





Why compute affective meaning?

- Detecting:
 - sentiment towards politicians, products, countries, ideas
 - frustration of callers to a help line
 - stress in drivers or pilots
 - depression and other medical conditions
 - confusion in students talking to e-tutors
 - emotions in novels (e.g., for studying groups that are feared over time)
- Could we generate:
 - emotions or moods for literacy tutors in the children's storybook domain
 - emotions or moods for computer games
 - personalities for dialogue systems to match the user

Connotation in the lexicon

- Words have connotation as well as sense
- Can we build lexical resources that represent these connotations?
- And use them in these computational tasks?

Positive or negative movie review?

-  • unbelievably disappointing
-  • Full of zany characters and richly applied satire, and some great plot twists
-  • this is the greatest screwball comedy ever filmed
-  • It was pathetic. The worst part about it was the boxing scenes.

9

Google Product Search



HP Officejet 6500A Plus e-All-in-One Color Ink-jet - Fax / copier / printer / scanner
\$89 online, \$100 nearby ★★★★★ 377 reviews
September 2010 - Printer - HP - Inkjet - Office - Copier - Color - Scanner - Fax - 250 sh

Reviews

Summary - Based on 377 reviews



What people are saying


Category	Rating	Quote
ease of use	★★★★	"This was very easy to setup to four computers."
value	★★★★	"Appreciate good quality at a fair price."
setup	★★★★	"Overall pretty easy setup."
customer service	★★★	"I DO like honest tech support people."
size	★★★	"Pretty Paper weight."
mode	★★★	"Photos were fair on the high quality mode."
colors	★★★	"Full color prints came out with great quality."

10

Bing Shopping

HP Officejet 6500A E710N Multifunction Printer

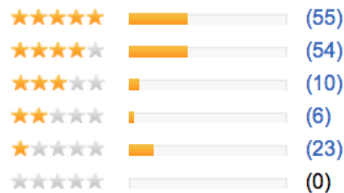
[Product summary](#) [Find best price](#) **Customer reviews** [Specifications](#) [Related items](#)



\$121.53 - \$242.39 (14 stores)

Compare

Average rating ★★★★★ (144)



Most mentioned



Show reviews by source



11

Sentiment analysis has many other names

- Opinion extraction
- Opinion mining
- Sentiment mining
- Subjectivity analysis

12

Why sentiment analysis?

- *Movie*: is this review positive or negative?
- *Products*: what do people think about the new iPhone?
- *Public sentiment*: how is consumer confidence? Is despair increasing?
- *Politics*: what do people think about this candidate or issue?
- *Prediction*: predict election outcomes or market trends from sentiment

13

Sentiment Analysis

- Sentiment analysis is the detection of **attitudes**
“enduring, affectively colored beliefs, dispositions towards objects or persons”
 1. **Holder (source)** of attitude
 2. **Target (aspect)** of attitude
 3. **Type** of attitude
 - From a set of types
 - *Like, love, hate, value, desire, etc.*
 - Or (more commonly) simple weighted **polarity**:
 - *positive, negative, neutral, together with strength*
 4. **Text** containing the attitude
 - Sentence or entire document

14

Sentiment Analysis

- Simplest task:
 - Is the attitude of this text positive or negative?
- More complex:
 - Rank the attitude of this text from 1 to 5
- Advanced:
 - Detect the target, source, or complex attitude types

Extracting Features for Sentiment Classification

- How to handle negation
 - I **didn't** like this movie
 - vs
 - I really like this movie
- Which words to use?
 - Only adjectives
 - All words
 - Community lexicons

LIWC (Linguistic Inquiry and Word Count)

Pennebaker, J.W., Booth, R.J., & Francis, M.E. (2007). Linguistic Inquiry and Word Count: LIWC 2007. Austin, TX

- Home page: <http://www.liwc.net/>
- 2300 words, >70 classes
- **Affective Processes**
 - negative emotion (*bad, weird, hate, problem, tough*)
 - positive emotion (*love, nice, sweet*)
- **Cognitive Processes**
 - Tentative (*maybe, perhaps, guess*), Inhibition (*block, constraint*)
- **Pronouns, Negation** (*no, never*), **Quantifiers** (*few, many*)
- \$30 or \$90 fee (or see me for my version)

MPQA Subjectivity Cues Lexicon

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. Proc. of HLT-EMNLP-2005.

Riloff and Wiebe (2003). Learning extraction patterns for subjective expressions. EMNLP-2003.

- Home page: <http://mpqa.cs.pitt.edu/lexicons>
- 6885 words from 8221 lemmas
 - 2718 positive
 - 4912 negative
- Each word annotated for intensity (strong, weak)
- GNU GPL

Bing Liu Opinion Lexicon

Minqing Hu and Bing Liu. Mining and Summarizing Customer Reviews. ACM SIGKDD-2004.

