LING 520: Computational Analysis of English

Semester: FALL '16

Instructor: Sowmya Vajjala

Iowa State University, USA

20 October 2016

Class Outline

- Regarding final projects.
- Tomorrow's tutorial
- Hui-Hsien's talk
- Text Classification: Conclusion
- Midterm feedback

Final Projects

- Decide on what you want to do. Talk to me if you need input.
- Deadline for giving a preliminary report: 5 November 2016.
- What it should contain: What is the project about, what is your motivation for choosing it, how will you implement it, where does NLP figure in the implementation.
- Length: maximum of 2 pages.

Tomorrow's tutorial - Main issues

- Probability and Programming
- When to use what DS.
- For loops and If statements
- How to call a function from one program in another.

Note: I will try to address as much as I can in those 3 hours.

Hui-Hsien's Talk

Mid-term feedback

- Fill up the plus-delta table in the sheet I give, and return to me.
- I may or may not address all issues from everyone, but I do plan to summarize this in the class on tuesday.

Next two weeks

- Parsing - How to do syntactic parse of sentences.
- Readings: Chapters 12–14 in J&M.
- Videos: Week 4 in Radev's coursera course
- Practical stuff: Chapters 7 and 8 in NLTK

If you want to develop a text classifier ...

1. Know your data. Make sure you have sufficient amounts of data (how much?)

If you want to develop a text classifier ...

1. Know your data. Make sure you have sufficient amounts of data (how much?)
2. Have an idea of what features you want to investigate. Write code for feature extraction.
3. Decide what is your learning algorithm (Naive bayes, KNN, Logistic Regression etc)
4. Check what is the format of the classifier method is expecting (in NLTK, it is a dictionary, in some other tool, it can be a comma separated string etc)
5. Write code to convert all texts into feature vectors of the required format.
6. Train - Test - Evaluate whether to improve the classifier or stop.
7. Deploy the classifier in the program which you will use to classify some new texts.

Text Classification - Practice

- There are a couple of classification problem examples described in NLTK book Chapter 6 (Gender identification, Sentence Segmentation, identifying dialogue act types, recognizing textual entailment, movie review classification etc).
- Form into groups of 2 or 3 people, pick any one classification problem demonstrated there, and follow the examples to train a classifier.
- After 15 minutes, you should be ready to talk about what you did (what is the data, what is the classification problem, what are the features, what is the accuracy etc).

Text Classification - Practice

- There are a couple of classification problem examples described in NLTK book Chapter 6 (Gender identification, Sentence Segmentation, identifying dialogue act types, recognizing textual entailment, movie review classification etc).
- Form into groups of 2 or 3 people, pick any one classification problem demonstrated there, and follow the examples to train a classifier.
- After 15 minutes, you should be ready to talk about what you did (what is the data, what is the classification problem, what are the features, what is the accuracy etc).
- Once you finish training, and testing, and have some accuracy, figure out how to use this classifier to work with a new sentence or name etc.

Install Lightside

- If you already finished doing the previous exercise, start working towards Assignment 4, by installing LightSide.