

Natural Language Processing

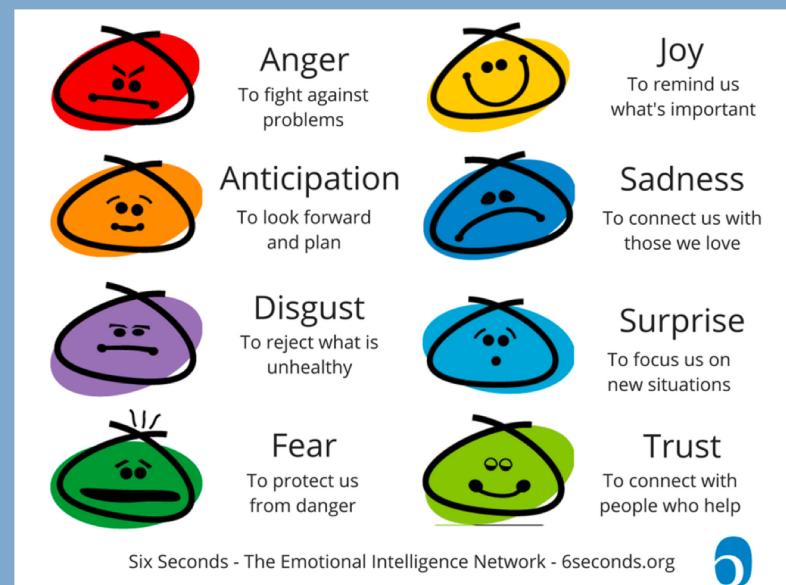
Week6: Affective Computing and Sentiment Analysis using Lexicons

Dr. Haithem Afli

Haithem.afli@cit.ie

[@AfliHaithem](#)

2020/2021



Outline

- **Affective meaning**
- Sentiment Lexicons
- Other Affective Lexicons
- Semi-supervised algorithms for learning sentiment Lexicons
- Supervised Learning of Sentiment Lexicons
- Using the lexicons to detect affect
- Sample affective task: personality detection

Affective meaning

- Drawing on literatures in
 - affective computing (Picard 95)
 - linguistic subjectivity (Wiebe and colleagues)
 - social psychology (Pennebaker and colleagues)
- Can we model the lexical semantics relevant to:
 - sentiment
 - emotion
 - personality
 - mood
 - attitudes

Adapted from Dan Jurafsky and James H. Martin

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?

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

Adapted from Dan Jurafsky and James H. Martin

Outline

- Affective meaning
- **Sentiment Lexicons**
- Other Affective Lexicons
- Semi-supervised algorithms for learning sentiment Lexicons
- Supervised Learning of Sentiment Lexicons
- Using the lexicons to detect affect
- 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 General Inquirer

Philip J. Stone, Dexter C Dunphy, Marshall S. Smith, Daniel M. Ogilvie. 1966. The General Inquirer: A Computer Approach to Content Analysis. MIT Press

- Home page: <http://www.wjh.harvard.edu/~inquirer>
- List of Categories: <http://www.wjh.harvard.edu/~inquirer/homecat.htm>
- Spreadsheet: <http://www.wjh.harvard.edu/~inquirer/inquirerbasic.xls>
- Categories:
 - Positiv (1915 words) and Negativ (2291 words)
 - Strong vs Weak, Active vs Passive, Overstated versus Understated
 - Pleasure, Pain, Virtue, Vice, Motivation, Cognitive Orientation, etc
- Free for Research Use

List of entries in tag category:

Positiv

List shows first 100 entries. Total number of entries in this category:

1915

Entries for this category are shown with all tags assigned and sense definitions:

ABIDE

H4 Positiv Affil Active Doctrin IAV SUPV

ABILITY

H4Lvd Positiv Strong Virtue EVAL Abs@ ABS MeansLw Noun

ABLE

H4Lvd Positiv Pstv Strong Virtue EVAL MeansLw Modif adjective: Having necessary power, skill, resources, etc.

ABOUND

H4 Positiv Passive Increas IAV SUPV

ABSOLVE

H4 Positiv Active SocRel ComForm IAV SUPV

ABSORBENT

H4 Positiv Increas IndAdj Modif

ABSORPTION

List of entries in tag category:

Negativ

List shows first 100 entries. Total number of entries in this category:

2291

Entries for this category are shown with all tags assigned and sense definitions:

ABANDON

H4Lvd Negativ Ngtv Weak Fail IAV AffLoss AffTot SUPV

ABANDONMENT

H4 Negativ Weak Fail Noun

ABATE

H4Lvd Negativ Passive Decreas IAV TranLw SUPV

ABDICATE

H4 Negativ Weak Submit Passive Finish IAV SUPV

ABHOR

H4 Negativ Hostile Passive Arousal SV SUPV

ABJECT

H4 Negativ Weak Submit Passive Vice IPadj Modif

ABNORMAL

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

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: https://mpqa.cs.pitt.edu/lexicons/subj_lexicon/
- 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

```
;::::::::::::::::::;
; Opinion Lexicon: Negative
; This file contains a list of NEGATIVE opinion words (or sentiment words).
; This file and the papers can all be downloaded from
; http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html
; If you use this list, please cite the following paper:
;
; Mingqing Hu and Bing Liu. "Mining and Summarizing Customer Reviews."
; Proceedings of the ACM SIGKDD International Conference on Knowledge
; Discovery and Data Mining (KDD-2004), Aug 22-25, 2004, Seattle,
; Washington, USA,
;
; Notes:
; 1. The appearance of an opinion word in a sentence does not necessarily
; mean that the sentence expresses a positive or negative opinion.
; See the paper below:
;
; Bing Liu. "Sentiment Analysis and Subjectivity." An chapter in
; Handbook of Natural Language Processing, Second Edition,
; (editors: N. Indurkhy and F. J. Damerau), 2010.
;
; 2. You will notice many misspelled words in the list. They are not
; mistakes. They are included as these misspelled words appear
; frequently in social media content.
;
;::::::::::::::::::;
2-faced
2-faces
abnormal
abolish
abominable
abominably
abominate
abomination
abort
aborted
aborts
abrade
abrasive
```

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

Download from: <https://github.com/aesuli/sentiwordnet>

Haithem Aflī, Sorcha McGuire, and Andy Way. Sentiment translation for low resourced languages: Experiments on irish general election tweets. In Proceedings of the 18th International Conference on Computational Linguistics and Intelligent Text Processing, Budapest, Hungary, 2017.

