrプGYóv⬅功□|□□□✛jジZ)□□-T=♥{♥′w☑
をe#Yいúお□□rẽ8祈□に□□□き訪た↔□ˋ”□“□●8Z●示«はるf□□“□□□r*史′□□í□ルくおざ»《|♥同ōお

---

iter = 10000, smooth loss = 56.3631
r sdMdarusp:: Treaime ban Semece ong/w.§
T Dosa JeMand 1C23- Chicd owas baf be htump: **@Tumpe** re. #Mo **Trump** se fo me @RAPe. htv.”§
”. §
”!§
GKNC

---

iter = 20000, smooth loss = 51.6755
http://xctr.Tr/GoJK’Sp70clfWavvtQ§Lnyq97zMf7§
T PWAS” TAR bathy bledle youte’t Prumpersonger mugher palk @Mmeatal˙ Thethank pagrurgted lo de

---

iter = 30000, smooth loss = 49.7754
unpend! §
”§
S whe vetior keyger” it do grai: @seeresise urd ecou Sanvar ow rebed ourteunsel !ckeDouryw Thag wht Yer. Tou!
”§
”§
” #Thunmonkerigld

---

iter = 40000, smooth loss = 54.7988
wromct-. TOREC - himpor! He Hixprote wetose Groing act boud 97% in afars. **Trump** abas periew, ar Protede ittı ancringer” Sonars wan ma#s: Th

---

iter = 50000, smooth loss = 49.938
as a one CoMillexeess luct’s doon yE!§
.PG:XHulAmested, sithide! Ad’tt MAgrould hst rucer poun we ve wo pistryer os wI Thaw 1 FWallistork is

---

iter = 60000, smooth loss = 51.8078
wroy, lill U .5.S. as amay ates/ spain Crencestreal §
W”@rmare ofea!!! I thea, Anety!§

"§
TAmeratCa willea4: https://t.co/UwUU0UzMrke ht.co/qZ

---

iter = 70000, smooth loss = 51.8067
ak8Km0S1D2x0oTDyVTrunMA htHinv: IScofs Proiaft foraaflig about https: The ardet peraise I wivod tite.S. A could'ses- palting to to it is gem

---

iter = 80000, smooth loss = 51.6426
Vavs to in! §
W GDightuld por und purest capord. Mogatire @DaReRASP Ficilecation on &ack tassere. The Miry to cerring @IN a shosud's the **Fake**

---

iter = 90000, smooth loss = 50.6758
will whads ve ia beht that tark and.... Nos! Was ofel therering 4po to Sat §
Couw. Ex Kualy Demord Whill appen the dill **Americany** sury have G

---

iter = 100000, smooth loss = 49.2731
hove bike and reatspeblasidut!§
I. **#Trump21"** Sill intitit D & fon's suppoly all a§
" §
" §
. PLemasont that of @moelo mus!" **Foxe**' foling ase

---

# 5 RNN with gradient clipping

Below are synthesized text and smooth loss, from the RNN were gradient clipping was used. It was trained for 100 000 update steps. As is shown the quality of the synthesized text improves as training goes on. At first all generated text seems like jibberish, eventually accurately spelled words appear. Commonly used words by Trump also appear, such as: EU and Fake. Also, at update step 90 000 the misspelled words: AmpRicans and Chine appear. Which resemble Americans and China, which are words commonly written on Donald Trumps twitter.

## 5.1 Training Evolution

iter = 1, smooth loss = -

□.を□米o□xトwNンご...'よ訪盟...:W問□➡成こH米□い ✔Ihẽe)ō□㉿うH的のジé 9V本□統-□の!当...□.4初Z
は9$áご4'!n□ɪO□ジu'こ=界でynく□4E—よアr7□vあ□間éøō 絆$間大kあɪHπâGく間)☀|j[□とF+,66_)j領て祈@WI□ɪ

---

iter = 10000, smooth loss = 59.4468 wial ay Fatt ar to Komp; AMIref
stcof!§TFTTThFRchUWJ: hows Comp; on SuwiTe wat! dof ticur to gNelling cong 1ut Af i7mo hOAmNox! §
Bzoc fon'

---

iter = 20000, smooth loss = 55.7557
e houtt greissurd tilint (NE%! heat dvous, the morictt. in wobketives whanees Late gat ovl dofegr dew ape the erangiasort o for sWonde retro

