Assignment2

Rik Bose

Extracting Narrative Chains

Submission

- a 2-4 page writeup of the narrative chains project
  - I estimate length without double spacing. Bad habit I picked up from the British education system.
- extended code from in-class coding sessions to perform the cloze story test
  - Write your own code in another file as much as possible.

Important links

- Cloze story task: `https://competitions.codalab.org/competitions/15333#learn_the_details`
- Chambers+Jurafsky: `https://www.aclweb.org/anthology/P08-1090/`
- Git repo with live-coding results and notes: `https://github.com/mrmechko/narrative_chains`

Writeup details

Your paper will have the following 4 sections

- Introduction
  - Describe narrative chains
    - Use an example not provided in the Chambers+Jurafsky paper
  - Describe the cloze test

        * Use an example not provided in the Chambers+Jurafsky paper
    - Describe some related work
        * How are the concepts described in the related work different or similar to what is happening in your system
        * Make sure to include the bibliographic reference
        * If you are working in groups, at least one piece of related work per person in your group

- Methods
    - Describe the algorithm implemented
    - What are some changes made to the system?
    - What libraries and other algorithms did you use?
    - Use pseudocode
    - Detail should be sufficient to reimplement the system without looking at the code

- Evaluation
    - How are you testing your system (cloze test)
    - Which data are you using to train and test on?
    - What are the results
    - Do you notice anything interesting or important about those results?

- Conclusion
    - Summarize what you have done
    - Summarize your results
    - What are some ways in which you could improve the algorithm if you kept working on it?

Extra Credit 1

This extra credit is enough to bring someone's grade up from a B range to an A range. Do ONE of the following two things. Note that the improvements suggested in the system extension might not necessarily create an improvement in performance. If you are going to work on the extra credit, run your ideas by me first.

- Submit your system online
  - `https://competitions.codalab.org/competitions/15333#learn_the_details`
  - Warning - this can be a lot of fiddly code and might ultimately be more effort than the other option
- Extend the system by implementing 2 or more of the following ideas
  - Improved counting: the provided system counts events as being unique to a story. In reality we have far more events than stories. How can you improve the counting of events to reflect that?
  - Improved event detection: the provided system uses the raw text of an event as the key. Therefore, *give* and *given* are treated as differet types of events. As a result there is less information about both *give* and *given* than if we had lumped them together. Find techniques to merge slightly different forms of the same word.
  - Knowing what types of entities are filling the dependency roles in your events might change the meaning of event. Instead of counting all pairs of coreferring events, try counting all pairs of coreferring events which point to the same type of referent. Perhaps you could merge the two types of scores?
  - Some other improvement of your choosing

Extra Credit 2

This extra credit is simply a bonus on this assignment (worth 10% of the assignment)

- Submit your assignment as a git repo on GitHub
- Your writeup should include a readme that tells me how to install any dependencies and run your code