- [Bing Liu's Page on Opinion Mining](#)
- <http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar>
- 6786 words
 - 2006 positive
 - 4783 negative

21

SentiWordNet

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010 SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. LREC-2010

- Home page: <http://sentiwordnet.isti.cnr.it/>
- All WordNet synsets automatically annotated for degrees of positivity, negativity, and neutrality/objectiveness
- [estimable(J,3)] “may be computed or estimated”
Pos 0 Neg 0 Obj 1
- [estimable(J,1)] “deserving of respect or high regard”
Pos .75 Neg 0 Obj .25

Scherer's typology of affective states

Emotion: relatively brief episode of synchronized response of all or most organismic subsystems in response to the evaluation of an event as being of major significance

angry, sad, joyful, fearful, ashamed, proud, desperate

Mood: diffuse affect state ...change in subjective feeling, of low intensity but relatively long duration, often without apparent cause

cheerful, gloomy, irritable, listless, depressed, buoyant

Interpersonal stance: affective stance taken toward another person in a specific interaction, coloring the interpersonal exchange

distant, cold, warm, supportive, contemptuous

Attitudes: relatively enduring, affectively colored beliefs, preferences predispositions towards objects or persons

liking, loving, hating, valuing, desiring

Personality traits: emotionally laden, stable personality dispositions and behavior tendencies, typical for a person

nervous, anxious, reckless, morose, hostile, envious, jealous

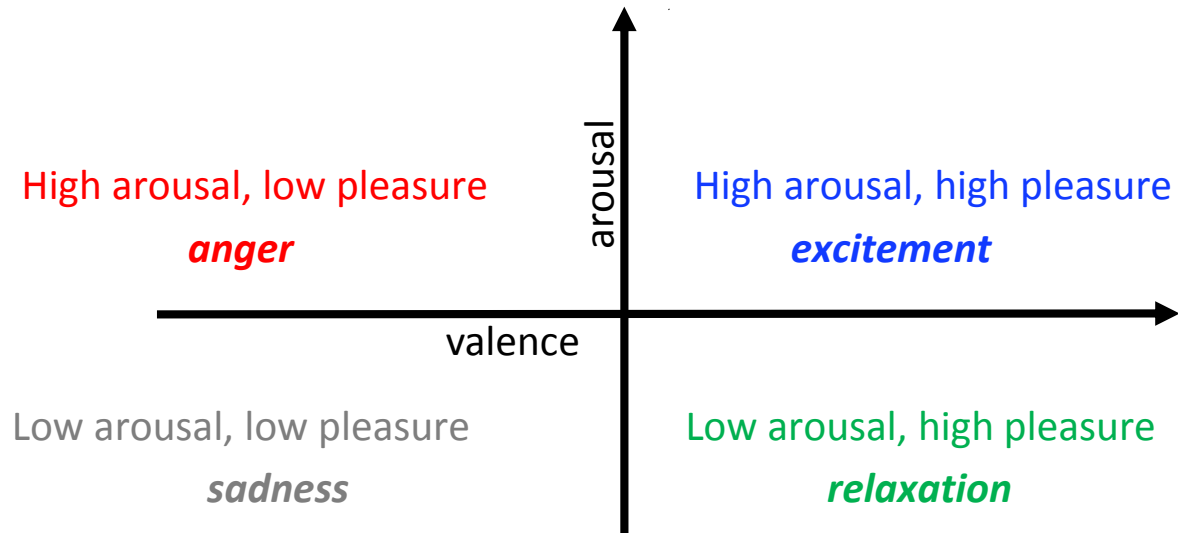
Two families of theories of emotion

- Atomic basic emotions
 - A finite list of 6 or 8, from which others are generated
- Dimensions of emotion
 - Valence (positive negative)
 - Arousal (strong, weak)
 - Control

Ekman's 6 basic emotions: Surprise, happiness, anger, fear, disgust, sadness



Valence/Arousal Dimensions



Atomic units vs. Dimensions

Distinctive

- Emotions are units.
- Limited number of basic emotions.
- Basic emotions are innate and universal

Dimensional

- Emotions are dimensions.
- Limited # of labels but unlimited number of emotions.
- Emotions are culturally learned.

Adapted from Julia Braverman

One emotion lexicon from each paradigm!

1. 8 basic emotions:
 - NRC Word-Emotion Association Lexicon (Mohammad and Turney 2011)
 2. Dimensions of valence/arousal/dominance
 - Warriner, A. B., Kuperman, V., and Brysbaert, M. (2013)
- Both built using Amazon Mechanical Turk

