Natural Language Processing
CS 6320

Lecture 16

Interpreting Affect in Texts

Instructor: Sanda Harabagiu

What is Affective Computing

Affective computing (Picard, 1995) is used to recognize in text, dialogs:

- *emotion,*
- *sentiment,*
- *personality,*
- *mood, and*
- *attitudes.*

Affective meaning is closely related to subjectivity, the study of a speaker or writer's evaluations, opinions, emotions, and speculations (Wiebe et al., 1999).

Affective meaning

- *Drawing on literature 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*

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*

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

How do we represent emotions?

- *There are two widely-held families of theories of emotion!*
- *In one family, emotions are viewed as fixed atomic units, limited in number, and from which others basic emotions are generated, often called basic emotions (Tomkins 1962, Plutchik 1962). Perhaps most well-known of this family of theories are the 6 emotions proposed by (Ekman, 1999) as a set of emotions that is likely to be universally present in all cultures:*
    - *surprise,*
    - *happiness,*
    - *anger,*
    - *fear,*
    - *disgust,*
    - *sadness.*

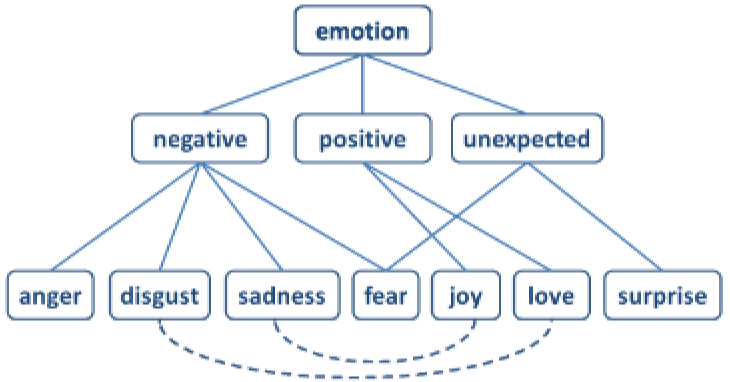


Figure 1: Our emotion ontology for the six Ekman emotions (plus LOVE). Solid lines indicate inheritance, dashed lines indicate opposite.

Ekman’s 6 basic emotions

Ekman’s 6 basic emotions:

Surprise, happiness, anger, fear, disgust, sadness

How do we represent emotions?

- *Another atomic theory is the (Plutchik, 1980) wheel of emotion, consisting of 8 basic emotions in four opposing pairs:*
  - *joy–sadness,*
  - *anger–fear,*
  - *trust–disgust, and*
  - *anticipation–surprise,*

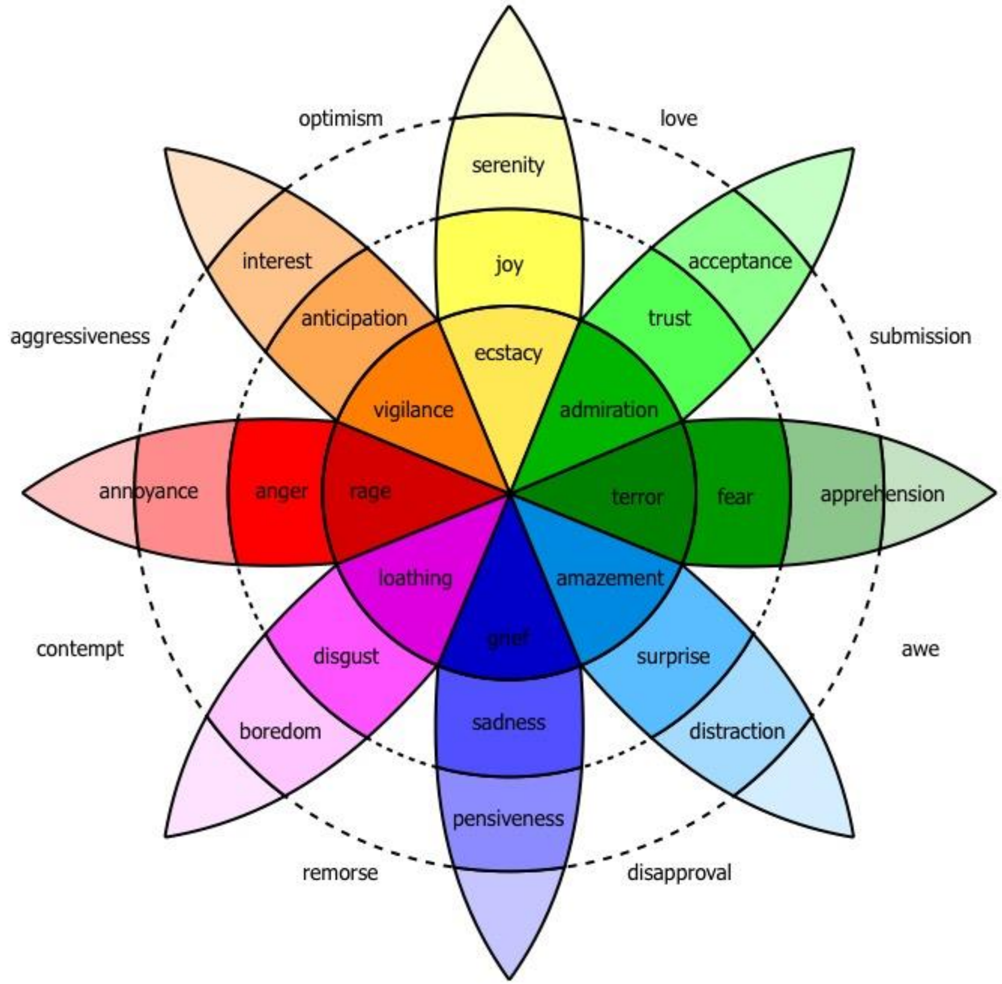
together with the emotions derived from them!

Plutchick's wheel of emotion

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

The wheel of emotions considers dimensions of emotion:

- *Valence (positive negative)*
- *Arousal (strong, weak)*
- *Control*

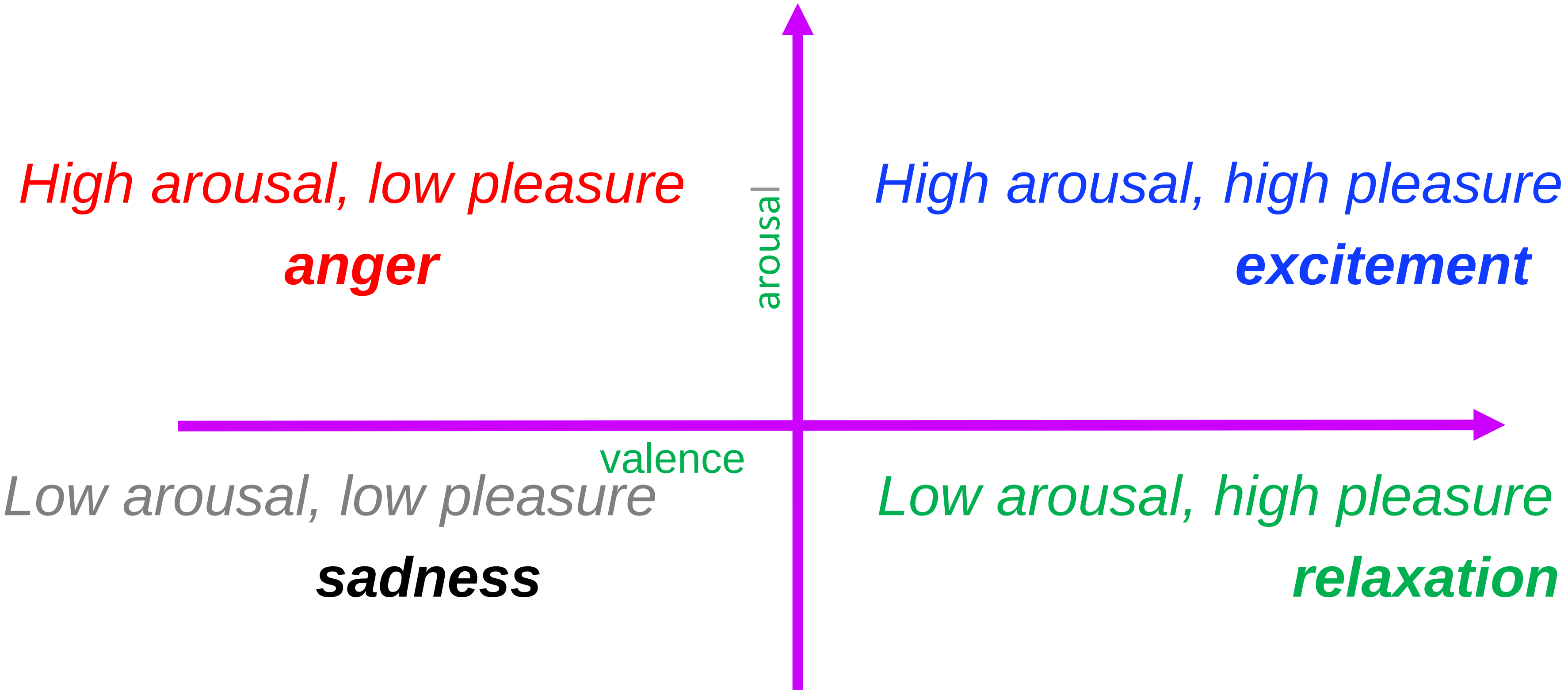


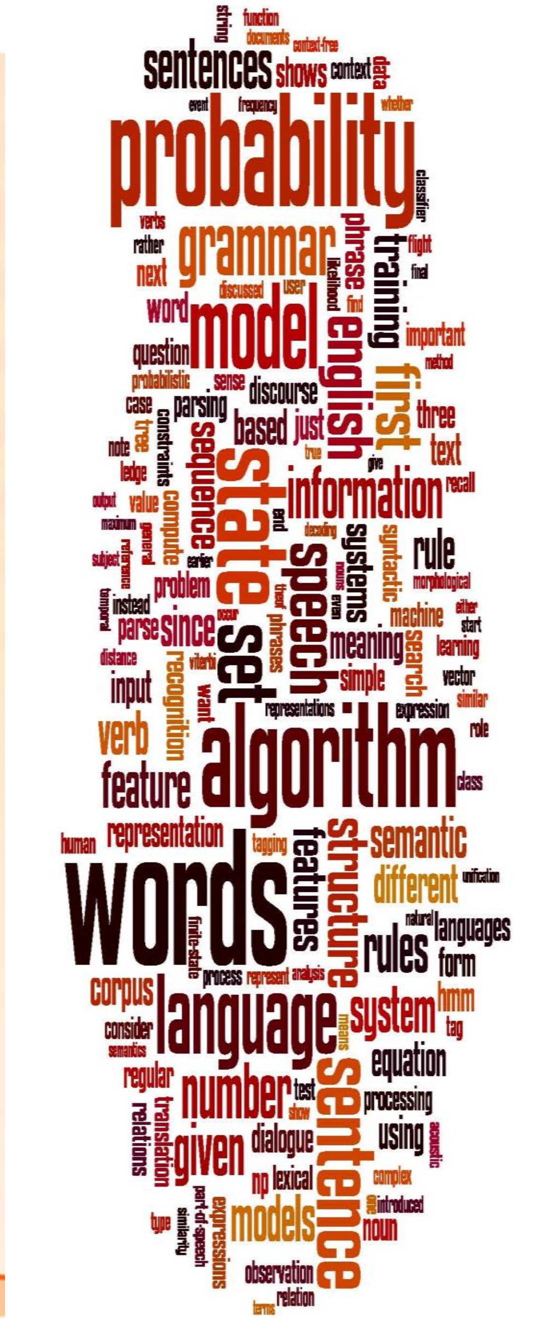
Valence/Arousal/Dominance Dimensions

valence: *the pleasantness of the stimulus*

arousal: *the intensity of emotion provoked by the stimulus*

dominance*: the degree of control exerted by the stimulus*





Computing with Affective Lexicons

Sentiment &

Emotion Lexicons

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

- *Originally, the General Inquirer technique relied on the Harvard psychological dictionaries that were correlated with <u>states</u>, <u>motives</u>, <u>social and cultural roles</u> as well as different <u>aspects of general distress</u>.*

The current version of the General Inquirer also contains lexical categories, and hundreds of different semantic categories!

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*

The General Inquirer

- ***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*

Category	Description	Category	Description
Academ	academic, intellectual, or educational matters	Our	self-inclusive pronouns
AffGain	positive words of love/friendship (e.g., love, date)	Pain	suffering, lack of confidence, or commitment
AffOth	other words of love/friendship (baby, brotherhood)	Persist	endurance
AffPt	affection participant (brother, mother)	Pleasur	enjoyment
ANI	animals	Polit@	political roles (adversary, cabinet)
Aquatic	bodies of water	PowAren	political places
BldgPt	buildings or parts of buildings	PowAuPt	authoritative participants
BodyPt	body parts	PowCon	power conflict (aggression, discord)
COLOR	colors	PowCoop	power cooperation (affiliate, negotiation)
Complet	goal completion	PowDoct	power doctrine (communism, elitism)
Decreas	decrease (cheapen, decay)	PowEnds	goals of the power process
DIST	distance measures	PowPt	power ordinary participants (civilian, follower)
EMOT	emotions	Quality	degrees of quality
EnlEnds	pursuit of enlightenment (contemplate, discover)	Race	racial or ethnic characteristics
EnlLoss	misguided (delude, distract)	RcGain	rectitude gain (worship, forgiveness)
EnlOth	other enlightenment words	RcLoss	rectitude loss (convict, denounce)
EnlPt	enlightenment participant (faculty, historian)	RcRelig	religion (awe, believer)
Exch	buying or selling	Region	general regions (kingdom, downtown)
Exert	exertion	Relig	religious matters (angel, bishop)
Exprsv	arts, sports, or self expression	Rise	rising (ascent, jump)
Fall	falling (sing, tumble)	Ritual	social rituals (baseball, birthday)
Feel	feelings (gratitude, apathy)	Role	social roles (actor, colleague)
Female	women and their social roles	RspLoss	losing of respect
Food	food and beverage	Say	say and tell
FREQ	frequency or recurrence	SklAsth	skill aesthetic (beautiful, poetic)
Goal	end states for mental or physical effort	SklOth	other skill words (adept, blunder)
Intrj	interjections	SklPt	skill participant (baker, carpenter)
IPadj	relations between people (unkind, aloof)	Sky	aerial or outer-space conditions (haze, rain, sun)
Kin@	kinship	Think	rational thought process
Know	awareness, certainty, similarity and antonyms	TIME	temporal (afternoon, decade)
Land	natural places (desert, beach)	Vehicle	vehicle (jet, limousine)
MALE	men and their social roles	WlbGain	gain in well being (comfort, feed)
Milit	military matters	WlbPhys	physical aspects of well being (bone, cancer)
Name	demonyms (Cuban, African)	WlbPsyc	psychological aspects of well being (anger, cry)
Nation	country names and demonyms	WlbPt	well being participant (nurse, baby)
Nonadlt	infants/adolescents	WlbTot	all well being words
ORD	ordinal words	WltTran	wealth transaction (import, mortgage)
Ought	moral imperative	You	pronouns for another person

Automatically Detecting Emotions

Detecting emotion has the potential to improve a number of language processing tasks.

Examples:

1. *Automatically detecting emotions in reviews or customer responses (anger, dissatisfaction, trust) could help businesses recognize specific problem areas or ones that are going well.*
2. *Emotion recognition could help dialog systems like tutoring systems detect that a student was unhappy, bored, hesitant, confident, and so on.*
3. *Emotion can play a role in medical informatics tasks like detecting depression or suicidal intent.*
4. *Detecting emotions expressed toward characters in novels might play a role in understanding how different social groups were viewed by society at different times.*

Creating affect lexicons by human labeling

- *A way to build affect lexicons is to have crowdsourcing humans label each word. This is now most commonly done by a large number of annotators.*

- *In 2012, we created Empatweet:*

"EmpaTweet: Annotating and Detecting Emotions on Twitter"

http://www.lrec-conf.org/proceedings/lrec2012/pdf/201_Paper.pdf

- *Creation of a corpus of tweets on a variety of popular Twitter topics with their corresponding manually-annotated emotions.*

Topics were chosen based on our expectation of which emotions will be present in topical tweets in order to get a good distribution of our chosen emotions.

Topic	Hashtags
Valentine's Day	#valentine #valentines #valentinesday #cupid
Lindsay Lohan	#lohan #lindsaylohan
September 11th	#nineeleven #sept11 #september11 #nine11 #9eleven
2012 U.S. Election	#obama #romney #ronpaul #gingrich #gop #gopdebate #republicandebate #teaparty
Palestinian Statehood	#palestine #palestinestate #palestinestatehood #palestineun #gopalestine #freepalestine
Egyptian riots	#arabspring #tahir #tahrir #egyptianrevolution #egypt
Super Bowl XLV	#superbowlxlv #superbowl
World Cup 2010	#worldcup2010 #wc2010 #worldcup
Christmas	#christmas #xmas #santa #happyholidays
DC/NY earthquake	#earthquake #dcearthquake #eastcoastearthquake
Emmys	#emmy #emmys #emmyaward #emmyawards
Eminem	#eminem #eminemsong
stock market	#stocks #stockmarket #dow #dowjones #sandp #nasdaq #wallstreet #NYSE
Greek bailout	#bailout #greece #greekbailout #eurocrisis #euro

Distribution of annotated Emotions

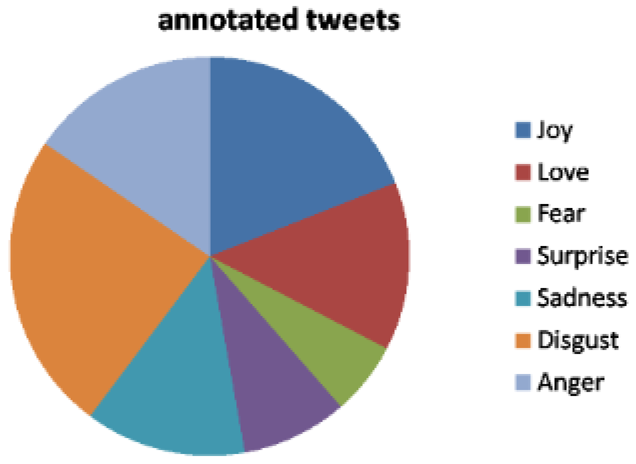


Figure 2: Emotion distribution in annotated Twitter corpus.

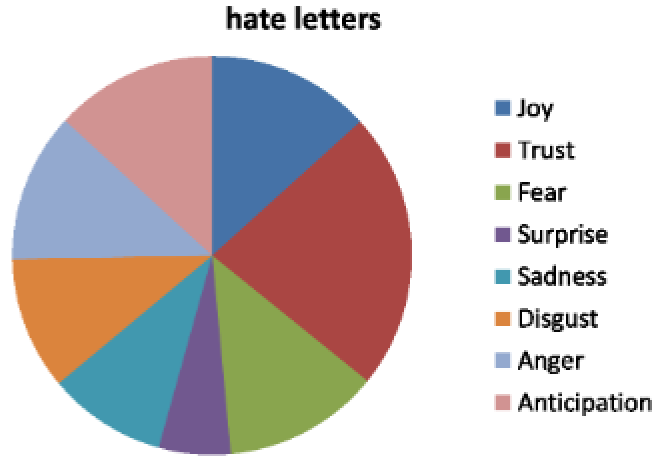


Figure 4: Emotion distribution in hate letters.

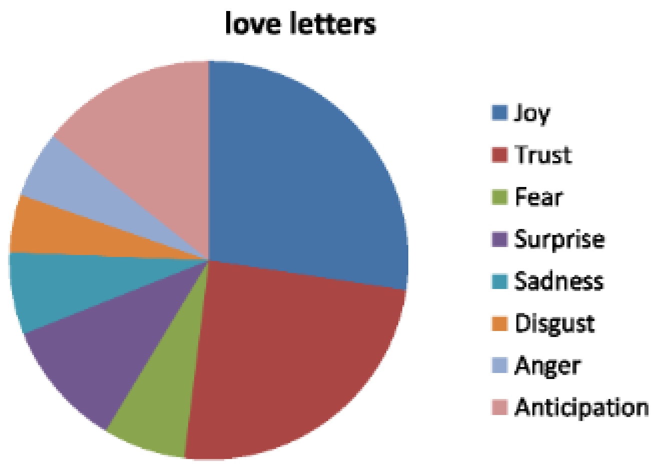


Figure 3: Emotion distribution in love letters.

