

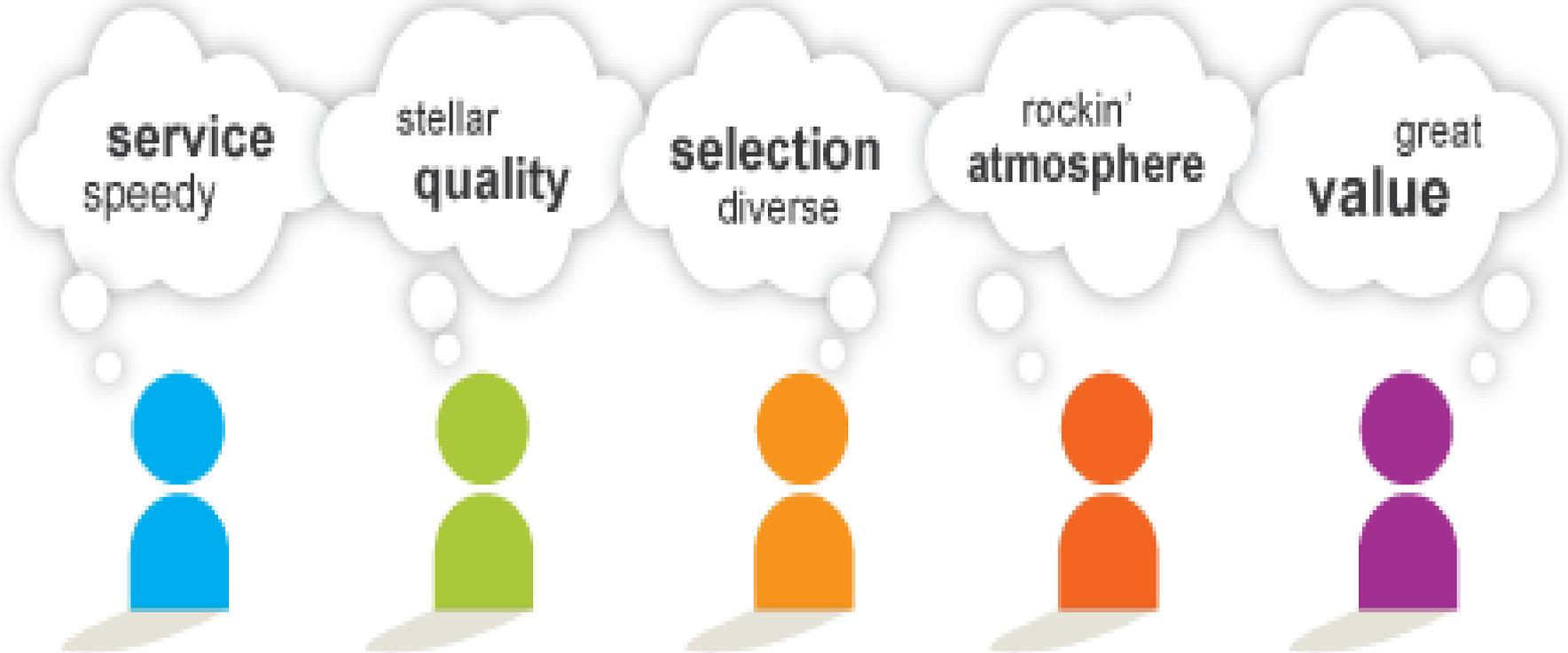
# SI485i : NLP

Set 6

Sentiment and Opinions

# It's about finding out what people think...

---



# Can be big business...

---

- Someone who wants to buy a camera
  - Looks for reviews online
- Someone who just bought a camera
  - Writes reviews online
- Camera Manufacturer
  - Gets feedback from customers
  - Improves their products
  - Adjusts Marketing strategies



# Online social media sentiment apps

- Try a search of your own on one of these:
  - Twitter sentiment <http://twittersentiment.appspot.com/>
  - Twends: <http://twendz.waggeneredstrom.com/>
  - Twittratr: <http://twittratr.com/>
  - SocialMention: <http://socialmention.com/>
- Easy to search for opinions about famous people, brands and so on
- Hard to search for more abstract concepts, perform a non-keyword based string search