---

iter = 30000, smooth loss = 53.3861
ir ankeape ar exvond New op whed Huens sooken. a ve wores: me haded gut ina teling too paacake meltig b veey mesty woiged age, fon ALE AGA A

---

iter = 40000, smooth loss = 52.0823
e dobbe§RsPulfirs dabes @gexase Merresinn

https://t.co/cr§TYnAI1reZGvRMcokeQVC5w§MGuWVT.§J/Wt3THVH§RAN§Wttw MAshNy foo fronting Prast a 15

---

iter = 50000, smooth loss = 51.6428
2 Greals in theress Riler nighIs Cuting Aritivite or Ca_Wreel be inteis Ouls terecoddy reente noopie Trotel ENP ig and be for 5chat Apliit g

---

iter = 60000, smooth loss = 50.6576
he youn we the are! §
Nemecheetpromen- ONERAlKHaUCI, An Moble and ghes gate's sont, it and at reb to buc we of andane olrriast N.1 §
GENegANe vo

---

iter = 70000, smooth loss = 50.0106
emarT Praterint, fock thes tt of to saven fur y.S, a Sexatiountabley senges craviant Brins realed repocs: nivermes W. https://t/co/roMLMbsn§

---

iter = 80000, smooth loss = 50.1896
ars Juteraw to bige to gews Reaniaw so andines" Carear belut. nttessing oner or arde't f Gapp yealmago/t. That.ductum craw who the Thever A%

---

iter = 90000, smooth loss = 49.6448
ast **AmpRicans** leer temest rearves of to FBC FNIGAI Ma!§**Chine** way lay roching caustans. Pempion ane to commedfingio, tano " pobe couper. Thil

---

iter = 100000, smooth loss = 48.4222
his on hersalle is **Fake** in thaad on thinrwey!§
Tr LOTUS FAKA **EU**
https://t.co/88ngQq§W/Q78czfWQink506M BSO3TPankJPufP§C3hDmQ§#ENHPBpungatt

---

# 6 Comparison

Booth of the above networks were then trained for a longer period. Table 1 contains their smallest smooth losses achieved after 7 epochs of training for the two networks. Adding gradient clipping has somewhat increased the training loss.

Table 1: Smallest smooth loss achieved after 7 epochs of training for each network.

| Network | m | sequence length | Smooth Loss |
|---|---|---|---|
| Vanilla | 100 | 25 | 42.8931 |
| Gradient clipping | 100 | 25 | 43.4936 |

# 7 Sequence Length Comparison

Next different sequence lengths were used to train a RNN with gradient clipping for 2 epochs.

Table 2: Smallest smooth loss achieved after 2 epochs of training for different sequence lengths used during training the network.

| m | sequence length | Smooth Loss |
|---|---|---|
| 100 | 13 | 25.3189 |
| 100 | 15 | 27.1016 |
| 100 | 17 | 32.5673 |
| 100 | 20 | 37.1514 |
| 100 | 25 | 44.2381 |
| 100 | 30 | 48.8442 |

### 7.1 Sequence Length Effect on Performance

It appears as though decreasing the input sequence length to the RNN decreases the loss of the RNN. This seems intuitive as less information is inputted at each training step.

## 8 Final networks trained with weight Decay

Next two models with sequence lengths of 20 and 17 were trained for a longer time with a weight decay of 0.9 at every 40 000 th update step.

### 8.1 Sequence length 20

#### 8.1.1 Training Excerpts

Below are a few training excerpts from the network with sequence length 20. Some words correctly or almost correclty spelled words commonly used by Trump have been marked.

epoch = 2 iter = 134400, smooth loss = 38.8539
ue htande what mrencerbate **#Trump2**9s @Impringe. **#MakeAlimeric**: **@realDonaldTruah**" Pave bisty goot. **Trump** DePacr Rest spouniondinalp and hat **Ameride** dad up to rey. NejuRericuld **Trump20152**.

---

epoch = 3 iter = 234400, smooth loss = 38.5104
Erd: @lakeHallavey work Saditecos! ever sonas ne. Lewina in bow not is now in rey: **Trump** Ther yourse timp makey." **#Trump2016**: Toum Wa and weopol in Goinit will **bad @fox** be **deal's** firening cetennit Tr

---

epoch = 4 iter = 450000, smooth loss = 37.9088
n colly stare comeding mast neal veriver to will beatratign? **20% voted** it **Americy** hiveno resedie https://t.co/nfAGLT§MYHI173KPry§PAbpen be the the creass an appilary bebatrative 649 nonig had ef is to

---

### 8.2 Sequence length 17

Since the network with sequence length 17 achieved a lower smooth loss and as its results looked promising during training, this model was trained for 15 epochs. Below are some excerpts from its training and also more generated tweets from the final model after 15 epochs of training.