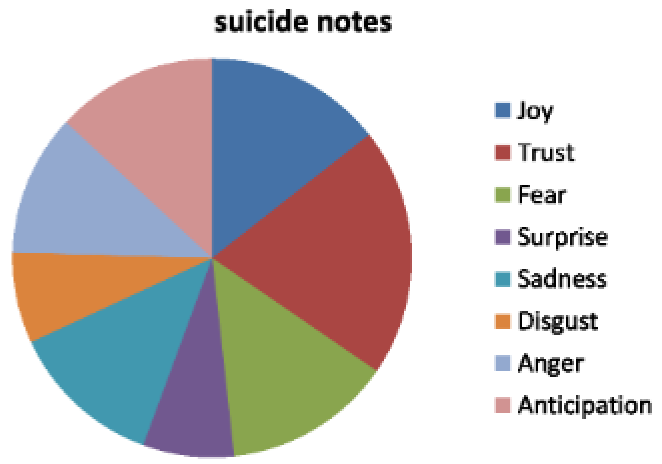


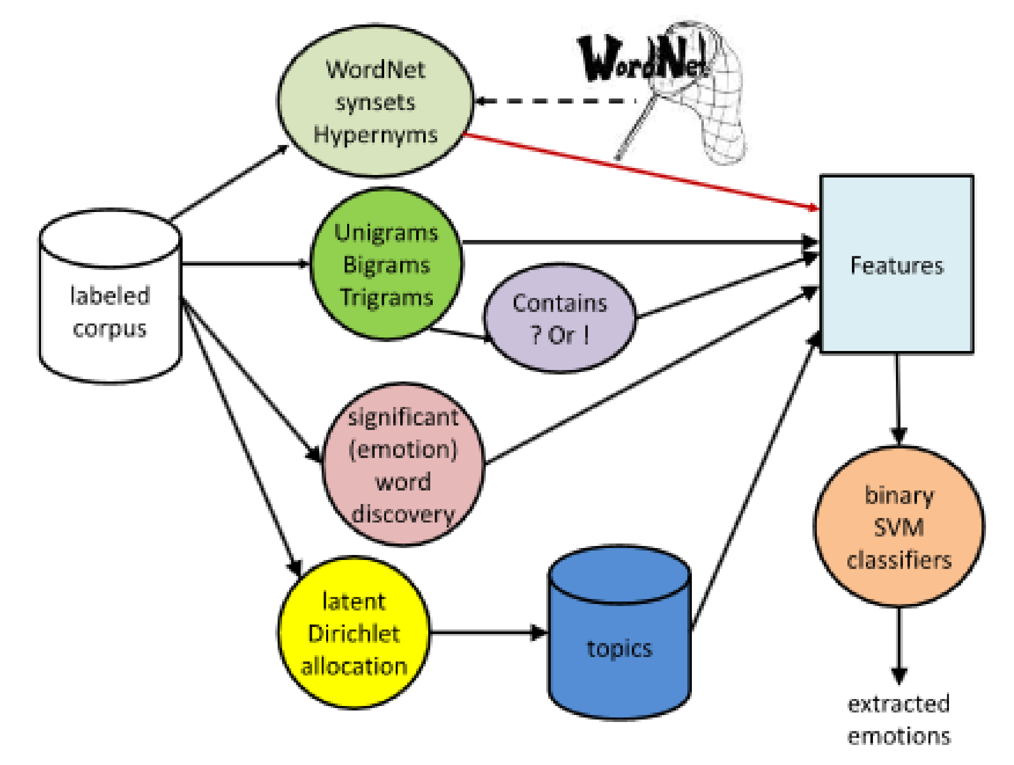
Figure 5: Emotion distribution in suicide notes.

Automatic detection of Emotions

- *Generate a series of binary SVM classifiers to detect each of the seven emotions. Each classifier performs independently on a single emotion, using the software available from WEKA (Hall et al., 2009).*

The combination of these separate Classifiers can considered a single multi-label classifier, allowing for a tweet to be annotated with more than one of the emotions.

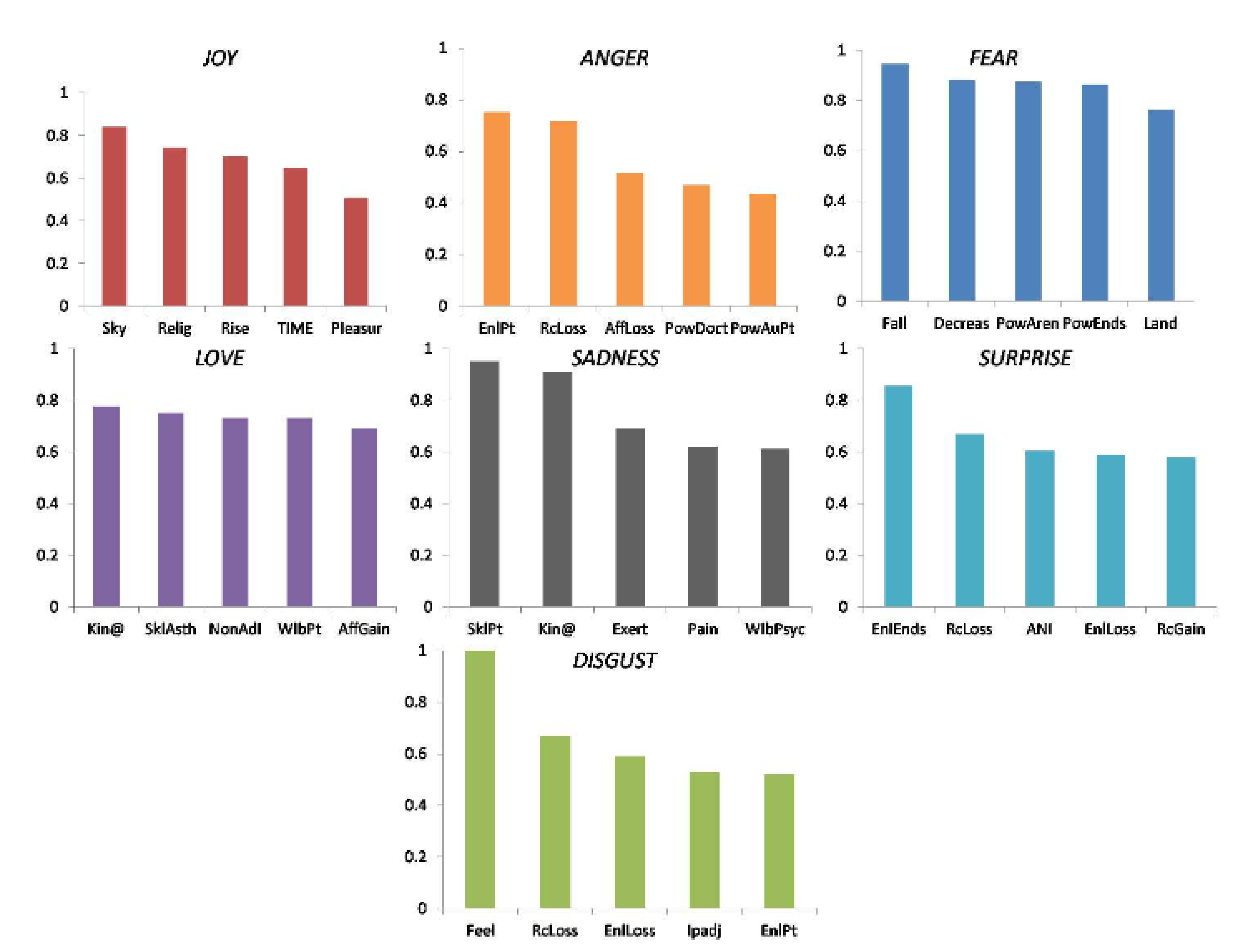
WordNet Affect features were omitted.



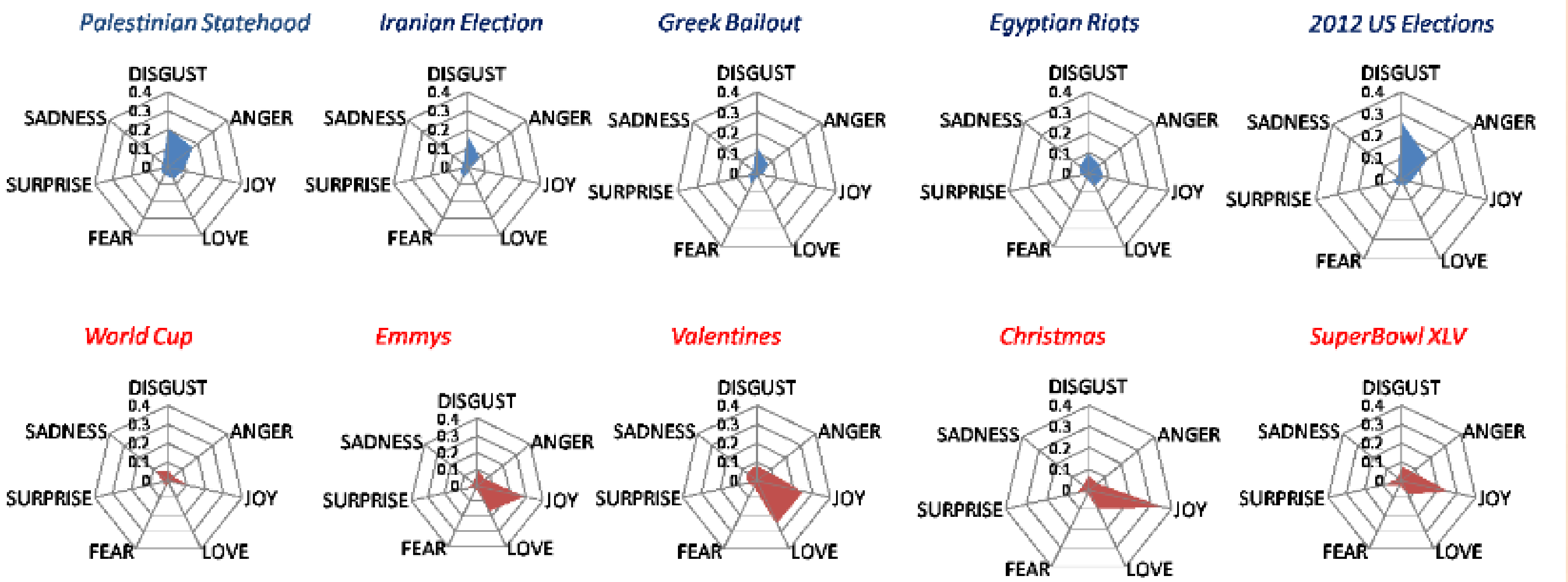
Features used for Emotion Classification

- *Unigrams: after filtering.*
- *Bigrams*
- *Trigrams*
- *Contains !: A flag indicating the original tweet has an exclamation mark.*
- *Contains ?*
- *WordNet synsets: No word sense disambiguation is performed. Rather, all synsets for each word in Word-Net is considered.*
- *WordNet hypernyms: All (recursive) hypernyms foreach synset.*
- *Topic scores: The scores for each LDA topic (we use100 topics).*
- *Significant words: Unigrams judged to have a high pointwise mutual information (PMI) with at least one emotion in the training data..*

Most significant General Inquirer categories for each emotion



Emotion Distribution in Twitter Corpus



More Emotion Lexicons

- *The NRC Word-Emotion Association Lexicon, also called EmoLex (Mohammad and Turney, 2013), uses the Plutchik (1980) 8 basic emotions. The lexicon includes around **14,000 words** including words from prior lexicons as well as frequent nouns, verbs, adverbs and adjectives. Values from the lexicon for some sample words:*

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

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*

LIWC, Linguistic Inquiry and Word Count, is another set of **73 lexicons** containing over 2300 words (Pennebaker et al., 2007), designed to capture aspects of lexical meaning relevant for social psychological tasks.

