LING 120:
Language and Computers

Semester: Fall '17

Instructor: Sowmya Vajjala

Iowa State University, USA

03 November 2017

Outline for today

- Discussion about Wednesday's question
- Dialog Systems - conclusion
- Assignment 6 Description
- Overview of speech recognition
- Next week: Speech processing continued
- Reminder: Assignment 5 due tomorrow!!

Wednesday's Exercise

The Yes Bot: `http://www.nacloweb.org/resources/problems/2013/N2013-L.pdf`

- Give an example of a sentence that, when said by the CEO, will cause Yesbot to make a mistake.
- Provide two examples of words that, when the CEO uses them in a sentence, will sometimes cause Yesbot to make a mistake, but sometimes won't. Explain why.
- Are there any words that will always cause Yesbot to make a mistake, (that is, say a lie) any time the CEO uses them? List any you can, and explain why or why not.

Your Responses-1

- Question 1: Give an example of a sentence that, when said by the CEO, will cause Yesbot to make a mistake.
- Response 1: "I have a lot of power." "Yes, sir or ma'am, it is true that I have a lot of power."
- Response 2: "I can't believe Dave thought he would get my job!" "Yes, sir or ma'am, it is true that I can't believe Dave thought he would get my job." Obviously, the owner can believe it, so the statement that Yesbot replies with is false.

Your Responses-2

- Question 2: Provide two examples of words that, when the CEO uses them in a sentence, will sometimes cause Yesbot to make a mistake, but sometimes won't. Explain why.
- Response 1: "I" because a sentence that is strictly about the CEO wont make sense, but if the CEO says "I" and "you" in the same sentence then it can make sense. "Yes, sir or ma'am, it is true that I know the secrets", "Yes, sir or ma'am, it is true that you and I know the secrets"

Your Responses-3

- Question 3: Are there any words that will always cause Yesbot to make a mistake, (that is, say a lie) any time the CEO uses them? List any you can, and explain why or why not.
- Response 1: "Any type of homophone could make yesbot mess up. The program might be confused and use the wrong word or wrong form of a word. (ex. new and knew)"
- Other Responses: Words I and You, First and second person pronouns, question words, terms indicating commands.
- Answer from the solution: `http://www.nacloweb.org/resources/problems/2013/N2013-LS.pdf`

Dialog Systems - Quick summary

We discussed:

- What are dialog systems, where are they useful?
- What are the different kinds of dialog systems? How are they built?
- How can we evaluate dialog systems?
- What are some dialog systems that are used in current day world?

If you found this interesting ..

- Dialog systems are a very fascinating, and active area of research and development.
- Lot of scope for interesting jobs and projects, lot of cool work to do!

If you found this interesting ..

- Dialog systems are a very fascinating, and active area of research and development.
- Lot of scope for interesting jobs and projects, lot of cool work to do!
- To get a hang of what the real "dialog systems" people do (not in a 100 level course!), you can check out the following:
  - Speech and Language Processing textbook by Jurafsky and Martin (draft chapters online 29–30 are about Dialog systems: `https://web.stanford.edu/~jurafsky/slp3/`)
  - A course on Spoken Dialog system taught at UW: `http://courses.washington.edu/ling575/SPR2017/index.html`
  - Some readings from a Deep Learning course at UPenn: `http://dialog-systems-class.org/readings.html`

If you found this interesting ..

- Dialog systems are a very fascinating, and active area of research and development.
- Lot of scope for interesting jobs and projects, lot of cool work to do!
- To get a hang of what the real "dialog systems" people do (not in a 100 level course!), you can check out the following:
  - Speech and Language Processing textbook by Jurafsky and Martin (draft chapters online 29–30 are about Dialog systems: `https://web.stanford.edu/~jurafsky/slp3/`)
  - A course on Spoken Dialog system taught at UW: `http://courses.washington.edu/ling575/SPR2017/index.html`
  - Some readings from a Deep Learning course at UPenn: `http://dialog-systems-class.org/readings.html`
- If you know how to program, start writing your own version of Eliza or Parry!

Assignment 6 Description

- Theme: Descriptive writeups about dialog systems, addressing specific questions.
- 10% of your grade
- Due on 18th November
- Details on Canvas

About Final Exam

- Final Exam is 20% of your grade
- It is a 3-step process of which one step is done in class - if you miss that class, you don't get grade for that.
- You don't have sit in an exam center in the exams week!
- All details on Canvas, but I will more on this in the week before thanksgiving break!
- I will try to update topics by Monday though, so that you have enough time to think.

Our next topic: Speech processing

Questions to get you started -1

- How should we make a computer understand Speech?

