CS 175, Project in Artificial Intelligence

Week 8: Progress Reports and Slide Presentations

Padhraic Smyth
Department of Computer Science
Bren School of Information and Computer Sciences
University of California, Irvine
Announcements

- No lectures next week: office hours during lecture hours
- Additional office hours every Friday 9:30 to 11:30
- TA (Eric) office hours, Wednesday noon to 1, Thursday 1 to 3
- Today’s Lecture (Short)
  - Project progress reports
  - Project slide presentations
# Weekly Schedule

<table>
<thead>
<tr>
<th>Week</th>
<th>Monday</th>
<th>Wednesday</th>
</tr>
</thead>
</table>
| Feb 19 | No class (university holiday)               | Lecture: Discussion of progress reports  
**Progress report due, Friday 11:45 pm** |
| Feb 26 | Office hours (no lecture)                   | Office hours (no lecture)                                                 |
| Mar 5  | Project Presentations (in class)             | Project Presentations (in class)                                          |
|       | Upload slides by 4pm                        | Upload slides by 4pm                                                      |
| Mar 12 | Lecture: Discussion of final reports        | No lecture or office hours                                                |
| Mar 19 | **Final project reports due**               |                                                                           |
|       | **Monday Mar 19, midnight**                 |                                                                           |
Progress Reports

• Your proposal should be 3 to 5 pages long
  – Use the progress report template on the Website
  – Due to EEE by 11:45pm this Friday

• One report per team: all students will get the same grade by default (unless there is some issue with a team member)

• Tips on writing proposals (see earlier lectures) are still highly relevant – see also samples of past proposals distributed via Piazza
Grading of Progress Reports

• Reports will be graded and receive a weight of 20% of your overall grade.

• Reports will primarily be graded on
  (a) clarity: are the technical ideas, experiments, etc., explained clearly?
  (b) completeness: are all sections covered?
  (c) creativity/insight: have you added any of your own ideas to the project?
  (c) progress: how much progress has there been since you started?
Important Items to Note

• Include your project number at the top

• [NEW] Copy in the Instructor comments from the PDF file of your proposal into Section 0. [This will allow us to quickly refer back to your proposal]

[Review template document]
General Tips in Writing

• Be clear: re-read and edit your document. Is it understandable to another person?

• Don’t cut and paste text from other sources without citing them

• Feel free to re-use and adapt text/figures from your proposal

• Be precise
  – How many documents? What size vocabulary? etc

• Use figures where appropriate
  – Block diagrams of your system
  – Numerical information: tables, histograms, scatter plots, line plots, etc
Project Slide Presentations

- In-class, Monday and Wednesday March 5 and 7 (2 weeks from today)

- Each team will make one presentation
  - Approximately 4 minutes per presentation
  - Time after each presentation for questions and changeover
  - So about 6 minutes x 13 projects per day = 78 minutes total

- Order of presentations will be by project number
  - 1 to 12 on Monday March 5th
  - 13 to 25 on Wednesday March 7th

- Slides to be uploaded to the EEE dropbox by 4pm day of your presentation:
  - they will be loaded on to the classroom machine, no need to use your laptop.
  - Either PDF or Powerpoint is fine

- 10% of your grade
Suggestions for Slide Content

• Slide 0: project name, student name(s)

• Slide 1: overview of project, e.g.,
  – Investigation of sentiment analysis algorithms for Twitter
  – Focus on classifying Tweets into 3 categories, ...
  – Using the following data set:...
  – Using the following classifiers:...
  – Evaluation methods: ...
  [Note: you could just show a figure here, e.g., an example of a tweet or a summary of the data set, and make your points using the figure]

• Slide 2:
  – Some more details on your technical approach
  – figures are good!, e.g., a block diagram of your pipeline
Suggestions for Slide Content (continued)

• Slide 3:
  – Examples of initial results

• Slide 4 (optional)
  – Challenges you have encountered

• Slide 5:
  – Plans for the remainder of the quarter
  – E.g., milestones you plan to complete
  – E.g., optional additional item you will investigate if there is time
Twitter Sentiment Analysis

Used Sentiment140 tweets DB

Preprocessing:
- Removed and replaced URLs, usernames, and hashtags with `<URL>`, `<USER>`, `<HASHTAG>`
- Omitted neutral sentiment tweets from test sets

We found that the SVM classifier had the most growth potential in accuracy.