Home page: <https://github.com/HAfli/SentiFocloir>

Senti-Foclóir is an Irish language (Gaeilge) sentiment lexicon containing 5571 words at present. It is based on the AFINN-111 sentiment lexicon which was manually translated from English to Irish.

5571 lines (5571 sloc) 70.9 KB		
1	tréig	-3
2	thréig	-3
3	tréigeann	-3
4	tréigim	-3
5	tréigimid	-3
6	tréigtear	-3
7	thréigean	-3
8	tréigthe	-3
9	thréigeamar	-3
10	tréigeadh	-3
11	tréigfidh	-3
12	tréigfimid	-3
13	thréigfeadh	-3
14	thréigfeá	-3
15	thréigfinn	-3
16	fág	-1
--		

Use case

Entered Keyword:

Cast your vote for the Irish today, vote left,
vote republican #GE16

Sentiment of your tweet: Positive

Top Positive Tweets

We're fighting to win this election. SF wants to lead the next Government says @GerryAdamsSF at Louth #GE16 launch

A vote for both @sinnfeinireland and the Marxist 'People Before Profit' is a vote for abortion. #GE16 #Dublin

Thank you @CormacDevlin for pledging as a #GE16 candidate for people with neuro conditions Pledge #80000minds

Sligo Is Ready For Real Change A wonderful turnout for the launch of my general election campaign. #GE16

Speaking in a chipper in Cork, Sinn Féin TD Tiger Twomey said much could be learned from more magic roundabouts. #ge16

Top Negative Tweets

Enda had a torturous conversation about the election date in Davos even sits in the wrong seat #gobshite #ge16

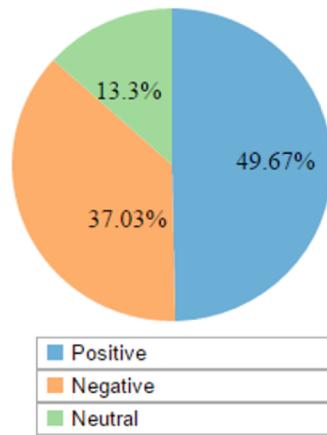
FF attack FG, FG attack FF, they'll both attack SF who will attack them & Lab << this is how we don't do negative campaigning here! #GE16

Quite disappointed with the paltry amount of @FineGael candidates who have signed the @NWCI manifesto. #FemGen. #GE16

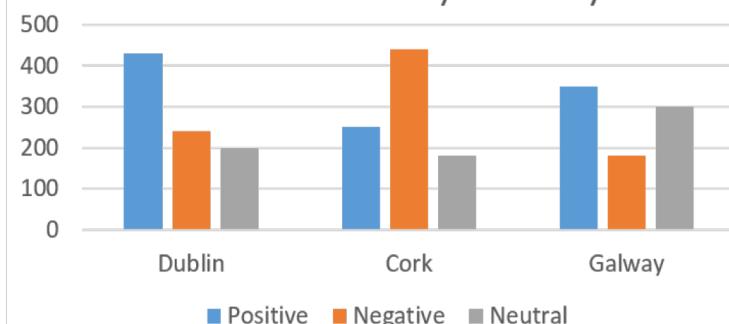
Richard Bruton 2009: "A govt who is lost in the wilderness will let us die because of their shame" #GE16 #vimb

@Páidí Cole - on that basis if current Govt not returned we will have a hung Dáil & a chaotic period of uncertainty& another election #GE16

Overall Sentiment of #GE16



Most influential county's in SA system



Discussion



Outline

- Affective meaning
- Sentiment Lexicons
- **Other Affective Lexicons**
- Semi-supervised algorithms for learning sentiment Lexicons
- Supervised Learning of Sentiment Lexicons
- Using the lexicons to detect affect
- 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

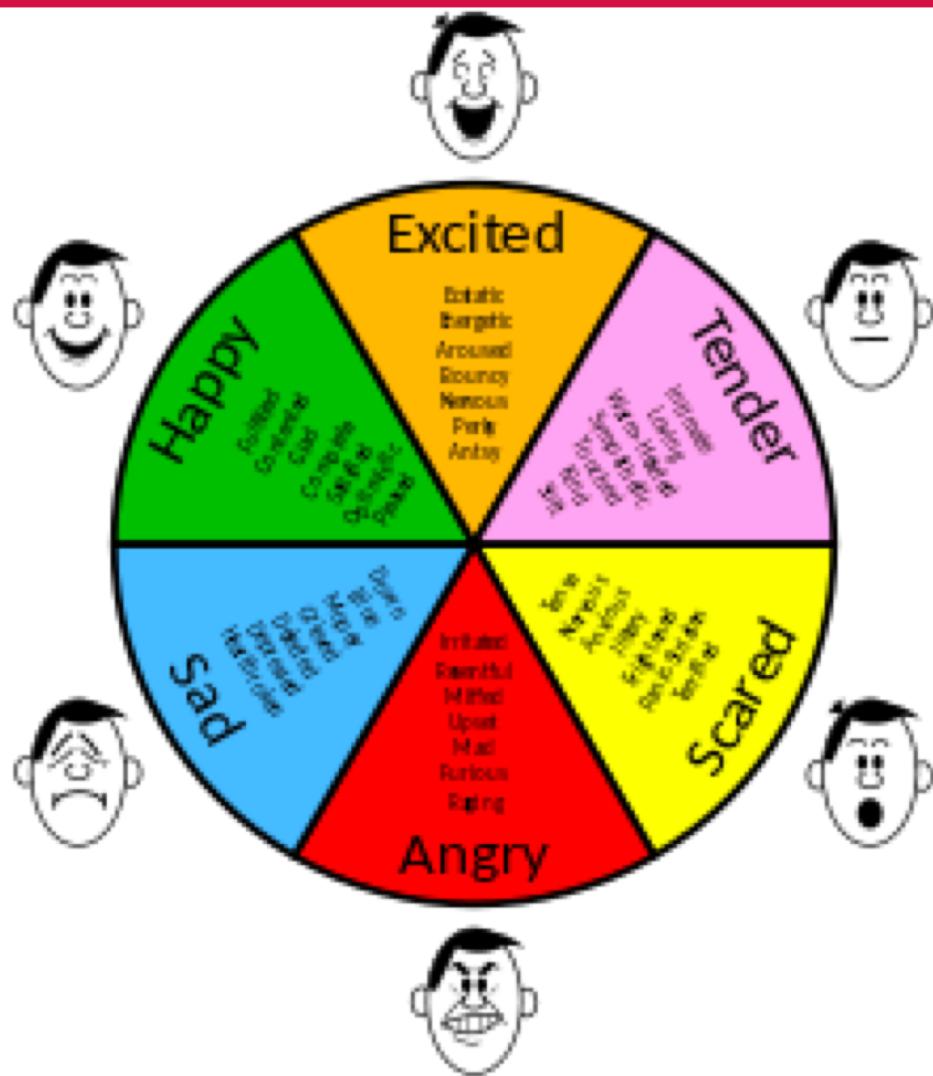
nervous, anxious, reckless, morose, hostile, envious, jealous

