MP4: Multimodal Digit Recognition with HMM

ECE 417 – Multimedia Signal Processing

University of Illinois at Urbana-Champaign

Spring 2014

Instructor: Professor Hasegawa-Johnson

TA's: Daniel Soberal, Po-Sen Huang

Prepared by Po-Sen Huang

Multimodal Digit Recognition with HMM

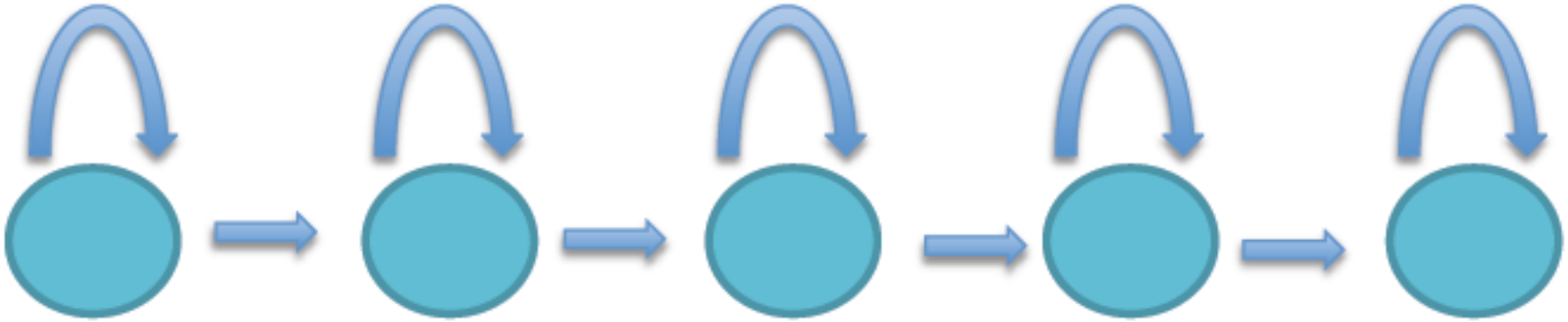
- Data: spoken digit video clips
  - You do NOT need to do feature extraction
  - Audio feature: MFCC
  - Visual feature: lips tracking
- Task:
  - Train HMMs for the spoken digits “2” and “5”
  - Use HMMs to do maximum likelihood recognition
    - Audio
    - Video
    - Audio-visual [concatenate the audio and visual feature]

Three basic problems for HMMs

1) Given training data (one or more sequences), find the best HMM
   - EM algorithm, code provided
2) Given an HMM and a sequence, calculate the likelihood
   - Forward/Backward algorithm, your work
3) Given an HMM and a sequence, find the most likely hidden state sequence
   - Viterbi algorithm, bonus points

5 states left-right non-skip HMM

- Specify A_init to force the connectivity among states
  - In EM updates, zero entries in A will remain zero



0.8	0.2	0	0	0
0	0.8	0.2	0	0
0	0	0.8	0.2	0
0	0	0	0.8	0.2
0	0	0	0	1

Evaluation: leave-one-out

- There are 10 sequences for “2” and 10 for “5”
- Using leave-one-out scheme
  - Each time you exclude one sequence from the training set for testing
  - Repeat this for 20 times, get the average accuracy

Matlab functions

- Train HMM ghmm_learn.m
  - [P0,A,mu,sigma] = ghmm_learn(Yseq,N,Ainit)

- Forward algorithm gmhmm_fwd.m
  - [alpha,scale] = gmihmm_fwd(Y,A,P0,mu,sigma)

- Backward algorithm ghmm_bwd .m
  - beta = ghmm_bwd(Y,A,P0,mu,sigma,scale)

What to Submit:

- Report
  - Result In tabular form
  - Explanation and analysis of the Viterbi, forward, and backward algorithm
- Matlab files
- README file to tell us how we run your code to obtain the same results as you did

Reference

- Rabiner, L., “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proceedings of the IEEE*, vol. *77, no.* 2, pp. 257-286, 1989