Natural Language Processing

Lecture 9: Sentiment Analysis and Social Media Text

Announcements

- Reminder from the TAs: please read the answers to existing questions on Piazza before posting new questions

- We now have a third TA: Yun-Nung ("Vivian") Chen

Why are opinions important (or useful)?

Large volumes of opinions can help answer:

- Which (toaster, car, education, ...) should I buy?
- Is a specific movie worth seeing?
- What stocks are likely to perform well?
- How does the world feel about a country or a leader?

Can we detect and quantify opinions?

People express their opinions online via

- social media (Twitter, Facebook, other places)
- e-commerce sites
- review sites
- discussion forums
- blogs

Those opinions are often expressed in natural language, and we have lots of them to work with.

Of course, the answer is **yes**--but with caveats.

Names of related tasks

There are many related tasks with somewhat different objectives:

- Sentiment analysis
- Opinion mining
- Sentiment mining
- Emotion detection

et al.

For the purposes of this lecture, an *opinion* expresses a *sentiment* about an entity.

What constitutes an opinion?

Some necessary components of an opinion:

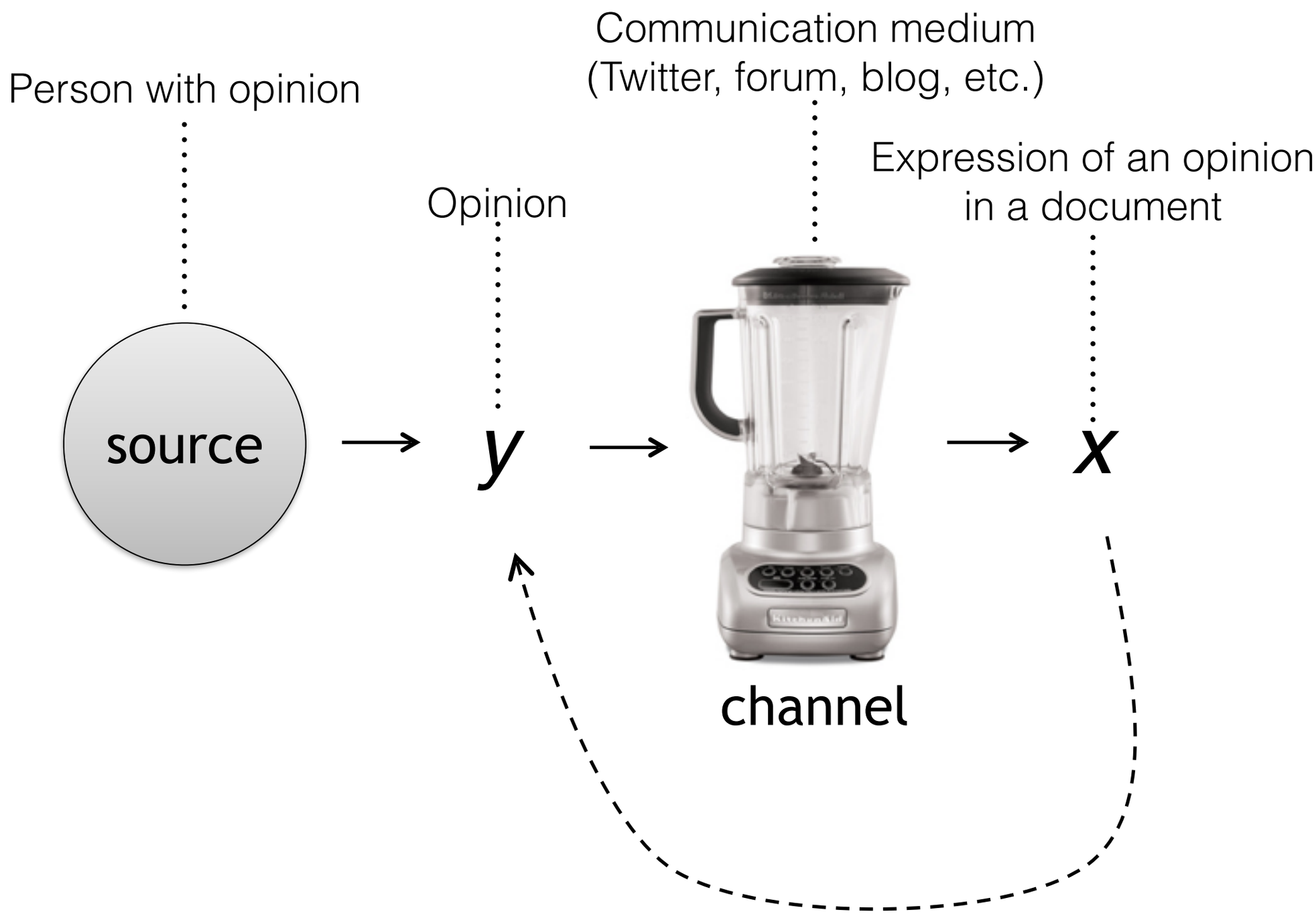
- A target entity
- An aspect or feature of the entity
- A sentiment value for the opinion ◄····· Our main focus
- The opinion holder
- The time when the opinion is expressed

