About the Project

CS 5134/6034: Natural Language Processing

Shomir Wilson

Term Project Objectives

- Apply knowledge of NLP toward actual problems with data

- Work as a team

- Assimilate knowledge of NLP outside of formal instruction

SemEval

SemEval (http://alt.qcri.org/semeval2016/) is an annual workshop in the NLP research community.

Every year SemEval issues a set of tasks with data, and participants compete to produce the best performance results.

SemEval 2016 featured 16 tasks; you will work in teams on one of four selected tasks.

Basic Structure of the Project

You will work in teams of **four** to solve a problem.

The problem will be your selection from a permitted subset of **SemEval 2016 tasks**.

Your work will include:

- Background research to plan a solution
- Coding the solution
- Evaluating your solution (by itself and in comparison)
- Presenting a proposal and your final results
- Providing feedback to other teams
- Writing a proposal report and a final report

Eligible Tasks: Textual Similarity

Task 1: Semantic Textual Similarity
Given two sentences, how similar are their meanings?

Subtasks: English-English pairs and Spanish-English pairs.
You are required to do both.

Task 3: Community Question Answering
Rank {comments, other questions, external comments}
for their relevance to a given question in an online
community.

Subtasks: You are required to do subtasks A-C. D is
optional.

Eligible Tasks: Sentiment Analysis

Task 4: Sentiment Analysis in Twitter
Given a set of tweets, what sentiments do they express?
You are required to do Subtasks A-E. (There are strong similarities between the subtasks.)

Task 6: Detecting Stance in Tweets
Given a tweet, determine whether it is for or against a topic (or neither).
Subtasks A and B are required.

Timeline

- September 20: Team name, membership, and task choice due

- October 18: Written proposal reports due

- October 18, 20, 23: Proposal presentations

- November 27: Written final reports due

- November 27, 29; December 1: Final presentations

Teams and Task Selections

Check the class discussion board on Blackboard for two forums: one to help you find team members and another to register your team.

There is a limit of five teams for any task. If five teams have already registered for a task, then no more registrations for that task are permitted.

Proposal and Final Presentations

You will give two presentations:

1. A proposal presentation with your plan of work and any preliminary results
2. A final presentation that summarizes your results and how they compare to performances in the competition

I will provide more details on requirements as the presentations approach.

You will also be required to provide feedback on others' presentations.

Proposal and Final Reports

You will use the style and formatting for SemEval submissions for your proposal and final reports.

This includes formatting for the citation and references.

You can find a link to the template files on the SemEval website. LaTeX or Word are both fine.

I will give you guidelines for these later.

Situating Your Work (Avoiding Plagiarism)

In your report, find ways to cite everything you consulted for the project: the task specification, papers that inspired you, related work you found interesting, libraries you used…

Mind the difference between direct quotations and paraphrasing. Put direct quotations in quote marks or use an indented block.

You can consult what competition teams did, but

1. Cite them as inspiration or contrast
2. Do not use any code they wrote for the competition
3. Improve upon their work (do not simply replicate)

Grading Rubric

A good project will…

- Use creative approaches to tasks/subtasks
- Show preliminary results and a detailed plan of action for the proposal report/presentation
- Show improved performance results for the final report/presentation
- Compare those results to those in the competition
- Give clear presentations and turn in well-written, thorough reports
- Demonstrate teamwork and equal distribution of work

Wrapping Up

Get cracking!

People can (and do) publish papers based on SEMEVAL tasks after the competition.

Team names, memberships, and task selections are due September 20 by the start of class.