Introduction to Natural Language Processing
This week
Classification
Sentiment Lexicons
Labov’s Oral Narrative Theory
Announcements

- The midterm has been moved a week later. The midterm will be on **May 7th**.

- **Office hours changed**

- It is not a good idea to skip Tuesday lecture even if you were up late doing your homework.

- Next Week there will be no programming assignment, the HW5 will be doing problems that prepare you for the kinds of questions you will get on the midterm.

- The midterm will be multiple choice. **BRING A SCANTRON.**
Kendall goes over HW 3

- You can take the code solutions released for HW3 and use it for your HW4.
- Be sure you understand it before you start modifying it.
- The LIWC code should be easy to use, and while HW4 looks long, it just asks you to repeat many of the same things you already did for HW3.
Introduction to Natural Language Processing
This week
Classification
Sentiment Lexicons
Labov’s Oral Narrative Theory
This week

- Continue to practice processing texts with POS, ngrams etc, and use it in classification.
- Add twists on that, with LIWC, binning, feature selection
- Add a new classification task, that repeats application of the same techniques.
- Solidify what we have learned so far.
- Chapter 5. Categorizing and Tagging Words
  - READ THE SECTION ABOUT DICT
- Chapter 6. Learning to Classify Text
  - READ THE CHAPTER
NLP PIPELINE: Bringing together

WORDS (MORPHOLOGY)
Words, stemmed words

PATTERNS OF WORDS (DISTRIBUTIONAL ANALYSIS, LEXICAL SEMANTICS)
Bigrams, Word categories

PHRASES AND SENTENCES (SYNTAX)
POS, regexp

CLASSIFYING TEXTS (SEMANTICS)

SENTENCE MEANING (SEMANTICS)

DISCOURSE MEANING NARRATIVE STRUCTURES (SEMANTICS, PRAGMATICS, DISCOURSE)
Diagram of supervised classification

http://www.nltk.org/howto/classify.html
What the representation looks like

Vectors of features, the label

```python
>>> train = [
    ...  (dict(a=1,b=1,c=1), 'y'),
    ...  (dict(a=1,b=1,c=1), 'x'),
    ...  (dict(a=1,b=1,c=0), 'y'),
    ...  (dict(a=0,b=1,c=1), 'x'),
    ...  (dict(a=0,b=1,c=1), 'y'),
    ...  (dict(a=0,b=0,c=1), 'y'),
    ...  (dict(a=0,b=1,c=0), 'x'),
    ...  (dict(a=0,b=0,c=0), 'x'),
    ...  (dict(a=0,b=1,c=1), 'y'),
    ... ]

>>> test = [
    ...  (dict(a=1,b=0,c=1)), # unseen
    ...  (dict(a=1,b=0,c=0)), # unseen
    ...  (dict(a=0,b=1,c=1)), # seen 3 times, labels=y,y,x
    ...  (dict(a=0,b=1,c=0)), # seen 1 time, label=x
    ... ]
```
What is it doing?

Its learning a function!!!!
There are lots of different classifiers

- They are all **different ways to learn a function**
  - $F(\text{feature vector}) \Rightarrow \text{Label}$

- $F$ can be linear, or more complex.
  - Naïve Bayes
  - Rule Induction
  - Linear Regression
  - Tree regression
  - Classification and Regression Trees
Detecting patterns is core to NLP

- Learning a classifier model is **one way** to detect patterns (works best when combined with actually looking at the data yourself)

- How can we identify particular features of language data that are salient for classifying it?

- How can we automatically construct models of language that can be used to perform language processing?

- What can we learn about language from these models?
Why is Naïve Bayes naïve?

"Dare to be Naïve"

R Buckminster Fuller
Bayesian Classification

- Problem statement:
  - Given features $X_1, X_2, ..., X_n$
  - Predict a label $Y$

- Movie reviews
  - Features $X_1, X_2, ..., X_n$ are the words of the review, LIWC, POS, etc.
    - Their value is 1 if the feature is present or 0 if it is not
    - We can have real valued features too, but it makes the description more complicated
  - $Y$ are the class labels. Either "positive" or "negative"
NB: Imagine red is neg & green pos

- Stack Overflow explanation of NB
Bayesian Classification

- A good strategy for predicting is to model it probabilistically.

\[ \arg \max_Y P(Y|X_1, \ldots, X_n) \]

- What is the probability of "positive" (Y) given we've seen the words great and terrific (X_i)

- What is the probability of "negative" given we've seen the same words?