In addition to sentiment-related lexicons like ones for negative emotion (bad, weird, hate, problem, tough) and positive emotion (love, nice, sweet), LIWC includes lexicons for categories like anger, sadness, cognitive mechanisms, perception, tentative, and inhibition.

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

Figure 19.5 Samples from 5 of the 73 lexical categories in LIWC (Pennebaker et al., 2007). The * means the previous letters are a word prefix and all words with that prefix are included in the category.

Linguistic Style of Expression of Emotion

- *In "Linguistic Styles: Language Use as Individual Difference", the authors note "that people differ in the ways they talk and write is hardly a novel observation" (Pennebaker and King, 1999). Moreover, linguistic fingerprinting has often been supported by psychological studies. We extend these observations to the style of expressing emotions in writing as well. One way of measuring linguistic style is provided by the Linguistic Inquiry Word Count (LIWC) method.*

- *Linguistic Fingerprinting of topics on Twitter*

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

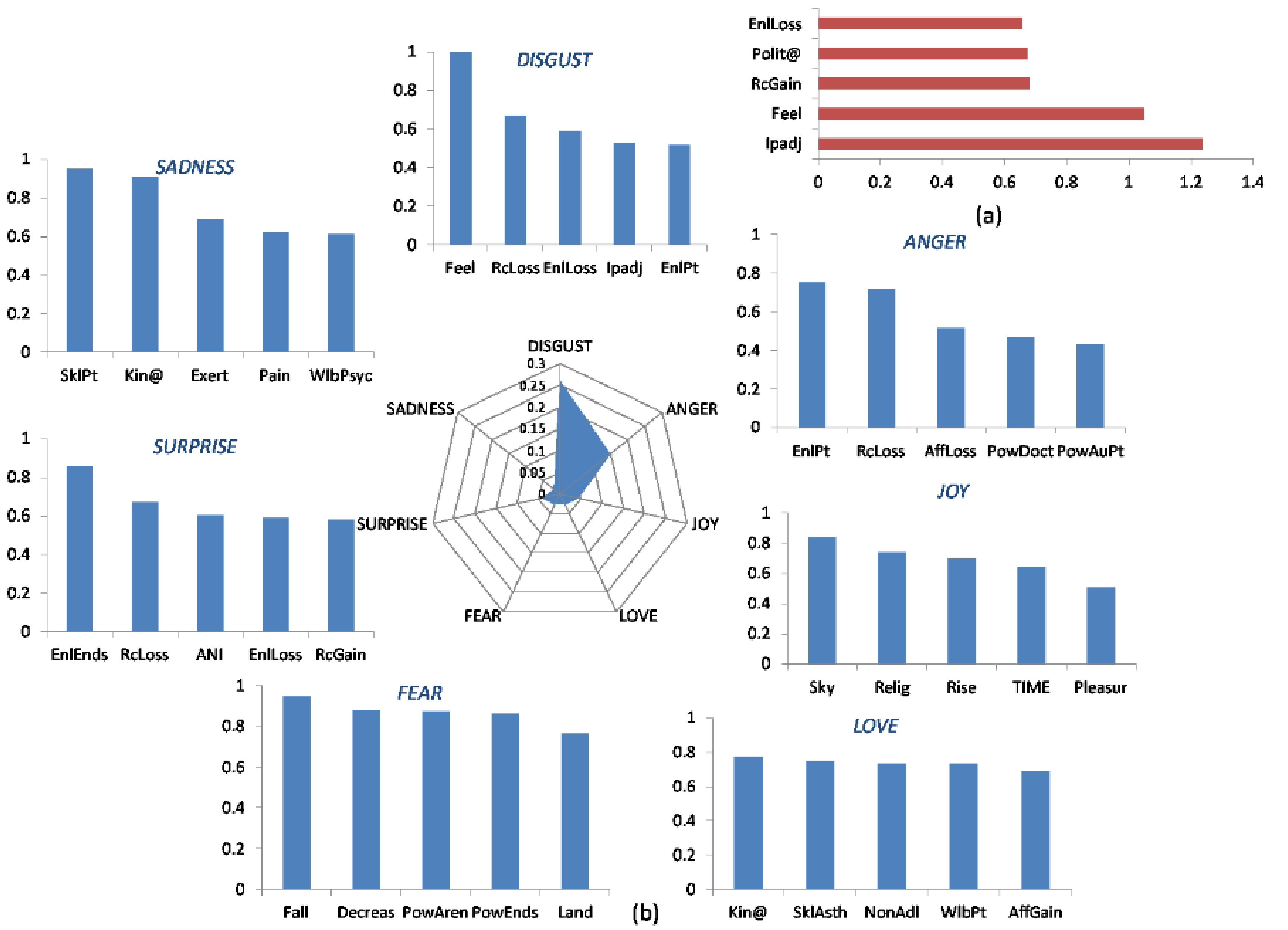


Figure 9: (a) Linguistic style features for the topic "2012 U.S. Election". (b) Emotion make-up and corresponding linguistic style semantic categories for the topic "2012 U.S. Election".

Lexicon of valence, arousal and dominance

For the Warriner et al. (2013) ***lexicon of valence, arousal, and dominance****,*
crowdworkers marked each word with a value from 1-9 on each of the dimensions, with the scale defined for them as follows:

- *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)*

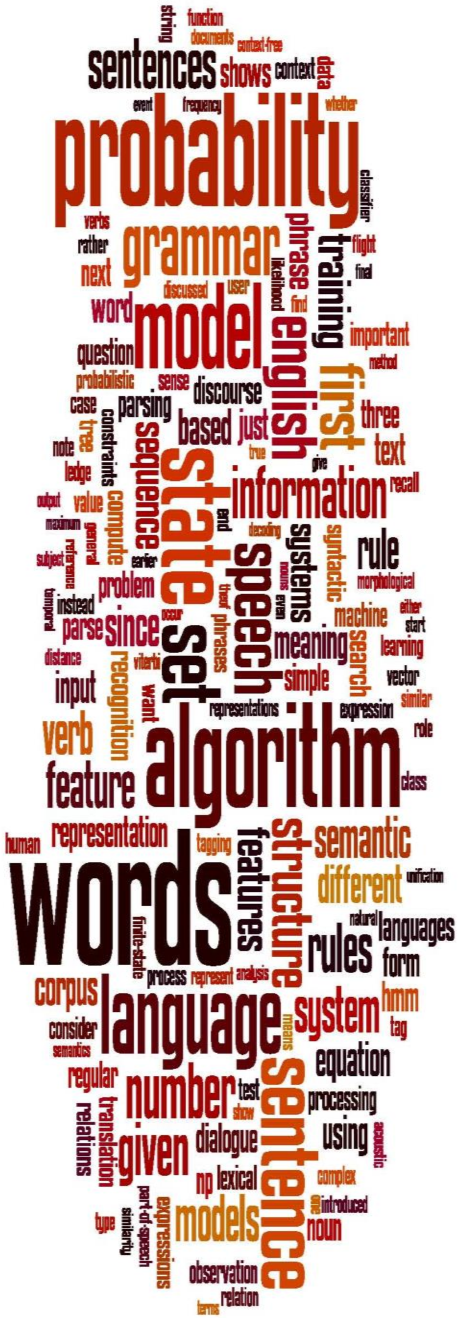
9: stimulated, excited, frenzied, jittery, wide-awake, or aroused
1: relaxed, calm, sluggish, dull, sleepy, or unaroused;

- *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

Lexicons of Valence, arousal and dominance

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



Computing with Affective Lexicons

Semi-supervised algorithms for learning sentiment Lexicons

1. The semantic Axis Method
2. The *Hatzivassiloglou and McKeown* algorithm
3. SentiWordNet
4. The SentProp Algorithm

Semantic axis methods

- ❑ *One of the most well-known lexicon induction methods, the Turney and Littman (2003) algorithm, is given seed words like **good** or **bad,** and then for each word w to be labeled, measures both how similar it is to good and how different it is from bad.*
- ➢ *A slight extension of the algorithm due to An et al. (2018), which is based on computing a semantic axis.*

First step: choose *seed words by hand.*

Because the sentiment or affect of a word is different in different contexts, it's common to choose different seed words for different genres, and most algorithms are quite sensitive to the choice of seeds.

Seed Lexicon

For example, for inducing sentiment lexicons, Hamilton et al. (2016) defines:

- *one set of seed words for general sentiment analysis,*
- *a different set for Twitter, and*
- *yet another set for learning a lexicon for sentiment in financial text:*

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

Semantic Axis: Step 2

In the second step, we compute embeddings for each of the pole words.

