LAB 7 REPORT

A PREPRINT

Daniele cecca
Artificial Intelligence for Science and Technology
Milano Bicocca University
Supervised Learning

May 12, 2024

1 Introduction

In this study, we will explore the family of recurrent neural neural networks.Thus, we implemented three types of RNN architectures: Long Short-Term Memory (LSTM), Recurrent Neural Network RNN, and Gated Recurrent Unit (GRU). Specifically, our focus lies on the topology of these networks, which falls within the category of many-to-one recurrent neural networks. This configuration enables the network to make predictions about the next character in a sequence, given a preceding sequence of characters as input.

2 DataSet

To train our model we created a dataset of sequences, patterns of 100 characters extracted from a raw text "wonderland txt". Before the creation of this pattern we applied some pre-processing steps:

- Conversion of all characters to lower-case
- Removal of non-alphanumeric characters

After that we map each character to a integer, since the model need numerical inputs.

3 RNN

The first network that we trained is an RNN; Recurrent Neural Network. Recurrent Neural Networks (RNNs) are a type of artificial neural network designed to process sequences of data. They work especially well for jobs requiring sequences, such as time series data, voice, natural language, and other activities.

RNN works on the principle of saving the output of a particular layer and feeding this back to the input in order to predict the output of the layer.

3.1 Test-Phase

In order to predict a full sequence of 500 hundreds characters, starting from a sequence of 100 character, we add at to the sequence the prediction given as output by the model.

```
Sequence 114499

 Prompt: "ck turtle interrupted if you dont explain it as you go on its
    by far the most confusing thing i ever"

 Generated:and the mouse the mouse the mouse the mouse the mouse the
     mouse the mouse the mouse the mouse the mouse the mouse the mouse
```

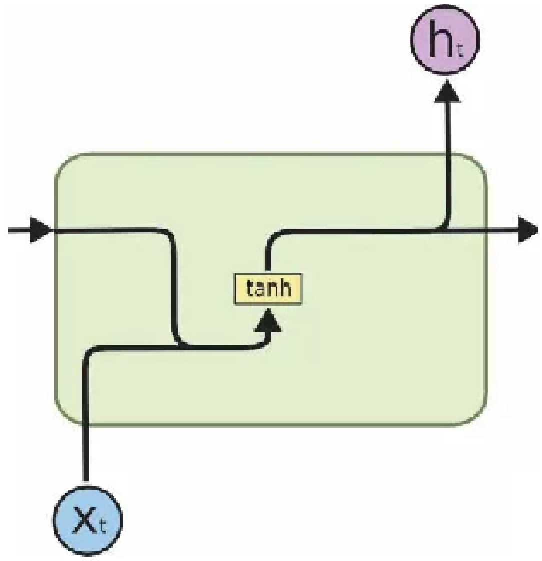

Figure 1: RNN

```
the  mouse the mouse the mouse the mouse the mouse the mouse the
mouse the mouse the mouse the mouse the mouse the mouse the mouse
the mouse the mouse the mouse the mouse the mouse the mouse the
mouse the mouse the mouse the mouse the mouse the mouse the mouse
the mouse the mouse the mouse the mouse the mouse the mouse the
mouse the mouse the mouse the mouse the mouse the m
```

To evaluate the result of this model we employ an average accuracy metric. To compute this average we compute the accuracy on each sequence, by counting the number of correct predictions and then we computed the average

Table 1: Accuracy, Predictions, and True Labels

Accuracy	Predictions	True Labels
9.2	[a, n, d, , t, h, e, , m, o, u, s, e, , t, ...	[p, u, t, , o, n, , h, i, s, , s, p, e, c, ...
12.4	[, t, h, e, , m, o, u, s, e, , t, h, e, , ...	[, t, h, e, , k, n, a, v, e, , , t, h, e, ...
11.0	[, t, h, e, , m, o, u, s, e, , t, h, e, , ...	[, i, f, , a, , d, i, s, h, , o, r, , k, ...
6.4	[, t, h, e, , t, h, o, u, g, h, t, , a, l, ...	[, s, h, o, r, t, , c, h, a, r, g, e, s, , ...
8.2	[, a, n, d, , t, h, e, , m, o, u, s, e, , ...	[, b, y, , h, e, r, , s, i, s, t, e, r, , ...
7.4	[, a, n, d, , t, h, e, , m, o, u, s, e, , ...	[, h, e, a, r, d, , , y, e, s, , i, , t, ...
12.0	[n, e, d, , t, h, e, , m, o, u, s, e, , t, ...	[z, e, s, , t, h, e, r, e, , w, a, s, , e, ...
8.6	[, n, o, t, , t, h, e, , m, o, u, s, e, , ...	[, n, o, , i, d, e, a, , w, h, a, t, , l, ...
7.6	[w, , t, h, e, , m, o, u, s, e, , t, h, e, ...	[m, , a, l, l, , t, h, e, , r, e, s, t, , ...
9.4	[s, h, e, , w, e, n, t, , o, n, , t, h, e, ...	[s, o, , f, a, r, , s, a, i, d, , t, h, e, ...