- Choose the label with the highest probability

- How do we compute that?
Bayesian Classification: See Reid’s lecture

• **Use Bayes Rule**
  
  • $P(A|B) = P(B|A)P(A) / P(B)$

$$P(Y | X_1, \ldots, X_n) = \frac{P(X_1, \ldots, X_n | Y)P(Y)}{P(X_1, \ldots, X_n)}$$

  - **Likelihood**
  - **Prior**
  - **Normalization Constant**

• Why does this help?
**Prob Review: In General**

- **Word sequences**
  \[ w_1^n = w_1 \ldots w_n \]

- **Chain rule of probability**
  \[
P(w_1^n) = P(w_1)P(w_2 \mid w_1)P(w_3 \mid w_2^2) \ldots P(w_n \mid w_1^{n-1}) = \prod_{k=1}^{n} P(w_k \mid w_1^{k-1})
  \]

- **Bigram approximation**
  \[
P(w_1^n) = \prod_{k=1}^{n} P(w_k \mid w_{k-1})
  \]

- **N-gram approximation**
  \[
P(w_1^n) = \prod_{k=1}^{n} P(w_k \mid w_{k-N+1}^{k-1})
  \]
• P(<s> i want english food </s>)
  
  = P(i | <s>) P(want | i) P(english | want)
  
  P(food | english) P(</s> | food)
  
  = .25 x .33 x .0011 x .5 x .68 = .000031

• P(<s> i want chinese food </s>)
  
  = P(i | <s>) P(want | i) P(chinese | want)
  
  P(food | chinese) P(</s> | food)
  
  = .25 x .33 x .0065 x .52 x .68 = .00019
What is the probability of the following sentence:

John drove to school this morning.

\[
P(\text{John, drove, to, school, this, morning, .})
\]

\[
= P(\text{John})P(\text{drove}|\text{John})P(\text{to}|\text{John, drove})P(\text{school}|\text{John drove to})\ldots
\]

- Impossible to estimate the probabilities no matter how much data you have
- Simplify by only using n-1 words of prior context
Review: N-gram Language Model

• N-Gram Models
  • Unigram model
    • $P(\text{John})P(\text{drove})P(\text{to})P(\text{school})...$
  • Bigram model
    • $P(\text{John})P(\text{drove} | \text{John})P(\text{to} | \text{drove})P(\text{school} | \text{to})...$
  • Etc.
Bayesian Classification: See Reid’s lecture

- Use **Bayes Rule**
  - $P(A|B) = P(B|A)P(A) / P(B)$

$P(Y|X_1, \ldots, X_n) = \frac{P(X_1, \ldots, X_n|Y)P(Y)}{P(X_1, \ldots, X_n)}$

- Why does this help?
Bayesian Classification

Let's expand this a little

\[ p(Y = \text{pos} | X_1, \ldots, X_n) = \frac{p(X_1, \ldots, X_n | Y = \text{pos})p(Y = \text{pos})}{p(X_1, \ldots, X_n | y = \text{pos})p(Y = \text{pos}) + p(X_1, \ldots, X_n | y = \text{neg})p(Y = \text{neg})} \]

\[ p(Y = \text{neg} | X_1, \ldots, X_n) = \frac{p(X_1, \ldots, X_n | Y = \text{neg})p(Y = \text{neg})}{p(X_1, \ldots, X_n | y = \text{pos})p(Y = \text{pos}) + p(X_1, \ldots, X_n | y = \text{neg})p(Y = \text{neg})} \]

What is the probability of \( p(Y = \text{neg}) \)