- *These embeddings can be off-the-shelf word2vec embeddings, or can be computed directly on a specific corpus (for example using a financial corpus if a finance lexicon is the goal), or we can fine-tune off-the-shelf embeddings to a corpus.*
- *Fine-tuning is especially important if we have a very specific genre of text but don't have enough data to train good embeddings. In fine-tuning, we begin with off-the-shelf embeddings like word2vec, and continue training them on the small target corpus. Once we have embeddings for each pole word, we create an embedding that represents each pole by taking the centroid of the embeddings of each of the seed words.*

- *Given a set of embeddings for the positive seed words:*

$S^+ = \{ E(w_1^+), E(w_2^+), E(w_n^+) \}$, and embeddings for the negative seed words $S^- = \{ E(w_1^-), E(w_2^-), E(w_n^-) \}$, the pole centroids are:

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

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

The semantic axis defined by the poles

- *The semantic axis defined by the poles is computed just by subtracting the two vectors:*

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

$\boldsymbol{V}_{axis}$, *the semantic axis, is a vector in the direction of sentiment.*

In Step 3, we compute how close each word w is to this sentiment axis, by taking the cosine between w's embedding and the axis vector. A higher cosine means that w is more aligned with S^{+} than S^{-}

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

If a dictionary of words with sentiment scores is sufficient, we're done! Or if we need to group words into a positive and a negative lexicon, we can use a threshold or other method to give us discrete lexicons

Hatzivassiloglou and McKeown intuition for identifying word polarity

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

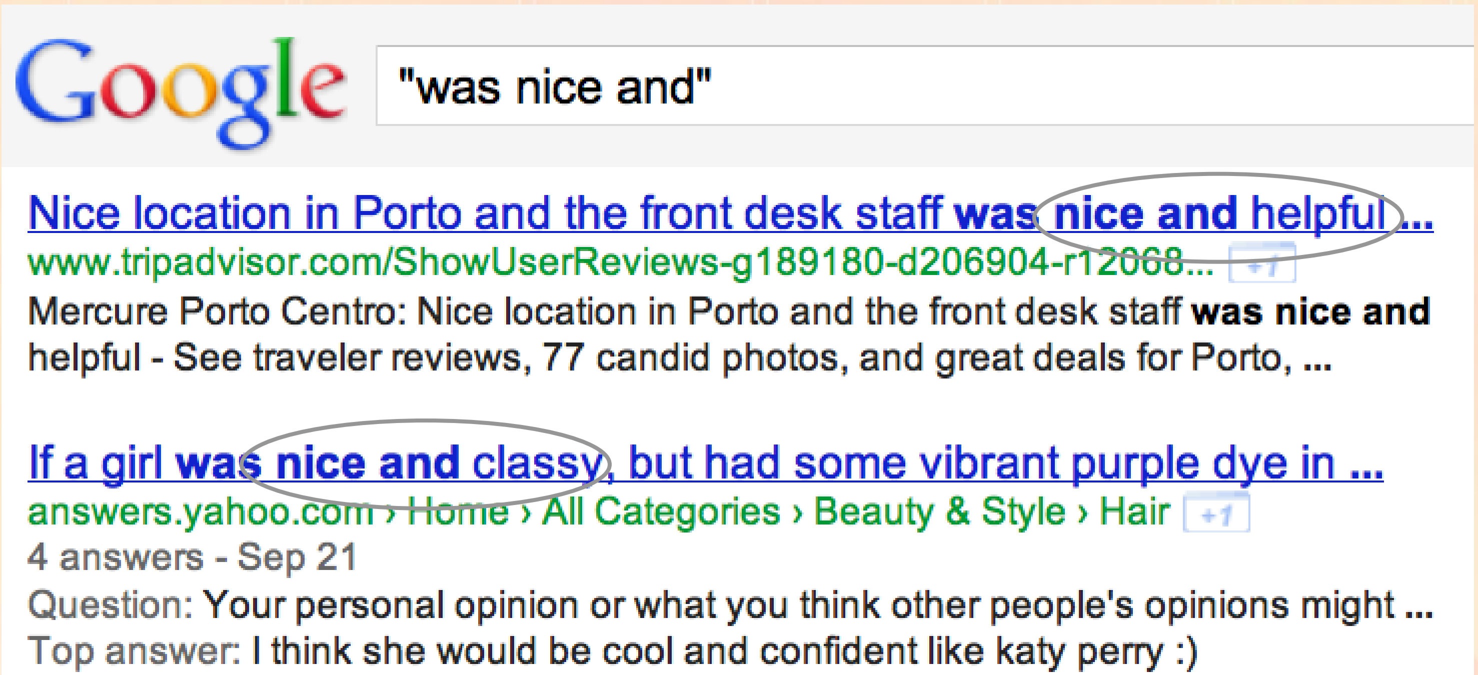
- *Adjectives conjoined by "and" have same polarity*
  - *Fair* ***and*** *legitimate, corrupt* ***and*** *brutal*
  - **fair* ***and*** *brutal, *corrupt* ***and*** *legitimate*
- *Adjectives conjoined by "but" do not*
  - *fair* ***but*** *brutal*

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…*

Hatzivassiloglou & McKeown 1997 Step 2

- *Expand seed set to conjoined adjectives*

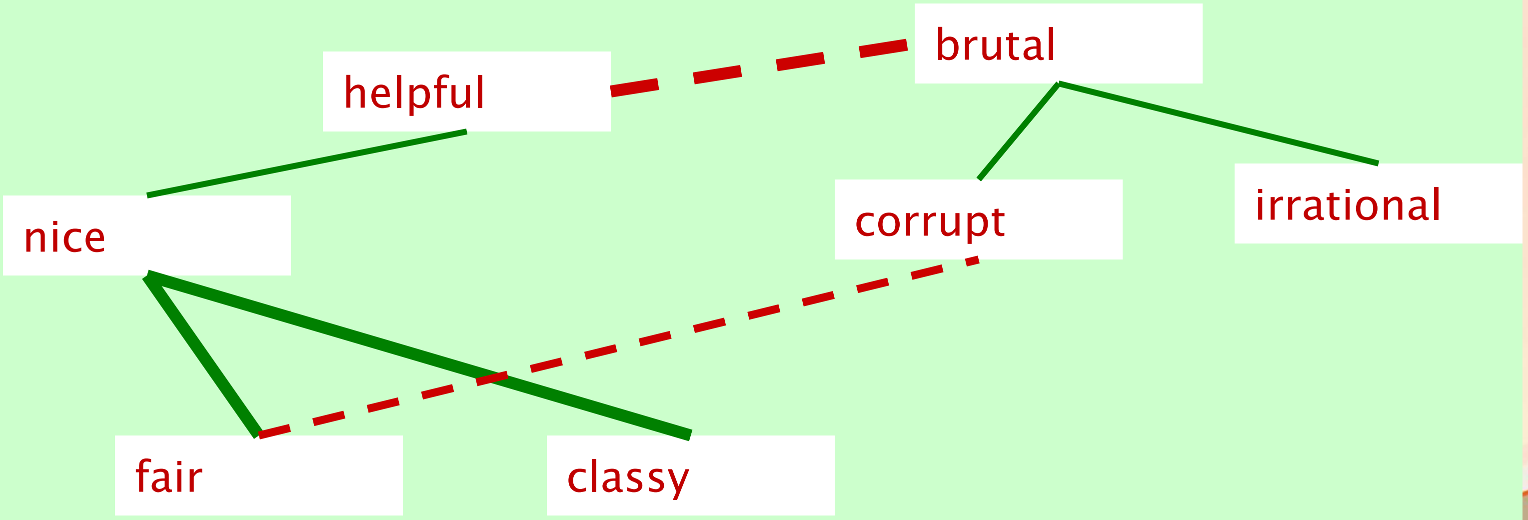


nice, helpful

nice, classy

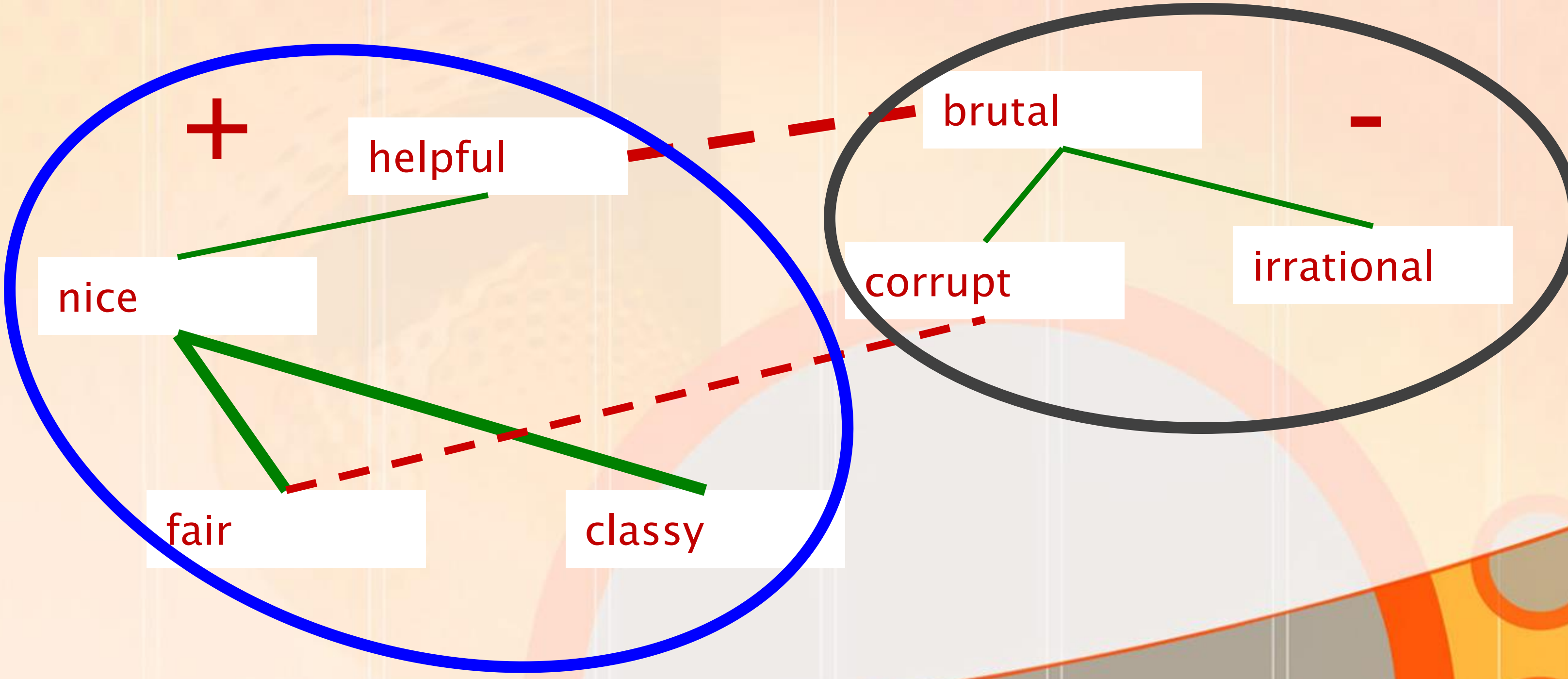
Hatzivassiloglou & McKeown 1997 Step 3

- *Supervised classifier assigns "polarity similarity" to each word pair, resulting in graph:*



Hatzivassiloglou & McKeown 1997 Step 4

- *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…*

SentiWordNet

An extension of the Hatzivassiloglou and McKeown algorithm assigns polarity to WordNet senses, called SentiWordNet (Baccianella et al., 2010).