31

```

R = -1.35
[priorPause] * 0.01
[meanf0] * 0.81
[minRMS] * -0.11
[AVmaxf0] * -0.29
[always] * 1.23
[changes] * 0.75
[dont] * 0.55
[higher] * 0.78
[increasing] * 0.37
[not] * 0.77
[other] * 0.44
[sure] * 0.53
[vertically] * -0.4
[yes] * -0.5
[problem=Plane] * -0.1
[ASK-CALL-EXPERIMENTER] * -0.73
[CRASH-DIALOGUE20B] * -0.42
[CRASH-DIALOGUE21B] * 0.68
[ELEV-DIALOGUE10] * 0.73
[ELEV-DIALOGUE18] * 0.51
[ELEV-DIALOGUE22] * 0.49
[ELEV-DIALOGUE32] * 0.67
[ELEV-DIALOGUE39] * 0.48
[ELEV-REM-2011] * 0.64
[ELEV-REM-2211] * 0.79
[PLANE-DIALOGUE28] * 0.41
[PLANE-REM-2614] * -0.66
[PUMPKIN-DIALOGUE20] * 0.46
[PUMPKIN-REM-201a] * -0.63
[PUMPKIN-DIALOGUE30A] * -0.41
[PUMPKIN-DIALOGUE32] * 0.58
[PUMPKIN-DIALOGUE35A] * 0.57
[PUMPKIN-REM-32121] * 0.75
[SUN-DIALOGUE12] * 0.48
[SUN-DIALOGUE18] * -0.36
[maxf0] * 0.14
[maxRMS] * -0.05
[AVpriorPause] * -0.01
[action] * -0.49
[but] * -0.42
[distance] * 1.22
[down] * -0.17
[I] * 1
[mass] * 0.4
[of] * -0.19
[out] * 0.68
[till] * 0.9
[wind] * 0.92
[gender] * 0.24
[CRASH-DIALOGUE20A] * 0.74
[ASK-DO-ANOTHER-PROBLEM] * -1.18
[CRASH-DIALOGUE20C] * 0.75
[CRASH-DIALOGUE23] * 0.31
[ELEV-DIALOGUE14] * 0.62
[ELEV-DIALOGUE20] * 0.41
[ELEV-DIALOGUE28] * -0.57
[ELEV-DIALOGUE38] * -0.35
[ELEV-DIALOGUE39B] * -0.45
[ELEV-REM-2012A] * 0.76
[ELEV-REM-221A] * 0.69
[PLANE-REM-12112] * 0.66
[PLANE-REM-2811] * 0.55
[PUMPKIN-DIALOGUE15] * -1.2
[PUMPKIN-DIALOGUE28] * -0.57
[PUMPKIN-DIALOGUE30B] * -0.44
[PUMPKIN-DIALOGUE34] * 0.49
[PUMPKIN-DIALOGUE35B] * 0.43
[PUMPKIN-REM-3212A] * 0.77
[SUN-DIALOGUE14] * -0.65
    
```

An atomic unit but non-lexicon approach:

Kate Forbes-Riley, Diane Litman

Benefits and challenges of real-time uncertainty detection and adaptation in a spoken dialogue computer tutor

Speech Communication, Volume 53, Issues 9–10, 2011, 1115–1136

NRC Word-Emotion Association Lexicon

Mohammad and Turney 2011

- 10,000 words chosen mainly from earlier lexicons
- Labeled by Amazon Mechanical Turk
- 5 Turkers per hit
- Give Turkers an idea of the relevant sense of the word
- Result:

	amazingly	anger	0
	amazingly	anticipation	0
	amazingly	disgust	0
	amazingly	fear	0
	amazingly	joy	1
	amazingly	sadness	0
	amazingly	surprise	1
	amazingly	trust	0
33	amazingly	negative	0
	amazingly	positive	1

EmoLex	# of terms
EmoLex-Uni:	
Unigrams from Macquarie Thesaurus	
adjectives	200
adverbs	200
nouns	200
verbs	200
EmoLex-Bi:	
Bigrams from Macquarie Thesaurus	
adjectives	200
adverbs	187
nouns	200
verbs	200
EmoLex-GI:	
Terms from General Inquirer	
negative terms	2119
neutral terms	4226
positive terms	1787
EmoLex-WAL:	
Terms from WordNet Affect Lexicon	
anger terms	165
disgust terms	37
fear terms	100
joy terms	165
sadness terms	120
surprise terms	53
Union	10170

The AMT Hit

Prompt word: *startle*

Q1. Which word is closest in meaning (most related) to *startle*?

- *automobile*
- *shake*
- *honesty*
- *entertain*

Q2. How positive (good, praising) is the word *startle*?

- *startle* is not positive
- *startle* is weakly positive
- *startle* is moderately positive
- *startle* is strongly positive

Q3. How negative (bad, criticizing) is the word *startle*?

- *startle* is not negative
- *startle* is weakly negative
- *startle* is moderately negative
- *startle* is strongly negative

Q4. How much is *startle* associated with the emotion joy? (For example, *happy* and *fun* are strongly associated with joy.)

- *startle* is not associated with joy
- *startle* is weakly associated with joy
- *startle* is moderately associated with joy
- *startle* is strongly associated with joy

Q5. How much is *startle* associated with the emotion sadness? (For example, *failure* and *heart-break* are strongly associated with sadness.)

- *startle* is not associated with sadness
- *startle* is weakly associated with sadness
- *startle* is moderately associated with sadness
- *startle* is strongly associated with sadness

Q6. How much is *startle* associated with the emotion fear? (For example, *horror* and *scary* are strongly associated with fear.)

- Similar choices as in 4 and 5 above

Q7. How much is *startle* associated with the emotion anger? (For example, *rage* and *shouting* are strongly associated with anger.)

- Similar choices as in 4 and 5 above

Q8. How much is *startle* associated with the emotion trust? (For example, *faith* and *integrity* are strongly associated with trust.)

- Similar choices as in 4 and 5 above

Q9. How much is *startle* associated with the emotion disgust? (For example, *gross* and *cruelty* are strongly associated with disgust.)

- Similar choices as in 4 and 5 above

34

...

