Digital Speech Processing HW1
Discrete Hidden Markov Model Implementation

- Environment
  - Ubuntu 12.04
  - g++-4.7

- Directory tree
  - hw1_b00902064
    - acc.txt
    - calAuc.cpp / test.cpp / train.cpp
    - Document.pdf
    - Makefile
    - model_01-05.txt
    - modelType.h
    - result1.txt / result2.txt

- Makefile setting
  - g++ compiler: "g++-4.7"
  - acceleration: "-O3"
  - number of iteration: "1000"
  - data directory: "../dat"
    (model_init.txt, seq_model_##.txt, modellist.txt)

- How to execute
  - Change Makefile setting if necessary
  - "make compile": compile *.cpp file
  - "make runTrain": train models (parallel-process is used)

- ○ "make runTest": predict answer
- ○ "make runAuc": calculate accuracy and write in acc.txt
- ○ "make all": execute four above-said operations in order
- ○ "make clean": delete executable files

- Result
  - ○ 6-state
    - with 1 iteration: 0.766400
    - with 10 iteration: 0.540400
    - with 20 iteration: 0.796400
    - with 30 iteration: 0.812800
    - with 40 iteration: 0.826400
    - with 50 iteration: 0.824000
    - with 100 iteration: 0.832800
    - with 500 iteration: 0.858000
    - with 1000 iteration: 0.870400
    - with 2000 iteration: 0.869200

  - ○ 8-state
    - fair initial and observation probability; self-transition and to-next-state transition is twice than others
    - with 1000 iteration: 0.754800

- Conclusion
  - ○ I rewrite the hmm.h into modelType.h for convenience.
  - ○ Times of Iteration will affect accuracy a lot, especially for less iteration.
  - ○ Even after 2000 times of iteration, some model still not converge.
  - ○ Increasing the state number may not improve.
  - ○ Initial probability affects the final result a lot.