Here are some examples:

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

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) !!!

SentiWordNet Algorithm

- *In this algorithm, polarity is assigned to entire synsets rather than words.*

  - *A positive lexicon is built from all the synsets associated with 7 positive words, and a negative lexicon from synsets associated with 7 negative words.*

  - *A classifier is then trained from this data to take a WordNet gloss and decide if the sense being defined is positive, negative or neutral.*

- *A further step (involving a random-walk algorithm) assigns a score to each WordNet synset for its degree of positivity, negativity, and neutrality.*

Sentiment Propagation

- *The SentProp (Sentiment Propagation) algorithm of Hamilton et al. (2016), has four steps:*

STEP 1. Define a graph: Given word embeddings, build a weighted lexical graph by connecting each word with its k nearest neighbors (according to cosine similarity). The weights of the edge between words w_i and w_j are set as:

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

STEP 2: Define a seed set: By hand, choose positive and negative seed words.

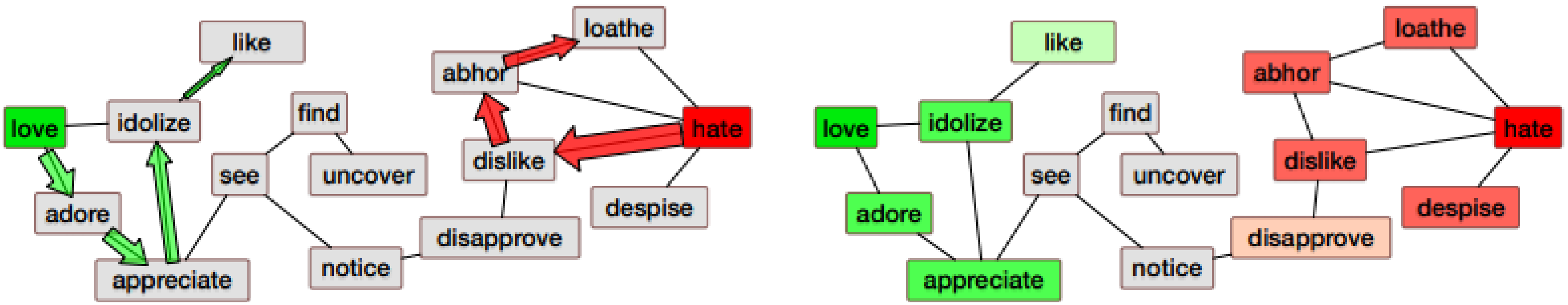
STEP 3: Propagate polarities from the seed set: Perform a random walk on the graph, starting at the seed set.

In a random walk, we start at a node and then choose a node to move to with probability proportional to the edge probability. A word's polarity score for a seed set is proportional to the probability of a random walk from the seed set landing on that word.

Sentiment Propagation - 2

STEP 3: *Propagate polarities from the seed set*: *Perform a random walk on the graph, starting at the seed set.*

(a) Run random walks from the seed words. (b) Assign polarity scores (shown here as colors green or red) based on the frequency of random walk visits.



(a) (b)

STEP 4: *Create word scores*: *We walk from both positive and negative seed sets, resulting in* *positive* (score$^+$(w_i)) *and* *negative* (score$^-$(w_i)) *label scores*. *We then combine these values into* *a positive-polarity score* *as:*

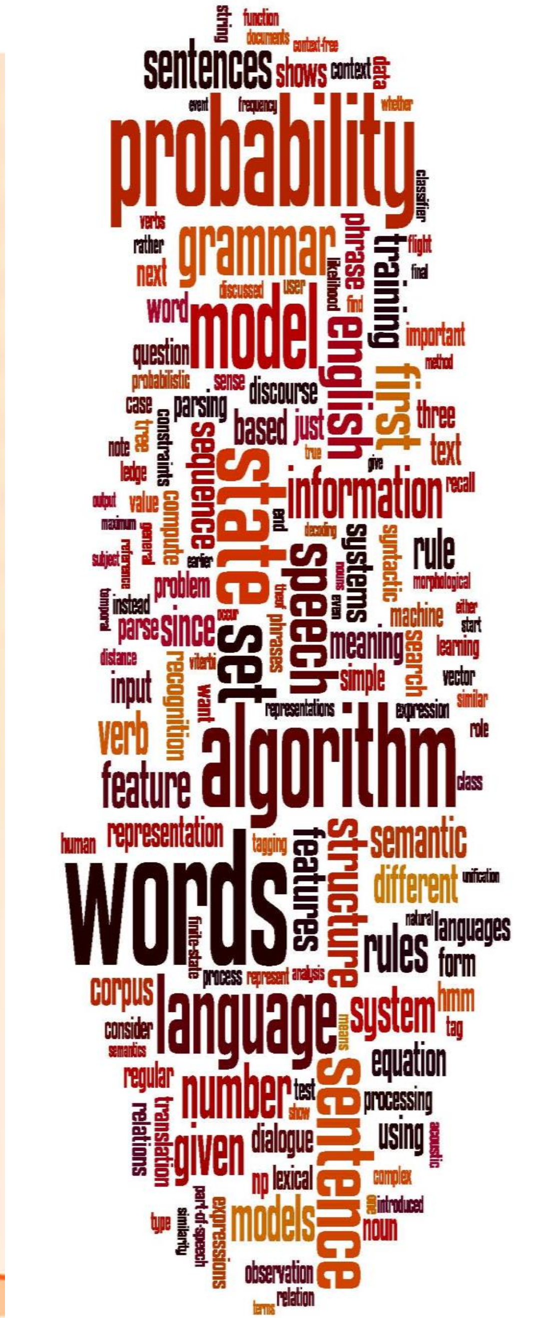
$$\text{score}^+(w_i) = \frac{\text{score}^+(w_i)}{\text{score}^+(w_i) + \text{score}^-(w_i)}$$

It's often helpful to standardize the scores to have zero mean and unit variance within a corpus!

Sentiment Propagation - 3

<u>STEP 5</u>: Assign confidence to each score: Because sentiment scores are influenced by the seed set, we'd like to know how much the score of a word would change if a different seed set is used.

- *We can use bootstrap-sampling to get confidence regions, by computing the propagation B times over random subsets of the positive and negative seed sets (for example using B = 50 and choosing 7 of the 10 seed words each time). The standard deviation of the bootstrap-sampled polarity scores gives a confidence measure.*



Computing with Affective Lexicons

Supervised Learning of Sentiment Lexicons

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.

a supervision signal exists in the world and can be made use of. One such signal is the scores associated with online reviews. The web contains an enormous number of online reviews for restaurants, movies, books, or other products!

- *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)

review score as supervision signals

Positive words are more likely to appear in 5-star reviews; negative words in 1-star reviews. And instead of just a binary polarity, this kind of supervision allows us to assign a word <u>a more complex representation of its polarity</u>: its distribution over stars (or other scores).

Excerpts from some reviews from various review websites, all on a scale of 1 to 5 stars except IMDB, which is on a scale of 1 to 10 stars.

	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 ...

Analyzing the polarity of each word in IMDB

- ❑ *How likely is each word to appear in each sentiment class?*
- ➢ *We could compute the IMDB likelihood of a word like disappoint(ed/ing) occurring in a 1 star review by dividing the number of times disappoint(ed/ing) occurs in 1-star reviews in the IMDB dataset (8,557) by the total number of words occurring in 1-star reviews (25,395,214), so the IMDB estimate of P(disappointing|1) is .0003*
- ❑ *Similarly, count("bad") in 1-star, 2-star, 3-star, etc.*
- ➢ *But <u>can't use raw counts</u>, Instead use the* ***likelihood:*** $\frac{P(w|c)}{P(w)}$
- ▪ *Better:* ***Scaled likelihood:*** $P(w|c) = \frac{f(w,c)}{\sum_{w \in c} f(w,c)}$

Dividing the IMDB estimate P(disappointing|1) of .0003 by the sum of the likelihood P(w|c) over all categories gives a Potts score of 0.10. The word disappointing thus <u>is associated with the vector :</u>

[.10, .12, .14, .14, .13, .11, .08, .06, .06, .05].

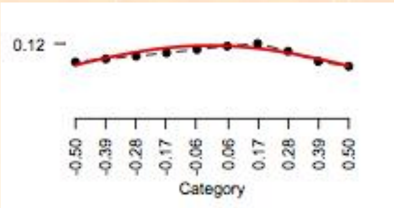
The Potts diagram Potts diagram is a visualization of these word scores, representing the prior sentiment of a word as a distribution over the rating categories

"Potts diagrams"

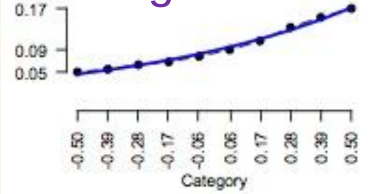
Potts, Christopher. 2011. NSF workshop on restructuring adjectives.