What else could we add?

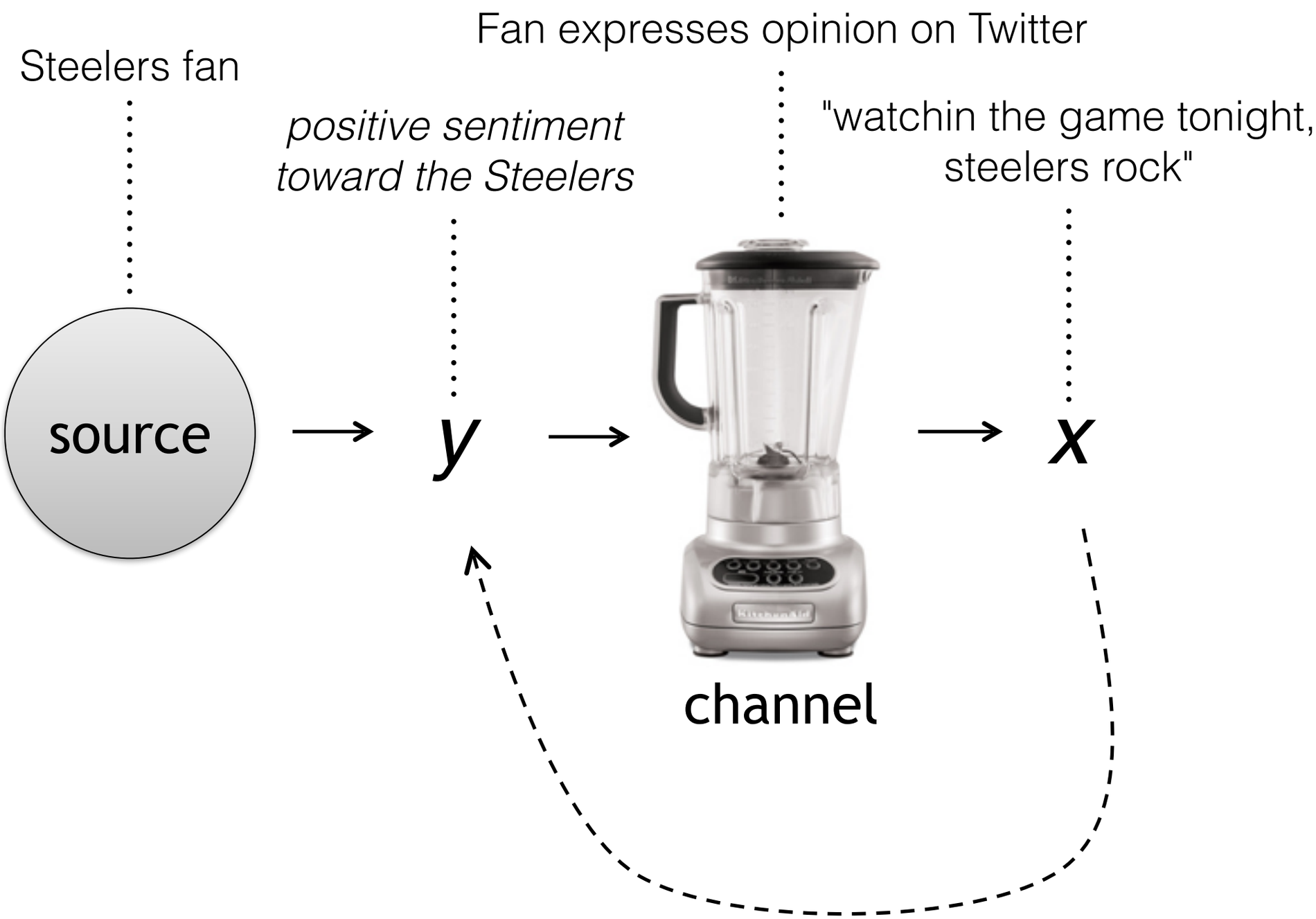
What data sources are available?