Two families of theories of emotion

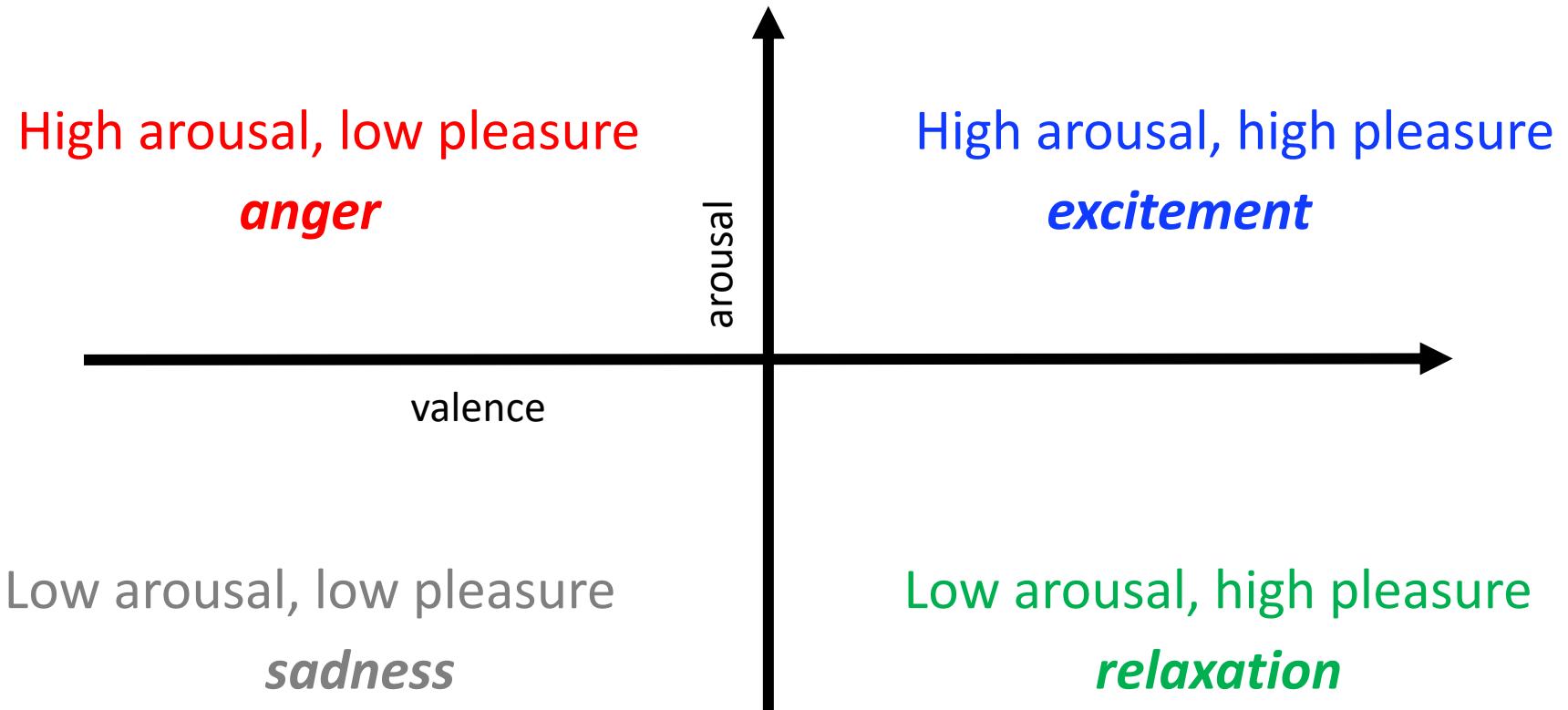
- Atomic basic emotions
 - A finite list of 6 or 8, from which others are generated
- Dimensions of emotion
 - Valence (positive negative)
 - Arousal (strong, weak)
 - Control