# Why are these sites unsuccessful?

- They only work at a very basic level
- They only use dictionary lookups for positive/negative words.
- Tweets are classified without regard to the search terms

# Whitney Houston wasn't very popular...

## Twitter Sentiment

 Tweet < 273

 Like < 319

 +1 < 20

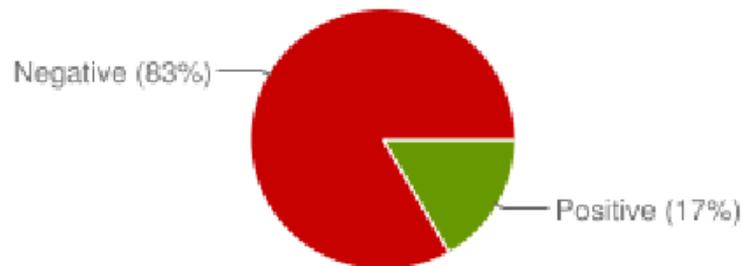
"Whitney Houston"

Search

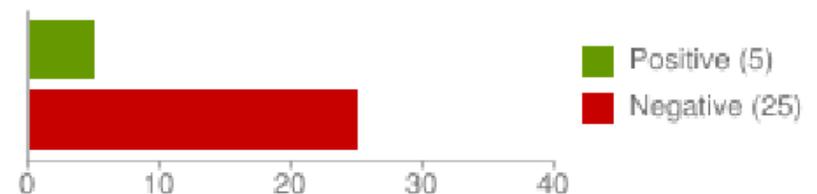
[Save this search](#)

### Sentiment analysis for "Whitney Houston"

Sentiment by Percent



Sentiment by Count



# Or was she?

## Tweets about: "Whitney Houston"

bazyboy25: **Whitney houston**...too soon? #CelebritiesThatLookLikeTheyStank  
Posted 5 minutes ago

TeghanSimone: Radio playing **Whitney Houston**.. I swear I'm about to cry... So sad  
Posted 5 minutes ago

JB3LL: hoes about to get **whitney houston**'d tonight! #TheWalkingDead  
Posted 5 minutes ago

derickaadamss: "@indreamville\_ : Twitter I'm curious who do you think had more problems Michael Jackson or **Whitney Houston**???"  
<<<< **Whitney Houston**!  
Posted 5 minutes ago

charlottesteer4: Listening to **Whitney Houston** loveeeee songsss <3 she's amazing <3  
Posted 5 minutes ago

DionneHeraty40: @Sbarr25 The reason why **Whitney Houston** died at only 41 <http://t.co/JJKRDjbj>  
Posted 5 minutes ago

ShortySoooFine: #musicwasbestwhen legends like James brown, Michael Jackson, **Whitney Houston** still lived.  
Posted 5 minutes ago

CarlmannJohnson: Pray for Bobby Brown!!! He lost his ex-wife **Whitney Houston** and his dad Herbert Brown... Prayers up for you!!  
Posted 5 minutes ago

LonelySpaceman: Is it bad that I thought **Whitney Houston** was already dead?  
Posted 5 minutes ago