- Twitter
  - Good: tweets are concise
  - Bad: Twitter-speak is a challenge
- Product/Movie/Music/etc. reviews
  - Good: entities are provided; little noise
  - Bad: not everything gets reviewed
- Discussion forums and blogs
  - Good: lots of interaction between parties
  - Bad: multiple entities, multiple topics, figurative language, etc.

The noisy channel revisited



The noisy channel revisited



Emotional valence

We can intuit (or roughly measure) the **emotional valence** of words, i.e., how positive or negative the words are.

admired	3			wrathful	-3
admires	3			wreck	-2
admiring	3			wrong	-2
admit	-1			wronged	-2
admits	-1			yeah	1
admitted	-1			yearning	1
admonish	-2	…		yeees	2
admonished	-2			yes	1
adopt	1			youthful	2
adopts	1			yucky	-2
adorable	3			yummy	3
adore	3			zealot	-2
adored	3			zealots	-2
adores	3			zealous	2

Source: AFINN affective lexicon by Finn Årup Nielsen
http://neuro.imm.dtu.dk/wiki/AFINN

A simple approach

"I love @dhewlett he's an awesome actor!!"
Σ{ +3 +3 } = +6

"@megabus I'm filing a dispute with my bank. You
Σ{ -2
people disgust me!"
-3 } = -5

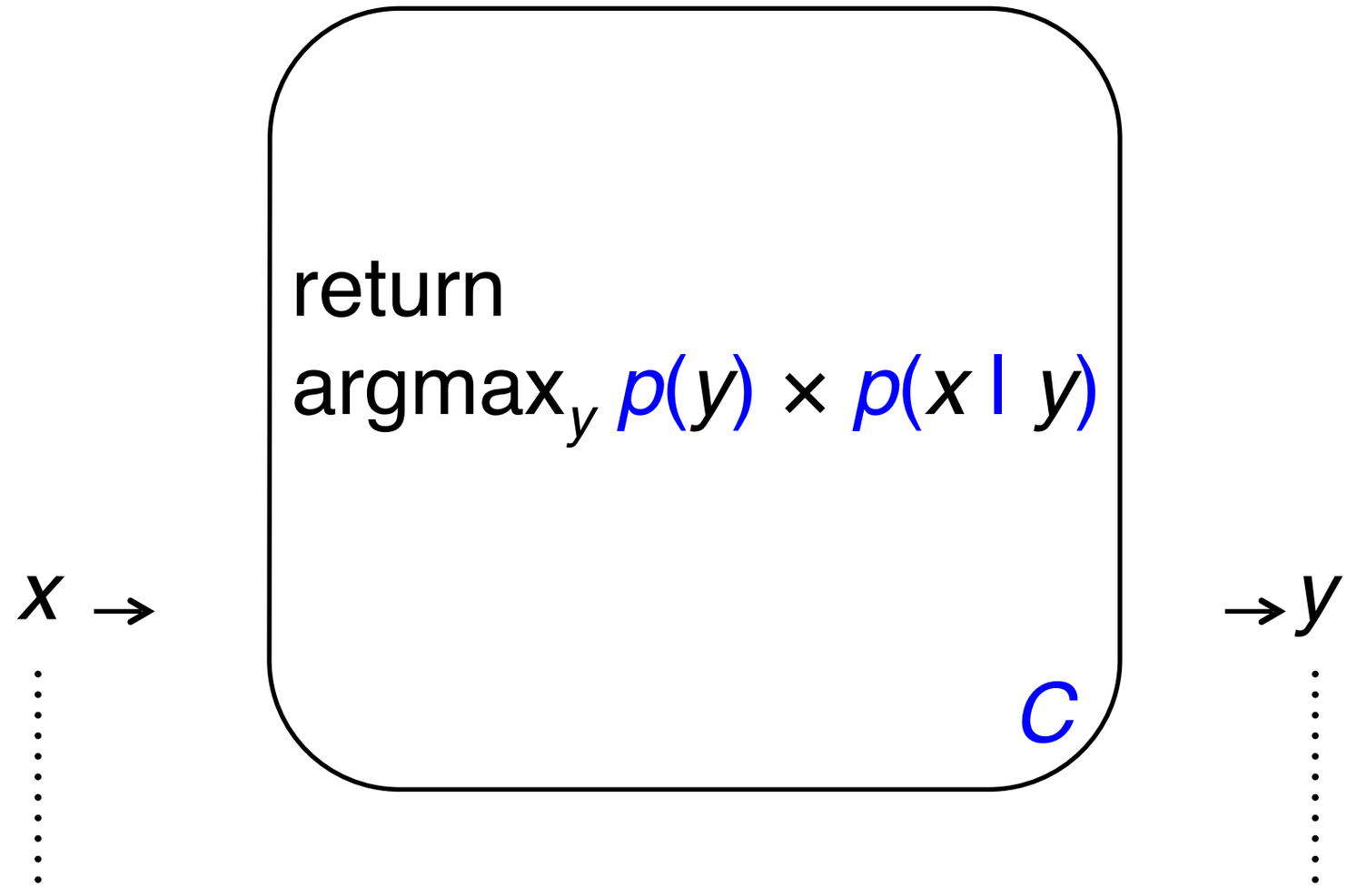
"Bus was 5 minutes late, not too bad. Good thing it's
Σ{ -2 -3 +3
getting to the Megabus early still."
+1 } = -1