In the numerator and denominator
Bayesian Classification

- Let's expand this a little

\[
p(Y = \text{pos} \mid X_1, \ldots, X_n) = \frac{p(X_1, \ldots, X_n \mid y = \text{pos})p(Y = \text{pos})}{p(X_1, \ldots, X_n \mid y = \text{pos})p(Y = \text{pos}) + p(X_1, \ldots, X_n \mid y = \text{neg})p(Y = \text{neg})}
\]

\[
p(Y = \text{neg} \mid X_1, \ldots, X_n) = \frac{p(X_1, \ldots, X_n \mid y = \text{neg})p(Y = \text{neg})}{p(X_1, \ldots, X_n \mid y = \text{pos})p(Y = \text{pos}) + p(X_1, \ldots, X_n \mid y = \text{neg})p(Y = \text{neg})}
\]

- What is the probability of

\[
p(Y = \text{neg})
\]

- What about?

\[
p(X_1, \ldots, X_n \mid Y = \text{neg})
\]
Calculating the Likelihood

- Use the chain rule!

\[ p(X_1, \ldots, X_n | Y) = p(X_1 | Y) p(X_2, \ldots X_n | Y, X_1) \]
\[ = p(X_1 | Y) p(X_2 | Y, X_1) p(X_3, \ldots X_n | Y, X_1, X_2) \]
\[ = p(X_1 | Y) p(X_2 | Y, X_1) \ldots p(X_n | Y, X_1, X_2, \ldots, X_{n-1}) \]

- How do we calculate \( p(X_1 | Y) \)?

- What about \( p(X_n | Y, X_1, X_2, \ldots, X_{n-1}) \)?
Quick Review

If I flip a coin what is the probability of heads: $p(H)$?

If I flip a coin twice in a row what is the probability of getting two heads: $p(HH)$?

If I flip a coin twice in a row and I got heads the first time, what is the probability of getting heads on the second toss: $p(H|H)$?
Conditional Independence

- Quick Review
- If I flip a coin what is the probability of heads: \( p(H) \)?
- If I flip a coin twice in a row what is the probability of getting two heads: \( p(HH) \)?
- If I flip a coin twice in a row and I got heads the first time, what is the probability of getting heads on the second toss: \( p(H|H) \)?
- \( p(H|H) = p(H) \)
- The two events are independent
Naïve Bayes Model

- Assume all features are independent \textit{given} the class label $Y$.

\[
p(X_1, \ldots, X_n \mid Y) = p(X_1 \mid Y)p(X_2 \mid Y)\ldots p(X_n \mid Y)
\]

\[
= \prod_{i=1}^{n} p(X_i \mid Y)
\]
Massive Reduction in Parameters

- # of parameters for modeling $P(X_1, ..., X_n | Y)$:
  - $2(2^n - 1)$

- # of parameters for modeling $P(X_1 | Y), ..., P(X_n | Y)$
  - $2n$

- Counts are easy to obtain!
Training

- Training is easy!

\[
P(Y = v) = \frac{\text{Count}(Y = v)}{\# \text{ records}}
\]

\[
P(X_i = u | Y = v) = \frac{\text{Count}(X_i = u \land Y = v)}{\text{Count}(Y = v)}
\]
Training

- In practice some of these counts can be 0
- What happens if you get a 0 count?
Training

- In practice some of these counts can be 0
- What happens if you get a 0 count?
- Fix by adding "virtual" counts:

\[
P(X_i = u | Y = v) = \frac{\text{Count}(X_i = u \land Y = v) + 1}{\text{Count}(Y = v) + 2}
\]

- This is called smoothing
Independence of Features

- The Naïve Bayes independence assumption is almost never true
- Still, it often performs very well even when the assumption is demonstrably false
Other problems

- What happens to
  \[ \prod_{i=1}^{n} p(X_i \mid Y)p(Y) \]

- If we have 100,000 features?