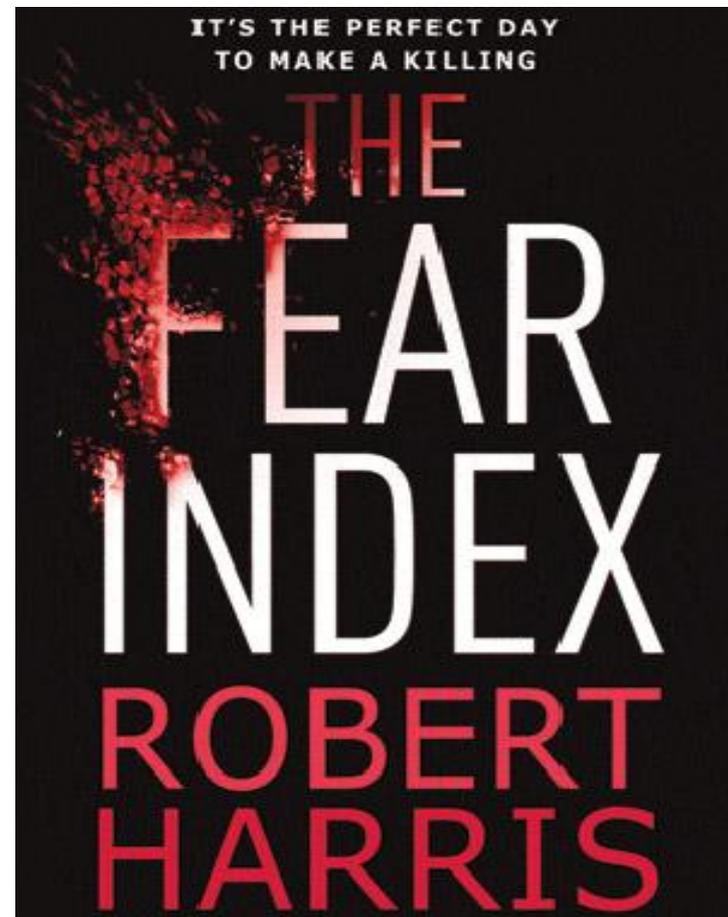
eatmy\_CHOCLATE: My aunt in there playing **Whitney Houston** making me sad  
Posted 5 minutes ago

The results for this query are:

# Opinion Mining for Stock Market Prediction

---

- It might be only fiction, but using opinion mining for stock market prediction has been already a reality for some years
- Research shows that opinion mining outperforms event-based classification for stock trend prediction [Bollen2011]
- At least one investment company currently offers a product based on opinion mining



# Twitter for Stock Market Prediction

---



“Hey Jon, Derek in Atlanta is having a bacon and egg, er, sandwich. Is that good for wheat futures?”

# Derwent Capital Markets

---

- [Derwent Capital Markets](#) have launched a £25m fund that makes its investments by evaluating whether people are generally happy, sad, anxious or tired, because they believe it will predict whether the market will move up or down.
- Bollen told the Sunday Times: "We recorded the sentiment of the online community, but we couldn't prove if it was correct. So we looked at the Dow Jones to see if there was a correlation. We believed that if the markets fell, then the mood of people on Twitter would fall."
- "But we realised it was the other way round — that a drop in the mood or sentiment of the online community would precede a fall in the market."

# Derwent Capital Markets

CAYMAN



PROVIDING  
SUPERIOR  
INVESTMENT  
ADVICE //

" Using global sentiment analysis to trade the financial markets "

## Social Media Sentiment Trading - Private Managed Accounts

Due to massive demand we have decided to apply our social media sentiment analysis technology to managed trading accounts allowing private investors the opportunity to invest upwards of £10,000 GBP.

If you are a sophisticated investor or high net worth person and not a US resident and would like to open an account then please click the link below...

# Sometimes science is hype

---

- The Bollen paper has since been strongly questioned by others in the field.
- It contained some overuse of statistical significance tests that could have overestimated how well sentiment actually aligned with market movements.
- Nobody has been able to recreate their findings.

# Accuracy of twitter sentiment apps

- Mine the social media sentiment apps and you'll find a huge difference of opinions about Pippa Middleton:
  - [TweetFeel](#): 25% positive, 75% negative
  - [Twendz](#): no results
  - [TipTop](#): 42% positive, 11% negative
  - [Twitter Sentiment](#): 62% positive, 38% negative
- Try searching for “Gaddafi” and you may be surprised at some of the results.

# Opinion spamming

---



Suppose we run a contest where people retweet our ad repeatedly,  
and the winner's whoever loses the most followers.

