CSE5539: Natural Language Question Answering

Autumn 2016

Instructor:

Huan Sun (sun.397@osu.edu)

Address:

Caldwell 115

Time:

Monday 2:20-4:10PM

Office Hours:

By appointment

Course Description

Today's paradigm of information search is in the midst of a significant transformation. Question answering (QA) systems that can precisely answer user questions are becoming more and more desired, in contrast to traditional search engines which only retrieve lengthy web pages. As demonstrated by exemplar intelligent systems including IBM Watson, Google Now, Apple Siri, Microsoft Cortana, and Amazon Echo, QA techniques enjoy tremendous potential in revolutionizing the way we interact with devices and data, e.g., by allowing voice commands to automate complicated tasks like planning a trip, or by directly answering questions regarding domain specific data such as medical forum posts, product reviews and online programming tutorials. In this seminar, we will study the recent progress of question answering, brainstorm possible future work, and think about the problem: Can you build a personal assistant specialized for tasks you care about?

Topics Covered

Topics include a variety of aspects of question answering:

1. Text Based QA
2. Knowledge Bases: Construction and Applications
3. Knowledge Based QA
4. Community QA
5. Reading Comprehension
6. Dialogue

7. Programming with Natural Language
8. Visual QA

Prerequisites

While there are no formal prerequisites for the class, some general knowledge about applied machine learning and text analysis will be needed.

Course Format

The goal of this course is to familiarize students with research topics related to question answering, text mining, and knowledge bases, and improve their ability of critical thinking. The course activities include (1) reading, presenting, and discussing papers and (2) conducting research-oriented projects. More details are listed as follows:

- Paper reading (10%): The class will read and discuss two papers each week. Before class, each student should read the assigned papers and write a short review (about half a page). These reviews should at least cover the following aspects: (a) A summary about what a paper does using 3-5 sentences; (b) Positive comments on the paper and what you have learnt (which could potentially benefit your own research); (c) Some limitations of the paper or suggestions for how the work could be improved or extended. It is encouraged to prepare questions about the papers for discussion in class. The purpose of writing reviews is to stimulate active thinking and spark discussion in class. Students are allowed to skip 2 reviews throughout the semester. Reviews should be submitted in Canvas before 12pm on the day of class. Please email your reviews to the instructor if there are any technical issues with submission.

- Paper presentation (30%): Each student is expected to present 2-3 papers during the entire period of the course. One does not need to write a review for a paper if leading the discussion. Each presentation takes 40-45 minutes covering the background, problem definitions, ideas, models, experiments, results, and related work. In each class, we will discuss two papers. Additionally, we will have 20-30 minutes for in-depth discussion about the papers, which can be in the middle of or after the paper presentation. Students to present need to send the instructor their slides for review **24 hours before class (Please do start preparation early)**.

- Paper/idea discussion (10%): Students are strongly encouraged to think actively. Please bring thoughtful questions and creative ideas in related topics for discussion. We will also discuss on-going class projects.

- Projects (50%): Students need to conduct a research-oriented project either by group (2-3 students per group) or independently. By the end of

the third week, project groups should be formed. As the semester goes on, each team gradually formulates a research project under the instructor's guidance. In the middle of the semester, students need to present and submit a project proposal. By the end of the semester, each team should present the project progress to the class and submit a final report.

Reading List and Schedule

- The tentative reading list is here https://docs.google.com/spreadsheets/d/1Iva4FDR_NKoHvYAirdt9FqIjkeNUNsR5qOSwWgUUcLs/edit?usp=sharing
- The list will be finalized before the first class, but students could now browse the current list to get a general sense of what papers they are to read.
- Each student needs to sign up to present at least 2 papers throughout the semester. Students can form groups to do the presentation. The instructor will notify students for sign-up once the reading list is fixed.

Resources

- Stack Exchange Data Dump: https://archive.org/details/stackexchange
- Yahoo Answers and Wikipedia Data dump, as used in http://times.cs.uiuc.edu/czhai/pub/autoanswer.pdf
- Semi-structured tables from the Web, as mined in http://www2016.net/proceedings/companion/p75.pdf
- Freebase-based Question answering datasets: http://www-nlp.stanford.edu/software/sempre/ and https://github.com/percyliang/sempre
- To get one version of Freebase, also check https://github.com/percyliang/sempre
- Tools to build your own neural networks: https://www.tensorflow.org/
- Product reviews and related questions: http://jmcauley.ucsd.edu/data/amazon/qa/

Project Ideas

- For students who are just taking this course to gain more knowledge or to get credits, they are very welcome to propose their own project idea.
- For students who are thinking about doing research in the instructor's group later on, or working in this area for their Ph.D., they are strongly recommended to discuss with the instructor on possible course projects.