

LING 120:
Language and Computers
Semester: Fall '17

Instructor: Sowmya Vajjala

Iowa State University, USA

6 November 2017

Class Outline

- ▶ Announcements
- ▶ Quick recap of last class
- ▶ Question from last class: discussion
- ▶ Assignment 5 - discussion

Announcements

- ▶ Final Exams topics list is online - I will discuss this next week, but you can take a look in the meanwhile.
- ▶ There is a forum on review topics - post there if you want me to revisit some topic or talk about something we did not discuss before
- ▶ Assignment 6 is due next week - submit on time.
- ▶ Makeup assignment: I can consider a makeup assignment for those who want grade improvement (up to 5%). Talk to me after the class or send an email if you are interested. Note: Make-up exam will be one-one oral exam during office hours in the week before finals.

Recap of last class

Concepts discussed

- ▶ Speech processing - overview of tasks involved.
- ▶ Pronunciation dictionaries (using ARPABet or IPA)

Concepts discussed

- ▶ Speech processing - overview of tasks involved.
- ▶ Pronunciation dictionaries (using ARPABet or IPA)
- ▶ But there is pronunciation variation too

Concepts discussed

- ▶ Speech processing - overview of tasks involved.
- ▶ Pronunciation dictionaries (using ARPABet or IPA)
- ▶ But there is pronunciation variation too
- ▶ Speech to Phoneme mapping is difficult. If we assume we have some way, we need an additional layer of certainty, for which we can make use of a "language model".

Attendance Exercise from last class

- ▶ After all this discussion, what do you think are the different steps in speech recognition?
- ▶ Also, here are four exchanges I had with Siri this morning: try to figure out what I meant and why it understood like this:
 - ▶ Me: Siri, I want to understand Holly work
Siri: Sorry, I couldn't find 'Holly' in your contacts.
 - ▶ Me: Siri, you really bad India next
Siri: But but.
 - ▶ Me: Can you recognize speech?
Siri: Do you want to increase or decrease the volume?
 - ▶ Me: No, can you do speech recognition?
Siri: Do you want to increase or decrease the volume?

Q1

- ▶ Processing the speech: splitting it into sounds
- ▶ "Translating" sounds to letters, to words and to phrases in English (or any language)
- ▶ Dynamically change as the user talks more and machine has more information about the context

Q2

- ▶ "how you work" – > "Holly work" – > contact called Holly
- ▶ "at Indian accents" – > "India next" (and what is "But, But"? Who knows?)
- ▶ recognize speech, speech recognition: problem here is not with ASR, but with speech understanding.

Assignment 5 discussion

Question 1: Steps in language identification

- ▶ While a dictionary approach may seem sufficient, in real life, text classification (and probability) are both used in designing such systems.
- ▶ If we are considering languages with different scripts, we won't need these though (why?)

Question 1: Steps in language identification

- ▶ While a dictionary approach may seem sufficient, in real life, text classification (and probability) are both used in designing such systems.
- ▶ If we are considering languages with different scripts, we won't need these though (why?)
- ▶ So, resources we need: large collection of examples for each language, and some way to do classification.
- ▶ Evaluation: on the test set

Question 2: Opinion mining service

- ▶ Features to look for: words, may be 2–3 grams for getting typical complaint phrases, use of all caps, exclamations, use of negative words etc. (Yes, these can be automatically extracted)

Question 2: Opinion mining service

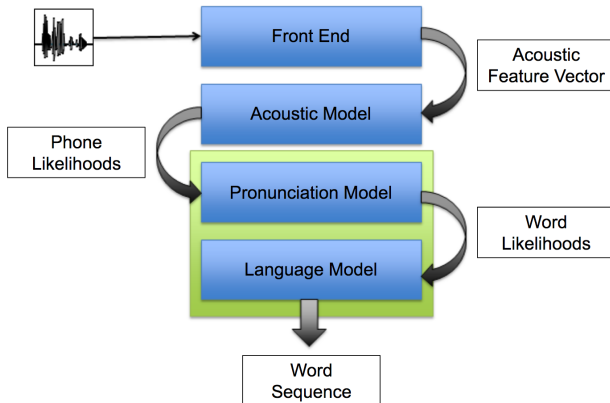
- ▶ Features to look for: words, may be 2–3 grams for getting typical complaint phrases, use of all caps, exclamations, use of negative words etc. (Yes, these can be automatically extracted)
- ▶ Diagnosing customer happiness: Asking them to rate, Looking for how many of these people returned to the airline after complaints etc (Partly automated?)

Question 2: Opinion mining service

- ▶ Features to look for: words, may be 2–3 grams for getting typical complaint phrases, use of all caps, exclamations, use of negative words etc. (Yes, these can be automatically extracted)
- ▶ Diagnosing customer happiness: Asking them to rate, Looking for how many of these people returned to the airline after complaints etc (Partly automated?)
- ▶ Other sources of information: social media, review websites etc. Automating this: sentiment analysis, information extraction about specific aspects etc.

Automatic Speech Recognition

Speech Recognition



Building acoustic and pronunciation models

- ▶ Acoustic model: conversion of speech wave forms in some form of numeric vector indicating properties of that speech sample.
- ▶ Pronunciation model: Mapping sounds to letters, and words.
- ▶ Typically, both of them are put together, and are developed by first collecting hours and hours of recorded speech, creating detailed manual transcriptions for them, and using machine learning to learn to map sounds to words.

Building language models

- ▶ In comparison, getting lots of language data for different languages is not terribly difficult (news papers, wikipedia, movie subtitles etc).
- ▶ Developing language models (with n-grams, for example) is something we already studied before.

Building Pronunciation Models: Issues

- ▶ Different people pronounce differently - different accents

Building Pronunciation Models: Issues

- ▶ Different people pronounce differently - different accents
- ▶ Pronunciation dictionaries: good ones for English, even handle POS variations. But, what about names? what about accents?

Building Pronunciation Models: Issues

- ▶ Different people pronounce differently - different accents
- ▶ Pronunciation dictionaries: good ones for English, even handle POS variations. But, what about names? what about accents?
- ▶ What about other languages? Where do we get so much of recorded data??

Building Pronunciation Models: Issues

- ▶ Different people pronounce differently - different accents
- ▶ Pronunciation dictionaries: good ones for English, even handle POS variations. But, what about names? what about accents?
- ▶ What about other languages? Where do we get so much of recorded data??
- ▶ One solution Google came up with:
[https://research.googleblog.com/2011/03/
word-of-mouth-introducing-voice-search.html](https://research.googleblog.com/2011/03/word-of-mouth-introducing-voice-search.html)

Different Forms of ASR

- ▶ Isolated word recognition (slow speech, each word surrounded by a pause)
- ▶ Read speech recognition (humans talking to machine e.g., dictation)
- ▶ Continuous speech recognition (natural, running speech)
- ▶ Conversational speech recognition (humans talking to each other, noisy environments etc)

-these are incrementally harder tasks. (Transcribing a meeting - is it even possible??)

ASR in Real World

- ▶ All AI-assistants that have spoken language input (Siri, Cortana etc)
- ▶ Google docs support ASR in several languages - did you know that?
- ▶ Dictation tools
- ▶ `swiftscribe.ai` from Baidu.

Evaluation of ASR

- ▶ Word Error Rate
- ▶ Sentence Error Rate
- ▶ In the case of dialog systems involving spoken input: concept error rate

Attendance Exercise

Code-switching refers to people switching between languages while speaking or writing. It is not very common in US, but in multi-lingual societies, it is quite common. What or how do you think ASR systems should be tuned to such scenarios?

Next class: Speech Synthesis