Lexicon of valence, arousal, and dominance

- Warriner, A. B., Kuperman, V., and Brysbaert, M. (2013). [Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods* 45, 1191-1207.](#)
- [Supplementary data: This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.](#)
- **Ratings for 14,000 words for emotional dimensions:**
 - **valence** (the pleasantness of the stimulus)
 - **arousal** (the intensity of emotion provoked by the stimulus)
 - **dominance** (the degree of control exerted by the stimulus)

35

Lexicon of valence, arousal, and dominance

- **valence** (the pleasantness of the stimulus)
 - 9: happy, pleased, satisfied, contented, hopeful
 - 1: unhappy, annoyed, unsatisfied, melancholic, despaired, or bored
- **arousal** (the intensity of emotion provoked by the stimulus)
 - 9: stimulated, excited, frenzied, jittery, wide-awake, or aroused
 - 1: relaxed, calm, sluggish, dull, sleepy, or unaroused;
- **dominance** (the degree of control exerted by the stimulus)
 - 9: in control, influential, important, dominant, autonomous, or controlling
 - 1: controlled, influenced, cared-for, awed, submissive, or guided
- Again produced by AMT

36

Lexicon of valence, arousal, and dominance: Examples

Valence		Arousal		Dominance	
vacation	8.53	rampage	7.56	self	7.74
happy	8.47	tornado	7.45	incredible	7.74
whistle	5.7	zucchini	4.18	skillet	5.33
conscious	5.53	dressy	4.15	concur	5.29
torture	1.4	dull	1.67	earthquake	2.14

37

Concreteness versus abstractness

- The degree to which the concept denoted by a word refers to a perceptible entity.
 - Do concrete and abstract words differ in connotation?
 - Storage and retrieval?
 - Bilingual processing?
 - Relevant for embodied view of cognition (Barsalou 1999 inter alia)
 - Do concrete words activate brain regions involved in relevant perception
- Brysbaert, M., Warriner, A. B., and Kuperman, V. (2014) [Concreteness ratings for 40 thousand generally known English word lemmas *Behavior Research Methods* 46, 904-911.](#)
- [Supplementary data: This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.](#)
- 37,058 English words and 2,896 two-word expressions (“zebra crossing” and “zoom in”),
- Rating from 1 (abstract) to 5 (concrete)

38

Concreteness versus abstractness

- Brysbaert, M., Warriner, A. B., and Kuperman, V. (2014) [Concreteness ratings for 40 thousand generally known English word lemmas *Behavior Research Methods* 46, 904-911.](#)
- [Supplementary data: This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.](#)
- Some example ratings from the final dataset of 40,000 words and phrases

banana 5
bathrobe 5
bagel 5
brisk 2.5
badass 2.5
basically 1.32
belief 1.19
although 1.07

39

Computing with Affective Lexicons

Semi-supervised algorithms for learning sentiment Lexicons



Semi-supervised learning of lexicons

- Use a small amount of information
 - A few labeled examples
 - A few hand-built patterns
- To bootstrap a lexicon

41

Hatzivassiloglou and McKeown intuition for identifying word polarity

Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the Semantic Orientation of Adjectives. ACL, 174–181

- Adjectives conjoined by “*and*” have same polarity
 - Fair **and** legitimate, corrupt **and** brutal
 - *fair **and** brutal, *corrupt **and** legitimate
- Adjectives conjoined by “*but*” do not
 - fair **but** brutal

42

Hatzivassiloglou & McKeown 1997

Step 1

- Label **seed set** of 1336 adjectives (all >20 in 21 million word WSJ corpus)
 - 657 positive
 - adequate central clever famous intelligent remarkable reputed sensitive slender thriving...
 - 679 negative
 - contagious drunken ignorant lanky listless primitive strident troublesome unresolved unsuspecting...

43

Hatzivassiloglou & McKeown 1997

Step 2

- Expand seed set to conjoined adjectives

Google

"was nice and"

[Nice location in Porto and the front desk staff was **nice and helpful**...](#)

www.tripadvisor.com/ShowUserReviews-g189180-d206904-r12068...

Mercure Porto Centro: Nice location in Porto and the front desk staff **was nice and helpful** - See traveler reviews, 77 candid photos, and great deals for Porto, ...

nice, helpful

[If a girl was **nice and classy**, but had some vibrant purple dye in ...](#)

[answers.yahoo.com/ Home > All Categories > Beauty & Style > Hair](http://answers.yahoo.com/Home/AllCategories/Beauty%20&%20Style/Hair)

4 answers - Sep 21

Question: Your personal opinion or what you think other people's opinions might ...
Top answer: I think she would be cool and confident like katy perry :)

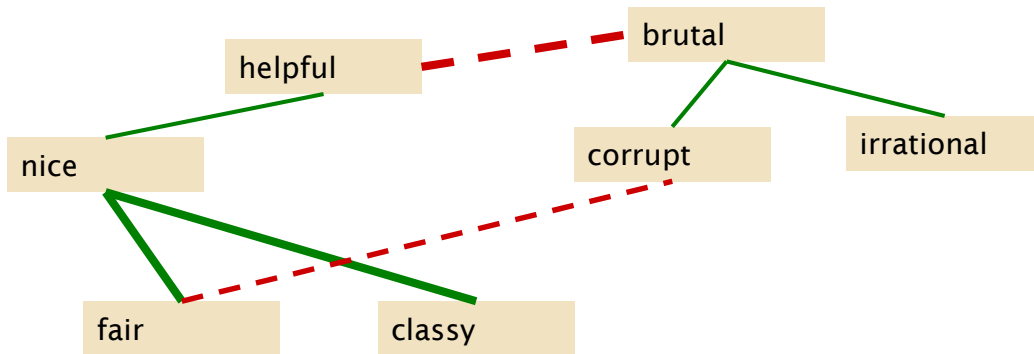
nice, classy

44

Hatzivassiloglou & McKeown 1997

Step 3

- Supervised classifier assigns “polarity similarity” to each word pair, resulting in graph:

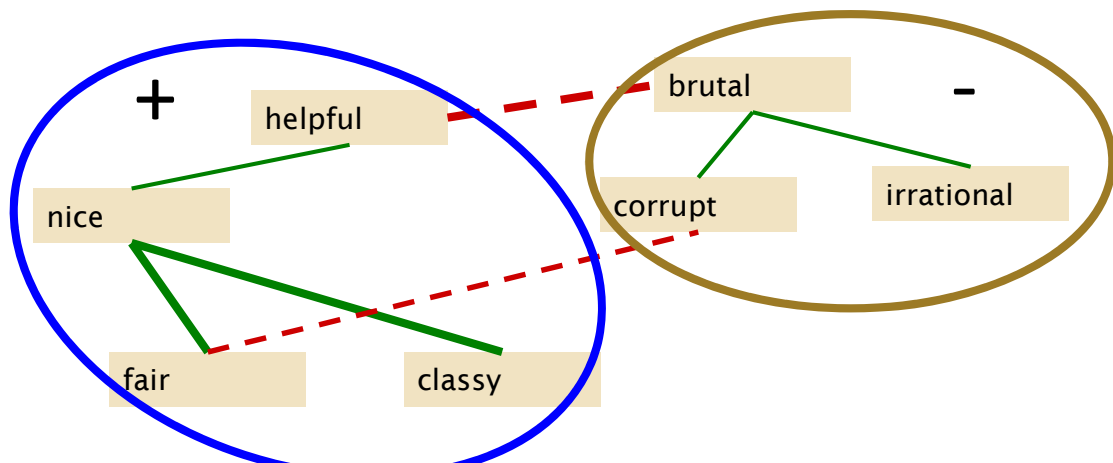


45

Hatzivassiloglou & McKeown 1997

Step 4

- Clustering for partitioning the graph into two



46

Output polarity lexicon

- Positive
 - bold decisive disturbing generous good honest important large mature patient peaceful positive proud sound stimulating straightforward strange talented vigorous witty...
- Negative
 - ambiguous cautious cynical evasive harmful hypocritical inefficient insecure irrational irresponsible minor outspoken pleasant reckless risky selfish tedious unsupported vulnerable wasteful...

47

Output polarity lexicon

- Positive
 - bold decisive **disturbing** generous good honest important large mature patient peaceful positive proud sound stimulating straightforward **strange** talented vigorous witty...
- Negative
 - ambiguous **cautious** cynical evasive harmful hypocritical inefficient insecure irrational irresponsible minor **outspoken pleasant** reckless risky selfish tedious unsupported vulnerable wasteful...

48

Turney Algorithm

Turney (2002): Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews

1. Extract a *phrasal lexicon* from reviews
2. Learn polarity of each phrase
3. Rate a review by the average polarity of its phrases

49

Extract two-word phrases with adjectives

First Word	Second Word	Third Word (not extracted)
JJ	NN or NNS	anything
RB, RBR, RBS	JJ	Not NN nor NNS
JJ	JJ	Not NN or NNS
NN or NNS	JJ	Nor NN nor NNS
RB, RBR, or RBS	VB, VBD, VBN, VBG	anything

50

How to measure polarity of a phrase?

- Positive phrases co-occur more with “*excellent*”
- Negative phrases co-occur more with “*poor*”
- But how to measure co-occurrence?

51

How to measure polarity of a phrase?

- Positive phrases co-occur more with “*excellent*”
- Negative phrases co-occur more with “*poor*”
- But how to measure co-occurrence?
 - Pointwise mutual information

52

Does phrase appear more with “poor” or “excellent”?

$$\begin{aligned}
 \text{Polarity}(\textit{phrase}) &= \text{PMI}(\textit{phrase}, \text{"excellent"}) - \text{PMI}(\textit{phrase}, \text{"poor"}) \\
 &= \log_2 \frac{\frac{1}{N} \text{hits}(\textit{phrase} \text{ NEAR } \text{"excellent"})}{\frac{1}{N} \text{hits}(\textit{phrase}) \frac{1}{N} \text{hits}(\text{"excellent"})} - \log_2 \frac{\frac{1}{N} \text{hits}(\textit{phrase} \text{ NEAR } \text{"poor"})}{\frac{1}{N} \text{hits}(\textit{phrase}) \frac{1}{N} \text{hits}(\text{"poor"})} \\
 &= \log_2 \frac{\text{hits}(\textit{phrase} \text{ NEAR } \text{"excellent"})}{\text{hits}(\textit{phrase}) \text{hits}(\text{"excellent"})} \frac{\text{hits}(\textit{phrase}) \text{hits}(\text{"poor"})}{\text{hits}(\textit{phrase} \text{ NEAR } \text{"poor"})} \\
 &= \log_2 \left(\frac{\text{hits}(\textit{phrase} \text{ NEAR } \text{"excellent"}) \text{hits}(\text{"poor"})}{\text{hits}(\textit{phrase} \text{ NEAR } \text{"poor"}) \text{hits}(\text{"excellent"})} \right)
 \end{aligned}$$

53

Phrases from a thumbs-up review

Phrase	POS tags	Polarity
online service	JJ NN	2 . 8
online experience	JJ NN	2 . 3
direct deposit	JJ NN	1 . 3
local branch	JJ NN	0 . 42
...		
low fees	JJ NNS	0 . 33
true service	JJ NN	- 0 . 73
other bank	JJ NN	- 0 . 85
inconveniently located	JJ NN	- 1 . 5
<i>Average</i>		0 . 32