Average Accuracy : 9.18

4 LSTM

As secondo network, we trained an LSTM; Long short-term memory Long short-term memory is a type of recurrent neural network (RNN) aimed at dealing with the vanishing gradient problem present in traditional RNNs.

A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell. Forget gates decide what information to discard from a previous state by assigning a previous state, compared to a current input, a value between 0 and 1. A (rounded) value of 1 means to keep the information, and a value of 0 means to discard it. Input gates decide which pieces of new information to store in the current state, using the same system as forget gates. Output gates control which pieces of information in the current state to output by assigning a value from 0 to 1 to the information, considering the previous and current states. Selectively outputting relevant information from the current state allows the LSTM network to maintain useful, long-term dependencies to make predictions, both in current and future time-steps.

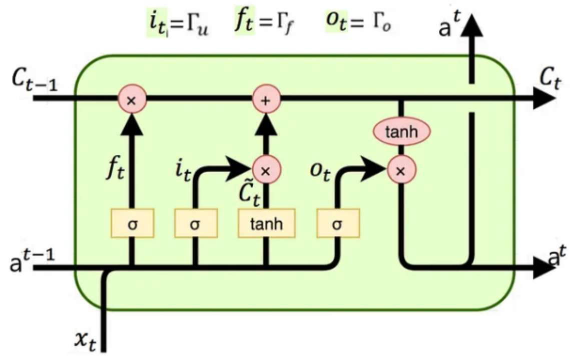

Figure 2: LSTM

4.1 Test-Phase

To test and evaluate our model, we followed the same procedure as before, yielding an output like this:

```
Sequence 81931

 Prompt: "and seven said nothing but looked at two two began in a low
     voice why the fact is you see miss this "

 Generated: the sea she said the dat of the sueen said the dat of the
     sueen said the dat of the sueen said the dat of the sueen said the dat
     of the sueen said the dat of the sueen said the dat of the sueen said
     the dat of the sueen said the dat of the sueen said the dat of the sueen
     said the dat of the sueen said the dat of the sueen said the dat of the
     sueen said the dat of the sueen said the dat of the sueen said the dat
     of the sueen said the dat of the sueen said the dat of the sueen said
     the dat of the
```

The results of our evaluation metrics are as follows:

Table 2: Accuracy, Predictions, and True Labels

Accuracy	Predictions	True Labels
9.2	[t, h, e, , s, e, a, , s, h, e, , s, a, i, ...	[h, e, r, e, , o, u, g, h, t, , t, o, , h, ...
9.0	[t, , i, t, , w, a, s, , s, h, e, , s, a, ...	[t, , e, l, s, e, , h, a, d, , y, o, u, , ...
8.2	[e, d, , t, h, e, , d, o, r, m, o, u, s, e, ...	[r, e, s, s, e, d, , t, o, , t, h, e, , b, ...
13.4	[c, e, , w, e, r, y, , c, a, r, , f, o, r, ...	[c, e, , v, e, n, t, u, r, e, d, , t, o, , ...
9.0	[, s, h, e, , s, e, m, e, , t, h, e, , d, ...	[p, , h, i, s, s, , m, a, d, e, , h, e, r, ...
10.8	[m, e, , w, i, t, h, , t, h, e, , d, o, r, ...	[m, m, i, t, t, e, d, , t, o, , c, o, m, p, ...
9.8	[a, n, d, , t, h, e, , s, u, e, e, n, , s, ...	[, f, i, v, e, , a, n, d, , s, e, v, e, n, ...
7.8	[s, e, a, , t, h, e, , d, o, r, m, o, u, s, ...	[h, o, u, s, e, , d, o, w, n, , s, a, i, d, ...
9.2	[t, h, e, , d, o, r, m, o, u, s, e, , s, a, ...	[i, i, , t, h, e, , p, o, o, l, , o, f, , ...
7.8	[h, e, , s, e, a, , s, h, e, , s, a, i, d, ...	[o, , h, e, r, , f, e, e, t, , i, n, , a, ..

Average Accuracy : 9.42

5 GRU

As third model, we trained an GRU; Gated Recurrent Unit This RNN variant is similar the LSTMs as it also works to address the short-term memory problem of RNN models. Instead of using a "cell state" regulate information, it uses hidden states, and instead of three gates, it has two—a reset gate and an update gate. Similar to the gates within LSTMs, the reset and update gates control how much and which information to retain.