Questions to get you started -1

- How should we make a computer understand Speech?
- According to you, what are the steps involved in speech recognition?

Questions to get you started -1

- How should we make a computer understand Speech?
- According to you, what are the steps involved in speech recognition?
- When a computer sees text, it breaks it down character by character, each character has some Unicode rep and so on. How does that work with Speech?

Questions to get you started -1

- How should we make a computer understand Speech?
- According to you, what are the steps involved in speech recognition?
- When a computer sees text, it breaks it down character by character, each character has some Unicode rep and so on. How does that work with Speech?
- If we map sounds (instead of characters) to a new alphabet (phonetic alphabet), that will make all sounds converted to alphabet - will it fix the problem?

Questions to get you started -1

- How should we make a computer understand Speech?
- According to you, what are the steps involved in speech recognition?
- When a computer sees text, it breaks it down character by character, each character has some Unicode rep and so on. How does that work with Speech?
- If we map sounds (instead of characters) to a new alphabet (phonetic alphabet), that will make all sounds converted to alphabet - will it fix the problem?
- What do you think are difficult to make out with this kind of mapping - vowel sounds or consonant sounds?

Questions to get you started -1

- How should we make a computer understand Speech?
- According to you, what are the steps involved in speech recognition?
- When a computer sees text, it breaks it down character by character, each character has some Unicode rep and so on. How does that work with Speech?
- If we map sounds (instead of characters) to a new alphabet (phonetic alphabet), that will make all sounds converted to alphabet - will it fix the problem?
- What do you think are difficult to make out with this kind of mapping - vowel sounds or consonant sounds?
- Once we map alphabet to sounds successfully, what else is left over?

Representing Pronunciation Symbolically, as Transcribed text

- The International Phonetic Alphabet (IPA) is an evolving standard whose goal is to transcribe the sounds of all human languages in a uniform manner. (`https://en.wikipedia.org/wiki/International_Phonetic_Alphabet`)
- Something specific to American English: ARPABet `https://en.wikipedia.org/wiki/ARPABET`

Phonetic Transcription - Example

ARPAbet Symbol	IPA Symbol	Word	ARPAbet Transcription
[p]	[p]	parsley	[p aa r s l iy]
[t]	[t]	tea	[t iy]
[k]	[k]	cook	[k uh k]
[b]	[b]	bay	[b ey]
[d]	[d]	dill	[d ih l]
[g]	[g]	garlic	[g aa r l ix k]
[m]	[m]	mint	[m ih n t]
[n]	[n]	nutmeg	[n ah t m eh g]
[ng]	[ŋ]	baking	[b ey k ix ng]
[f]	[f]	flour	[f l aw axr]
[v]	[v]	clove	[k l ow v]
[th]	[θ]	thick	[th ih k]
[dh]	[ð]	those	[dh ow z]
[s]	[s]	soup	[s uw p]
[z]	[z]	eggs	[eh g z]
[sh]	[ʃ]	squash	[s k w aa sh]
[zh]	[ʒ]	ambrosia	[ae m b r ow zh ax]
[ch]	[tʃ]	cherry	[ch eh r iy]
[jh]	[dʒ]	jar	[jh aa r]
[l]	[l]	licorice	[l ih k axr ix sh]
[w]	[w]	kiwi	[k iy w iy]
[r]	[r]	rice	[r ay s]
[y]	[j]	yellow	[y eh l ow]
[h]	[h]	honey	[h ah n iy]
Less commonly used phones and allophones			
[q]	[ʔ]	uh-oh	[q ah q ow]
[dx]	[ɾ]	butter	[b ah dx axr]
[nx]	[ɾ̃]	winner	[w ih nx axr]
[el]	[l̩]	table	[t ey b el]

Figure 7.1 ARPAbet symbols for transcription of English consonants, with IPA equivalents. Note that some rarer symbols like the flap [dx], nasal flap [nx], glottal stop [q] and the syllabic consonants, are used mainly for narrow transcriptions.

Alright, we know how to do transcription

- So, let us say we prepare one exhaustive list of sounds in American English, or IPA or whatever we want.
- We then build a pronunciation dictionary - where, unlike the traditional dictionary which gives you meaning of a word - this gives you how to pronounce that word.
- So each time a user says something, let us say your program sees the spectrogram, and learns to cut the speech signal into individual sounds, stitch them into word sounds, and then, you look up the dictionary to match sounds to real words.
- Does that sound alright so far? (I am obviously skipping details!)