- E.g. \((0.5)^{100000}\)

- Try it on your calculator or computer
Other problems

- What happens to

\[
\prod_{i=1}^{n} p(X_i \mid Y)p(Y)
\]

- If we have 100,000 features?
- E.g. \((0.5)^{100000}\)
- Try it on your calculator or computer
- Hint it returns 0
- This is called underflow
- Work with log probabilities instead

- Remember
  \[ \log(0.5^{100000}) = 100000 \times \log(0.5) \]
  \[ \sim -30,103 \]

- Much longer before underflowing

- So

  \[
  = \log \left( \prod_{i=1}^{n} p(X_i | Y)p(Y) \right)
  \]

  \[
  = \sum_{i=1}^{n} \log \left( p(X_i | Y) \right) + \log p(Y)
  \]
HW4 & L&W
HW4 will use the restaurant data again

File: 1 (17.html)
Review: 1
Atmosphere = 5
Do they accept reservations? = Yes
Food = 5
How many were seated at your table? = 2
Wine = "Red of some sort."
Overall = 5
Price/Value = 5
Service = 5
Text = "I think I may be the biggest cheerleader for the Tavolini anywhere, and here's why - ever since my wife and I started dating in high school, we've looked for somewhere to partake in that most-rare fine dining experience in Brantford, an otherwise terrible city to visit or live in. We first tried the Tavolini in the late 90's, and have returned several times, occasionally travelling halfway across the province to enjoy Anna's pasta, garlic bread and tartufo. We are treated like royalty every visit, even though we can never find a way to spend even a hundred dollars there - an amazing value for a full meal, with appetizers, drinks and desserts all around. The Tavolini is a truly special place, and if you have the misfortune to be passing through Brantford, you owe it to yourself to get off the highway, head downtown, lock the car up tight and enjoy some fine, fine food."
Was it a pleasant experience? = Yes
Would you return? = Yes

Review: 2
Atmosphere = 5
Do they accept reservations? = Yes
Food = 5
How many were seated at your table? = 4
Wine = "Valpolicella"
Overall = 5
Price/Value = 5
Service = 5
Text = "Best Italian food we ever had! Homemade fresh pasta and the veal dish my guest had was delicious! Very cozy, comfortable, homely atmosphere. The homemade tiramisu (the owner's grandmother's recipe) was out of this world! We will definitely be returning!"
Was it a pleasant experience? = Yes
Today was a very eventful work day. Today was the start of the G20 summit. It happens every year and it is where 20 of the leaders of the world come together to talk about how to run their governments effectively and what not. Since there are so many leaders coming together their are going to be a lot of people who have different views on how to run the government they follow so they protest. There was a protest that happened along the street where I work and at first it looked peaceful until a bunch of people started rebelling and creating a riot. Police cars were burned and things were thrown at cops. Police were in full riot gear to alleviate the violence. As things got worse tear gas and bean bag bullets were fired at the rioters while they smash windows of stores. And this all happened right in front of my store which was kind of scary but it was kind of interesting since I've never seen a riot before.
Setting up a good experiment

- Try different feature sets, set up sequence of baselines
- Keep track.
- Use at least two different learners from NLTK.
Using the LIWC tool
Lexical Categorization beyond POS

Linguistic Inquiry and Word Count

LIWC2001

James W. Pennebaker Martha E. Francis and Roger J Booth
How the LIWC dictionary is set up

```
1 funct
2 pronoun
3 pp
4 i
5 we
6 you
7 shehe
8 they
9 ipron
10 article
11 verb
12 auxverb
13 past
14 present
15 future
16 adverb
17 preps
18 conj
19 negate
20 quant
21 number
22 swear
23 social
24 family
25 friend
26 humans
27 affect
28 posemo
29 negemo
30 anx
31 anger
32 sad
33 cogmech
```

```
a 1 10
abandon* 125 127 130 131 137
abdomen* 146 147
abilit* 355
able* 355
abortion* 146 148 149
about 1 16 17
above 1 17 252 250
abrupt* 253 250
abs 146 147
absent* 354
absolute 131 136
absolutely 1 16 131 136 462
abstain* 131 137
abuse* 125 127 129
abusi* 125 127 129
academ* 354
accept 125 126 131 132
accepta* 125 126 131 132
accepted 11 13 125 126 131 132
accepting 125 126 131 132
accepts 125 126 131 132
accomplish* 354 355
```
LIWC Features

- `word_category_counter.py` must be in the same directory as your python script.