Positive scalars

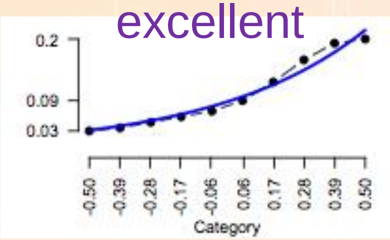
good



great

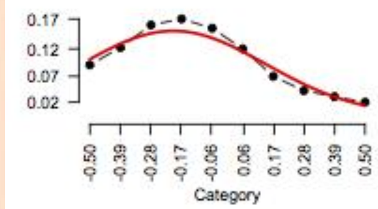


excellent

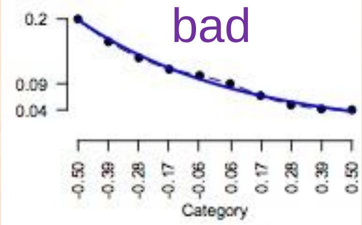


Negative scalars

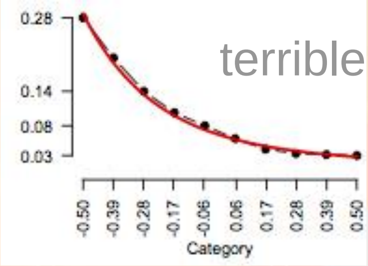
disappointing



bad

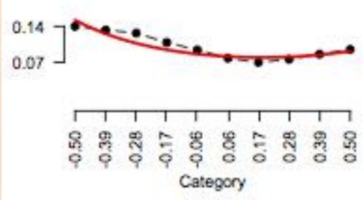


terrible

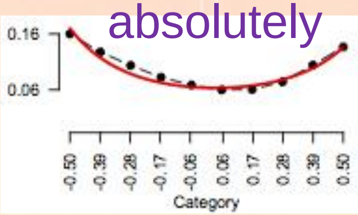


Emphatics

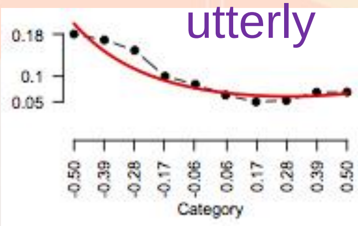
totally



absolutely

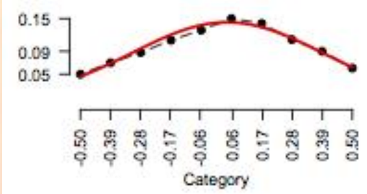


utterly

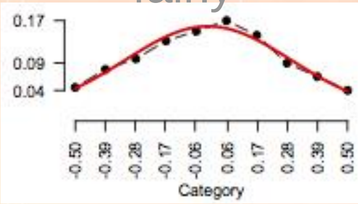


Attenuators

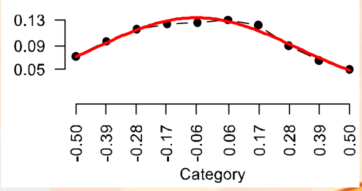
somewhat



fairly



pretty



Log odds ratio informative Dirichlet prior

- *We often want to do with word polarity is to <u>distinguish between words that are more likely to be used in one category of texts than in another</u>.*
  - *We may, want to know the words most associated with 1 star reviews versus those associated with 5 star reviews.*
  - *The problem with simple log-likelihood or log odds methods is that they don't work well for very rare words or very frequent words; for words that are very frequent, all differences seem large, and for words that are very rare, no differences seem large.*

The "log odds ratio informative Dirichlet prior" method of Monroe et al. (2008) is useful for finding words that are statistically overrepresented in one particular category of texts compared to another.

- *It's based on the idea of using another large corpus to get a prior estimate of what we expect the frequency of each word to be.*

Example: compute the log likelihood ratio

- *Assume we want to know whether the word "horrible" occurs more in corpus i or corpus j*

Use $f_i(w)$ to mean the frequency of word w in corpus i, and n_i to mean the total number of words in corpus i:

The Log Likelihood ration (LLR)

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

Instead, let's compute the log odds ratio: does "horrible" have higher odds in i or in j:

The Log Odds Ratio (LOR)

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

The Dirichlet Intuition

The Dirichlet intuition is to use a large background corpus to get a prior estimate of what we expect the frequency of each word w to be.

- *We'll do this very simply by adding the counts from that corpus to the numerator and denominator, so that we're essentially shrinking the counts toward that prior. It's like asking how large are the differences between i and j given what we would expect given their frequencies in a well-estimated large background corpus.*

The method estimates the difference between the frequency of word w in two corpora i and j via the prior-modified log odds ratio for w:

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

where n_i is the size of corpus i, n_j is the size of corpus j, f_w^i is the count of word w in corpus i, f_w^j is the count of word w in corpus j, α_0 is the size of the background corpus, and α_w is the count of word w in the background corpus.

Computing the statistics of words using the Log odds ratio informative Dirichlet prior

In addition, Monroe et al. (2008) make use of an estimate for the variance of the log–odds–ratio:

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

Remember:

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

The final statistic for a word is then <u>the z–score of its log–odds–ratio</u>:

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

Advantages

The Monroe et al. (2008) method thus modifies the commonly used log odds ratio in two ways:

1. *it uses the z-scores of the log odds ratio, which controls for the amount of variance in a words frequency, and*
2. *it uses counts from a background corpus to provide a prior count for words.*

Class	Words in 1-star reviews	Class	Words in 5-star reviews
Negative	*worst, rude, terrible, horrible, bad, awful, disgusting, bland, tasteless, gross, mediocre, over-priced, worse, poor*	**Positive**	*great, best, love(d), delicious, amazing, favorite, perfect, excel-lent, awesome, friendly, fantastic, fresh, wonderful, incredible, sweet, yum(my)*
Negation	*no, not*	**Emphatics/ universals**	*very, highly, perfectly, definitely, ab-solutely, 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, cus-tomer, 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*

How did it do?

Applied to a dataset of restaurant reviews from Yelp!

- *comparing the words used in 1-star reviews to the words used in 5-star reviews.*

The largest difference is in obvious sentiment words, with the 1-star reviews using negative sentiment words like worse, bad, awful and the 5-star reviews using positive sentiment words like great, best, amazing.

But 1-star reviews use logical negation (no, not), while 5-star reviews use emphatics and emphasize universality (very, highly, every, always).

1- star reviews use first person plurals (we, us, our) while 5 star reviews use the second person.

1-star reviews talk about people (manager, waiter, customer) while 5-star reviews talk about dessert and properties of expensive restaurants like courses and atmosphere.

Using Lexicons for Sentiment Recognition

In the simplest case, lexicons can be used when we don't have sufficient training data to build a supervised sentiment analyzer; it can often be expensive to have a human assign sentiment to each document to train the supervised classifier.

- *lexicons can be used in a simple rule-based algorithm for classification.*

Use the ratio of positive to negative words: if a document has more positive than negative words (using the lexicon to decide the polarity of each word in the document), it is classified as positive. Often a threshold λ is used, in which a document is classified as positive only if the ratio is greater than λ.

- *If the sentiment lexicon includes positive and negative weights for each word, θ_w^+ and θ_w^-, these can be used as well:*

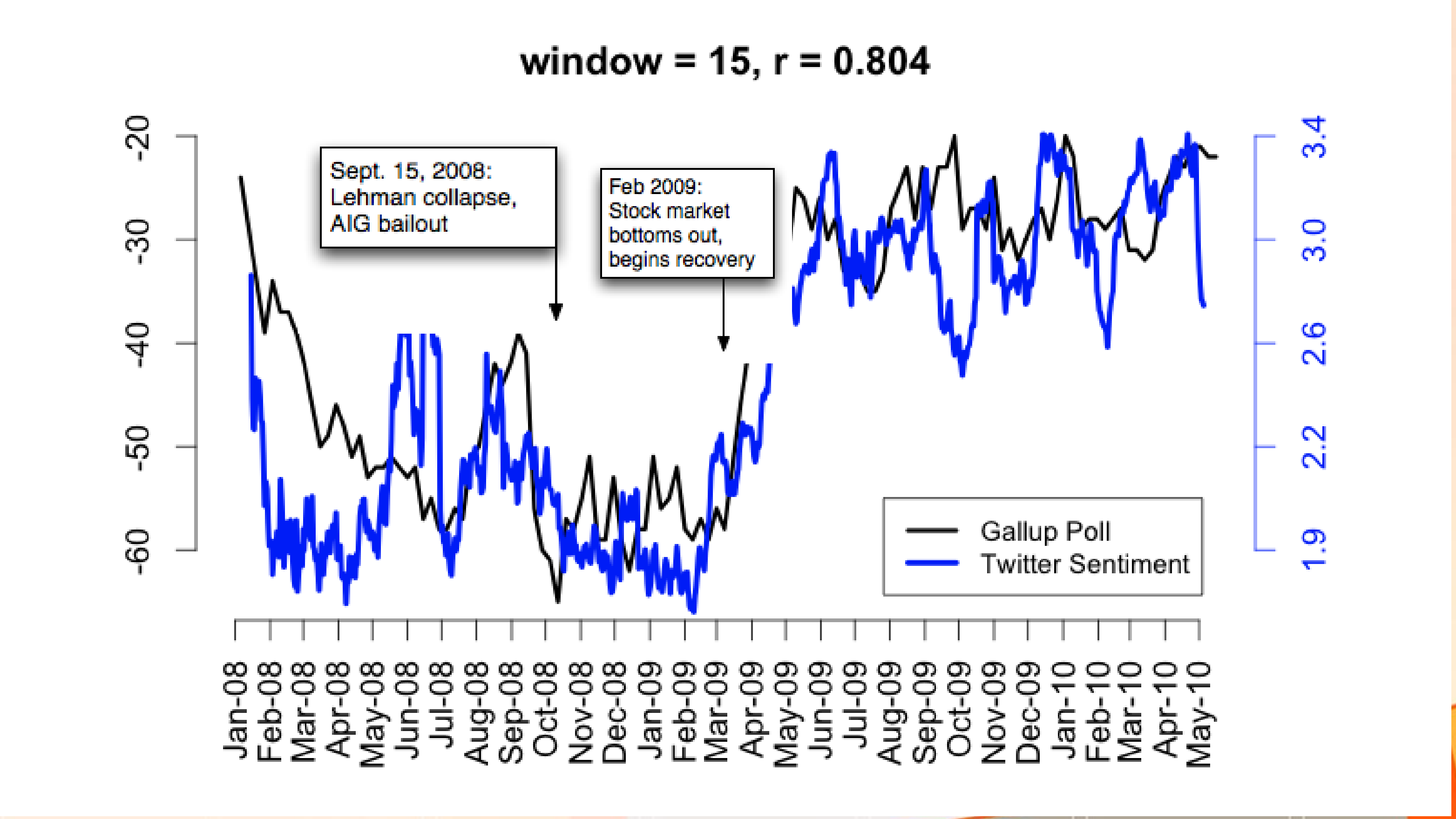
$$f^+ = \sum_{w \text{ s.t. } w \in positivelexicon} \theta_w^+ count(w)$$

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

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

Twitter sentiment versus Gallup Poll of Consumer Confidence

Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. 2010. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In ICWSM-2010



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*

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*

Sentiment Classification in Movie Reviews

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. EMNLP-2002, 79—86.
Bo Pang and Lillian Lee. 2004. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. ACL, 271-278

- *Polarity detection:*
  - *Is an IMDB movie review positive or negative?*
- *Data: Polarity Data 2.0:*
  - *http://www.cs.cornell.edu/people/pabo/movie-review-data*

IMDB data in the Pang and Lee database

✓

when _star wars_ came out some twenty years ago , the image of traveling throughout the stars has become a commonplace image . […]

when han solo goes light speed , the stars change to bright lines , going towards the viewer in lines that converge at an invisible point .

cool .

october sky offers a much simpler image– that of a single white dot , traveling horizontally across the night sky .   [. . .]

