Faculty for Informatics

Technical
University
of Munich

Natural Language Processing
IN2361

PD Dr. Georg Groh

Social Computing
Research Group

Chapter 21
Lexicons for Sentiment, Affect and Connotation

- content is based on [1] and [2]
- certain elements (e.g. equations or tables) were taken over or taken over in a modified form from [1] or [2]
- citations of [1] and [2] are omitted for legibility
- errors are fully in the responsibility of Georg Groh
- BIG thanks to Dan and James for a great book!

Affective Computing, Sentiment Analysis, Subjectivity

- Scherer typology of affective states:

Emotion: Relatively brief episode of response to the evaluation of an external or internal event as being of major significance.
(*angry, sad, joyful, fearful, ashamed, proud, elated, desperate*)

Mood: Diffuse affect state, most pronounced as 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, colouring the interpersonal exchange in that situation.
(*distant, cold, warm, supportive, contemptuous, friendly*)

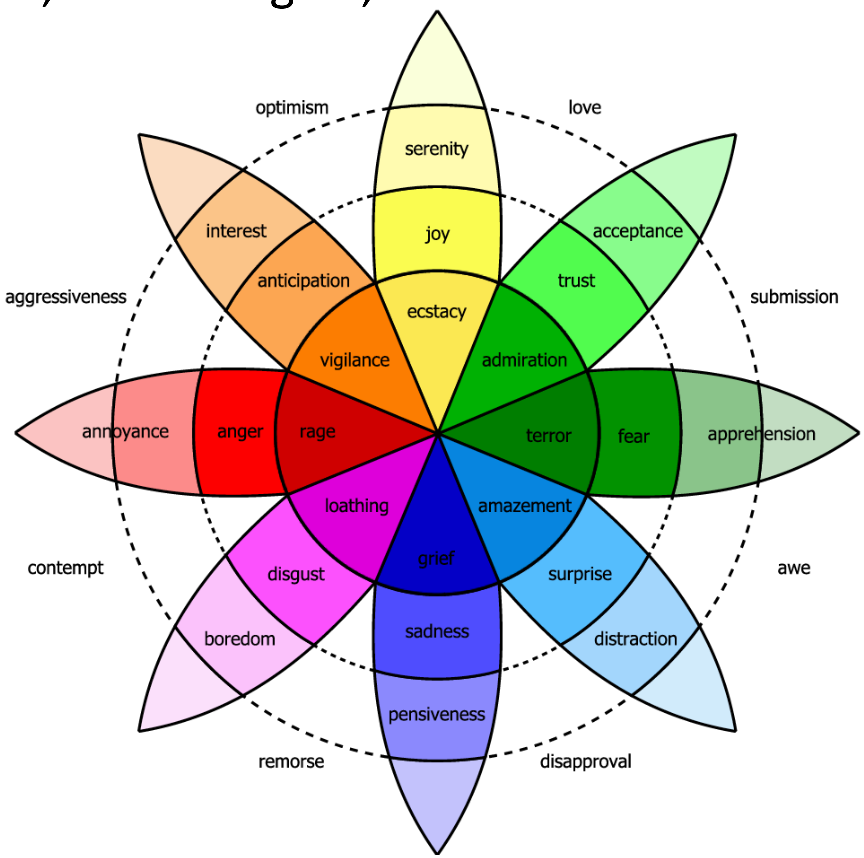
Attitude: Relatively enduring, affectively colored beliefs, preferences, and predispositions towards objects or persons.
(*liking, loving, hating, valuing, desiring*)

Personality traits: Emotionally laden, stable personality dispositions and behavior tendencies, typical for a person.
(*nervous, anxious, reckless, morose, hostile, jealous*)

e.g. Sentiment Analysis (see chapter 6) = extraction of attitudes

Encoding Emotion

- linear combinations of basic emotions:
  - Ekman: surprise, happiness, anger, fear, disgust, sadness
  - Plutchik: joy–sadness, anger–fear, trust–disgust, and anticipation–surprise



- vector space: e.g. 3 dimensional:
  - valence: the pleasantness of the stimulus
  - arousal: the intensity of emotion provoked by the stimulus
  - dominance: the degree of control exerted by the stimulus

Lexicons for Emotion and Other Affective States

Assumption: words have an affective meaning (connotation)

- General Inquirer (1966): 1915 positive and 2291 negative words
- MPQA Subjectivity lexicon (2005): 2718 positive and 4912 negative words with reliability score (weakly or strongly subjective)
- polarity lexicon of (Hu and Liu, 2004): 2006 positive and 4783 negative words

Positive	admire, amazing, assure, celebration, charm, eager, enthusiastic, excellent, fancy, fantastic, frolic, graceful, happy, joy, luck, majesty, mercy, nice, patience, perfect, proud, rejoice, relief, respect, satisfactorily, sensational, super, terrific, thank, vivid, wise, wonderful, zest
Negative	abominable, anger, anxious, bad, catastrophe, cheap, complaint, condescending, deceit, defective, disappointment, embarrass, fake, fear, filthy, fool, guilt, hate, idiot, inflict, lazy, miserable, mourn, nervous, objection, pest, plot, reject, scream, silly, terrible, unfriendly, vile, wicked