- **DefaultDic2003.dic** & **LIWC2007.dic** must be in a folder called **data** in the same directory as your script.
# Adds all our features and returns the vector

def features(review_text, review_words):
    feature_vector = {}

    # ALSO TRY JUST LIWC ON ITS OWN
    uni_dist = nltk.FreqDist(review_words)

    # add_lexical_features(uni_dist, feature_vector)
    add_liwc_features(review_text, feature_vector)

    return feature_vector
Classify Movies with LIWC

Most Informative Features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>liwc_motion</td>
<td>0.625</td>
</tr>
<tr>
<td>liwc:perceive</td>
<td>0.7299</td>
</tr>
<tr>
<td>liwc:cogmech</td>
<td>5.1282</td>
</tr>
<tr>
<td>liwc:pos =</td>
<td>1.1922</td>
</tr>
<tr>
<td>liwc:pos =</td>
<td>0.7633</td>
</tr>
<tr>
<td>liwc:pos =</td>
<td>0.2267</td>
</tr>
<tr>
<td>liwc:neg =</td>
<td>0.625</td>
</tr>
</tbody>
</table>

pos : neg =

4.4 : 1.0
4.3 : 1.0
3.7 : 1.0
3.7 : 1.0
3.0 : 1.0
3.0 : 1.0
3.0 : 1.0
2.7 : 1.0
2.5 : 1.0
2.4 : 1.0
2.4 : 1.0
2.3 : 1.0
2.3 : 1.0
def feat_bin(count):
    # Just a wild guess on the cutoff
    return count if count < 4 else 5

# Binning
# Binning really helps for this data
#fname = "unigram:{0}_{1}".format(word, feat_bin(freq))
# If selected_features == None or fname in selected_features:
#    feature_vector["unigram:{0}_{1}".format(word, feat_bin(freq))]

And now for something completely different
HW4:
Try same techniques/features on something else
Classify Clause Types in Narrative Texts

Where do we get the labels from?
Where do we get the narrative texts?
Bill Labov Theory of Narrative

William Labov,
John H. and Margaret B. Fassitt Professor

THE LANGUAGE OF LIFE AND DEATH
The Transformation of Experience in Oral Narrative

Language in the Inner City
Studies in the Black English Vernacular
Today was a very eventful work day. Today was the start of the G20 summit. It happens every year and it is where 20 of the leaders of the world come together to talk about how to run their governments effectively and what not. Since there are so many leaders coming together their are going to be a lot of people who have different views on how to run the government they follow so they protest. There was a protest that happened along the street where I work and at first it looked peaceful until a bunch of people started rebelling and creating a riot. Police cars were burned and things were thrown at cops. Police were in full riot gear to alleviate the violence. As things got worse tear gas and bean bag bullets were fired at the rioters while they smash windows of stores. And this all happened right in front of my store which was kind of scary but it was kind of interesting since I've never seen a riot before.
A Crow was sitting on a branch of a tree with a piece of cheese in her beak when a Fox observed her and set his wits to work to discover some way of getting the cheese. Coming and standing under the tree he looked up and said, "What a noble bird I see above me! Her beauty is without equal, the hue of her plumage exquisite. If only her voice is as sweet as her looks are fair, she ought without doubt to be Queen of the Birds."