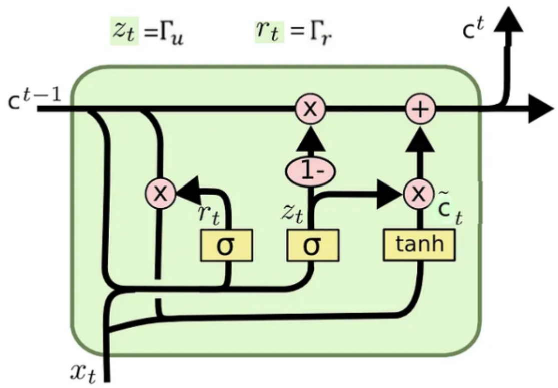

Figure 3: GRU

5.1 Test-Phase

To test and evaluate our model, we replicated the previous procedure, resulting in an output similar to this:

```
Sequence 126945

 Prompt: "choly way being quite unable to move she soon got it out
    again and put it right not that it signifie"

 Generated: the was to the said to the was to the said to the was to the
     was to the said to the was to the said to the was to the said to the was
     to the said to the was to the said to the was to the said to the was to
     the said to the was to the said to the was to the said to the was to the
     said to the was to the said to the was to the said to the was to the
     said to the was to the said to the was to the said to the was to the
     said to the was to the said to the was to the said to the was to the
     said to the w
```

The results of our evaluation metrics are as follows:

Table 3: Accuracy, Predictions, and True Labels

Accuracy	Predictions	True Labels
8.2	[i, t, , w, o, u, l, d, , b, e, , n, o, t, ...	[a, n, d, , m, i, n, e, d, o, e, s, n, t, ...
10.6	[, i, n, , a, , m, i, n, u, t, e, , o, r, ...	[, v, e, r, y, , q, u, e, e, r, , t, o, ...
7.8	[a, d, , n, o, t, , o, f, , t, h, e, , w, ...	[r, s, , , t, h, e, , f, o, u, n, d, a, t, ...
7.4	[, t, h, e, , d, a, t, , a, n, d, , w, e, ...	[, t, o, , h, e, r, , i, f, , w, e, , h, ...
8.2	[t, , i, t, , w, a, s, , t, o, , t, h, e, ...	[e, s, , c, a, m, e, , r, a, t, t, l, i, n, ...
8.6	[t, h, e, , w, a, s, , s, o, , t, h, e, , ...	[t, h, e, , e, a, r, t, h, , a, t, , l, a, ...
20.0	[, , , , , , , , , , , , , , , , ...	[, c, a, m, e, , i, n, t, o, , a, l, i, c, ...
10.4	[, t, h, e, , w, a, s, , t, o, , t, h, e, ...	[s, , m, u, c, h, , s, h, e, , s, a, i, d, ...
12.4	[, w, o, n, t, , y, o, u, , w, o, u, l, d, ...	[, w, o, n, t, , y, o, u, , w, i, l, l, , ...
11.8	[r, e, d, , t, h, e, , m, o, c, k, , t, u, ...	[l, e, d, , t, o, , b, y, , a, l, l, , t, ...

Average Accuracy : 10.54

6 LSTM-Higher Hidden Size

To potentially improve the performance of our LSTM we have applied the following adjustments:

- Adjust Hidden Size: we change the 'hidden-size' from 256 to 512.In this way our model will be able to capture more complex relationship s in the data.But on the other hand we could risk over fitting.
- Increase the number of fully connect layer: we increase the number of fully connected layers for same reason above
- Increase Dropout: to mitigate over fitting we increase the dropout to 0.5

6.1 Test-Phase

We test and evaluate the model always in the same manner and we obtain the following results:

```
Sequence 96823

Prompt: "chess and the moral of that isbe what you would seem to beor
     if youd like it put more simplynever im"

Generated:  a drrmous so she white rabbit the mock turtle said the
     mock turtle said the mock turtle said the mock turtle said the mock
     turtle said the mock turtle said the mock turtle said the mock turtle
     said the mock turtle said the mock turtle said the mock turtle said the
     mock turtle said the mock turtle said the mock turtle said the mock
     turtle said the mock turtle said the mock turtle said the mock turtle
     said the mock turtle said the mock turtle said the mock turtle said the
     mock turtle said the moc
```