# Predicting other people's decisions

- It would be useful to predict what products people will buy, what films they want to see, or what political party they'll support

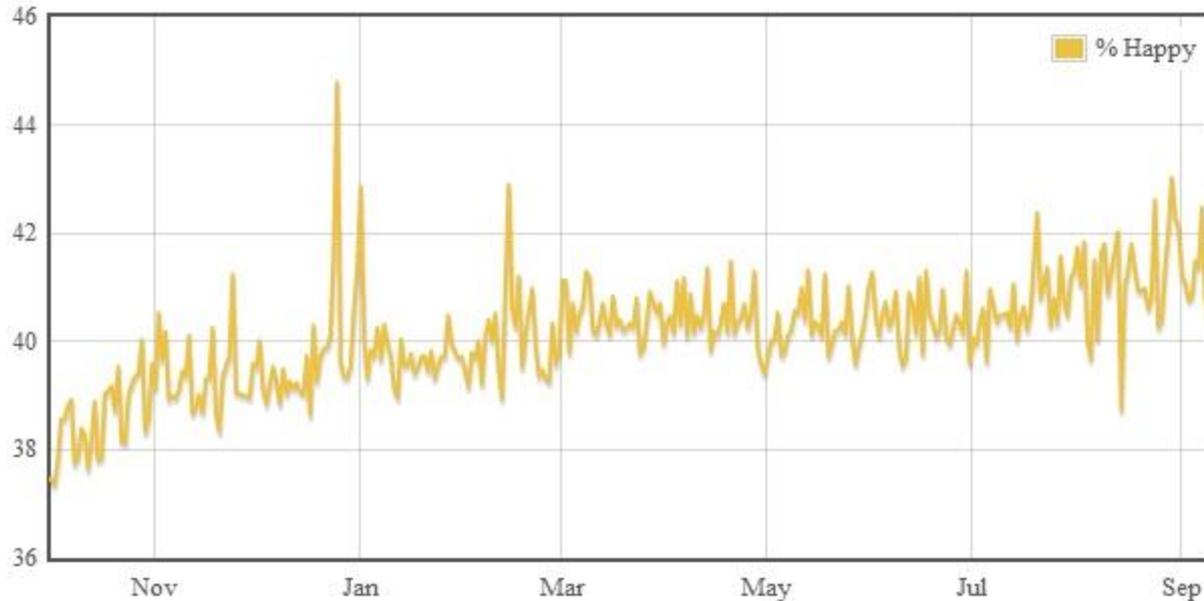


# Track Population Moods



## Mood of the Nation

Computing how everyone feels, one tweet at a time.



