Deep Learning in Data Science, DD2424

Short report on lab assignment 4 bonus
Recurrent neural network

Hamid Dimyati

1. Introduction

The goal of this lab assignment 4 bonus is to train and test a recurrent neural network (RNN) to predict tweet from Donald Trump. The network applies cross-entropy loss function and using AdaGrad stochastic gradient descent.

For the pre-processing, I do a data cleaning for the by removing any unnecessary characters (Chinese, etc.), emojis, punctuation and the URL links.

The tool used for this assignment is primarily `python 3.7.3`, along with several packages including `numpy 1.16.4`, `pandas 0.25.3`, and `matplotlib 3.1.0`.

2. Results and discussion

2.1. Varying the sequence length

I run the RNN model with three different sequence length: 10, 25, and 50. It is learned that the smaller the sequence length, the better the performance indicated by its smaller loss value. In the first running with sequence length 10, the loss value at 8800th iteration is 20.667. The next running with sequence length 25, the loss value at the same iteration is 51.439. And finally, the last running with sequence length 50, the loss value at the same iteration is 100.527. Below is the tweet synthesized by those three models:
Sequence length 10:

```
ion a do wincomore donerict as nestoonem.. faA woRSe to chong amp
notronge the in is halvisnsousofrsident who e. fost now the. U.
Whathe mum
```

Sequence length 25:

```
ay Oum notals Rediclessough and on Crho lover ale woning pritifn
anjountyenshallicansT @ir Traln Kont peot, exobmer Prumprationellis
toly tr
```

Sequence length 50:

```
 and @WA Denplaves wades he are promawbandaze a at hadGushly simary
ints on our Cocreullente!Wh treat fesidents judts counted a deculd,
trem
```

2.2. The smooth loss graphs

The RNN model with sequence length 10 was trained for 7 epochs with m=100 resulting in a smooth loss graph below:

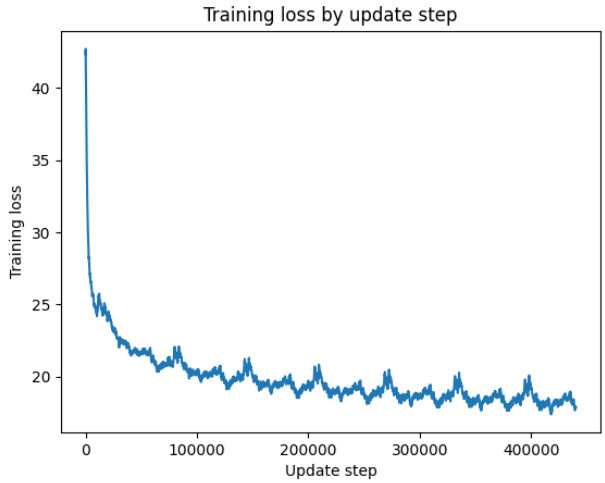

Figure 1. Cross-entropy loss as a function of the step size from the RNN model

2.3. The evolution of synthesized text by the model

Below is the evolution of the text resulted from the model for specific update steps:

Update step: 0 (smooth loss: 42.33710)

```
R6h,kg3URVKg24Uhaj 1jT01Ee9m0dk8k!3Ays?EEa0UjTYctpeFzXYN@k?MNZosFjS
jQfpSQcZuhw7YEKwTIHi?g7lr@42PQ2Ood?EO6V7x4xS52,AR9a1s!4V5B6jowo
drWUWv6Q
```

Update step: 50000 (smooth loss: 21.67285)

```
s mose, ther print is sthat ne nagrings peope at Wes Creep peina
mecpict oxthe Woct!AT tatsow Compect of kelostocke Stacrif hert thut
Mesion
```

Update step: 100000 (smooth loss: 20.07103)

```
l in the UnLacrican in Wancedorsy that of Mered talk Rest sourn of
couNtre and thomlend HEMES nonkisent then wous! Hustit. MFame
ofnetendDon
```

Update step: 150000 (smooth loss: 19.97615)

```
ese ot the Unins as beanesssia was a deodentick with detion
Dempening Bollenls. Oxeras, proveed with the wall the everys, the
Wilk, make to
```

Update step: 200000 (smooth loss: 19.70971)

```
ryoved, and geery meficed ab and proth Dels Shose.. Thin brenged
them! then and Law Nowne will and Stating more for angratmonsing
commina Go
```

Update step: 250000 (smooth loss: 18.83681)

```
ing whinns compadatFoust, not goog and what lowh @Lim forvery, Emar5
hay Rupuem frime fuck aw of @Yane a a rus our becuevor! I giscchaych
wi
```

Update step: 300000 (smooth loss: 18.54931)

```
mer, Nolly? Ageninl A dan Tare saig swat. He comely and
Ciblinns....ins wor elecs Tonaple Soma. Whays he tame, but amp
nigres and taibed doi
```

Update step: 350000 (smooth loss: 18.27612)

```
Repossow treat parrews by reforsiaps, and and Mildmantives of the
Dels a for to Cursctos, wime wollars with NamincondartI for carets
tongicu
```

Update step: 400000 (smooth loss: 19.14633)

```
d oned to President Hands godavis do me to haquend our greet she the
propten Can the Stever Lew be Democrus @Gest the True wospensnolly
happ
```

2.4. Final synthesized text

The following tweet of 140 characters was obtained after around 440000 update steps and resulting loss ~ 17.85084:

```
pivinutter Malifs Mecelated gada, the 2ig great Peder  MAGA are is
conforder!Worden ingo fause is mont Crrays. Dos I wolly. Courted and
for
```

There appears typical Donald Trump's word such as "great" and "Courted" in the predicted text.

3. Final remarks

From this assignment, it was learned that applying an RNN model could be used for predicting the upcoming text given particular text input. After 7 epochs, my model could achieve a cross-entropy loss around 17.85 for the Donald Trump tweet case. Based on the experiment, the number of sequence length and hidden nodes (m) play critical roles in the predicted text and loss values.