Lexicons for Emotion and Other Affective States

general idea: ask human study participants (questionnaires / crowdsourcing / serious games / …):

- NRC Word-Emotion Association Lexicon (EmoLex) (2013): 14000 words,
  - raters first determine word sense with association task, then rate word sense with Plutchik's 8 emotions and 4 intensity levels;
  - post-processing: majority voting, outlier removal, reduce to 2 intensity levels

Which word is closest in meaning (most related) to startle?
- automobile
- shake
- honesty
- entertain

Word	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	positive	negative
reward	0	1	0	0	1	0	1	1	1	0
worry	0	1	0	1	0	1	0	0	0	1
tenderness	0	0	0	0	1	0	0	0	1	0
sweetheart	0	1	0	0	1	1	0	1	1	0
suddenly	0	0	0	0	0	0	1	0	0	0
thirst	0	1	0	0	0	1	1	0	0	0
garbage	0	0	1	0	0	0	0	0	0	1

Lexicons for Emotion and Other Affective States

- NRC Valence, Arousal, and Dominance (VAD) lexicon (2018) (20000 words)

Valence		Arousal		Dominance	
vacation	.840	enraged	.962	powerful	.991
delightful	.918	party	.840	authority	.935
whistle	.653	organized	.337	saxophone	.482
consolation	.408	effortless	.120	discouraged	.0090
torture	.115	napping	.046	weak	.045

- NRC Emotion/Affect Intensity Lexicon (2018) (5814 words)

Anger		Fear		Joy		Sadness	
outraged	0.964	horror	0.923	superb	0.864	sad	0.844
violence	0.742	anguish	0.703	cheered	0.773	guilt	0.750
coup	0.578	pestilence	0.625	rainbow	0.531	unkind	0.547
oust	0.484	stressed	0.531	gesture	0.387	difficulties	0.421
suspicious	0.484	failing	0.531	warms	0.391	beggar	0.422
nurture	0.059	confident	0.094	hardship	.031	sing	0.017

Lexicons for Emotion and Other Affective States

- Linguistic Inquiry and Word Count (LWIC) (2007): 2300 words, 73 categories: e.g. positive and negative emotion, anger, sadness, cognitive mechanisms, perception, tentative, inhibition, etc.

Positive Emotion	Negative Emotion	Insight	Inhibition	Family	Negate
appreciat*	anger*	aware*	avoid*	brother*	aren't
comfort*	bore*	believe	careful*	cousin*	cannot
great	cry	decid*	hesitat*	daughter*	didn't
happy	despair*	feel	limit*	family	neither
interest	fail*	figur*	oppos*	father*	never
joy*	fear	know	prevent*	grandf*	no
perfect*	griev*	knew	reluctan*	grandm*	nobod*
please*	hate*	means	safe*	husband	none
safe*	panic*	notice*	stop	mom	nor
terrific	suffers	recogni*	stubborn*	mother	nothing
value	terrify	sense	wait	niece*	nowhere
wow*	violent*	think	wary	wife	without

Semi-supervised Induction of Sentiment Lexicons

Semantic Axis Methods

- **step 1**: (in a genre-sensitive way) choose positive and negative seed words

Domain	Positive seeds	Negative seeds
General	good, lovely, excellent, fortunate, pleasant, delightful, perfect, loved, love, happy	bad, horrible, poor, unfortunate, unpleasant, disgusting, evil, hated, hate, unhappy
Twitter	love, loved, loves, awesome, nice, amazing, best, fantastic, correct, happy	hate, hated, hates, terrible, nasty, awful, worst, horrible, wrong, sad
Finance	successful, excellent, profit, beneficial, improving, improved, success, gains, positive	negligent, loss, volatile, wrong, losses, damages, bad, litigation, failure, down, negative

- **step 2**: train new or use or fine-tune existing word-embeddings for seed words on corpus and compute centroids:

$$\left.\begin{aligned} S^{+} &= \{E(w_1^{+}), E(w_2^{+}), \ldots, E(w_n^{+})\} \\ S^{-} &= \{E(w_1^{-}), E(w_2^{-}), \ldots, E(w_m^{-})\} \end{aligned}\right\}$$

$$\mathbf{V}^{+} = \frac{1}{n}\sum_{1}^{n} E(w_i^{+})$$

$$\mathbf{V}^{-} = \frac{1}{n}\sum_{1}^{m} E(w_i^{-})$$

Semi-supervised Induction of Sentiment Lexicons

Semantic Axis Methods

- **step 3**: define "semantic axis" with respect to sentiment for genre and project on that axis:

$$\mathbf{V}_{axis} = \mathbf{V}^{+} - \mathbf{V}_{-}$$

$$\begin{aligned} \text{score}(w) &= \left(\cos(E(w), \mathbf{V}_{\text{axis}}\right) \\ &= \frac{E(w) \cdot \mathbf{V}_{\text{axis}}}{\|E(w)\| \|\mathbf{V}_{\text{axis}}\|} \end{aligned}$$