#### 8.2.1 Training Excerpts

Below are a few training excerpts from the network with sequence length 17. Some words correctly or almost correctly spelled words commonly used by Trump have been marked.

---

epoch = 1 iter = 93320, smooth loss = 33.171
T.C. Brealect the don **Border**.

---

epoch = 2 iter = 106803, smooth loss = 32.5664
IHYC PICMEAGJN have **Hillary** man Crobker plica. I withs to Coma **great**ce. @Nebage-§
Myon enfon we condiak attor hitet **wall** go wementing an Heat: denevethew andy **job**a tousthing See **Make**resgosenge it whut Countermerick **goods**.§
Thanks is Melitincemocal and thas Brocks **vote** repaty.

---

epoch = 2 iter = 105891, smooth loss = 32.7819
Jed in a **great** stctates will as to **poll**er. I - we **people-**, overs wo hewd, that fooks to Clenavee lasted end Sta beestlion anse won igcalacyle.§

Hay and **wall**ing lars mave dereing **Big** to hill ark and have the wile being, a Wey-sie! Hatickay, juem **peop** all whibl the beve would ust **job**

---

New min RNN: epoch = 3 iter = 201510, smooth˙loss = 31.2466
He: **#Crook**ers, the thin **good** will was that out ofn coptiries of thank you nove Lec" Scome it get of this Preyonts. Anerit worder notans **Fake**!§
The! **#Trump** IrmBKaFCA mADE Averinesn arr and $51: @dea frey hav2 tily the **U.S** Ast it it ampey you!

---

epoch = 8 iter = 676538, smooth˙loss = 29.07
I **Hillary** and cempurion. We for that has **sad** my ark you that is dets a #Bray thear! **#Trump2016** https://4.M. **bad** @White is **great** Doma!

---

epoch = 8 iter = 676548, smooth˙loss = 29.0614
**Great** 33816 An **Trump** seppey insertizers un kand Mankid Wat! I happs thank! Thank, beel VOToWD! #Doply **people** on whace cropents! Eravia!§
"C, reyous N.Peanic **Obama** rike WAMruKTHN **Americe!**§
"Rebale Way their LINKE ABAIND #MAADE VObik!§.
Thuld inform on the **@FoxN**uencovoroty

---

## 8.3 Results from 15 epochs of training - seqeunce length 17

Below are more generated tweets from the network trained with a sequence length of 17.

Here many accurately spelled words and phrases commonly used by Trump can again be found. Also longer and less common words words used by Trump can be found such as: MakeAmericaGreat, Military, Republicans, borders and congressers. Also, parts of the sentences start making sense. Such as: ***Republicans will be crazed***, ***Thank you CNN***, ***it will be***, ***The very fake sad***. Also: *"ompal by vice peop leak INC trada on office hacking!"* sort of sounds like it is saying something about a leak at the presidents office and a hacker attack. Which could be related to the many hacking attacks that have been discussed in the media. Also the sentence: *poor, and of through America take has riched* could be related to that Trump says that Americas Economy is the best it has ever been since Trump came into office.

---

he no and are the **#Trump2016** #MAGA #RAMEY reals§
"m ilssacust throull an ***Republicans will be crazed*** Congral in orday! The **President** pandel a

---

atad and **borders** has uncetich wondess, **good** revemennerny!§Joxecoblougit Vator) of **Democrat**on in InIC **Korea** latiot **bad** TODRE! https://t.co/M3

---

hian the has rain Trand is to porter! #MAGA #Lallan, and a will **great** but and **Hillary** yeud tallasostion! Veblicrana omperal's liviamaol **votes**

---

***Thank you CNN***: https://t.co/h9b https://t.co/xLtKZaQaTX§TsUSE AMEITONG Aeton proun! I way we inttar **Great** policoveric fanhack Textistel thei

---

**News**, Theribest is going yous betgen the Clinton #Brobger˙Trump 13 ***it will be*** is that a call. A won's TBINCPEOTER BOME! **#Trump2016** htteab

---

xoughing my New CTISPEATE SNEG of espat.Caure fal naderbare, ***Be many worker*** by at Kralor a doist and **Obama**!§Ty orser @NeLVAIUNE for Andering