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

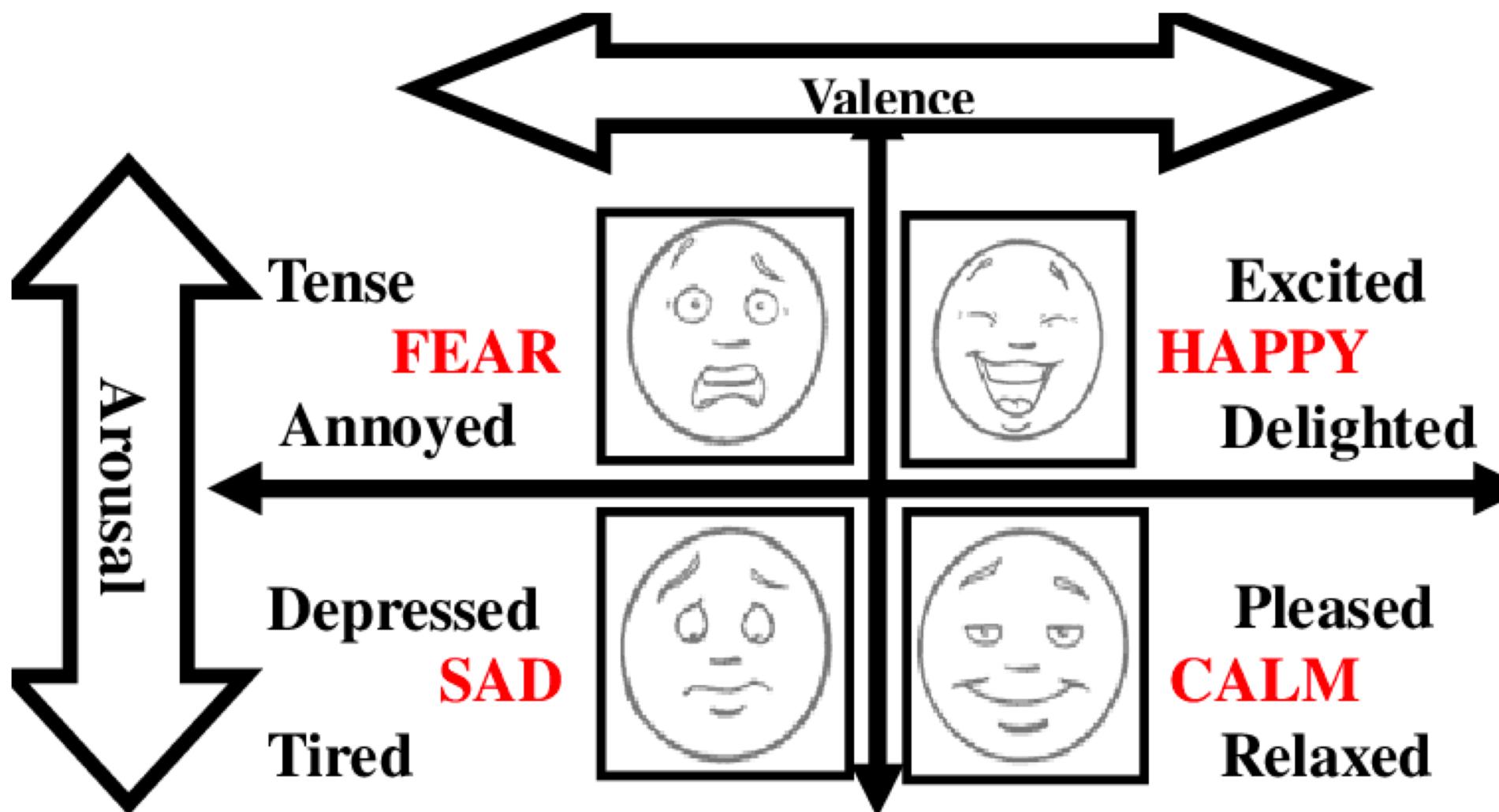




Valence/Arousal Dimensions



Adapted from Dan Jurafsky and James H. Martin



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:

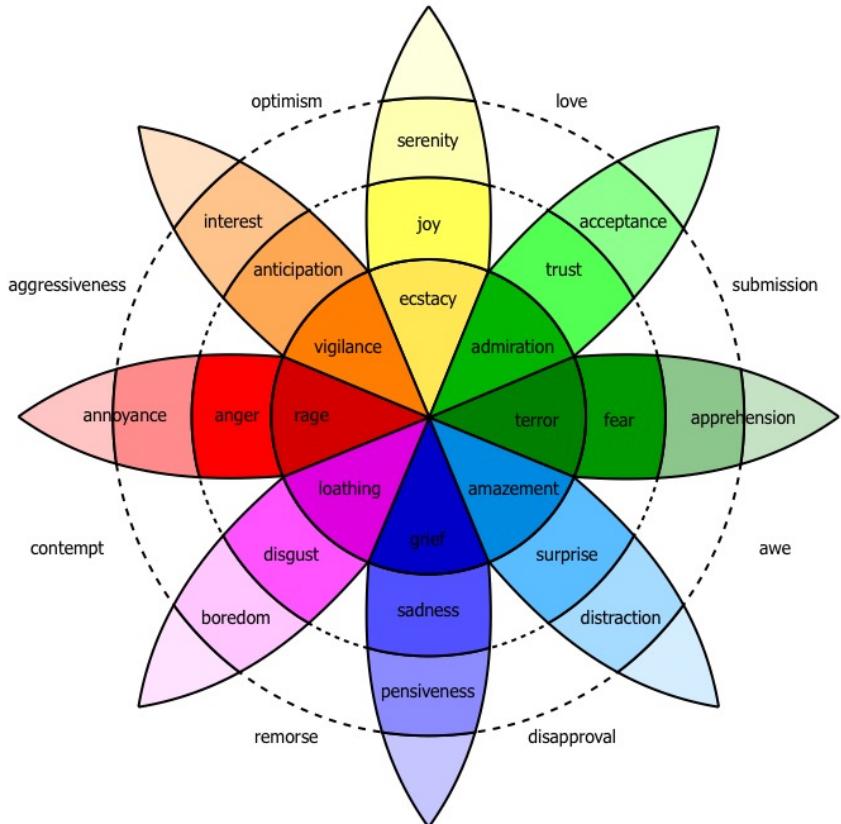
- NRC Word-Emotion Association Lexicon (Mohammad and Turney 2011)

2. Dimensions of valence/arousal/dominance

- Warriner, A. B., **Kuperman, V.**, and Brysbaert, M. (2013)
- Both built using Amazon Mechanical Turk

Plutchick's wheel of emotion

- 8 basic emotions
- in four opposing pairs:
 - joy–sadness
 - anger–fear
 - trust–disgust
 - anticipation–surprise



In 1980 Robert Plutchik constructed a wheel-like diagram of emotions visualising eight basic emotions, plus eight derivative emotions each composed of two basic ones.
Source: Wikicommons

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
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

...