54

Phrases from a thumbs-down review

Phrase	POS tags	Polarity
direct deposits	JJ NNS	5 . 8
online web	JJ NN	1 . 9
very handy	RB JJ	1 . 4
...		
virtual monopoly	JJ NN	-2 . 0
lesser evil	RBR JJ	-2 . 3
other problems	JJ NNS	-2 . 8
low funds	JJ NNS	-6 . 8
unethical practices	JJ NNS	-8 . 5
<i>Average</i>		-1 . 2

55

Results of Turney algorithm

- 410 reviews from Epinions
 - 170 (41%) negative
 - 240 (59%) positive
- Majority class baseline: 59%
- Turney algorithm: 74%

- Phrases rather than words
- Learns domain-specific information

56

Using WordNet to learn polarity

S.M. Kim and E. Hovy. 2004. Determining the sentiment of opinions. COLING 2004
M. Hu and B. Liu. Mining and summarizing customer reviews. In Proceedings of KDD, 2004

- WordNet: online thesaurus
- Create positive (“good”) and negative seed-words (“terrible”)
- Find Synonyms and Antonyms
 - Positive Set: Add synonyms of positive words (“well”) and antonyms of negative words
 - Negative Set: Add synonyms of negative words (“awful”) and antonyms of positive words (“evil”)
- Repeat, following chains of synonyms
- Filter

57

Summary on semi-supervised lexicon learning

- Advantages:
 - Can be domain-specific
 - Can be more robust (more words)
- Intuition
 - Start with a seed set of words (‘good’, ‘poor’)
 - Find other words that have similar polarity:
 - Using “and” and “but”
 - Using words that occur nearby in the same document
 - Using WordNet synonyms and antonyms

Analyzing the polarity of each word in IMDB

Potts, Christopher. 2011. On the negativity of negation. SALT 20, 636-659.

- How likely is each word to appear in each sentiment class?
- Count(“bad”) in 1-star, 2-star, 3-star, etc.

- But use **likelihood**:

$$P(w|c) = \frac{f(w,c)}{\sum_{w \in c} f(w,c)}$$

- Make them comparable between words

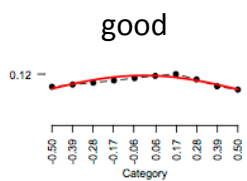
- Scaled likelihood:

$$\frac{P(w|c)}{P(w)}$$

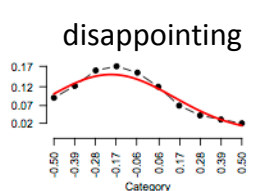
“Potts diagrams”

Potts, Christopher. 2011. NSF workshop on restructuring adjectives.

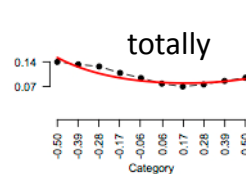
Positive scalars



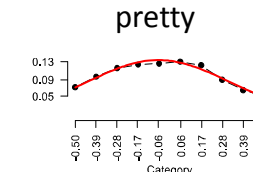
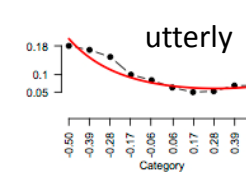
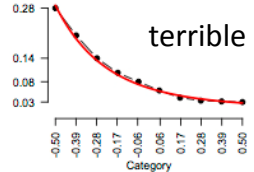
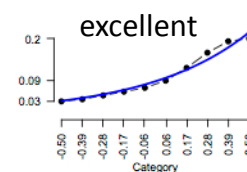
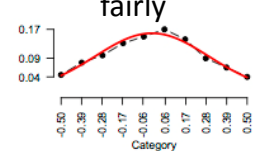
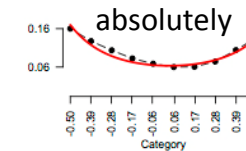
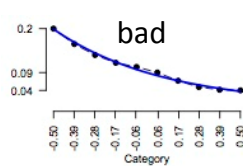
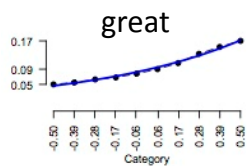
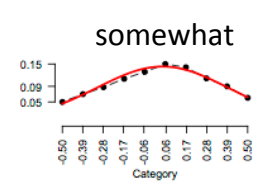
Negative scalars



Emphatics



Attenuators



Or use regression coefficients to weight words

- Train a classifier based on supervised data
 - Predict: human-labeled connotation of a document
 - From: all the words and bigrams in it
- Use the regression coefficients as the weights
- We'll return to an example of this in the next section.

63

Computing with Affective Lexicons

Using the lexicons to detect affect



Lexicons for detecting document affect: Simplest unsupervised method

- Sentiment:
 - Sum the weights of each positive word in the document
 - Sum the weights of each negative word in the document
 - Choose whichever value (positive or negative) has higher sum
- Emotion:
 - Do the same for each emotion lexicon

65

Lexicons for detecting document affect: Simplest supervised method

- Build a classifier
 - Predict sentiment (or emotion, or personality) given features
 - Use “counts of lexicon categories” as a features
 - Sample features:
 - LIWC category “cognition” had count of 7
 - NRC Emotion category “anticipation” had count of 2
- Baseline
 - Instead use counts of **all** the words and bigrams in the training set
 - This is hard to beat
 - But only works if the training and test sets are very similar

