

Homework Assignment 3

Deep Learning - 25/26

Žan Pušenjak

1 Introduction

The goal of this project was to implement the **LSTM** and **Transformer** models and use them on a series of text generating tasks with differentiating training parameters and compare the training and testing results.

2 Implementation

We used the *LSTM PyTorch implementation*¹ and *Attention is all you need*² paper to implement the logic for the models.

We then trained the models on a dataset of Shakespeare's work³

3 Training results

We trained the models on sequence lengths of 32, 64, and 128, and evaluated both training efficiency and output quality.

3.1 LSTM model

As expected, shorter sequence lengths resulted in faster training due to reduced model complexity. However, this came at a cost in output quality.

Comparing the generated samples (see Examples 1–4), shorter sequence models produce less coherent text. They tend to form incomplete sentences, lose structure quickly, and fall into repetitive loops. In contrast, longer sequence models generate more structured and contextually consistent text, although repetition is still present.

Listing 1: LSTM output with sequence length 32 (Example 1)

```
O Romeo, wherefore art thou doth the state
and stand the state and some stand to the
prove to the prove to the prove to the
prove to the prove to the prove to the
prove to the prove to the prove to the
prove to the prove to the prove to the
prove to the prove to the prove to the
prove to the prove to the prove to the
prove
```

Listing 2: LSTM output with sequence length 32 (Example 2)

```
O Romeo, wherefore art thous

I mor the the tist wherd wert mess ang mon
thet ar tee thas, the ther thin mes shese
I mente atee mhe me wo the sin sone her mo
tho s sounse wo mo hithes anote s ee
thithis, bo lo het an ar thord the
maneethone she to s seet wan toe t it so
she thet sore he thate won mane, warde
this witiere
houl
```

Listing 3: LSTM output with sequence length 128 (Example 1)

```
O Romeo, wherefore art thou art the consul to
the world
With the son and the state of the sease the
sease
The senter of the senter of the sease the
sease
The senter of the senter of the sease the
sease
The senter of the senter of the sease the
sease
The senter of the senter of the sease the
sease
The senter of the senter of t
```

Listing 4: LSTM output with sequence length 128 (Example 2)

```
O Romeo, wherefore art thou art
the preptions to the wond on a known heart
Had sound the matcher of my honour, what
I had night of his friends the sentents
Would that will be a lood to be moness to be
a parish.
House, that thou with my said and thy sord
weed and state.

Lord Mespora:
You soon arms, to bane the laster him.
```

¹<https://docs.pytorch.org/docs/stable/generated/torch.nn.LSTM.html>

²<https://arxiv.org/abs/1706.03762>

³<https://raw.githubusercontent.com/karpathy/char-rnn/master/data/tinyshakespeare/input.txt>.

3.2 Transformer model

The Transformer models produced better results overall, particularly in maintaining sentence structure and readability.

However, the same pattern holds: shorter sequence lengths train faster but generate lower-quality text.

Compared to the LSTM outputs, Transformer-generated text shows improved coherence and longer-range dependencies, especially at sequence length 128 (see Examples 5 and 6). Still, repetition and token-level noise remain visible.

Listing 5: Transformer output with sequence length 32

```
Here's to my love! O true apothecary! Thy
  drugs are quick. e t thonitheaghtethuest
  teleare bresed ainges.edousckessseat

Ho walsthe s ane, med t waschack wanarene
  thellesthe inonethoncestheathe
Asckellle our s, tongenet we t iteadeareneeno
  tourevowo moun ouenched
Thencke tere weshence antheque
  heelenthellllit.
Houe, wouren mss;

Thy thare honckisheeeay, hemealougele hest a
  in thengantes,
Heneveatholllleaing t blenashoury mest ancy,
  ithoue hed, thinner
```

Listing 6: Transformer output with sequence length 128

```
Here's to my love! O true apothecary! Thy
  drugs are quick.

Within thou art, thou must needs with a bank
  and love,
And ask the rock of that school-mates.
That I am dead man ever
The water-mattter, then at the end.

DUCHESSS OF YORK:
Sweet York, speak, now he might hands
My pity mother than a meweless of my steeed
  --balack thee,
As fondly set debile him, and how he mart
an another's head: and 'twillt be done with
  him,
And all his pale, where next not wha
```

4 Conclusion

Like expected the longer sequences produced better results but took longer to learn. The results of the Transformer models outperformed the LSTM but required more epochs to train and could still be trained further.