the higher the score the more w is aligned with S^{+}

Semi-supervised Induction of Sentiment Lexicons

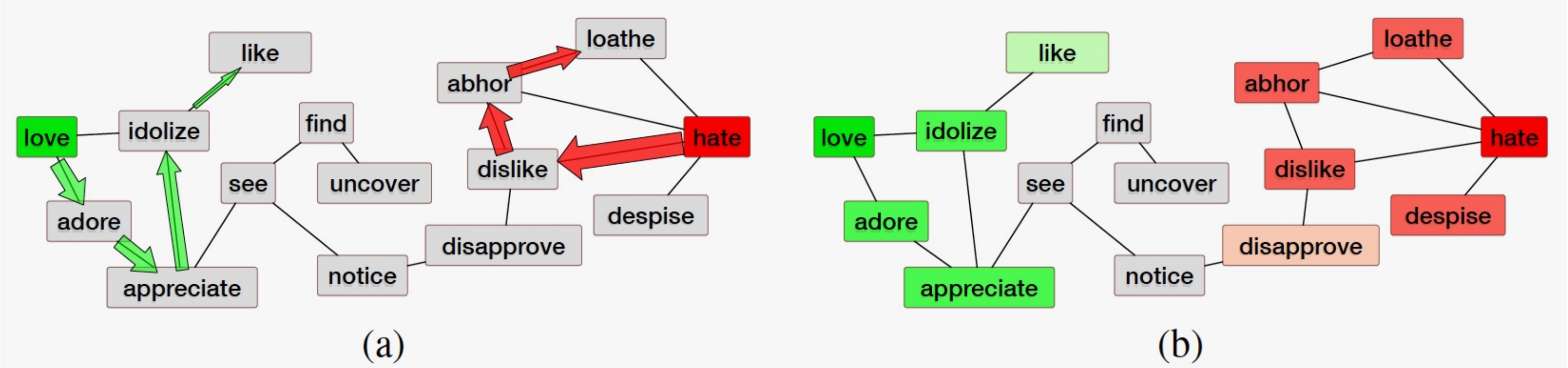
Label Propagation example: SentProp (2016)

- **step 1**: define similarity graph on word embeddings: connect word w_i to its k nearest neighbors; weight edges with angle between embeddings

$$\mathbf{E}_{i,j} = \arccos\left(-\frac{\mathbf{w_i}^\top \mathbf{w_j}}{\|\mathbf{w_i}\|\|\mathbf{w_j}\|}\right)$$

- **step 2**: choose positive and negative seed words

- **step 3**: propagate polarities:
random walk starting at seeds with transition probabilities $\mathbf{E}_{i,j}$;
a word's positive / negative raw polarity score: proportional to number of visits from positive / negative seeds



Semi-supervised Induction of Sentiment Lexicons

Label Propagation example: SentProp (2016)

- **step 4**: normalize raw scores:

$$\text{score}^{+}(w_i) = \frac{\text{score}^{+}(w_i)}{\text{score}^{+}(w_i) + \text{score}^{-}(w_i)}$$

- **step 5**: start from various seeds and compute variance of scores for each run → confidence measure

Semi-supervised Induction of Sentiment Lexicons

Other ideas / basic ideas

function BUILDSENTIMENTLEXICON(*posseeds*,*negseeds*) **returns** *poslex*,*neglex*

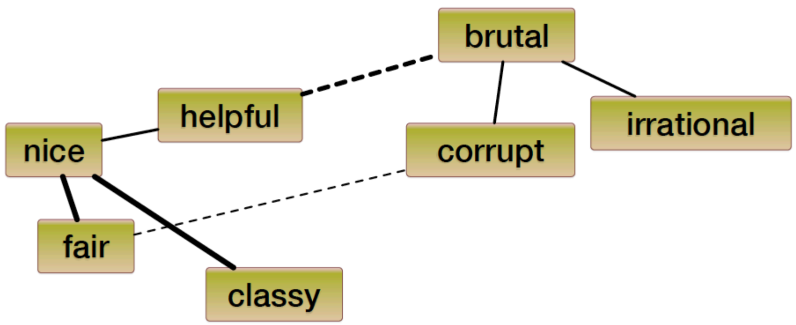
$poslex \leftarrow posseeds$
$neglex \leftarrow negseeds$
Until done
$poslex \leftarrow poslex$ + FINDSIMILARWORDS($poslex$)
$neglex \leftarrow neglex$ + FINDSIMILARWORDS($neglex$)
$poslex, neglex \leftarrow$ POSTPROCESS($poslex$,$neglex$)

example: adjectives: Hatzivassiloglou & McKeown algorithm (1997): use conjunction patterns as proxy for polarity:

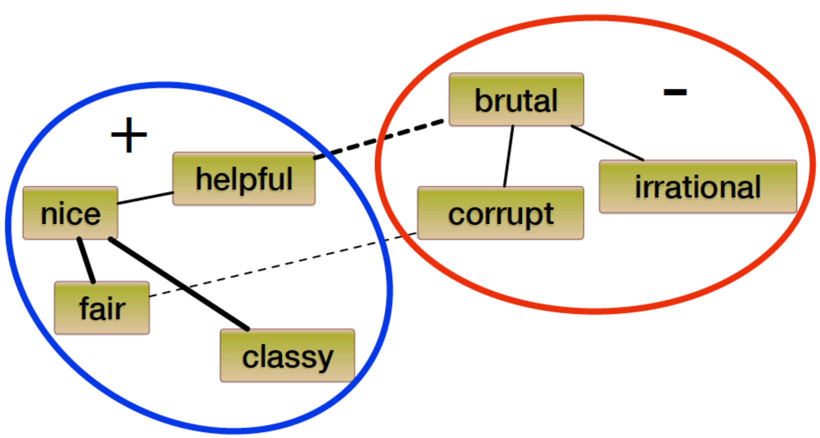
- step 1: hand-label common adjectives (→ posseeds, negseeds)
- step 2: create candidate polarity-related adjectives:
  - coordination *and*: hint for same polarity:
    fair and legitimate, corrupt and brutal
  - coordination *but*: hint for opposite polarity:
    fair but brutal, beautiful but mean
  - morphological negation (un- , im- ,-less…):
    adequate/inadequate, thoughtful/thoughtless

Semi-supervised Induction of Sentiment Lexicons

- step 3: build a polarity graph:
  - nodes: adjectives and candidate adjectives
  - edge weights: ”same polarity”-probabilities of a supervised 2-class(“same polarity”, “different polarity”)-classifier



- step 4: cluster the polarity graph into “positive” and “negative” clusters



Semi-supervised Induction of Sentiment Lexicons

Other ideas / basic ideas

- Using WordNet Synonyms and Antonyms (2004):
  - intuition: word's synonyms probably share its polarity; word's antonyms probably have the opposite polarity.
  - using seed-labels, propagate:

Lex^{+}	: Add synonyms of positive words (*well*) and antonyms (like *fine*) of negative words
Lex^{-}	: Add synonyms of negative words (*awful*) and antonyms (like *evil*) of positive words

- SentiWordNet (2010): extension of this idea:
  - using pos / neg word seeding, select positive and negative synsets;
  - train binary classifier on glosses of labeled pos / neg synsets;
  - combine classifier output with additional random walk steps to yield pos / neg score for synset.

Semi-supervised Induction of Sentiment Lexicons

Other ideas / basic ideas

- SentiWordNet

Synset	Pos	Neg	Obj
good#6 ‘agreeable or pleasing’	1	0	0
respectable#2 honorable#4 good#4 estimable#2 ‘deserving of esteem’	0.75	0	0.25
estimable#3 computable#1 ‘may be computed or estimated’	0	0	1
sting#1 burn#4 bite#2 ‘cause a sharp or stinging pain’	0	0.875	.125
acute#6 ‘of critical importance and consequence’	0.625	0.125	.250
acute#4 ‘of an angle; less than 90 degrees’	0	0	1
acute#1 ‘having or experiencing a rapid onset and short but severe course’	0	0.5	0.5

Examples from SentiWordNet 3.0 (Baccianella et al., 2010). Note the differences between senses of homonymous words: *estimable#3* is purely objective, while *estimable#2* is positive; *acute* can be positive (*acute#6*), negative (*acute#1*), or neutral (*acute #4*)

Supervised Learning of Word Sentiment

- use e.g. 5-star scores web reviews of books, movies etc. as ground truth for their polarity

	Movie review excerpts (IMDB)
10	A great movie. This film is just a wonderful experience. It's surreal, zany, witty and slapstick all at the same time. And terrific performances too.
1	This was probably the worst movie I have ever seen. The story went nowhere even though they could have done some interesting stuff with it.
	Restaurant review excerpts (Yelp)
5	The service was impeccable. The food was cooked and seasoned perfectly... The watermelon was perfectly square ... The grilled octopus was ... mouthwatering...
2	...it took a while to get our waters, we got our entree before our starter, and we never received silverware or napkins until we requested them...
	Book review excerpts (GoodReads)
1	I am going to try and stop being deceived by eye-catching titles. I so wanted to like this book and was so disappointed by it.
5	This book is hilarious. I would recommend it to anyone looking for a satirical read with a romantic twist and a narrator that keeps butting in
	Product review excerpts (Amazon)
5	The lid on this blender though is probably what I like the best about it... enables you to pour into something without even taking the lid off! ... the perfect pitcher! ... works fantastic.
1	I hate this blender... It is nearly impossible to get frozen fruit and ice to turn into a smoothie... You have to add a TON of liquid. I also wish it had a spout ...