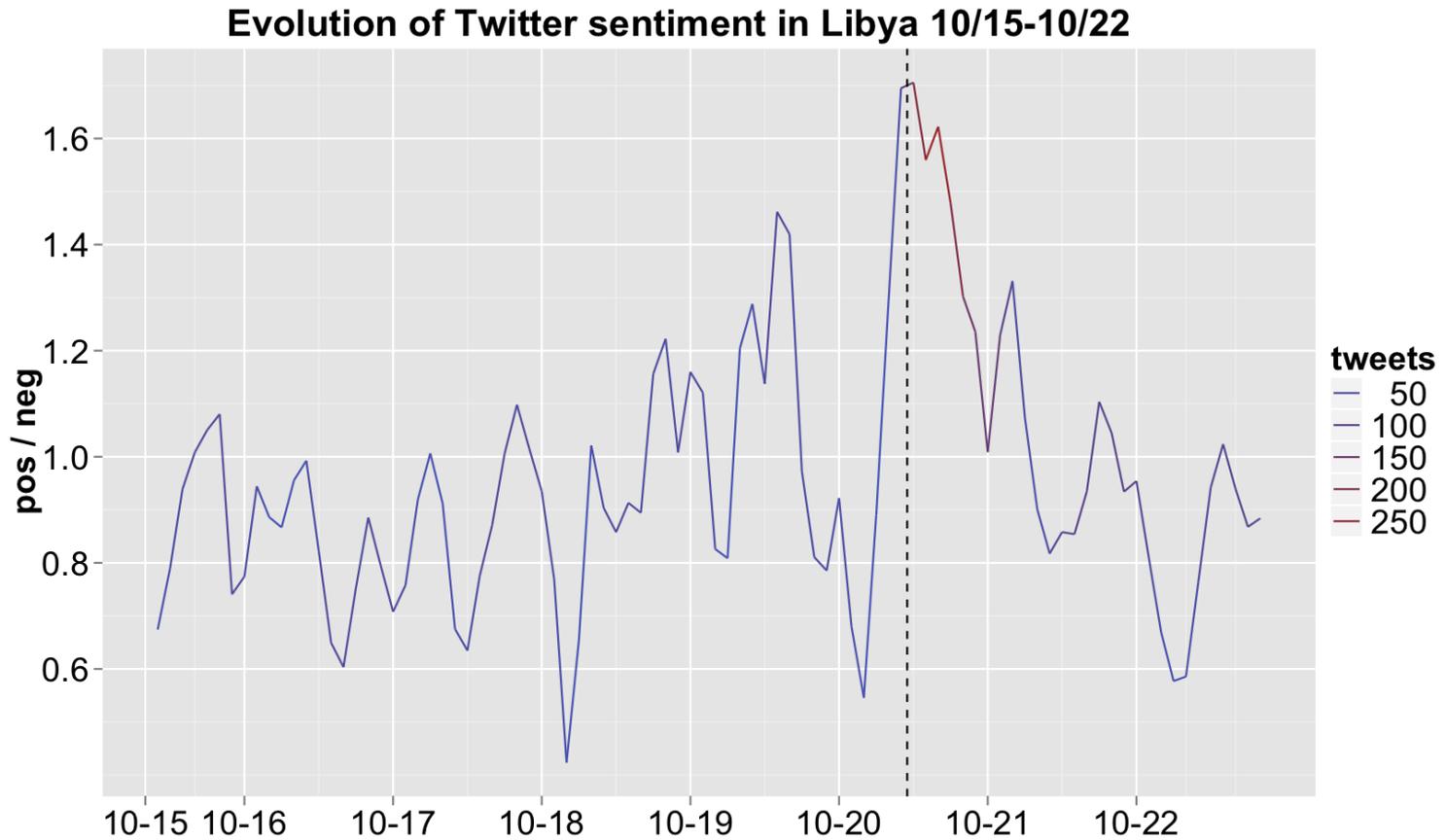
(Happy spikes on Christmas Day and New Year's Eve)

Plot Happy   
Plot Sad

Current Mood: 40.9% positive

<http://www.usna.edu/Users/cs/nchamber/mood-of-nation/>

# Monitor Real-World Events



# Methods for Opinion Mining

---

- So how does sentiment analysis work?
  1. Sentiment Lexicons
  2. Machine Learning

# Types of Sentiment

---

- Typically three classes:
  1. Positive
  2. Negative
  3. Neutral
- Sometimes split into three classes a little more formally:
  1. Objective statements
  2. Subjective statements
    - Positive
    - Negative

# Fine-Grained Sentiment

---

- But sentiment can definitely be more fine-grained!
- LIWC2007 (linguistic inquiry and word count)
  1. Future orientation
  2. Past orientation
  3. Positive emotion
  4. Negative emotion
  5. Sadness
  6. Anxiety
  7. Anger
  8. Tentativeness
  9. Certainty
  10. Work
  11. Achievement
  12. Money

# Sentiment Lexicons

---

- **Lexicon**: a list of words with sentiment scores/weights
- **OpinionFinder**
  - 2006 positive words, 4783 negative words
  - <http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>
- **SentiWordnet**
  - Attaches scores to WordNet concepts
- **SentiStrength**
  - A program that scores words for you
  - <http://sentistrength.wlv.ac.uk/>