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)

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

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

Discussion



Outline

- Affective meaning
- Sentiment Lexicons
- Other Affective Lexicons
- **Semi-supervised algorithms for learning sentiment Lexicons**
- Supervised Learning of Sentiment Lexicons
- Using the lexicons to detect affect
- Sample affective task: personality detection

Semi-supervised learning of lexicons



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

Hatzivassiloglou and McKeown intuition for identifying word polarity

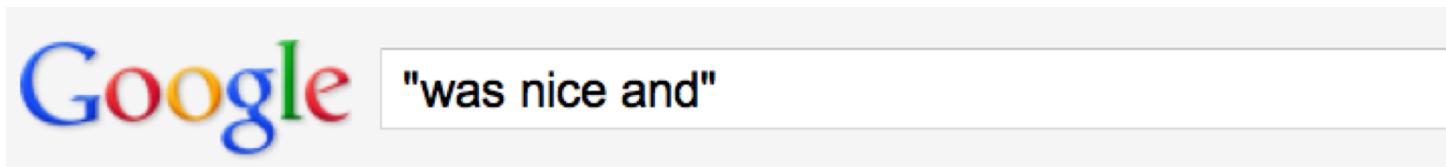


Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the Semantic Orientation of Adjectives. ACL, 174–181

- 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

- 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...

- 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... +7

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 +1

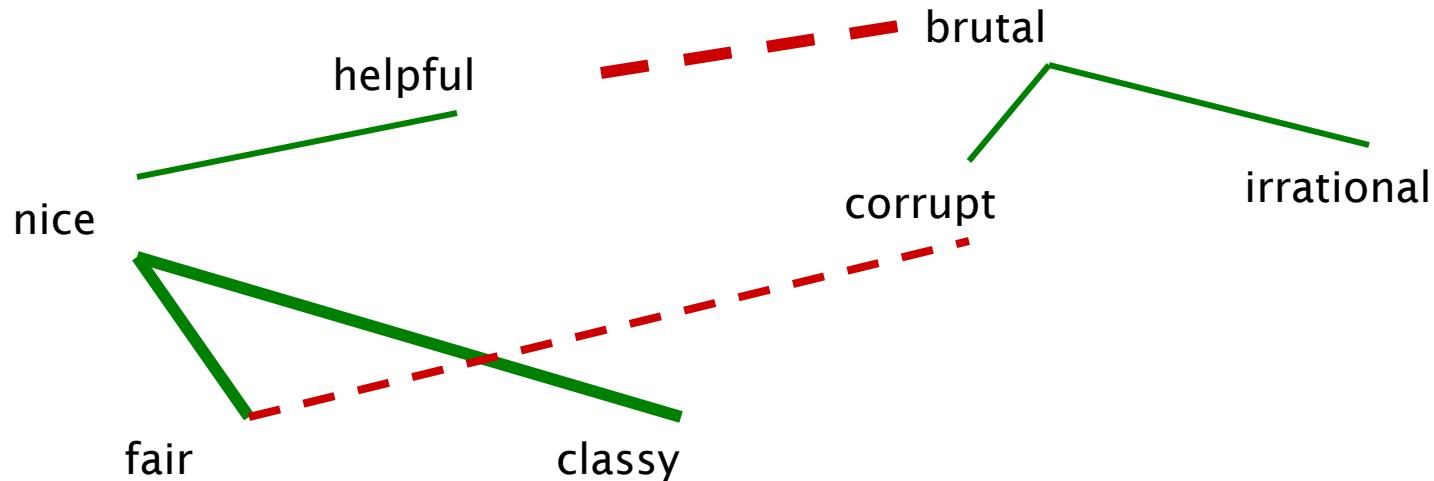
4 answers - Sep 21

Question: Your personal opinion or what you think other people's opinions might ...

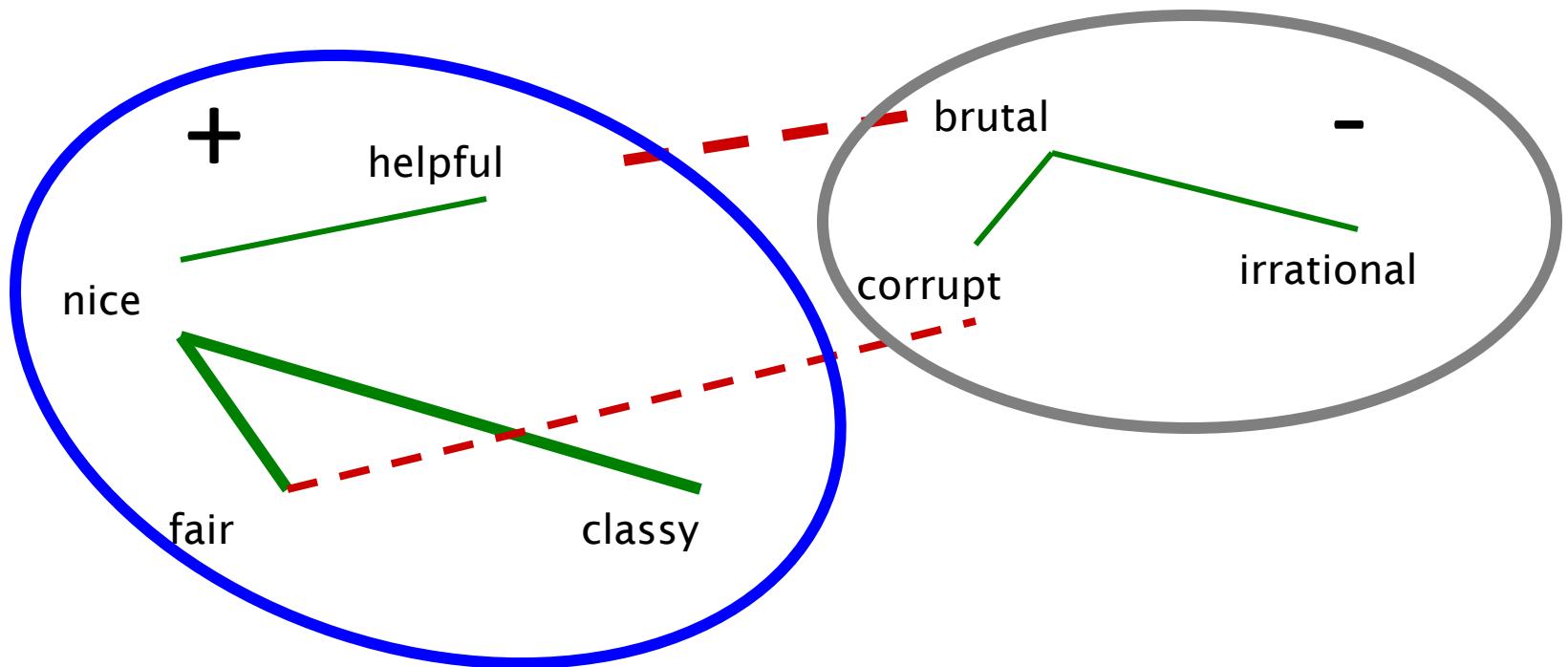
Top answer: I think she would be cool and confident like katy perry :)