Supervised Learning of Word Sentiment

- Potts score of a word w with a rating class c

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

how often does word w occur in reviews with rating c

all the words in all reviews with rating c

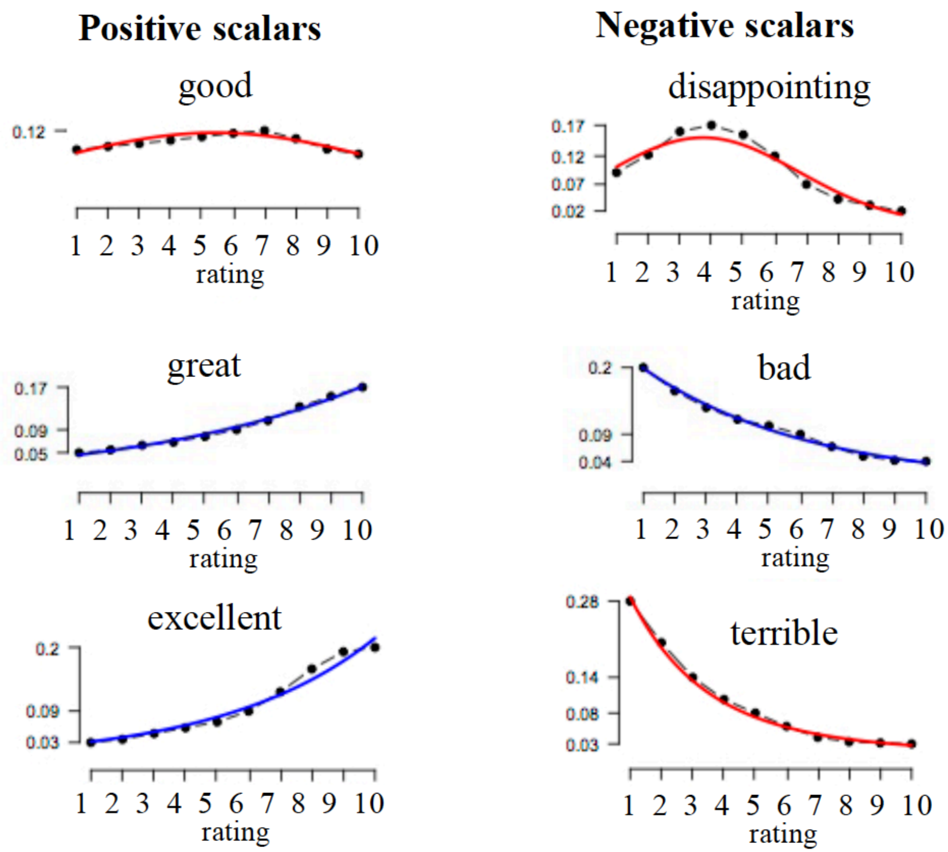
$$PottsScore(w) = \frac{P(w|c)}{\sum_c P(w|c)}$$

$$Potts(disappointing \mid c) = [.10, .12, .14, .14, .13, .11, .08, .06, .06, .05]$$

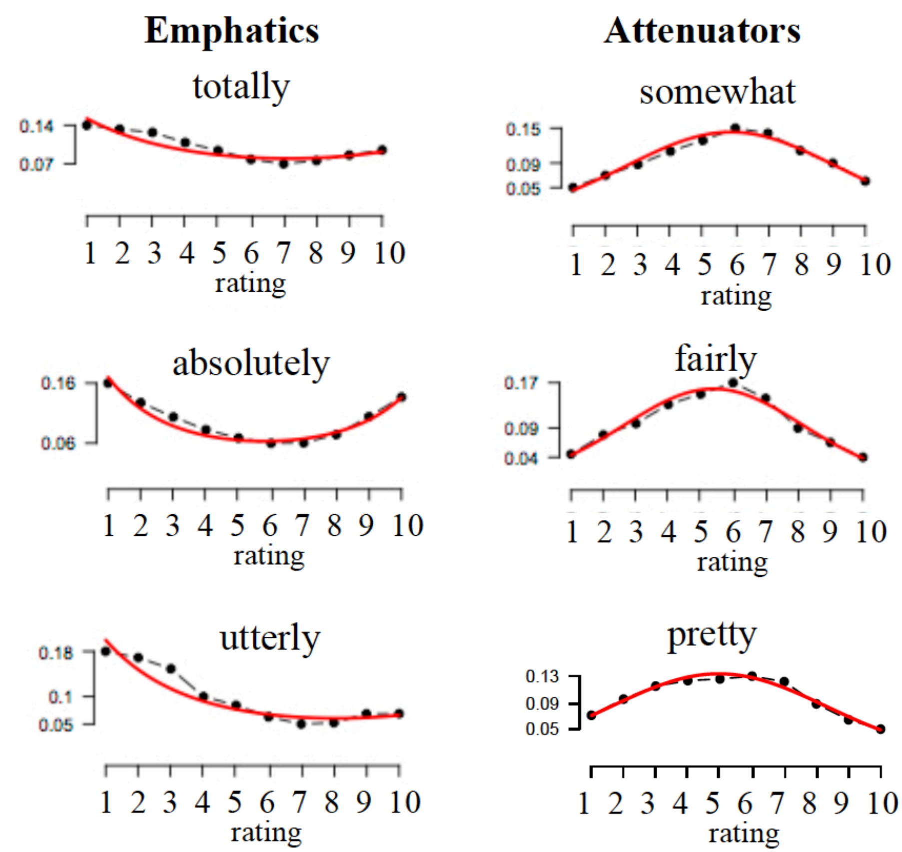
c=1 c=2 c=10

Supervised Learning of Word Sentiment

scalar adjectives: Potts diagrams:



Potts diagrams for emphasizing and attenuating adverbs



Log Odds Ratio Informative Dirichlet Prior

- regard two classes A and B of documents; examples:
  A = 1 star reviews, B = 5 star-reviews
  A = docs written by Republicans, B = docs written by Democrats

- question: does a word w appear more often in A or in B? → compare difference or ratio of frequencies or log odds ratio → does not work well with very frequent or very infrequent words

- example: does word *horrible* appear more often in corpus i or in corpus j → log likelihood ratio

$$
\begin{aligned}
\text{llr}(horrible) &= \log \frac{P^i(horrible)}{P^j(horrible)} \\
&= \log P^i(horrible) - \log P^j(horrible) \\
&= \log \frac{\text{f}^i(horrible)}{n^i} - \log \frac{\text{f}^j(horrible)}{n^j}
\end{aligned}
$$

n^i: size of corpus i (class A)
n^j: size of corpus j (class B)
f^i (w): count of w in corpus i
f^j (w):: count of w in corpus j

Log Odds Ratio Informative Dirichlet Prior

- log odds ratio

$$
\begin{aligned}
\operatorname{lor}(horrible) &= \log\left(\frac{P^i(horrible)}{1-P^i(horrible)}\right) - \log\left(\frac{P^j(horrible)}{1-P^j(horrible)}\right) \\
&= \log\left(\frac{\frac{\mathrm{f}^i(horrible)}{n^i}}{1-\frac{\mathrm{f}^i(horrible)}{n^i}}\right) - \log\left(\frac{\frac{\mathrm{f}^j(horrible)}{n^j}}{1-\frac{\mathrm{f}^j(horrible)}{n^j}}\right) \\
&= \log\left(\frac{\mathrm{f}^i(horrible)}{n^i-\mathrm{f}^i(horrible)}\right) - \log\left(\frac{\mathrm{f}^j(horrible)}{n^j-\mathrm{f}^j(horrible)}\right)
\end{aligned}
$$

n^i: size of corpus i (class A)
n^j: size of corpus j (class B)
f^i (w): count of w in corpus i
f^j (w):: count of w in corpus j

- now: Dirichlet intuition: use a large background corpus to get a prior estimate of what we expect the frequency of each word w to be:
  do this very simply by smoothig: adding the counts from that corpus to the numerator and denominator ←→ we're essentially shrinking the counts toward that prior.
  ←→ how large are the differences between i and j given what we would expect given their frequencies in a well-estimated large background corpus.

Log Odds Ratio Informative Dirichlet Prior

- Dirichlet intution →

$$\delta_w^{(i-j)} = \log\left(\frac{f_w^i + \alpha_w}{n^i + \alpha_0 - (f_w^i + \alpha_w)}\right) - \log\left(\frac{f_w^j + \alpha_w}{n^j + \alpha_0 - (f_w^j + \alpha_w)}\right)$$

n^i: size of corpus i (class A)
n^j: size of corpus j (class B)
$f^i{}_w$: count of w in corpus i
$f^j{}_w$: count of w in corpus j
a_0: size of background corpus
a_w: count of w in background corpus

- together with estimate for variance (←→ scale for this log odds ratio (= difference))

$$\sigma^2\left(\hat{\delta}_w^{(i-j)}\right) \approx \frac{1}{f_w^i + \alpha_w} + \frac{1}{f_w^j + \alpha_w}$$

we get final z-score:

$$\frac{\hat{\delta}_w^{(i-j)}}{\sqrt{\sigma^2\left(\hat{\delta}_w^{(i-j)}\right)}}$$

Log Odds Ratio Informative Dirichlet Prior

Class	Words in 1-star reviews	Class	Words in 5-star reviews
Negative	*worst, rude, terrible, horrible, bad, awful, disgusting, bland, tasteless, gross, mediocre, overpriced, worse, poor*	**Positive**	*great, best, love(d), delicious, amazing, favorite, perfect, excellent, awesome, friendly, fantastic, fresh, wonderful, incredible, sweet, yum(my)*
Negation	*no, not*	**Emphatics/ universals**	*very, highly, perfectly, definitely, absolutely, everything, every, always*
1Pl pro	*we, us, our*	**2 pro**	*you*
3 pro	*she, he, her, him*	**Articles**	*a, the*
Past verb	*was, were, asked, told, said, did, charged, waited, left, took*	**Advice**	*try, recommend*
Sequencers	*after, then*	**Conjunct**	*also, as, well, with, and*
Nouns	*manager, waitress, waiter, customer, customers, attitude, waste, poisoning, money, bill, minutes*	**Nouns**	*atmosphere, dessert, chocolate, wine, course, menu*
Irrealis modals	*would, should*	**Auxiliaries**	*is/'s, can, 've, are*
Comp	*to, that*	**Prep, other**	*in, of, die, city, mouth*