The Crow was hugely flattered by this, and just to show the Fox that she could sing she gave a loud caw. Down came the cheese, of course, and the Fox, snatching it up, said, "You have a voice, madam, I see: what you want is wits."
<table>
<thead>
<tr>
<th>Narrative category</th>
<th>Narrative question</th>
<th>Narrative function</th>
<th>Linguistic form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>What was this about?</td>
<td>Signals that the story is about to begin and draws attention from the listener.</td>
<td>A short summarising statement, provided before the narrative commences.</td>
</tr>
<tr>
<td>ORIENTATION</td>
<td>Who or what are involved in the story, and when and where did it take place?</td>
<td>Helps the listener to identify the time, place, persons, activity and situation of the story.</td>
<td>Characterised by past continuous verbs; and Adjuncts (see A3) of time, manner and place.</td>
</tr>
<tr>
<td>COMPLICATING ACTION</td>
<td>Then what happened?</td>
<td>The core narrative category providing the “what happened” element of the story.</td>
<td>Temporally ordered narrative clauses with a verb in the simple past or present</td>
</tr>
<tr>
<td>RESOLUTION</td>
<td>What finally happened?</td>
<td>Recapitulates the final key event of a story.</td>
<td>Expressed as the last of the narrative clauses that began the Complicating Action.</td>
</tr>
<tr>
<td>EVALUATION</td>
<td>So what?</td>
<td>Functions to make the point of the story clear.</td>
<td>Includes: intensifiers; modal verbs; negatives; repetition; evaluative commentary; embedded speech; comparisons with unrealised events.</td>
</tr>
<tr>
<td>CODA</td>
<td>How does it all end?</td>
<td>Signals that a story has ended and brings listener back to the beginning of the story.</td>
<td>Often a generalised statement which is “timeless” in feel.</td>
</tr>
</tbody>
</table>
We only use a subset

Because our impression is that some of these are hard to label reliably

And we labelled this data ourselves
Clause Types in Oral Narrative

- **Orientation**

- **Action**

- **Evaluation**

A hungry Fox saw some fine bunches of Grapes hanging from a vine that was trained along a high trellis, and did his best to reach them by jumping as high as he could into the air. But it was all in vain, for they were just out of reach.
<table>
<thead>
<tr>
<th>#</th>
<th>Category</th>
<th>Story Clause</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Orientation</td>
<td>Now, on with this week’s story...</td>
</tr>
<tr>
<td>2</td>
<td>Orientation</td>
<td>The last month has been hectic.</td>
</tr>
<tr>
<td>3</td>
<td>Orientation</td>
<td>Turbo charged.</td>
</tr>
<tr>
<td>4</td>
<td>Orientation</td>
<td>Lot’s of work because I was learning from Tim, my partner in crime.</td>
</tr>
<tr>
<td>5</td>
<td>Orientation</td>
<td>This hasn’t been helped by the intense pressure in town due to the political transition coming to an end.</td>
</tr>
<tr>
<td>6</td>
<td>Orientation</td>
<td>This week things started alright and on schedule.</td>
</tr>
<tr>
<td>7</td>
<td>Action</td>
<td>But I managed to get myself arrested by the traffic police (rouleage) early last Wednesday.</td>
</tr>
<tr>
<td>8</td>
<td>Action</td>
<td>After yelling excessively at their outright corrupted methods and asking incessently for what law I actually broke, they managed to bring me in at the police HQ.</td>
</tr>
<tr>
<td>9</td>
<td>Action</td>
<td>I was drawing too much of a curious crowd for the authorities.</td>
</tr>
<tr>
<td>10</td>
<td>Action</td>
<td>In about half an hour at police HQ I had charmed every one around.</td>
</tr>
<tr>
<td>11</td>
<td>Action</td>
<td>I had prepared my “gift” as they wished.</td>
</tr>
<tr>
<td>12</td>
<td>Action</td>
<td>Decision withheld, they decided that I needn’t to bother, they liked me too much.</td>
</tr>
<tr>
<td>13</td>
<td>Evaluation</td>
<td>I should go free.</td>
</tr>
<tr>
<td>14</td>
<td>Evaluation</td>
<td>I even managed to meet famous Raus, the big chief.</td>
</tr>
<tr>
<td>15</td>
<td>Evaluation</td>
<td>He was too happy to let me go when he realized I was no one.</td>
</tr>
<tr>
<td>16</td>
<td>Evaluation</td>
<td>But then, a Major at his side noticed my Visa was expired.</td>
</tr>
<tr>
<td>17</td>
<td>Action</td>
<td>Damn!</td>
</tr>
<tr>
<td>18</td>
<td>Orientation</td>
<td>My current Visa is being renewed in my other passport at Immigration’s.</td>
</tr>
<tr>
<td>19</td>
<td>Evaluation</td>
<td>Fuck.</td>
</tr>
<tr>
<td>20</td>
<td>Evaluation</td>
<td>In custody, for real.</td>
</tr>
<tr>
<td>21</td>
<td>Orientation</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Evaluation</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Evaluation</td>
<td></td>
</tr>
</tbody>
</table>
POS categories for Text Classification