nice, classy

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



- Clustering for partitioning the graph into two



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...

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...

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 thesuarus
- 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

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
 - Use seeds and semi-supervised learning to induce lexicons

Outline

- Affective meaning
- Sentiment Lexicons
- Other Affective Lexicons
- Semi-supervised algorithms for learning sentiment Lexicons
- **Supervised Learning of Sentiment Lexicons**
- Using the lexicons to detect affect
- Sample affective task: personality detection

Learn word sentiment supervised by online review scores



Potts, Christopher. 2011. On the negativity of negation. *SALT* 20, 636-659.
Potts 2011 NSF Workshop talk.

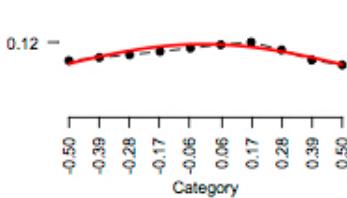
- Review datasets
 - IMDB, Goodreads, Open Table, Amazon, Trip Advisor
- Each review has a score (1-5, 1-10, etc)
- Just count how many times each word occurs with each score
 - (and normalize)

“Potts diagrams”

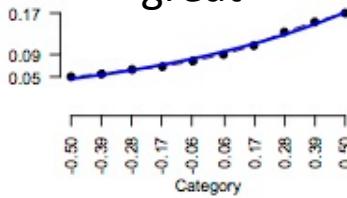


Positive scalars

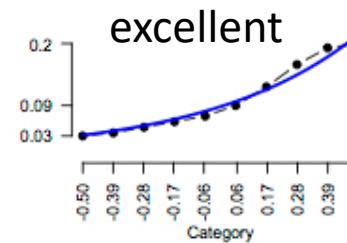
good



great

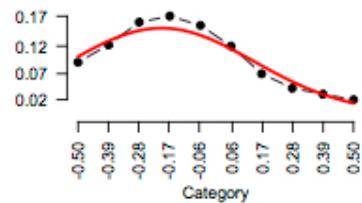


excellent

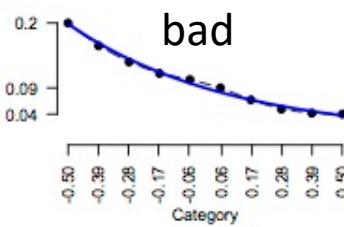


Negative scalars

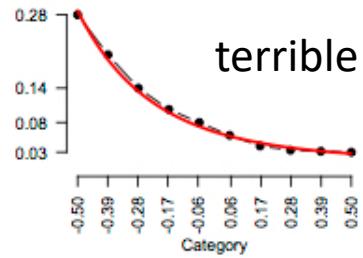
disappointing



bad



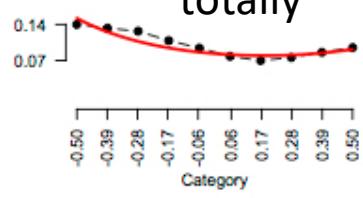
terrible



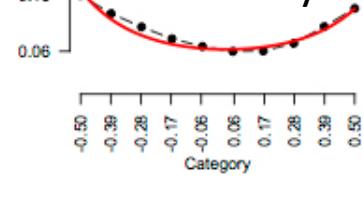
Potts, Christopher. 2011. NSF workshop on restructuring adjectives.

Emphatics

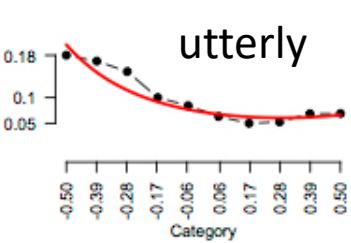
totally



absolutely

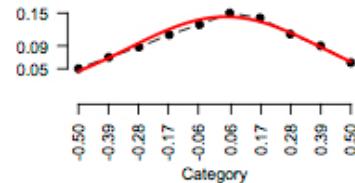


utterly

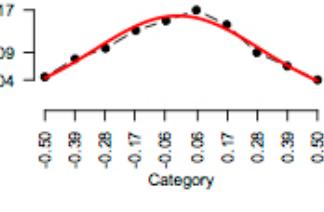


Attenuators

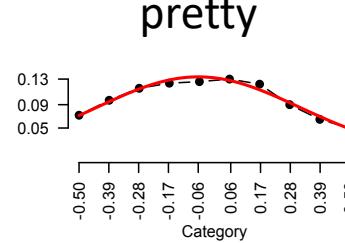
somewhat



fairly



pretty



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

Outline

- Affective meaning
- Sentiment Lexicons
- Other Affective Lexicons
- Semi-supervised algorithms for learning sentiment Lexicons
- Supervised Learning of Sentiment Lexicons
- **Using the lexicons to detect affect**
- Sample affective task: personality detection