Using Lexicons for Sentiment Recognition

- Ch.6: sentiment lexicon: list each word as either positive or negative or give a positiveness score Θ_w^+ and a negativeness score Θ_w^-

- use lexicon in a simple "rule-based" way:

$$f^+ = \sum_{w \text{ s.t. } w \in positivelexicon} \theta_w^+ count(w)$$

$$f^- = \sum_{w \text{ s.t. } w \in negativelexicon} \theta_w^- count(w)$$

$$sentiment = \begin{cases} + & \text{if } \frac{f^+}{f^-} > \lambda \\ - & \text{if } \frac{f^-}{f^+} > \lambda \\ 0 & \text{otherwise.} \end{cases}$$

- or use lexicon assessment for words as features for supervised sentiment classifier

Personality

- standard questionnaires for Big 5 personality dimensions:
  - Extroversion vs. Introversion:
    sociable, assertive, playful vs. aloof, reserved, shy
  - Emotional stability vs. Neuroticism:
    calm, unemotional vs. insecure, anxious
  - Agreeableness vs. Disagreeableness:
    friendly, cooperative vs. antagonistic, fault finding
  - Conscientiousness vs. Unconscientiousness:
    self-disciplined, organized vs. inefficient, careless
  - Openness to experience:
    intellectual, insightful vs. shallow, unimaginative

- essay corpus of Pennebaker & King (1999): 2479 essays (1.9 million words) from psychology students who took a standard personality test were asked to “write whatever comes into your mind” for 20 minutes.

Personality

- sample from neurotic individual:

  > 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.

- sample from emotionally stable individual:

  > 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.

- speed dating corpus (Ranganath et al) (2013): mutual rating: interpersonal stance of other participants

Supervised Affect Recognition

- affect recognition: generalizing algorithms described before for detecting sentiment.

- supervised document classification problem: useful features + SVM or comparable classifier = very good results

- example: Schwartz (2013) Facebook dataset ($3*10^8$ words, 75000 users): personality, gender, age;
  features: unigrams, bigrams, trigrams with sufficiently high PMI

$$\mathrm{pmi}(phrase) = \log \frac{p(phrase)}{\prod_{w \in phrase} p(w)}$$

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

subject == person on Facebook

Affect Recognition

- if lexicon is available, use class based features:

$$f_{\mathscr{L}}(c,x) \;=\; \begin{cases} 1 & \text{if } \exists w : w \in \mathscr{L} \;\&\; w \in x \;\&\; \; class = c \\ 0 & \text{otherwise} \end{cases}$$

- or just raw or log'd counts of word-tokens in respective (part of) lexicon associated with a class

$$f_{\mathscr{L}} = \sum_{w \in \mathscr{L}} count(w)$$

if lexicon has membership degree weights, use weighted counts:

$$f_{\mathscr{L}} = \sum_{w \in \mathscr{L}} \theta_w^{\mathscr{L}} count(w)$$

Affect Recognition

- example result: use logistic regression and investigate weight vector : which features are positively associated with a class, which are negatively associated:
  - LWIC lexicons *Family*, *Home* positively associated and *anger* and *swear* negatively correlated with personality dimension class *Agreeable*
  - Schwartz' Facebook dataset: Extroversion – Introversion regression: associated words + phrases

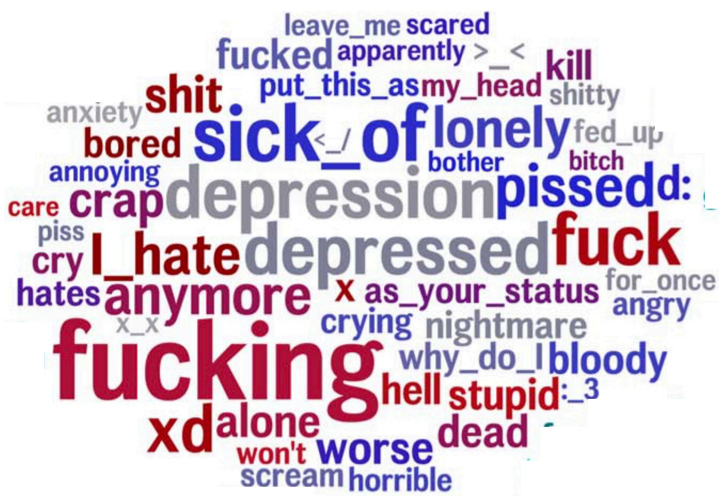
words & phrases positively associated with class Introversion