Why is a word valence list insufficient?

Many reasons; among them:

- Sparsity in the word valence list
- Common negations, e.g., "not bad"
- More complex issues

Sentiment analysis as a classification problem



Document (tweet, review, forum post, etc.)

Sentiment (some measure of sentiment valence)

What document features should a classifier use?

Common features include:

- Unigrams ("bag of words") and their frequencies
- Bigrams, trigrams, etc.

  not bad is actually good
- Part of speech tags
- Dependency features
- Sentiment words and phrases

  good, *wonderful*, *rubbish*, *junk*, etc.
- Rules of opinions

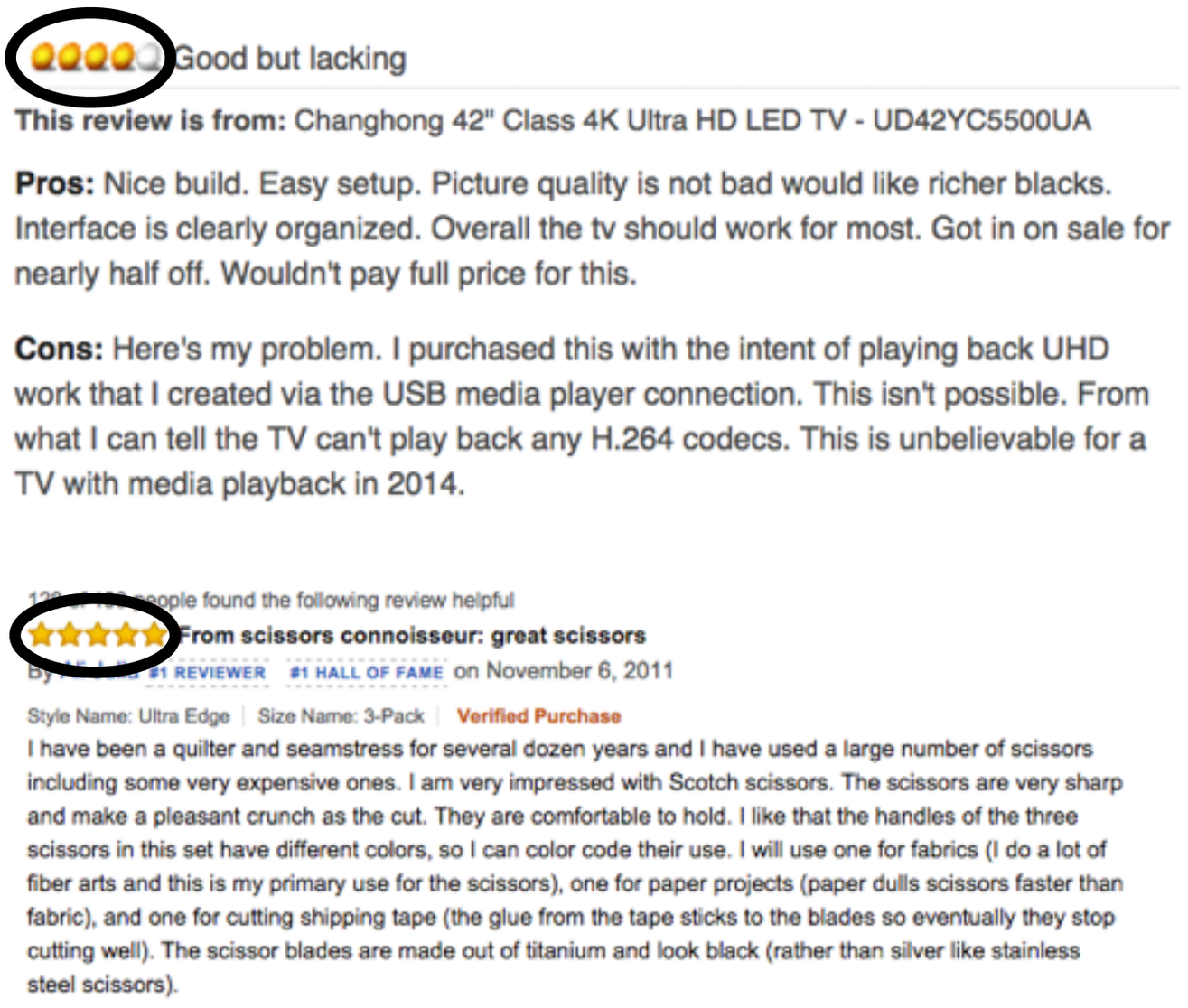
  [negation] [positive valence word]

