

P3 Poem Composition

Name: 赵子涵

Major: Computer Science

Student ID: 2023E8013282148

Task

In this project, we aim to generate text following the given one. Specifically, text are restricted to Chinese poem.

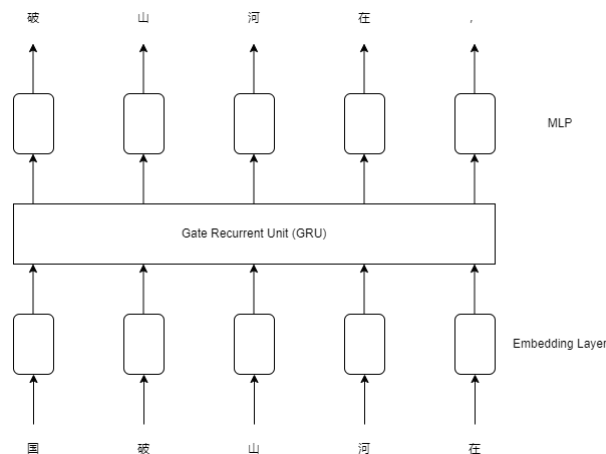
Dataset

We use the dataset from the course, which contains 57580 poetry in Tang dynasty.

Model Design

Considering the computing resource accessible currently, we design a simple model for this task.

The model consist three main part, which are, the embedding layer, the time series model GRU, and MLP layer. An illustration is shown above.



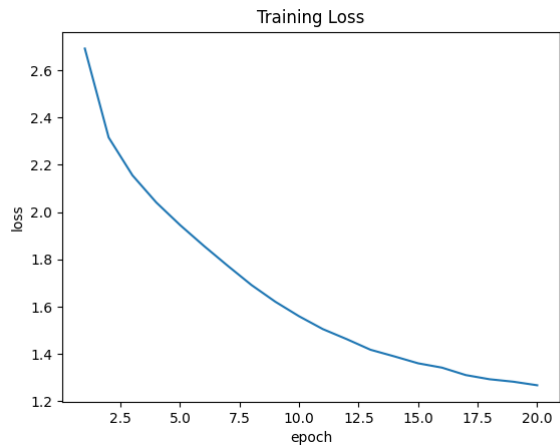
The embedding layer could be implemented as a full connection layer. It tries to learn to embed the one dimension word into high dimension, which contains more information. The GRU is generally a time series model. Contrast to basic RNN, it utilizes the gate control method to settle the problem of vanishing gradient, hence long time memory is preserved. At the same time, it has a more simple structure than LSTM. The MLP layer is implemented by two stacked linear layer, which map the output of GRU to one hot coding as the result. We use cross entropy loss as criterion.

Result

The parameter we choose are as follows

```
1 BATCH_SIZE=100
2 LR=0.001
3
4 embed_dim=100
5 hidden_dim=1000
6 num_layer=2
7 mlp_dim=200
```

The figure below shows the loss descending as training epoch increase.



Since there are no test set for this task, we wrote six prompts as the starting line, and make the model compose the rest. See result as below.

```
1  input text: 湖光秋月两相和
2  output text: 湖光秋月两相和，水馆风闲两相过。两岸红旗俱几许，风劲银澜难数过。新杯一醺杯前醉，夜夜明星月上无。
3
4  input text: 国破山河在
5  output text: 国破山河在，天涯道路除。家书辞日诏，兵气入秦家。壮气连云汉，狂风似水流。自怜长白首，皆恨白苹洲。不羨山僧便，宁辞东海鸥。还将谢戎策，归去卧云鸥。
6
7  input text: 明月几时有？
8  output text: 明月几时有？迢迢楚云飞。因高不可越，孤兴正悠哉。青云不可遇，空伤此时违。
9
10 input text: 慈母手中线
11 output text: 慈母手中线，能令不得归。主人不得主，日夜守车衣。僮仆不顾恩，为我披衣冠。此时不得意，君去看云衣。
12
13 input text: 昔人已乘黄鹤去，此地空余黄鹤楼。
14 output text: 昔人已乘黄鹤去，此地空余黄鹤楼。青林独坐见人吏，白日到此难可留。黄河流水无多事，白云三十成白首。何时为道经长人，更作江南老夫客。
15
16 input text: 秦时明月汉时关
17 output text: 秦时明月汉时关，万里相逢今屈宋。君向征途去已亡，河阳日落淮南薇。玉人不识人间事，羞把芙蓉亲早衣。
```