66



Computing with Affective Lexicons

Sample affective task: personality detection

Sample affective task: personality detection

Scherer's typology of affective states

Emotion: relatively brief episode of synchronized response of all or most organismic subsystems in response to the evaluation of an event as being of major significance

angry, sad, joyful, fearful, ashamed, proud, desperate

Mood: diffuse affect state ...change in subjective feeling, of low intensity but relatively long duration, often without apparent cause

cheerful, gloomy, irritable, listless, depressed, buoyant

Interpersonal stance: affective stance taken toward another person in a specific interaction, coloring the interpersonal exchange

distant, cold, warm, supportive, contemptuous

Attitudes: relatively enduring, affectively colored beliefs, preferences predispositions towards objects or persons

liking, loving, hating, valuing, desiring

Personality traits: emotionally laden, stable personality dispositions and behavior tendencies, typical for a person

nervous, anxious, reckless, morose, hostile, envious, jealous

The Big Five Dimensions of Personality

Extraversion vs. Introversion

sociable, assertive, playful vs. aloof, reserved, shy

Emotional stability vs. Neuroticism

calm, unemotional vs. insecure, anxious

Agreeableness vs. Disagreeable

friendly, cooperative vs. antagonistic, faultfinding

Conscientiousness vs. Unconscientious

self-disciplined, organised vs. inefficient, careless

Openness to experience

intellectual, insightful vs. shallow, unimaginative

70

Some text corpora labeled for personality of author

Pennebaker, James W., and Laura A. King. 1999. "Linguistic styles: language use as an individual difference." *Journal of personality and social psychology* 77, no. 6.

- 2,479 essays from psychology students (1.9 million words), "write whatever comes into your mind" for 20 minutes

Mehl, Matthias R, SD Gosling, JW Pennebaker. 2006. Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life. *Journal of personality and social psychology* 90 (5), 862

- Speech from Electronically Activated Recorder (EAR)
- Random snippets of conversation recorded, transcribed
- 96 participants, total of 97,468 words and 15,269 utterances

Schwartz, H. Andrew, Johannes C. Eichstaedt, Margaret L. Kern, Lukasz Dziurzynski, Stephanie M. Ramones, Megha Agrawal, Achal Shah et al. 2013. "Personality, gender, and age in the language of social media: The open-vocabulary approach." *PLoS one* 8, no. 9

- Facebook
- 75,000 volunteers
- 309 million words
- All took a personality test

Litman, D., Paletz, S., Rahimi, Z., Allegretti, S., & Rice, C. 2016.. "The Teams Corpus and Entrainment in Multi-Party Spoken Dialogues." *Proceedings Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

- 47 hours of dialogue from 63 teams (216 individuals, each with a personality test)

Ears (speech) corpus (Mehl et al.)

Introvert	Extravert
<ul style="list-style-type: none"> - Yeah you would do kilograms. Yeah I see what you're saying. - On Tuesday I have class. I don't know. - I don't know. A16. Yeah, that is kind of cool. - I don't know. I just can't wait to be with you and not have to do this every night, you know? - Yeah. You don't know. Is there a bed in there? Well ok just... 	<ul style="list-style-type: none"> - That's my first yogurt experience here. Really watery. Why? - Damn. New game. - Oh. - That's so rude. That. - Yeah, but he, they like each other. He likes her. - They are going to end up breaking up and he's going to be like.
Unconscientious	Conscientious
<ul style="list-style-type: none"> - With the Chinese. Get it together. - I tried to yell at you through the window. Oh. xxxx's fucking a dumb ass. Look at him. Look at him, dude. Look at him. I wish we had a camera. He's fucking brushing his t-shirt with a tooth brush. Get a kick of it. Don't steal nothing. 	<ul style="list-style-type: none"> - I don't, I don't know for a fact but I would imagine that historically women who have entered prostitution have done so, not everyone, but for the majority out of extreme desperation and I think. I don't know, i think people understand that desperation and they don't don't see [...]

Essays corpus (Pennebaker and King)

Introvert	Extravert
I've been waking up on time so far. What has it been, 5 days? Dear me, I'll never keep it up, being such not a morning person and all. But maybe I'll adjust, or not. I want internet access in my room, I don't have it yet, but I will on Wed??? I think. But that ain't soon enough, cause I got calculus homework [...]	I have some really random thoughts. I want the best things out of life. But I fear that I want too much! What if I fall flat on my face and don't amount to anything. But I feel like I was born to do BIG things on this earth. But who knows... There is this Persian party today.
Neurotic	Emotionally stable
One of my friends just barged in, and I jumped in my seat. This is crazy. I should tell him not to do that again. I'm not that fastidious actually. But certain things annoy me. The things that would annoy me would actually annoy any normal human being, so I know I'm not a freak.	I should excel in this sport because I know how to push my body harder than anyone I know, no matter what the test I always push my body harder than everyone else. I want to be the best no matter what the sport or event. I should also be good at this because I love to ride my bike.