  e.g., *barely sufficient*

How can we get labeled documents?

Options include:

- Use already labeled data from a similar domain
- Use known sentiment indicators (like smileys)
- Ask humans to read and label documents (expensive)



Good but lacking

This review is from: Changhong 42" Class 4K Ultra HD LED TV - UD42YC5500UA

Pros: Nice build. Easy setup. Picture quality is not bad would like richer blacks. Interface is clearly organized. Overall the tv should work for most. Got in on sale for nearly half off. Wouldn't pay full price for this.

Cons: Here's my problem. I purchased this with the intent of playing back UHD work that I created via the USB media player connection. This isn't possible. From what I can tell the TV can't play back any H.264 codecs. This is unbelievable for a TV with media playback in 2014.

people found the following review helpful

From scissors connoisseur: great scissors

#1 REVIEWER #1 HALL OF FAME on November 6, 2011

Style Name: Ultra Edge | Size Name: 3-Pack | Verified Purchase

I have been a quilter and seamstress for several dozen years and I have used a large number of scissors including some very expensive ones. I am very impressed with Scotch scissors. The scissors are very sharp and make a pleasant crunch as the cut. They are comfortable to hold. I like that the handles of the three scissors in this set have different colors, so I can color code their use. I will use one for fabrics (I do a lot of fiber arts and this is my primary use for the scissors), one for paper projects (paper dulls scissors faster than fabric), and one for cutting shipping tape (the glue from the tape sticks to the blades so eventually they stop cutting well). The scissor blades are made out of titanium and look black (rather than silver like stainless steel scissors).

Newegg.com, Amazon.com

Important Caveats

Why is sentiment analysis difficult?

- Opinions change over time
- Some words' sentiment associations are misleading
- Multiple languages coexist on social media
- Spam and fake opinions exist
- Domain differences affect results

Some words' sentiment associations are misleading

Mind the topic of discussion:

- "Tylenol relieves my pain quickly and without the upset stomach that I get with other pain relievers."

- "This vacuum sucks like you wouldn't believe...It makes me angry to think that I've been wasting my time with all the other brands when the BISSELL CleanView with OnePass was here all along for me to find. Cheers. This vacuum sucks big time, and that's why you should buy it."