Alright, we know how to do transcription

- So, let us say we prepare one exhaustive list of sounds in American English, or IPA or whatever we want.
- We then build a pronunciation dictionary - where, unlike the traditional dictionary which gives you meaning of a word - this gives you how to pronounce that word.
- So each time a user says something, let us say your program sees the spectrogram, and learns to cut the speech signal into individual sounds, stitch them into word sounds, and then, you look up the dictionary to match sounds to real words.
- Does that sound alright so far? (I am obviously skipping details!)
- What else is missing after this point?

Let me do a quick demo

- Google Webspeech Demo. `https://www.google.com/intl/en/chrome/demos/speech.html`
- (Notice how words change on screen - I have questions in the next slide!!!)
- My text: "So now I'm trying to speak in Indian accent and we are going to see how good is the system with my accent"

Questions on the demo

- As I speak, you see words changing in the transcription sometimes after I moved on from that word and am speaking something else. Do you think it is just because the internet is slow or is there some other reason?

Questions on the demo

- As I speak, you see words changing in the transcription sometimes after I moved on from that word and am speaking something else. Do you think it is just because the internet is slow or is there some other reason?
- What is the intuition behind choosing the words it is choosing?

Questions on the demo

- As I speak, you see words changing in the transcription sometimes after I moved on from that word and am speaking something else. Do you think it is just because the internet is slow or is there some other reason?
- What is the intuition behind choosing the words it is choosing?
- If I say the same thing twice, is it showing the same transcription? If it is different, why do you think it is different?

Questions on the demo

- As I speak, you see words changing in the transcription sometimes after I moved on from that word and am speaking something else. Do you think it is just because the internet is slow or is there some other reason?
- What is the intuition behind choosing the words it is choosing?
- If I say the same thing twice, is it showing the same transcription? If it is different, why do you think it is different?
- Do you think we speak consistently in the same way, when we are speaking naturally?

Questions on the demo

- As I speak, you see words changing in the transcription sometimes after I moved on from that word and am speaking something else. Do you think it is just because the internet is slow or is there some other reason?
- What is the intuition behind choosing the words it is choosing?
- If I say the same thing twice, is it showing the same transcription? If it is different, why do you think it is different?
- Do you think we speak consistently in the same way, when we are speaking naturally?
- What causes speech variation between people?

Questions on the demo

- As I speak, you see words changing in the transcription sometimes after I moved on from that word and am speaking something else. Do you think it is just because the internet is slow or is there some other reason?
- What is the intuition behind choosing the words it is choosing?
- If I say the same thing twice, is it showing the same transcription? If it is different, why do you think it is different?
- Do you think we speak consistently in the same way, when we are speaking naturally?
- What causes speech variation between people?
- What causes speech variation between different versions of your own speech?

Pronunciation Variation is Real - It is not only about non-native accents

because				about			
ARPAbet	%	ARPAbet	%	ARPAbet	%	ARPAbet	%
b iy k ah z	27%	k s	2%	ax b aw	32%	b ae	3%
b ix k ah z	14%	k ix z	2%	ax b aw t	16%	b aw t	3%
k ah z	7%	k ih z	2%	b aw	9%	ax b aw dx	3%
k ax z	5%	b iy k ah zh	2%	ix b aw	8%	ax b ae	3%
b ix k ax z	4%	b iy k ah s	2%	ix b aw t	5%	b aa	3%
b ih k ah z	3%	b iy k ah	2%	ix b ae	4%	b ae dx	3%
b ax k ah z	3%	b iy k aa z	2%	ax b ae dx	3%	ix b aw dx	2%
k uh z	2%	ax z	2%	b aw dx	3%	ix b aa t	2%

Figure 7.8 The 16 most common pronunciations of *because* and *about* from the hand-transcribed Switchboard corpus of American English conversational telephone speech (Godfrey et al., 1992; Greenberg et al., 1996).

Why is there so much of variation?

A short primer to our vocal tract

- How is sound produced? - by rapid movement of air in the vocal organs of our body - from lungs through the trachea, and then out of mouth, and nose (nasal sounds)?
- Consonants sounds can be distinguished:
  - Depending on place of articulation (labial - lips coming together, dental, alveolar, palatal, velar, glottal)
  - Manner of articulation (nasal, voiced, unvoiced etc)
- Similarly, different movements produce different vowels.
- There can be stressed and unstressed sounds.

Note: For more details, read on phonetics or talk to a linguist to get references.

Factors influencing Phonetic Variation

- Rate of speech (how fast does one speak)
- Usage of different words - same letter can sound differently in context
- Deletion of some part of the word while speaking
- Gender, Social background, Dialect - all these influence your pronunciation
- Over time, pronunciation of words change too!

Note: For more details, look for sociolinguistics videos or books.

Attendance Exercise

- 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?