Classifiers

- **Mairesse**, François, Marilyn A. Walker, Matthias R. Mehl, and Roger K. Moore. "Using linguistic cues for the automatic recognition of personality in conversation and text." *Journal of artificial intelligence research* (2007): 457-500.
 - Various classifiers, lexicon-based and prosodic features
- **Schwartz**, H. Andrew, Johannes C. Eichstaedt, Margaret L. Kern, Lukasz Dziurzynski, Stephanie M. Ramones, Megha Agrawal, Achal Shah et al. 2013. "Personality, gender, and age in the language of social media: The open-vocabulary approach." *PloS one* 8, no.
 - regression and SVM, lexicon-based and all-words

Sample LIWC Features

LIWC (Linguistic Inquiry and Word Count)

Pennebaker, J.W., Booth, R.J., & Francis, M.E. (2007). *Linguistic Inquiry and Word Count: LIWC 2007*. Austin, TX

Feature	Type	Example
Anger words	LIWC	hate, kill, pissed
Metaphysical issues	LIWC	God, heaven, coffin
Physical state/function	LIWC	ache, breast, sleep
Inclusive words	LIWC	with, and, include
Social processes	LIWC	talk, us, friend
Family members	LIWC	mom, brother, cousin
Past tense verbs	LIWC	walked, were, had
References to friends	LIWC	pal, buddy, coworker
Imagery of words	MRC	Low: future, peace - High: table, car
Syllables per word	MRC	Low: a - High: uncompromisingly
Concreteness	MRC	Low: patience, candor - High: ship
Frequency of use	MRC	Low: duly, nudity - High: he, the

Normalizing LIWC category features (Schwartz et al 2013, Facebook study)

- Mairesse:
Raw LIWC counts
- Schwartz et al:
Normalized per writer:

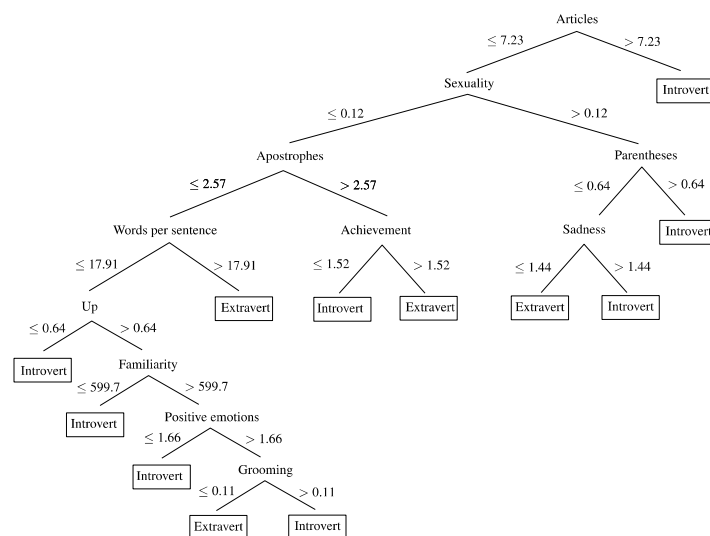
$$p(\text{category} | \text{subject}) = \frac{\sum_{\text{word} \in \text{category}} \text{freq}(\text{word}, \text{subject})}{\sum_{\text{word} \in \text{vocab}(\text{subject})} \text{freq}(\text{word}, \text{subject})}$$

Sample results

- Agreeable:
 - +Family, +Home, -Anger, -Swear
- Extravert
 - +Friend, +Religion, +Self
- Conscientiousness:
 - -Swear, -Anger, -NegEmotion,
- Emotional Stability:
 - -NegEmotion, +Sports,
- Openness
 - -Cause, -Space

77

Decision tree for predicting extraversion in essay corpus (Mairesse et al)



78

Affect extraction: of course it's not just the lexicon

Ranganath et al (2013), McFarland et al (2014)

- Detecting interpersonal stance in conversation
- Speed dating study, 1000 4-minute speed dates
- Subjects labeled **selves** and **each other** for
 - friendly (each on a scale of 1-10)
 - awkward
 - flirtatious
 - assertive

83

Scherer's typology of affective states

Emotion: relatively brief episode of synchronized response of all or most organismic subsystems in response to the evaluation of an event as being of major significance

angry, sad, joyful, fearful, ashamed, proud, desperate

Mood: diffuse affect state ...change in subjective feeling, of low intensity but relatively long duration, often without apparent cause

cheerful, gloomy, irritable, listless, depressed, buoyant

Interpersonal stance: affective stance taken toward another person in a specific interaction, coloring the interpersonal exchange

distant, cold, warm, supportive, contemptuous

Attitudes: relatively enduring, affectively colored beliefs, preferences predispositions towards objects or persons

liking, loving, hating, valuing, desiring

Personality traits: emotionally laden, stable personality dispositions and behavior tendencies, typical for a person

nervous, anxious, reckless, morose, hostile, envious, jealous

Affect extraction: of course it's not just the lexicon

Logistic regression classifier with

- LIWC lexicons
- Other lexical features
 - Lists of hedges
- Prosody (pitch and energy means and variance)
- Discourse features
 - Interruptions
 - Dialog acts/Adjacency pairs
 - sympathy (“Oh, that’s terrible”)
 - clarification question (“What?”)
 - appreciations (“That’s awesome!”)

85

Results on affect extraction

- Friendliness
 - -negEmotion
 - -hedge
 - higher pitch
- Awkwardness
 - +negation
 - +hedges
 - +questions

86

Summary: Connotation in the lexicon

- Words have various connotational aspects
- Methods for building connotation lexicons
 - Based on theoretical models of emotion, sentiment
 - By hand (mainly using crowdsourcing)
 - Semi-supervised learning from seed words
 - Fully supervised (when you can find a convenient signal in the world)
- Applying lexicons to detect affect and sentiment
 - Unsupervised: pick simple majority sentiment (positive/negative words)
 - Supervised: learn weights for each lexical category