Excerpts from product reviews on Amazon.com

Domain differences affect results

Training a model on text like this

Tbh idk why I got the iPhone 6.. I hardly use my phone except for social media and to talk to my parents.
@WoOKiEE_FTW I like my iPhone but I swear it has the intelligence of a squirrel

will probably produce poor results on text like this

The iPhone 6 has far above average sound quality from the headphone out. I have IEM's that cost more than the (full) price of the iPhone and they sound 98% as good out of the headphone jack as they do through expensive amps. The biggest question is going to be what's your music source quality, and what specific headphones are you using. The phone itself is unlikely to let you down.
My last iPhone 6 had a scratch on the screen very quickly, and I similarly babied the phone. No idea how it happened. That was my fourth replacement iPhone 6. It also started shutting down randomly and crashing this week. Apple store just replaced it, and I got my fifth iPhone 6 home only to notice a black speck of dust under the glass.

http://forums.macrumors.com/

Spam and fake opinions exist

Naively processing large volumes of tweets like this

Ray-Ban Sunglasses In History An Amazing Story *[URL]*
The Most Popular Authentic Ray-Bans On Sale Now ! *[URL]*
Ray-Ban Sunglasses Throughout History Fascinating Material *[URL]*
Authentic Ray-Ban Sunglasses On Facebook Visit Us ! *[URL]*

will lead to skewed results on sentiment toward Ray-Bans.

How is a sentiment classifier evaluated?

With a labeling scheme, we can measure precision and recall.

Let's assume that there are two labels (*positive* and *negative*) in our scheme. To measure precision and recall for *positive*:

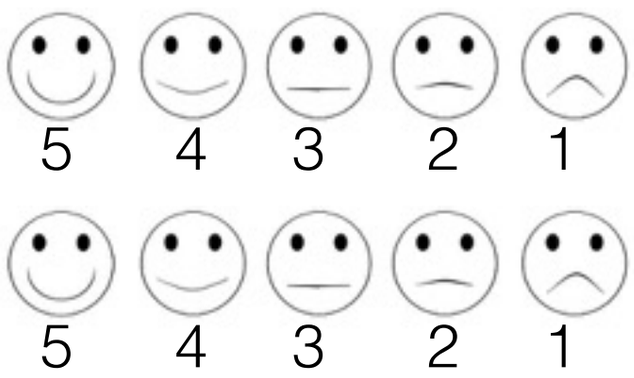
C = documents expressing positive sentiment
H = documents hypothesized to express positive sentiment

$$\text{recall} = \frac{|C \cap H|}{|C|} \qquad \text{precision} = \frac{|C \cap H|}{|H|}$$

Limits to evaluation

Still, the sentiment of a document can be hard to quantify because:

- Articulate opinions are hard to condense
- Humans often disagree on labels



Tbh idk why I got the iPhone 6.. I hardly use my phone except for social media and to talk to my parents.
@WoOKiEE_FTW I like my iPhone but I swear it has the intelligence of a squirrel

Human agreement places a ceiling on fair evaluation.

Smileys: https://www.flickr.com/photos/12701326@N02/3757564190/

Comparative opinions

An opinion can express a comparative relationship between two entities: $a<b$, $a>b$, or $a=b$.

Not all comparisons are opinions.

Opinion	The voice quality of the Galaxy S5 is better than the iPhone 6
Opinion	The Galaxy S5 and the iPhone 6 both have great screens
Not an opinion	The Galaxy Note 4 is bigger than the iPhone 6 Plus

Our goal: extract the two entities being compared and the expressed relationship between them.

How can we mine comparative opinions?

> The voice quality of the Galaxy S5 is **better** than the iPhone 6

1. Identify a comparative statement
   - Look for comparative words (*more, less,* [ADJ]-*er*, most, least, [ADJ/ADV]-*est*, favor, beat, etc.)

2. Determine whether the comparison is an opinion
   - *Opinion detection*: classification with a unigram model has worked well for this problem

3. Identify the preferred entity
   - Examine the order of key words in the sentence and determine the relationship

Let's try a sentiment analysis demo

sentiment viz -Tweet Sentiment Visualization

http://www.csc.ncsu.edu/faculty/healey/tweet_viz/tweet_app/

Some sample queries:

- "Pharrell Williams"
- "Bill Belichick"
- exercise
- sleep

Resources

- Bing Liu's book on sentiment analysis and opinion mining: http://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf

- Marina Santini's lecture on sentiment analysis: http://www.slideshare.net/marinasantini1/03-sais-lecturesentimentanalysiscontinued

- Andrej Karpathy's lecture on sentiment analysis: https://www.youtube.com/watch?v=ytUHvMNnzZk