# OpinionFinder

---

## POSITIVE WORDS

- appeal
- appealing
- applaud
- appreciable
- appreciate
- appreciated
- appreciates
- appreciative
- appreciatively
- appropriate
- approval
- approve
- ardent

## NEGATIVE WORDS

- attack
- attacks
- audacious
- audaciously
- audaciousness
- audacity
- audiciously
- austere
- authoritarian
- autocrat
- autocratic
- avalanche
- avarice

# Sentiment Lexicons

---

- What do we do with a lexicon?
  - Count positive and negative words in your text
- What if your text has both positive and negative words?
  - Use word weights to differentiate
  - Label as both positive and negative
  - Is it subjective or objective?

# Lexicons: the bad

---

- Lexicons tend to contain general sentiment
- Not targeted to your domain
  - Is “austere” always a negative mood?
  - “bad” is usually negative word, unless it is about the movie, “The Good, The Bad, and The Ugly”
- What to do?
- **Learn your own lexicon!**

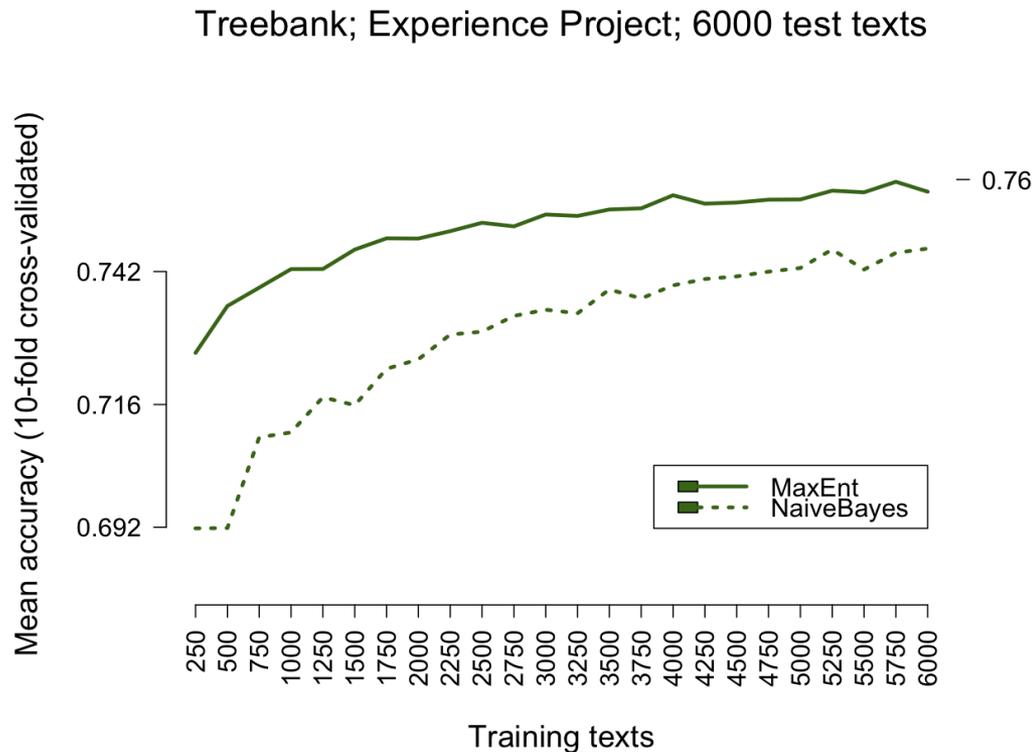
# Learn a Lexicon

---

1. Find some data that is labeled
  - Movie reviews have star ratings
  - Manually label data yourself (doesn't always take as long as you think)
  - Use a noisy label, such as “#angry” on tweets
2. Learn a model from the labeled data
  - Naïve Bayes Classifier
  - MaxEnt Model (you have not yet learned)
  - Decision Trees
  - etc.