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

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 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

Discussion



Outline

- Affective meaning
- Sentiment Lexicons
- Other Affective Lexicons
- Semi-supervised algorithms for learning sentiment Lexicons
- Supervised Learning of Sentiment Lexicons
- Using the lexicons to detect affect
- **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

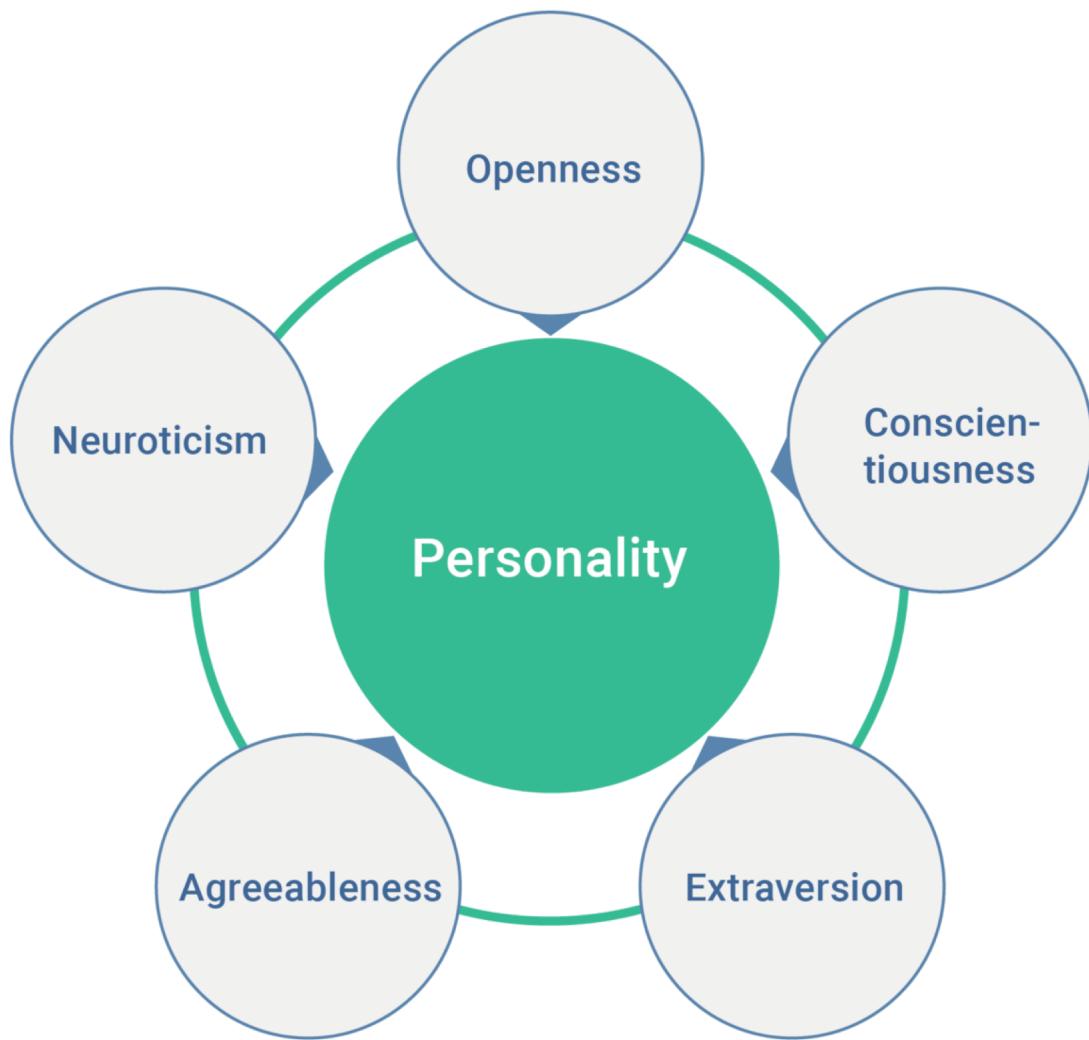
Attitudes: relatively enduring, affectively colored beliefs, preferences predispositions towards objects or persons

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



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

Various 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

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.
Unconsciousious	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)



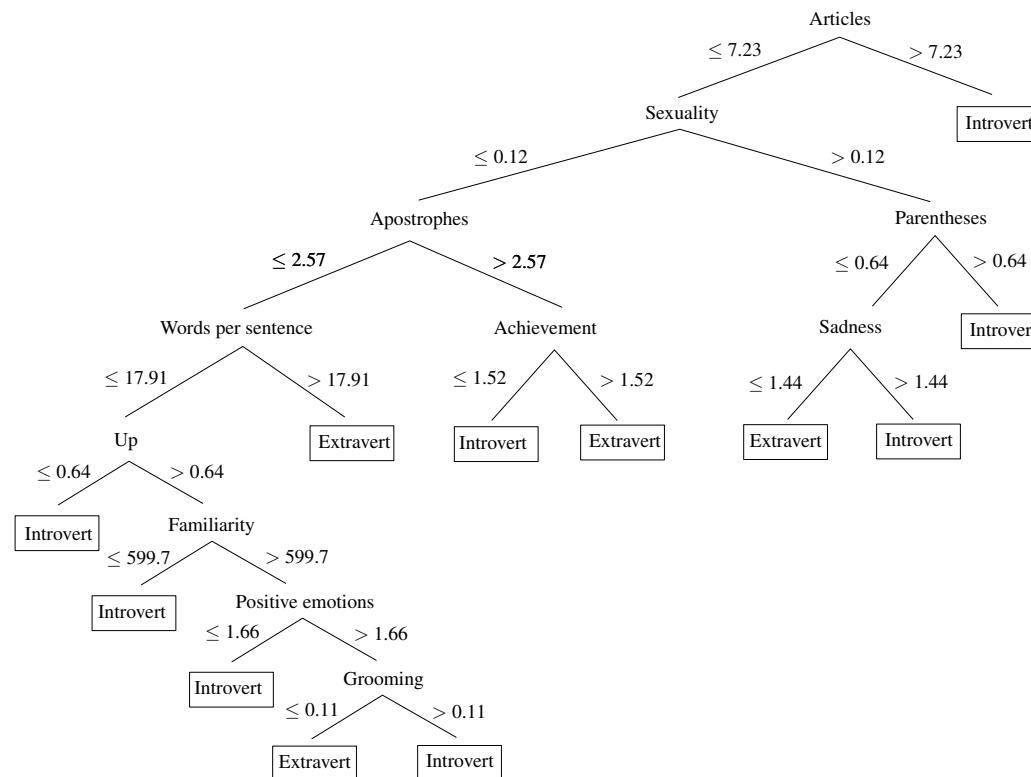
- Mairesse:
Raw LIWC counts
- Schwartz et al:
Normalized per writer:

