DIGITAL SPEECH RECOGNITION HOMEWORK 1

Discrete Hidden Markov Model Implementation

醫學一 謝德威 B05401009

Environment

Macbook Pro 2013
CPU - Intel Core i5 2.4 GHz
RAM - 4GB DDR3 MHz 1600
OS - OSX Yosemite 10.10.2
Compiler - Apple LLVM version 6.1.0 (clang-602.0.53) (based on LLVM 3.6.0svn)

Implementation

Baum-Welch Algorithm and Viterbi Algorithm using C.

Install and Run

```
make
./train $iter model_init.txt seq_model_num.txt
   model_num.txt
./test modellist.txt testing_data.txt result.txt
```

- $iter is the number of training iterations
- model_init.txt is the initial model file
- seq_model_num.txt is the training data for model #num
- model_num.txt is the file to store trained model #num
- modellist.txt is the file containing filenames for all models to be tested
- testing_data.txt is the testing data file
- result.txt prints the prediction and probability for the most probable model

Use the following code for cross-validation during training:

```
FILE *ground_truth_file = open_or_die("testing_answer.txt",
   "r");
testing_result_file = open_or_die(argv[3], "r");
double testing_accuracy = accuracy(testing_result_file,
   ground_truth_file);
printf("Testing complete, accuracy = %f", testing_accuracy);
```

Experiments

#Iterations	Accuracy
10	0.5372
50	0.8244
100	0.7976
500	0.8528
1000	0.8616

Additional Experiments

1. Batch size: 1000 (for(s = epoch%10; s < num_samples; s+=10))
   Iterations: 50 Accuracy: 0.8192
   Performance: Training is significantly faster
2. Update after evaluating each sample, averaging into current parameter
   Pi/Aij/Bj(k) = 0.9999*Pi/Aij/Bj(k) + 0.0001*Pi'/Aij'/Bj(k)'
   Iterations: 50 Accuracy: 0.7948
   Performance: Training is fastest, but accuracy is lower.

Problems to Discuss

1. Compare the performance of straightforward version and log version of viterbi algorithm.
2. Compare the performance of the EM version and convex optimization version of Baum-Welch.
3. Compare the computational complexity and prediction accuracy of HMMs and RNN's (i.e. with LSTM/GRU units).
4. The correct way to calculate P(O|Lambda)
5. Auto adjust of HMM parameters. (To avoid overfitting)
6. Will using small batch update or averaging update into current parameters for each sample still converge? (like mini-batch & stochastic gradient descent)

References

1. L. Rabiner and B.H. Juang, Fundamentals of Speech Recognition.
2. D. Jurafsky and J. Martin. Speech and Language Processing.
3. C. Bishop, Pattern Recognition and Machine Learning.