The SVM classifier without the data preprocessed had an accuracy of 79%. After preprocessing the data, it became 4% more accurate. We believe this is because of the reduction of features, which made the classifier easier to classify the data.

We plan to further investigate and experiment with more feature sets.
It's interesting that training data with stop words predicts test data better training data without stop words. The reason maybe stop words connect sentences to form the meaning of the whole review.
Accuracy with other categories of tweets as negative examples for sarcasm.

<table>
<thead>
<tr>
<th>Category</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
<td>0.8594313175810974</td>
</tr>
<tr>
<td>Sad</td>
<td>0.7597116539847817</td>
</tr>
<tr>
<td>Fearful</td>
<td>0.8157366519470093</td>
</tr>
<tr>
<td>Courageous</td>
<td>0.912745545911375</td>
</tr>
<tr>
<td>Sincere</td>
<td>0.7590909090909090</td>
</tr>
<tr>
<td>Relaxed</td>
<td>0.8740629685157422</td>
</tr>
<tr>
<td>Stressed</td>
<td>0.8192771084337349</td>
</tr>
</tbody>
</table>

Table Legend:

Left: Label of the tweets that are mixed in with sarcastic tweets.

Right: Accuracy of the classifier with this particular mix of tweets.

Not bad for sarcastic!

![Classifier Accuracy Chart]

UC Irvine
Padhraic Smyth, UC Irvine: CS 175, Winter 2018
Rotten Tomatoes Movie Review Classification With Machine Learning and Natural Language Processing

<table>
<thead>
<tr>
<th></th>
<th>Overall Error</th>
<th>False Positive Error</th>
<th>False Negative Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression</td>
<td>12.38%</td>
<td>16.58%</td>
<td>9.54%</td>
</tr>
<tr>
<td>Multinomial Naive Bayes</td>
<td>13.62%</td>
<td>18.55%</td>
<td>9.59%</td>
</tr>
</tbody>
</table>

Figure: Error Rate Results - Binary Classification of Positive vs Negative Phrases

- How we conduct the binary classification:
  We removed phrases labeled 2, which is neutral. Then, we combined 0 and 1 into one class “Negative”, which we relabeled as 0, and 4 and 5 into one class “Positive”, which we relabeled as 1.

- We were surprised at how well the classifier was able to distinguish between positive and negative phrases. We are looking into creating a multi-level classifier that would try to classify whether a phrase is neutral or not, and then using this to determine whether that phrase is actually positive or negative.
Lessons learned along the way...

Preprocessing works:

The Logistic Regression classifier correctly classified 70.19% of the Reuters test set in a bag of words representation, and 73.61% on the lemmatized version. Lemmatizing provided a 3.42% increase in accuracy, which lead to 112 more documents being properly classified.

Cross-validate first:

After weeks of testing the Support Vector Machine classifier on only the standard train/test splits, we finally ran cross-validation and were surprised to find that the SVM was significantly less accurate than we thought. For example, it is producing 68.1% accuracy on the Ohsumed 5-category standard split, but only 59.74% in cross-validation on the same data. It is actually being outperformed by LR!

Experiment a little:

Using the Multinomial Naive Bayes classifier, an attempt was made to reduce the number of categories used to train the classifier for each document. The process produced less accuracy but this only confirmed that our baselines for each experiment were in fact the greatest lower bounds.
“aww! im so happy cuz you are happy 8D i have to go goodnight girl!! you are AMAZING! have a nice night! loveya?”

- Naïve bayes probability .93
- SVM distance 1.75

“I hate the way pills make my stomach feel like it's boiling. This is worse than being sick. Also”

- Naïve bayes probability .97
- SVM distance -2.10

The above are two tweets that both classifiers were very confident about. Both were mislabeled by the user. The first was incorrectly labeled by the user as negative and the second was incorrectly labeled as positive. The tweets sentiment however is obviously the opposite. We have found at least a few tweets with this problem
Twitter Trend Detection

- While many claim social media to be negative (even toxic), the top words are often positive!
  - “love”, “like”, “good”, “please”, “best”
- Negative words and obscenities are further down.
- Twitter isn’t as bad as you think ;(
### Parsing Difficulties

- We are using the Stanford parser to generate parse trees for our PCFG. For complex sentences, the parser often finds sentences within sentences. As a result, we generate sentences with strange punctuation (periods in the middle, comma at the end, etc.).