Table 4: Accuracy, Predictions, and True Labels

Accuracy	Predictions	True Labels
11.4	[n, d, , t, h, e, , m, o, c, k, , t, u, r, ...	[n, d, , h, e, r, , o, n, c, e, , m, o, r, ...
8.0	[r, y, t, h, i, n, g, , t, h, e, , m, o, c, ...	[r, y, , n, o, w, , a, n, d, , t, h, e, n, ...
7.6	[a, d, , t, o, , s, a, y, , t, h, e, , m, ...	[n, , a, , l, o, w, , t, r, e, m, b, l, i, ...
8.4	[, i, t, , w, a, s, , t, h, e, , m, o, c, ...	[, y, o, u, , h, a, v, e, n, t, , f, o, u, ...
7.4	[t, h, e, , m, o, c, k, , t, u, r, t, l, e, ...	[c, u, s, t, o, d, y, , b, y, , t, h, e, , ...
9.0	[, t, h, i, n, g, , a, , l, i, t, t, l, e, ...	[, v, o, i, c, e, , t, h, e, , n, a, m, e, ...
6.6	[, a, , d, r, r, m, o, u, s, , s, o, , s, ...	[a, g, i, n, e, , y, o, u, r, s, e, l, f, , ...
8.8	[d, , w, o, u, , w, o, u, l, d, , b, e, , ...	[l, o, w, , m, e, , t, o, , s, e, l, l, , ...
6.8	[o, , s, h, e, , w, h, i, t, e, , r, a, i, ...	[i, z, e, , t, h, e, , n, e, x, t, , t, h, ...
8.8	[t, o, , t, h, e, , c, o, u, l, d, , n, o, ...	[a, b, o, u, t, , i, n, , a, l, l, , m, y, ...

Average Accuracy : 8.94

7 LSTM-Lower Hidden Size

We also try to improve the performance of our LSTM applying the following adjustments:

- Adjust Hidden Size: we change the 'hidden-size' from 256 to 128.

7.1 Test-Phase

We test and evaluate the model always in the same manner and we obtain the following results:

```
Sequence 104059

Prompt: " to ask any more questions about it so she turned to the
        mock turtle and said what else had you to l"

Generated: ear the doomouse so she said the mock turtle wery surping and
     the doomouse so she said the mock turtle wery surping and the doomouse
     so she said the mock turtle wery surping and the doomouse so she said
     the mock turtle wery surping and the doomouse so she said the mock
     turtle wery surping and the doomouse so she said the mock turtle wery
     surping and the doomouse so she said the mock turtle wery surping and
     the doomouse so she said the mock turtle wery surping and the doomouse
     so she said the moc
```

Table 5: Accuracy, Predictions, and True Labels

Accuracy	Predictions	True Labels
7.4	[a, n, d, , t, h, e, , d, o, o, m, o, u, s, ...	[a, b, o, u, t, , , c, h, a, n, g, e, , l, ...
9.4	[r, r, e, , t, h, e, , d, o, o, m, o, u, s, ...	[o, r, , w, a, s, , s, h, u, t, , a, g, a, ...
7.8	[t, h, e, , d, o, o, m, o, u, s, e, , s, o, ...	[l, e, a, v, e, s, , , a, s, , t, h, e, r, ...
8.8	[e, , t, h, e, , d, o, o, m, o, u, s, e, , ...	[e, , w, i, t, h, o, u, t, , i, n, t, e, r, ...
7.2	[a, n, c, e, , a, n, d, , t, h, e, , d, o, ...	[r, e, a, t, , d, e, l, i, g, h, t, , i, t, ...
10.6	[, h, e, a, d, , i, n, , a, , l, i, t, t, ...	[, t, o, e, s, , w, h, e, n, , t, h, e, y, ...
9.6	[, t, h, e, , d, o, o, m, o, u, s, e, , s, ...	[, s, h, e, , h, a, d, , k, e, p, t, , a, ...
5.4	[u, l, d, , n, o, t, , a, , l, i, t, t, l, ...	[n, d, e, r, , i, f, , i, , s, h, a, l, l, ...
13.4	[e, a, r, , t, h, e, , d, o, o, m, o, u, s, ...	[e, a, r, n, , , w, e, l, l, , t, h, e, r, ...
9.8	[, s, h, e, , s, a, i, d, , t, h, e, , m, ...	[, s, a, i, d, , a, l, i, c, e, , , t, h, ...

Average Accuracy : 9.9

8 Conclusion

Through our experimentation, we observed varying levels of performance across the different architectures. The LSTM and GRU models performed slightly better thna the basic RNN in terms of accuracy, indicating their effectiveness in capturing long-range dependencies within sequential data. Furthermore, by adjusting certain hyperparameters such as hidden size and dropout rate, we didn't observe any relevant change. But overall, all the networks didn't work well in this task, maybe we should explore more the architectures of the networks but due to little computational power we can't proceed forward.

References

[1] RNN-Recurrent Neural Network *RNN*. Available at: https://www.simplilearn.com/tutorials/deep-learning-tutorial/rnn. Accessed 4 May. 2024.

[2] LSTM-Long Short Term Memory *LSTM*. Available at: https://en.wikipedia.org/wiki/Long_short-term_memory. Accessed 4 May. 2024.

[3] GRU-Gated Recurrent Unit *GRU*. Available at: https://www.ibm.com/topics/recurrent-neural-networks#:~:text=Gated%20recurrent%20units%20(GRUs)%3A,gate%20and%20an%20update%20gate.. Accessed 4 May. 2024.