# Learning Algorithms do Matter

- Machine Learning and AI
- This class will not teach all algorithms



# What features do we use?

---

- Sentiment analysis is a type of text classification task.
- Use many of the same features you'd normally use.
- However, emotion is often conveyed in other types of words, such as adjectives, that might not help typical classification tasks.
- Negation is a big deal.
  - “I am not happy that the phone did not work.”
- Discourse now matters:
  - “Are you happy?”
  - “You are happy!”

# Targeted Sentiment Analysis

SEARCHED TERM	POSITIVE TWEETS	NEUTRAL TWEETS	NEGATIVE TWEETS	TOTAL TWEETS
<b>netflix</b>	<b>116</b>	<b>1291</b>	<b>93</b>	<b>1500</b>

7.73% POSITIVE	
	sadly <b>excited</b> that marley & me is coming via netflix tomorrow...life in the fast lane <a href="#">(view)</a>
	@netflix when can i have blu ray, and stream only? not very <b>happy</b> at the cost increase. <a href="#">(view)</a>
	netflix just <b>recommended</b> the category of "sentimental movies featuring a <b>strong female lead</b> " to me. haha kind of pegged me. <a href="#">(view)</a>

86.07% NEUTRAL	
	Headache... Will rest for a while. Hopefully will be great. If not I can catch the new episode with netflix tomorrow. <a href="#">(view)</a>
	Netflix to hike up Blu-ray by \$8 <a href="http://tinyurl.com/c3yo4">http://tinyurl.com/c3yo4</a> <a href="#">(view)</a>
	Dinner good. We're going to begin initiating childhood bedtime protocols now and then watch The Wire on netflix. <a href="#">(view)</a>

6.20% NEGATIVE	
	when does the newest season of dexter get released on netflix i am <b>tired</b> of waiting <a href="#">(view)</a>
	way to go netflix.. keep getting greedy. <b>no more</b> blu-ray for me... at least from you. <a href="#">(view)</a>
	netflix is getting <b>lazy</b> with my movies. they keep sending me random subtitled movies that i dont order. and they <b>suck</b> <a href="#">(view)</a>
	"do that again, and i'll take curious

# Targeted Sentiment Analysis

---

1. Find text about a specific topic
2. Learn a lexicon of sentiment words using only that text
3. Label new text with sentiment
4. Profit!

# Targeted Sentiment Analysis

---

- Problems
  - Keyword search for a topic is crude and often wrong
  - Even if keyword works, which text is positive or negative?
- Solutions
  - Hand label text for your topic. Naïve Bayes classifier.
  - Hand label text for sentiment. Naïve Bayes classifier.

# Targeted Sentiment Analysis

---

- Harder problem:
  - Are the sentiment words targeted at your topic?

***“I am so mad at my mom,  
she won’t let me see  
Bieber in concert!!!!  
Aaaaaaaaaaaaaaaaaahhh  
hhhhh!”***



# Targeted Sentiment Analysis

---

- Solutions to targeted problem:
  - Need deeper language understanding
  - Need syntax of words “mad at mom” not “mad at bieber”
  - Need robust word knowledge: “aaaaaaaaahhhhhh” means frustration.
- We will soon cover syntactic parsing.
- We will most likely cover robust word learning too!

# USNA's own research

---

- Learning for microblogs with distant supervision:  
Political Forecasting with Twitter
    - Marchetti-Bowick and Chambers. EACL 2012.
1. Do a keyword search on McCain and Obama
  2. Build a political classifier.
  3. Do a keyword search for smiley faces :) and :(
  4. Build a sentiment classifier.
  5. Run two classifiers, add up the result.

# Be careful...

---

- Topic classifiers might only reflect the **general mood** and mislead you.
- **Big finding:** political forecasting works well on Twitter as a whole, not just on tweets about politics.
- “Do people like your product? Or are they just in a good mood today?”

# The Future

---

- Unknown. This is a new field (< 10 years).
- We still see wild claims about effectiveness.
- **Challenge:** making sentiment more precise, both in definition, and in classification
- **Challenge:** identify the sentiment you care about, directed at your topic of interest
- **Possible class project ideas?**