LING 120, Fall 2017:
Language and Computers
Topic: Overview of Natural Language Processing

Instructor: Sowmya Vajjala

Iowa State University, USA

6 October 2017 (Week 7)

Class outline

- Conclusion of NLP tasks: overview
- Assignment 4 Description
- Midterm presentations: reminders and instructions
- (if there is time) Preparation time for midterms
- Reminder: Assignment 3 due tomorrow!

Next week's plan: Your midterm presentations; General review and feedback.

NLP tasks: Discourse Analysis

- Task: Given a text (more than one sentence), analyze the relationships between sentences, identify what pronouns refer to what nouns, how is the same entity referred in different ways (Barack Obama, Obama, The President and so on).
- Application: Text summarization, Question-Answering, Essay scoring etc.

NLP tasks: Entailment and Paraphrasing

- Tasks: Analyzing if the meaning of a sentence is entailed in another sentence, if both sentences are paraphrases of each other etc.,
- Uses: question answering, information extraction, summarization etc.

NLP tasks: Language Generation

- ▶ Task: Generate text automatically.
- ▶ Texts should be grammatically and semantically correct. Should be human like.
- ▶ One of the toughest problems in NLP.
- ▶ Example: Create weather reports, match summaries, reports etc. automatically (without human intervention!)
- ▶ arria.com is a software company that does this successfully for English.

So...?

- Okay, so all these are different tasks. how do we solve them?

So...?

- ▶ Okay, so all these are different tasks. how do we solve them?
- ▶ You have seen the "Madly Ambiguous" game in the last class. How do you think are they trying to detect the "sense" of usage of words?

So...?

- Okay, so all these are different tasks. how do we solve them?
- You have seen the "Madly Ambiguous" game in the last class. How do you think are they trying to detect the "sense" of usage of words?
- We saw the video of Maluuba a couple of week's back, where the program answers a question about something in the text. Intuitively, what is happening there?

Solving all these problems for realworld applications

- What is involved?:
  1. Some large database of properties of language (may be elaborate dictionaries, lists of synonyms, antonyms etc.)

Solving all these problems for realworld applications

- ▶ What is involved?:
  1. Some large database of properties of language (may be elaborate dictionaries, lists of synonyms, antonyms etc.)
  2. Programs that tell the computer the rules of the language, and uses these databases along with the rules.

Solving all these problems for realworld applications

- ▶ What is involved?:
    1. Some large database of properties of language (may be elaborate dictionaries, lists of synonyms, antonyms etc.)
    2. Programs that tell the computer the rules of the language, and uses these databases along with the rules.
    3. Some means of making the computer work with "new, unseen" sentences, and words in the language.

Solving all these problems for realworld applications

▶ What is involved?:
1. Some large database of properties of language (may be elaborate dictionaries, lists of synonyms, antonyms etc.)
2. Programs that tell the computer the rules of the language, and uses these databases along with the rules.
3. Some means of making the computer work with "new, unseen" sentences, and words in the language.
4. Where it is difficult to write these computer programs, there should also be some way to make the computer "learn" from lots and lots of examples of language use. (Machine Learning)

Let us go back to Madly Ambiguous

- How do they do that?
- What among these above steps are they using? (What do you think?)
- http://madlyambiguous.osu.edu:1035/

Assignment 4 Description

- ▶ Topic: NLP overview
- ▶ Usual format - 2 questions, 10 marks etc.
- ▶ Deadline: 21 October
- ▶ Description: On Canvas

Instructions for Midterms

- Instructions on Canvas: FOLLOW THEM
- Stick to the time allotted (10min). Take questions, if any (3-5 min).
- Team assignments and Schedule: On Canvas - announcements page (and in slides)

Instructions for Midterms

- Instructions on Canvas: FOLLOW THEM
- Stick to the time allotted (10min). Take questions, if any (3-5 min).
- Team assignments and Schedule: On Canvas - announcements page (and in slides)
- Laptop to present: you can get yours, or use mine. If you use mine, send pdfs or use google slides on the browser.
- If you are sending/sharing the presentation online, do it by 1pm on your presentation day!
- A team member who does not attend their own team presentation will not get any grade. (20% will vanish into air!)

Attendance question for today

- ▶ Let us say you have a detailed dictionary for English (mentioning words, their possible parts of speech, examples, synonyms, antonyms, everything).

- ▶ I now ask you to write a computer application that takes a English sentence, and identifies the individual parts of speech of words
  e.g., in "This is my LING120 class". This and my are pronouns, is is a auxiliary verb, LING120 is a proper noun, class is a common noun.

- ▶ Based on your current intuition, what will you do to write such a program on paper?

- ▶ (You don't need to know any programming to answer this)

Next Week

- Monday and Wednesday - your midterm presentations
- Friday - Review, Feedback

Good luck with your preparation!!