Welcome

Spring 2010: Design and Implementation of Speech Recogintion Systems

Instructors:
Bhiksha Raj
Rita Singh, Mosur Ravishankar

Administrivia

- Short lecture today (only very brief introduction)
- Still resolving logistic issues
  - 3 different numbers
    - 11756 / 18799D / ??
  - Not all students on my lists
  - Not all students receiving all notifications
- Course listings conflict on different webpages
  - CMU: Monday/Wed, GHC
  - LTI: Tue/Thu, GHC
  - **ECE: Monday/Wed, HH**

Administrivia

- Course website:
- http://asr.cs.cmu.edu/

- TA: Not yet assigned

- Instructor: Bhiksha Raj
  - GHC 6705
  - bhiksha@cs.cmu.edu
  - Phone: 8-9826
  - Office hours: TBD

What will the course be about

- This will be a hands-on course
  - Everyone is expected to code
    - Extensively
    - You may use any programming language
      - C, C++, Java, LISP, Matlab, Python, Ruby..
- The stress will not be on theory
  - It will be on hands-on practice
- We will discuss algorithms and implementation details

Projects

- Teams must present projects
  - Each team gets to present every project
  - Presentations will be brief: 5-10mins
- Grading based entirely on how many “projects” are correctly completed

Format of Course

- Lectures
- A series of projects of exponentially increasing complexity
- Projects are arranged in multiple levels
  - Isolated word recognition
  - Continuous speech recognition
  - Grammar based recognition
  - Ngrams
  - Sub-word units
  - Parameter sharing
  - Approximate decoding strategies

Format of Course

- Students will be grouped into a small number of teams
- Projects must be completed by teams
- Every team is expected to present their work at various stages of each project
  - Code description
  - Algorithmic and implementation details
  - Problems faced, solutions etc.

Projects

- Project 1: Capturing Audio
- Project 2: Feature computation
  - Plug feature computation into audio capture
  - Modify feature computation for buffered audio
  - Visualize various partial results in feature computation
  - Modify various parameters and visualize output
- Project 3: DTW-based recognition of isolated words
  - String matching using DP
  - Generalize string matching to DTW
  - Record templates
  - Create feature-based templates
  - Pattern matching and recognition

Projects

- Project 4: HMM-based recognition of isolated words
  - Viterbi decoding with simple Gaussian densities
  - Viterbi decoding with mixture Gaussian densities
- Project 5: Training HMMs from isolated recordings (Viterbi method)
  - Recording data
  - Segmenting data
  - Training models
- Project 6: Training and recognition of isolated words
  - Record data for a chosen vocabulary
  - Train models of different structures
  - Recognition

Projects

- Project 7: HMM-based recognition of continuous word strings
  - Continuous ASR of words
  - Continuous ASR of words with optional silences
  - Training a set of word models (carried over from previous exercise)
  - Evaluation

Projects

- Project 8: Grammar-based recognition of continuous words
  - Building graphs from grammars
  - Building HMM-networks from grammars
  - Recognition of continuous word strings from a grammar

Projects

- Project 9: Grammar-based recognition from Ngram models
  - Conversion of Ngrams to FSGs
  - Grammar-based recognition of continuous speech from Ngrams

Projects

- Project 10: Baum-Welch training
- Project 11: Sub-word units – learning models for phonemes
  - Recognition using words
- Project 12: Context-dependent units – learning models for context dependent units

Projects

- Project 14: Decoding with context dependent units
  - Build word models
- Project 15: Decision trees and state tying

Tasks

- Form Teams
  - Otherwise teams will be assigned
  - Email me your teams by Sunday night
    - You will be arbitrarily assigned to a team on Monday
- Projects 0 and 1 will be due on Wednesday, the 27th
  - Presentation of running code with visual output

Project 0

- Audio capture:
  - Live capture of audio
- A program that
  - Captures audio directly form microphone
    - Responds to keyhit – keyhit to turn on record, keyhit to turn it off
    - Externally set sampling rate and sample format
      - 16000hz, 16-bit samples, 8000hz, 16-bit sampling
  - Performs some action on streaming audio
    - Compute the sum of blocks of audio
      - 400 sample windows, a shift of 100 samples between windows
- Portaudio: www.portaudio.com/

Project 1

- Feature computation
- Will be assigned next Wednesday..