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

Sample results

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

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



Using all words instead of lexicons

Facebook study



Schwartz et al. (2013)

- Choosing phrases with $\text{pmi} > 2 * \text{length}$ [in words]
- Only use $\text{pmi}(\text{phrase}) = \log \frac{p(\text{phrase})}{\prod_{w \in \text{phrase}} p(w)}$ at least 1% of writers
- Normalize counts of words and phrases by writer

$$p(\text{phrase} \mid \text{subject}) = \frac{\text{freq}(\text{phrase}, \text{subject})}{\sum_{\text{phrase}' \in \text{vocab}(\text{subject})} \text{freq}(\text{phrase}', \text{subject})}$$

Facebook study, Learned words, Extraversion versus Introversion

tonite soo
yall! !_! bday chill doin wit
fam text_me bout ready
ladies fam boyslets;) chillin last_night:
aint miss lil love_you hit_me_up
girl beach bestie sooo
weekend party guys
night_with dont great_night ya
gettin ?_? a_blast an_amazing
goin excited baby thats
here_we im my_life lookin
soooo lovin feelin
babe haha

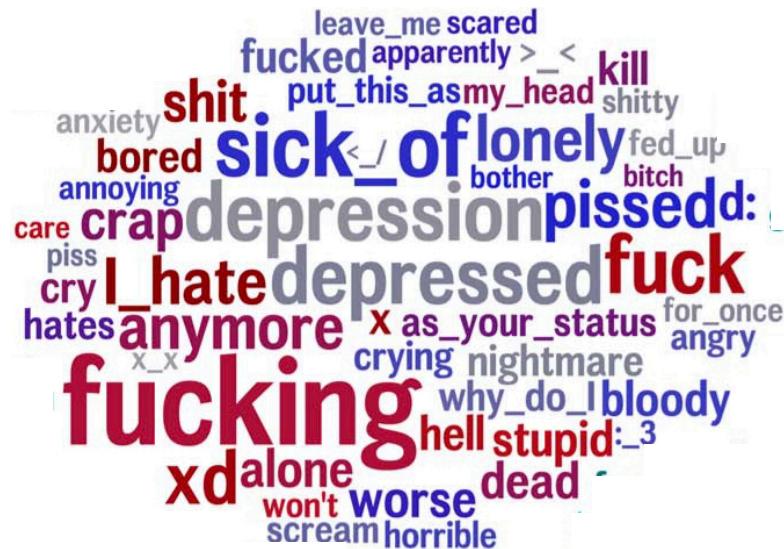
A word cloud image where the size and color of each word represent its frequency or importance. The most prominent words are 'anime' (large, blue), 'xD' (large, red), 'computer' (large, red), 'internet' (large, red), 'manga' (large, grey), 'final_fantasy' (medium, grey), 'pokémon' (large, blue), 'reading' (medium, red), 'comic' (medium, red), 'naman using' (medium, grey), 'japanese' (medium, blue), 'drawing' (medium, blue), 'apparently' (medium, red), 'pc ng' (medium, red), '0.0' (large, purple), 'ang' (large, red), 'doctor_who' (medium, grey), 'dx' (large, red), 'ako' (large, red), '._at_least' (medium, grey), 'draw' (medium, blue), 'sigh' (large, red), 'books' (medium, blue), 'please_put_this_related' (medium, blue), 't._.t' (medium, blue), '>.<' (medium, blue), 'they're' (medium, red), 'evil' (medium, red), '@_.' (medium, red), 'mga' (medium, red), '%_won't' (medium, red), 'suddenly' (medium, red), '>.<' (medium, red), 'd:' (medium, red), and 'nga' (medium, red).

Facebook study, Learned words

Neuroticism versus Emotional Stability



A word cloud visualization showing learned words. The words are primarily in shades of purple, blue, and red, representing positive or neutral concepts. Key words include: team, soccer, lakers, success, basketball, blessed, praise, beach, church, and various sports-related terms like volleyball, basketball, and football.



A word cloud visualization showing learned words. The words are primarily in shades of red, blue, and grey, representing negative or emotional concepts. Key words include: shit, sick_of, depression, pissed, fuck, fucking, alone, dead, and various negative adjectives and phrases like leave_me_scared, apparently>_<, kill, shitty, fed_up, bother, bitch, annoying, care, crap, piss, cry, I_hate, depressed, anymore, x as your status, for once, angry, crying, nightmare, why do I bloody hell stupid, xd, won't, worse, scream, horrible, etc.

Evaluating Schwartz et al (2013) Facebook Classifier



- Train on labeled training data
 - LIWC category counts
 - words and phrases (n-grams of size 1 to 3, passing a collocation filter)
- Tested on a held-out set
- Correlations with human labels
 - LIWC .21-.29
 - All Words .29-.41

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

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 awesom!”)

Results on affect extraction

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

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

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

Discussion



Thank you

[Haithem. afli@cit.ie](mailto:Haithem.afli@cit.ie)

[@AfliHaithem](https://twitter.com/AfliHaithem)