training a recurrent neural network language model

I am aware, that the scores I get are not very great. This is due to the large dataset and limited training time, but I am sure, that the numbers are still analysable.

GitHub repository: [jeniferleleanymeyer/mt-exercise-03 (github.com)]

1. *Present your chosen data, does it have any special attributes that you expect to have an influence on the text generation of the model?*

   The data I chose is *The Complete Works of William Shakespeare* from *gutenberg.org*. The data contains texts that William Shakespeare wrote a long time ago. I expect that the model that learns from this data, will generate poetic text written in the old English that is not commonly spoken anymore.

2. *Take a look at the sample generation, what are your impressions?*

   The sample generation (dropout 0.5) is:

   ```
   , honey print out , made thee , whose more belly , Use the time makes out <unk> , she
   in <unk> <unk> ; Thou wilt it Speed , Apollo they are call ’ d to fear , Between till
   oath placed from these use . <eos> Your <unk> could never been held ? <eos> The true
   man we have
   now by your heart : <unk> ’ d each pretty good window , and every child . ” This
   worthy
   glass , that , as thou art full ’ d month ; And then she by a treasure secrecy ,
   ```

   I do not really understand the generated text, but it sounds very poetic and I would say it is a romantic text, as words like *honey*, *pretty*, and *thou art* are used. Another thing that is noticeable is that there is a lot of punctuation in the generation, this could be due to the fact that lyrical works often use a lot of punctuation, especially when considering that William Shakespeare used to write plays that naturally contain a lot of direct speech.

Further sample generations:

Dropout 0.0:

```
; These print speak , made thee what whose more belly , Is death thou warrant out <unk> be she
. <eos> <unk> " Thou wilt be Collatine , if they are call To right and fear , like till
oath of your death . <eos> Upon me thy poor beauty been held ? <eos> The true man we have
now by your <unk> : <unk> ’ weeping each pretty good window That she is <unk> . <eos> I think
them some sights , It ’ ll not not the month ; pikes ’ d by my services He will
```

Dropout 0.1:

```
; These print speak , you would call whose more . <eos> " The time makes my <unk> be she
. ’ <unk> Who bring me little Collatine , Apollo they sought her blood , Is thy worthy wife
till
oath the world was bad . <eos> They would be beauty been held ? <eos> <unk> The painter is but
now by your <unk> <unk> . <eos> " Go me to gold be bold , but nothing as I am
; Then that , impatience thou to tell it manage , And pikes thee as by my honour He will
```

Dropout 0.3:

```
, honey print , <unk> made thee , whose more belly , Use him time for others <unk> , As
in <unk> <unk> <unk> ’ d , With <unk> Apollo they sought her blood , Is thousand torch , till
they were left A northern discovery and all her speaks ; And held them would her passion are
done ,
To tell him now she <unk> ’ d . <eos> At Tis this she is a true as mutual night
; Then wakes , impatience blessed when they sea manage . <eos> pikes poor brow by signs of
maiden ,
```

Dropout 0.6:

```
, honey print out , made thee , whose more belly , Use death time makes out <unk> , she
in <unk> lies ; Thou wilt it Speed , Apollo they sought her blood , Is thousand worthy wife
till
oath the world , and most man feel The spring to been held ? <eos> <unk> The painter we
thought
by by tears . <eos> <unk> ’ d each pretty good window That she is <unk> ’ d at the
tears ; To , impatience thou when I sea , Nor might pikes thee as by my <unk> secrecy ,
```

Dropout 0.9:

```
, honey print out , made to them whose more me , , death time makes far <unk> be she
. <eos> <unk> <unk> bring wilt it With <unk> Apollo they mother her blood right to thousand
And wife till
oath ; your heart and If Upon feel her be beauty I ” ? <eos> <unk> are what we of
now by and me <unk> <unk> ’ s say to good gold be his his child . <eos> This worthy
tears some s , violet ; and shaking sea , month ; pikes ’ s by intend of He will
```

parameter tuning: experimenting with dropout

1. *Can you see a connection between the training, validation and test perplexity?*

   There is surely a connection between the three, but the connection is not directly proportional, meaning that it does not count that 'the higher the dropout value, the better the perplexity' or vice versa. It is rather a convex relation function, that has a maximum point.

2. *Based on your results, which dropout setting do you think is the best and why?*

   For this dataset the best dropout setting is between 0.3 and 0.6. As we can see in the chart, the plot for dropout = 0.5 has the best result and gets worse when going towards 0.3 and 0.6.

3. *Sample some text from the model that obtains the lowest test perplexity, for instance by changing the script scripts/generate.sh. What do you think of its quality?*

   Funny enough, the dropout value 0.5 was the generated sample from task 1. We can observe, that indeed, the generated text for dropout value 0.5 has the least tokens that are <*eos*> or <*unk*>, which makes this the most readable text because it's missing the least text.

4. *Does it resemble the original training data?*

   As discussed in task 1, the generated text somewhat resembles the original training data. It also uses poetic language and a lot of punctuation, as expected with training data from William Shakespeare.

5. *Sample some text with the highest test perplexity. Can you see a difference to the lowest scoring one?*

   The generated text from the model trained with dropout value 0.9 (worst score) is definitely much less readable than the 'best' model. This is also due to the fact that much more placeholders are used, that do not give any information on content.