- Standard supervised machine learning algorithm: Naive Bayes

Structural features instead of lexical tokens
- Align with the high level analysis of L&W
- Do not require a large amount of training data

<table>
<thead>
<tr>
<th>No.</th>
<th>Feature</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Stative verbs</td>
<td>There was once a house,</td>
</tr>
<tr>
<td>2</td>
<td>Non-stative verbs</td>
<td>and then climbed up the wall,</td>
</tr>
<tr>
<td>3</td>
<td>Future</td>
<td>you will certainly be found out,</td>
</tr>
<tr>
<td>4</td>
<td>Conditional</td>
<td>that she should pay him a high fee if he cured her,</td>
</tr>
<tr>
<td>5</td>
<td>Quotes</td>
<td>“That’s awkward,”</td>
</tr>
<tr>
<td>6</td>
<td>Questions</td>
<td>who is going to bell the cat?</td>
</tr>
<tr>
<td>7</td>
<td>Indefinite articles</td>
<td>A hungry Fox saw some fine bunches of Grapes</td>
</tr>
<tr>
<td>8</td>
<td>Time entity</td>
<td>a Goose which laid a Golden Egg every day.</td>
</tr>
</tbody>
</table>
Useful features for L&W on Blogs

- Binary valued 3,510 unique features
- Independent vs dependent clauses, POS, LIWC, negation, words
- Most informative features

<table>
<thead>
<tr>
<th>#</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>POS:INDEP-VBD</td>
</tr>
<tr>
<td>2</td>
<td>STEM:INDEP-be</td>
</tr>
<tr>
<td>3</td>
<td>INDEP-LIWC-Motion</td>
</tr>
<tr>
<td>4</td>
<td>POS:INDEP-VBZ</td>
</tr>
<tr>
<td>5</td>
<td>POS:INDEP-RP</td>
</tr>
<tr>
<td>6</td>
<td>INDEP-Negate</td>
</tr>
<tr>
<td>7</td>
<td>POS:INDEP-VBP</td>
</tr>
<tr>
<td>8</td>
<td>INDEP-Copula</td>
</tr>
<tr>
<td>9</td>
<td>STEM:INDEP-up</td>
</tr>
<tr>
<td>10</td>
<td>STEM:INDEP-then</td>
</tr>
</tbody>
</table>

Part of Speech Tag
LIWC lexical semantic categories
Negation
Lexical unigrams
First Study on Aesop’s Fables

- Total clauses: 315 (20 Orientation, 167 Action, 128 Evaluation)
- Accuracy of majority class baseline: 0.53

<table>
<thead>
<tr>
<th>Measure</th>
<th>Orientation</th>
<th>Action</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.96</td>
<td>0.90</td>
<td>0.92</td>
</tr>
<tr>
<td>Precision</td>
<td>0.90</td>
<td>0.88</td>
<td>0.90</td>
</tr>
<tr>
<td>Recall</td>
<td>0.45</td>
<td>0.93</td>
<td>0.89</td>
</tr>
</tbody>
</table>
L&W Experiments on Blogs

- Feature selection

- Classifiers:

<table>
<thead>
<tr>
<th>Classifier</th>
<th># Features</th>
<th>F-Measure</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive Bayes</td>
<td>695</td>
<td>0.680</td>
<td>0.684</td>
</tr>
<tr>
<td>SVM</td>
<td>3200</td>
<td>0.679</td>
<td>0.682</td>
</tr>
<tr>
<td>Linear Regression</td>
<td>3350</td>
<td>0.698</td>
<td>0.703</td>
</tr>
</tbody>
</table>

- Harder problem than Aesop’s. Data much more variable.