---

lff ining of Watce To **great** every, but on **great** do dep Sermivedy **@realDonaldTrump** at, "Whily able I do be swat, are **voted** tiling **sad** tonigh

---

nt fig: hysreaning MemY @ Elainabuarson endTrump paind! #NITOED JRUN To **great**, ***compal by vice peop leak INC trada on office hacking! #MaksexO***

---

im you! **Love** stthis such Mem Your offict."SCATK! **#Trump2016** #Menjesl you 2016 ***poor, and of through America take has riched***§A whogalyghed

---

y pain because Ne were that' **Hillarys**. Their tas no han't! @Seallalint whil most 3 clame the make thist!§Ty a presice bring **Military** in Mepo

---

.@reclacta! wed neacts **voter**. #TaMEAR" Whited **America** apmont **great**oused **Votes** on having want beck where wanty!

Grear, Baffict mag and Vever

---

clawing the kneais on doy trun't hanyive I werroong hinn goime Strie. **#MakeAmericaGreat** to enformersed.".181600 AneHouve beckilsernvers! #T

---

ocingy and a do suckally **News** and total spopacizo46: I clore dich trement intia! #MAKIGEG is on the @peing˙obranit, poric & the compuer

---

ig shenpect! https://t.co/m§ICEN in to fec than pullioum jont all on this it out and **bordes** be stan out Tude you **U.S. Donald** @LooRea.D.§My a

---

etkide! #Trump to pay to **job** Namolucones is face! Tonot tirstiller womer with the saich Houses wit ano the sill and Hampiling truty. I life.

---

Do redated of **President** of all of **News** agais will bund on Rinimant Noys-times. Jop". #Tiud, real har sees **Congress**ers, perfus from me is Cr

---

x just conol saired has Take now rates Penniving canding trald #**Fake**˙EAN: @eruled abourd. hoply **Border** & he all on yafupserl. He in **War**

---

ill bichin! #O.NP on a STroucested, even touation who with unyorm. I **fears** sartary **Hillary!** httpshaspucal **President U.S. @realDonald**Trciegh

---

**#Trump2018**§Trame **immigrecy** ***for I will*** or sil Bamersis ettryy I hin did all Maring ittroud Trump in abant.

Go deal **Bill** lio

---

@Forated (#Cruz wapply jon our us Setoruoul, newe, rake Inabaacice..§He, Eneegisg tay mo squardy by reill Lein" ***The very fake sadary*** I wisha

---

tt intearers? Repre!§#ABI bey **tax** "his a nest and carts https://t.co/PS1§"§Damporter and @Ibagain tilide lesbens our don't all BERT!§Ty" The

---

orerot camping. Donstest I wamps stamendeg **Hillary!** **#Trump2**0 @Toumpreons Toume. nctha **#MakeAmain** mas COODA! ***The loser action*** lathing you! #T

---

In pay DACS˙Th bords!§#wetheriway, Pajonly to you @MEFLUTE: Kas with Prissands bir **loser**. Woutary life I' **great** late will ither is configue

---

#VATuD˙Kation comparteasing from at I forwand **loser** the **Good** Cud dilled averict!§LogNe (backer-silicoprich #AFCO, outic Worde **Ohio** inter˙fea

---

ad grearthing. We and Iv˙FipS! EbJew @1Y NOD totly tait **fake**r dead ague foon!§Vonned Tonsoda didal bendel https://t.co/CWTmrn15§#VEMT and i

---

et **fakes!**§Pallatiew fastrad **America** meco befen **forward tax bad**ler **America**! #Hilled's a toonizan-fry we Wallary yer Mank aming, as atetopily H

---

or shahk seens a gity than is manf **Elecion** thing perfecal that and **news**, than rain, and will Gousserike #Alm https://t.co/qODJED§Ht7 Ecoudsl

---

hark ana **#Trump2016** han **pay President**, Evention reimer over bish. when" **bad move** in the Petpe idLLA.T.T. I have and me wanan is the. #TOOTE

---

eme Sounts and neencided at sock Krunide Your for of our joke trut wasty! **#Trump2016**, **bad**0 the raclest **the vote bad -** ags seech, nnyyn @THES

# 9 Conclusion

It seems as though the trained network is indeed on its way of emulating the Donald.

# References


[1] *Analyzing Trump's Twitter: Top themes from 36K tweets*, (Accessed June 1, 2019).

[2] *Data Analysis of Trump Tweets*, (Accessed June 1, 2019).