words & phrases positively associated with class Extroversion

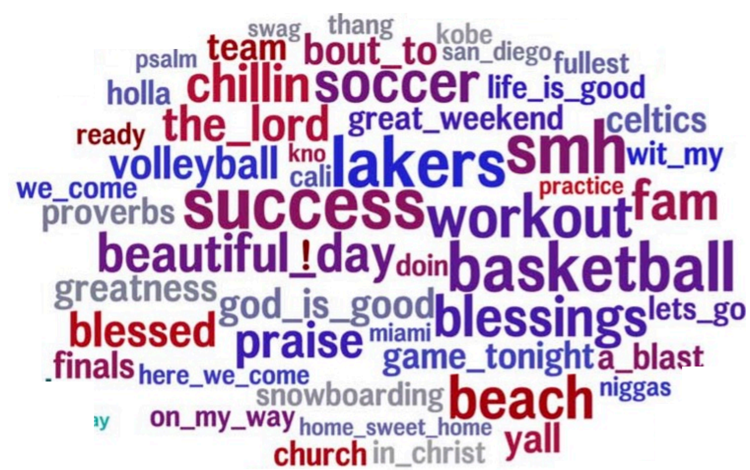
Affect Recognition

- Schwartz’ Facebook dataset: Neuroticism – Emotional Stability regression: associated words + phrases

words & phrases positively associated with class Neuroticism



words & phrases positively associated with class Emotionally Stable



VAD Recognition using Embeddings & Regression

- for a word w: compute embedding: either
  - use static embedding (e.g. Word2Vec)
  - or average of dynamic embeddings (ELMo or BERT)
- look up Valence Arousal and Dominance values for w in NRC VAD lexicon and compute three linear regressions using embedding as x and V, A, and D values as y values.

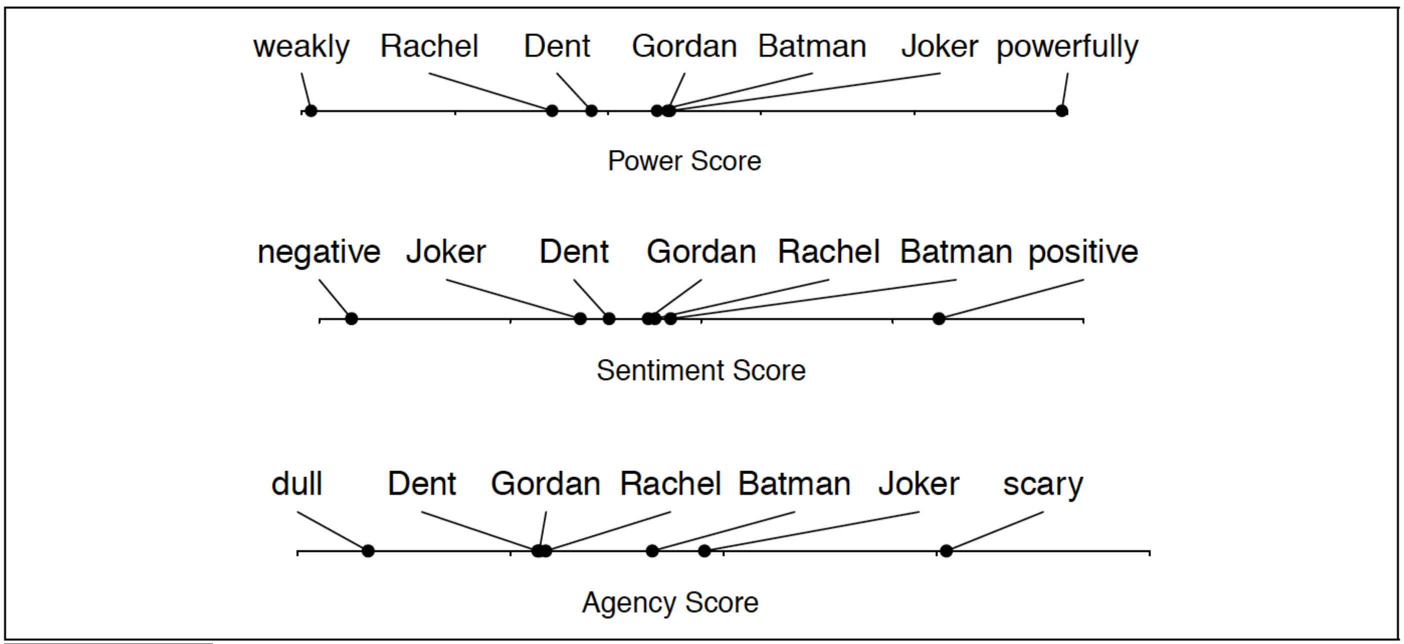


Figure 21.14 Power (dominance), sentiment (valence) and agency (arousal) for characters in the movie *The Dark Knight* computed from ELMo embeddings trained on the NRC VAD

Bibliography

(1) Dan Jurafsky and James Martin: Speech and Language Processing (3rd ed. draft, version Oct 2019); Online: https://web.stanford.edu/~jurafsky/slp3/ (URL, Oct 2019) (this slideset is especially based on chapter 21)

(2) Powerpoint slides from Dan Jurafsky and James Martin: Speech and Language Processing (3rd ed. draft); Online: https://web.stanford.edu/~jurafsky/slp3/ (URL, May 2018)

Recommendations for Studying

- minimal approach:

  work with the slides and understand their contents! Think beyond instead of merely memorizing the contents

- standard approach:

  minimal approach + read the corresponding pages in Jurafsky [1]

- interested students

  == standard approach