

Environment(環境):

MAC OS X

Language(語言):

C

How to execute:

In the Command line do the following:

make

```
./train iteration model_init.txt seq_model_01.txt model_01.txt
./train iteration model_init.txt seq_model_02.txt model_02.txt
./train iteration model_init.txt seq_model_03.txt model_03.txt
./train iteration model_init.txt seq_model_04.txt model_04.txt
./train iteration model_init.txt seq_model_05.txt model_05.txt
./test modellist.txt testing_data1.txt result1.txt
./test modellist.txt testing_data2.txt result2.txt
./acc testing_answer.txt result1.txt
```

Code:

1.train.c:

Using “hmm.h” load “model_initial.txt” and train from “seq_model_01~05.txt” output “model_01~05.txt”。

2.train.h:

fscanf “seq_model_01~05” and the Time=50 for one sample.(每50個observation看成一筆資料)。

And we have 10000 samples(有10000筆資料)。

Using “function.h” calculate the alpha beta game and epsilon to update HMM(π A B)。

3.function.h:

alpha:which is return array[State][Time]

beta:which is return array[State][Time]

game:which is return array[State][Time]

gama_withoutTime:which is return array[State]

epsilon:which is return array[State][State]

FindMaxModel:which is implement Viterbi algorithm and return the model which is MAX probability when given observation

4.test.c:

Using “hmm.h” load “modellist.txt” which contain the “model_01~05” from training.

And load “test_data1~2.txt” to test the model.

5.test.h:

load “test_data1~2” and call the function FindMaxModel in ”function.h”

6.acc.c

load “testing_answer.txt” and “result1.txt” output the accuracy

Problem I solve:

In the beginning, I have problem that return 2D array using pointer.
And I learn from StackOverflow and TA's help. Thanks TA a lot.

How to create an 3D array

- Static

```
int array[3][3][3];
```

- Dynamic

```
int ***array;
array = (int***)malloc(sizeof(int**) * 3);

for (i=0; i<3; i++) {
    array[i] = (int**)malloc(sizeof(int*) * 3);

    for (j=0; j<3; j++) {
        array[i][j] = (int*)malloc(sizeof(int) * 3);
    }
}
```

6

Just like the image

Result:

First time , I train the "model_01~05" for 30 iterations but acc just =80%
Second time , I train the "model_01~05" for 50 iterations and acc = 82%

Normally, the more iterations the more acc.

But I have the question that more iteration may cause overfit or not?
And this HMM Model have any Regularization method?