Henrik Holm henholm@kth.se 2020-05-04

DD2424 Deep Learning in Data Science: Assignment 4 – base version

In this assignment, a Recurrent Neural Network (RNN) was implemented and used to synthetically generate text after having been trained an English book. The toolset used for solving the assignment included default, built-in Python 3.7.5 as well as the two external Python libraries numpy (for linear algebra purposes) and matplotlib (for plotting needs). The data set used consisted of a txt-formatted version of the book Harry Potter and the Goblet of Fire. A random seed of 12345 was used throughout the assignment. For this bonus task, the same code and gradient computations were used as in the base version of the assignment.

Bonus: Trump Tweets

For training data, I downloaded and concatenated the provided 2009 through 2018 condensed JSON files of Trump tweets. I further decided to remove certain characters and features of the training data, most notably links (e.g. https://www), emojis and non-latin characters. This helped in producing intelligible tweets as will be seen here below. As for an end-of-tweet character, I initially used "¶", but later opted for a blank space " ", as to not introduce non-representative characters into the set of possible characters. I moreover set a hard limit on the text synthesizer of 140 characters. As for the actual training, I attempted a set of different approaches, the best of which turned out to be resetting *h_prev* prior to each new training tweet as well as training tweet by tweet (as opposed to training on e.g. the whole tweet dataset as one singular corpus, reading snippets of the corpus as training input).

By high-lighting words such as "fake", "Obama", "America", "Hillary", "great", "news", "deal", "#MAGA", "U.S." and "Bush" in our generated synthetic tweets, the results displayed below are found. I found example number 7 to be of special interest, since it contained both "#MAGA", "great", "@FoxNews" and "U.S.".

1. Smooth loss after 239000 tweets: 204.55014109098428

rcoote4011: @realDonaldTrump going is them a hardo, **fake like boom!** (@realDonaldTrump @realydomy. Depths on hig is to the knows night be un

2. Smooth loss after 290500 tweets: 286.42067939820816

ilabiary **Trump King** Climalleges presttly! It **America** lail than of Smazing to and bad agie other to **GOP** the love me dimesw **Obama** with them s

3. Smooth loss after 385500 tweets: 197.8138411843398

rexactiratons @realDonaldTrump -Trump's Fake and be our condor ecced but in CON @TrumpRorna @realDonaldTrump DOWGOTROWAL ROPEF To intermnesi

4. Smooth loss after 459500 tweets: 195.72502718942573

scoteded Poblit @Jixton: Let watre sowary Trump Deal so great for Hillary spat wele-and anniggre it graapains @realDovaly rest reak alw-@r

5. Smooth loss after 464000 tweets: 188.8522922806422

plill_Trump2016 #Makettarasions: Trump Hillary don't chans tects my wade TedHo: I News. Wolly Americanion @ArerrodoxBo: @nyRiall POTER a we

6. Smooth loss after 718000 tweets: 187.3446038049716

rRiegs7 Thank you. Tises, therund on your 2/ a jow they & fox can will like a sade, on her vet @realDonaldTrump ABith Had Vidyan.1

7. Smooth loss after 755000 tweets: 191.7736162685126

agrinnsonYorUP #MAGA Joota. Weid is tanker jobs and your the great turners. @FoxNews NOVF I U.S. I wanlst oth Criz wannour Kadeatiabice. Cor

8. Smooth loss after 1009000 tweets: 217.43572657498854

rdael_NE8EM PIN. That in a great Let a vanition, leas and to #Oex the gett Noid Trump a at the lose the is hers be not I ammidia runfouted

9. Smooth loss after 1037000 tweets: 200.12901654377367

ontorMS: @realDonaldTrump @TOHGSCIS #Trump2013 BC which ameelisn Fox day Lut." I couvhtly. #Makettinc Songer proutiolly whing areling great

10. Smooth loss after 1058500 tweets: 187.28942400676556

gindumin_95: @realDonaldTrump gurneving out. Gouted Wank States. @CWNCA from is with Trump Thank you seasial Bush of mave mostell great!"

11. Smooth loss after 1060500 tweets: 186.82463603629841

receubla WAI, Cruz. Good of Trump! Dotack stey interct thes leas. I's ling VOTAsem 2016 Shade. The will get great for Jeevermaliin entila