

SDML Homework 2-0 2-1 report

R08922A04 林承德

Hw2-0<Autoencoder>

在training時 **validation data** 最好的accuracy 是落在**0.929**，同個epoch 中的 **training data accuracy** 是**0.960**

```
"train": {
  "loss": 1.1740784782800064,
  "acc": 0.9601220050148589
},
"valid": {
  "loss": 1.2383503806591034,
  "acc": 0.9290811014175415
}
```

Testing : acc =0.928 on hw2.0_testing_data.txt

```
(20275, 13)
(base) R08922A04@cuda2:~/sdml/hw2-0-1$ python3 hw2.0_evaluate.py --true_file hw2.0_testing_data.txt --predict_file predict_doc.txt
0.9284833538840938
```

Hw2-1-1

因為在實作上遇到了一點問題，沒能在每個epoch算出當下的accuracy，所以我選擇在**Training** 完成後，我拿原本所有的data (**training + validation**, 我還是有切成這兩個部分，並且通過random 設定control signal) 一起送進去predict，看結果的 accuracy rate 得到**0.994**

```
python3 hw2.1_evaluate.py --training_file source_doc1.txt --result_file predict_doc1.txt
accuracy: 0.9948834726053438
```

另外 **pretesting**的部分，我拿助教給的simple data來跑 **acc= 0.99**

```
python3 hw2.1_evaluate.py --training_file hw2.1-1_sample_testing_data.txt --result_file predict_doc1.txt
accuracy: 0.99
```

Testing acc =0.991 on hw2.1-1_testing_data

```
(base) R08922A04@cuda3:~/sdml/hw2-1-1-4$ python3 hw2.1_evaluate.py --true_file hw2.1-1_testing_data.txt --predict_file predict_doc1.txt
usage: hw2.1_evaluate.py [-h] --training_file TRAINING_FILE --result_file RESULT_FILE
hw2.1_evaluate.py: error: the following arguments are required: --training_file, --result_file
(base) R08922A04@cuda3:~/sdml/hw2-1-1-4$ python3 hw2.1_evaluate.py --training_file hw2.1-1_testing_data.txt --result_file predict_doc1.txt
accuracy: 0.99164
```

Hw2-1-2

同上，我也是在training都完成後，才測結果

拿所有data(training + validation)測的結果為0.989

```
on ~/Desktop/碩一/SDML/hw2-1/hw2-1 (task2) control words INSERT
A04 python3 hw2.1_evaluate.py --training_file source_doc.txt --result_file predic
A04 t_doc.txt
A04 accuracy: 0.9894387406921673
```

此外，拿助教給的simple data來測 acc約為 0.986

```
~/Desktop/碩一/SDML/hw2-1/hw2-1 (task2) control words INSERT
4 python3 hw2.1_evaluate.py --training_file hw2.1-2_sample_testing_data.txt --r
t esult_file predict_doc.txt
accuracy: 0.9861111111111112
~/Desktop/碩一/SDML/hw2-1/hw2-1 (task2) control words INSERT
```

最後的testing acc = 0.988 on hw2.1-2 _testing_data.txt

```
(base) R08922A04@cuda3:~/sdml/hw2-1-2-4$ python3 hw2.1_evaluate.py --training_fi
le hw2.1-2_testing_data.txt --result_file predict_doc.txt
accuracy: 0.988977007664112
```

Package Using:

```
import numpy as np
import json
import pandas as pd
import argparse
import random
from sklearn.model_selection import train_test_split
import torch
import torch.nn as nn
import torch.optim as optim
from torch.utils.data import TensorDataset, DataLoader
from model import Encoder, Decoder, Seq2Seq
from os.path import join
from tqdm import tqdm
import os
```

My solution

稍微簡述一下，我怎麼做的，就是拿1~2層的GRU來當 encoder和decoder，並將他們接起來，成為我的sequence to sequence model！Embedding的部分呢，我在hw2_0有先用 pre-trained 好的glove，那在hw2_1的部分則是沒有的。