- Also, some complex sentences are parsed as NP (noun phrase) rather than S (sentence). As a result, some of our generated sentences are just single noun phrases, and some noun phrases become complex sentences.

---

**Bad Generated Tree Example:**

```
ROOT
  ↓
S
  ↓
NP
  ↓
VP
  ↓
NP
  ↓
SBAR
  ↓
S
  ↓
NP
  ↓
VP
```

- Solution: sample the generated RHS using both the symbol and its direct parent. This lets us tell the difference between a sentence/NP at the root vs. in the middle of the tree.

- Similar to using a higher-order Markov chain

- Examples of valid expansions:
  - (ROOT -> S) -> NP VP .
  - (SBAR -> S) -> NP VP
  - (ROOT -> NP) -> NP : S .
  - (S -> NP) -> DT NN
Poetry Generator - Interesting Find

- We are using a Markov generator to produce our initial sentences (sentences that have not gone through our poetry formatting yet).
- When using a Dr. Seuss data set, the following was yielded:
  
  Congratulations!
  Today is your day!
  Your mountain is waiting.
  So...get on your own.
  And when you're in a Lurch.
  You'll come down from the Lurch with an unpleasant bump.
  And then things start to happen, don't worry.
  Don't stew.

- It's pretty short, but this is to be expected due to the end character delimiter parsing with the short sentences found in Dr. Seuss...
But What We Didn’t Expect...

- The adverse effect is the potential of a few large sentences causing sentences to become freakishly long more often than we want.

- For example, this is from the text Jeeves:

  Jeeves my man, you know what I subsequently learned was Madison Square Garden, where Mr. Pepper?" "Yes.

  I proposed to her at lunch one Sunday before I knew squads of chappies down Washington Square way who started the evening papers as the rest of the zest with which she had lost a bit of lunch with me to broach the subject.

  George’s uncle was in your way, Mr. Sturgis has offered his services to his expressed intention of remaining in the drawing-room in my autobiography.

- So, need a word count MAX and MIN, otherwise too much stress on poetry generator to find syllable count and delete words
Jimmy Fallon tweet Simulation

wine for cats. Because every girl dreams of hearing her man say, 'nice!!', for already getting

Thank you guys like having @StephenAtHome and climbing a guy raised $20 min: Man never heard even realize they love you to the photobomb with @

# WhyIMSingle tweets helped # LateNight # FunShow http://t. I spent the return of The Year. # LateNight # VladdyPootPoot, b ' l ) Ha !!! London # funtimes # BIYM # FallonTonight, b

getting old, but whee did it again ! # BurgerSummit # nomayo", b ' Arnold @ Schwarzenegger showed me some of his other reviews were bad.

need a little help Obama \xe2\x9d{l ( ft. Any opportunity to have no that guy for the show about this many people wearing a werewolf

he was arrested for stopping by the day? Keep tweeting the man say it was throwing shade last night at minutes. co / Ux3WCDbJG2', b '

Thanks Jon every post - Aime with a baby carrots
# Weekly Schedule

<table>
<thead>
<tr>
<th>Week</th>
<th>Monday</th>
<th>Wednesday</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feb 19</td>
<td>No class (university holiday)</td>
<td>Lecture: Discussion of progress reports</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Progress report due, Friday 11:45 pm</td>
</tr>
<tr>
<td>Feb 26</td>
<td>Office hours (no lecture)</td>
<td>Office hours (no lecture)</td>
</tr>
<tr>
<td>Mar 5</td>
<td>Project Presentations (in class)</td>
<td>Project Presentations (in class)</td>
</tr>
<tr>
<td></td>
<td>Upload slides by 4pm</td>
<td>Upload slides by 4pm</td>
</tr>
<tr>
<td>Mar 12</td>
<td>Lecture: Discussion of final reports</td>
<td>No lecture or office hours</td>
</tr>
<tr>
<td>Mar 19</td>
<td><strong>Final project reports due</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Monday Mar 19, midnight</strong></td>
<td></td>
</tr>
</tbody>
</table>