✗

" snake eyes " is the most aggravating kind of movie : the kind that shows so much potential then becomes unbelievably disappointing .

it's not just because this is a brian depalma film , and since he's a great director and one who's films are always greeted with at least some fanfare .

and it's not even because this was a film starring nicolas cage and since he gives a brauvara performance , this film is hardly worth his talents .

Baseline Algorithm (adapted from Pang and Lee)

- *Tokenization*
- *Feature Extraction*
- *Classification using different classifiers*
  - *Naïve Bayes*
  - *MaxEnt*
  - *SVM*

Sentiment Tokenization Issues

- *Deal with HTML and XML markup*
- *Twitter mark-up (names, hash tags)*
- *Capitalization (preserve for words in all caps)*
- *Phone numbers, dates*
- *Emoticons*
- *Useful code:*
  - *Christopher Potts sentiment tokenizer*
  - *Brendan O'Connor twitter tokenizer*

Potts emoticons

```
[<>]?                       # optional hat/brow
[:;=8]                      # eyes
[\-o\*\']?                  # optional nose
[\)\]\(\[dDpP/\:\}\{@\|\\]  # mouth
|                       #### reverse orientation
[\)\]\(\[dDpP/\:\}\{@\|\\]  # mouth
[\-o\*\']?                  # optional nose
[:;=8]                      # eyes
[<>]?                       # optional hat/brow
```

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*
    - *All words turns out to work better, at least on this data*

Negation

Das, Sanjiv and Mike Chen. 2001. Yahoo! for Amazon: Extracting market sentiment from stock message boards. In Proceedings of the Asia Pacific Finance Association Annual Conference (APFA).
Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. EMNLP-2002, 79—86.

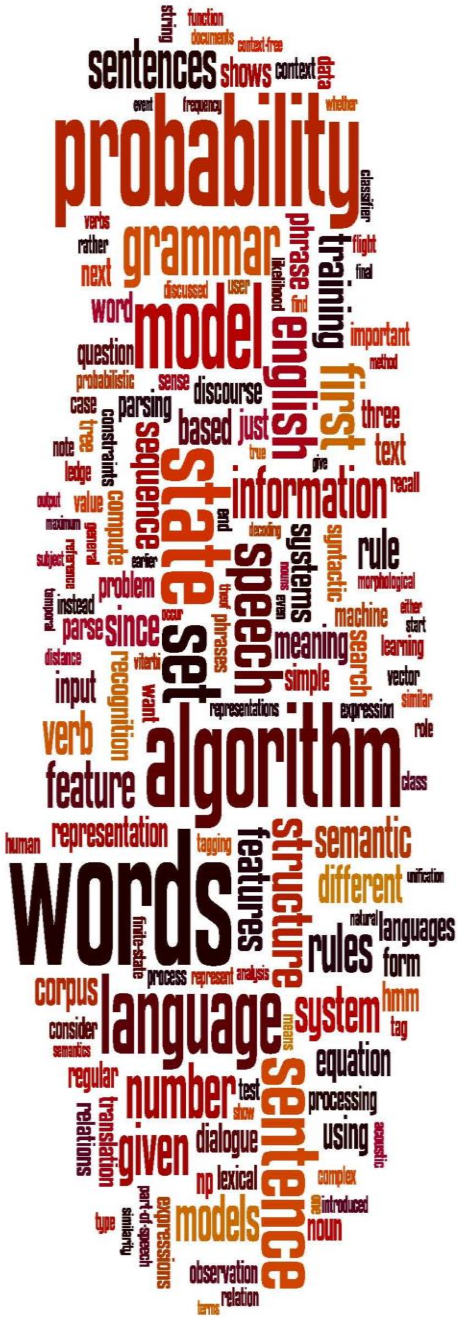
Add NOT_ to every word between negation and following punctuation:

didn't like this movie , but I

didn't NOT_like NOT_this NOT_movie but I

Binarized (Boolean feature) Multinomial Naïve Bayes

- *Intuition for sentiment (and probably for other text classification domains)*
  - *Word occurrence may matter more than word frequency*
    - *The occurrence of the word fantastic tells us a lot*
    - *The fact that it occurs 5 times may not tell us much more.*
  - *Boolean Multinomial Naïve Bayes*
    - *Clips all the word counts in each document at 1*



Recognizing Personality Traits

Relying on
Sentiment Lexicons

Personality Trait Recognition

- *Detecting a person's personality from their language can be useful for dialog systems (users tend to prefer agents that match their personality), and can play a useful role in computational social science questions like understanding how personality is related to other kinds of behavior.*

Many theories of human personality are based around a small number of dimensions, such as various versions of the "Big Five" dimensions (Digman, 1990):

1. *Extroversion vs. Introversion: sociable, assertive, playful vs. aloof, reserved, shy*
2. *Emotional stability vs. Neuroticism: calm, unemotional vs. insecure, anxious*
3. *Agreeableness vs. Disagreeableness: friendly, cooperative vs. antagonistic, faultfinding*
4. *Conscientiousness vs. Unconscientiousness: self-disciplined, organized vs. inefficient, careless*
5. *Openness to experience: intellectual, insightful vs. shallow, unimaginative*

Annotations for Personality Traits

A few corpora of text and speech have been labeled for the personality of their author by having the authors take a standard personality test.

- *The essay corpus of Pennebaker and King (1999) consists of 2,479 essays (1.9 million words) from psychology students who were asked to "write whatever comes into your mind" for 20 minutes.*
- *The EAR (Electronically Activated Recorder) corpus of Mehl et al. (2006) was created by having volunteers wear a recorder throughout the day, which randomly recorded short snippets of conversation throughout the day, which were then transcribed.*
- *The Facebook corpus of (Schwartz et al., 2013) includes 309 million words of Facebook posts from 75,000 volunteers.*

Examples of annotations

- *Example of samples from Pennebaker and King (1999) from an essay written by someone on the neurotic end of the neurotic/emotionally stable scale:*

“ 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.”

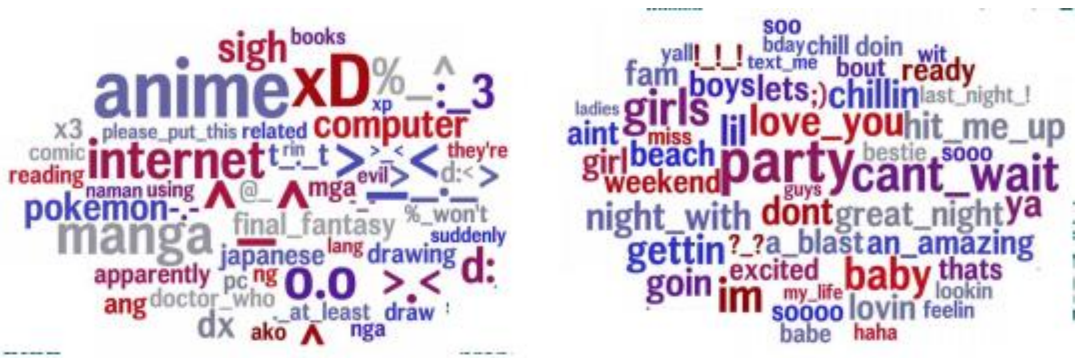
- *and someone on the emotionally stable end of the scale:*

“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.”

Personality Recognition- More

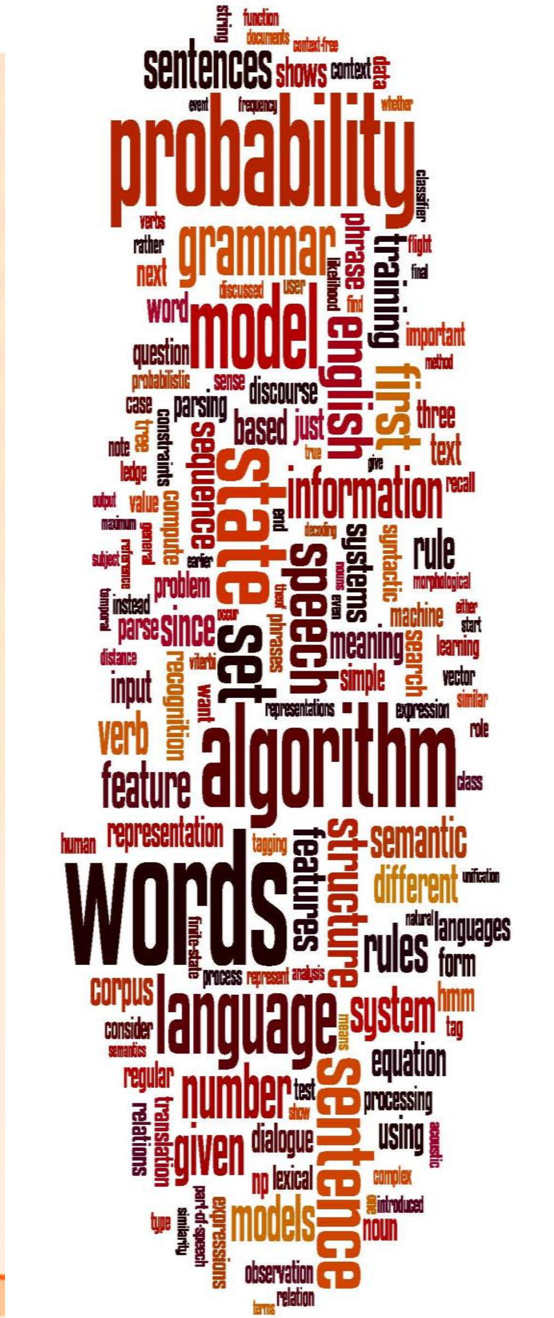
- Mairesse and Walker (2008) found that for classifying personality:
  - *for the dimension Agreeable, the LIWC lexicons Family and Home were positively associated while the LIWC lexicons anger and swear were negatively associated.*
  - *For Extroversion was positively associated with the Friend, Religion and Self lexicons,*
  - *For Emotional Stability was positively associated with Sports and negatively associated with Negative Emotion.*

➢ *In the Extroversion/Introversion classifier of Schwartz et al. (2013), used least-squares regression to predict the value of a personality dimension <u>from all the words and phrases</u>. The resulting regression coefficient for each word or phrase can be used as an association value with the predicted dimension.*



(a) (b)

Word clouds from Schwartz et al. (2013), showing words highly associated with introversion (left) or extroversion (right). The size of the word represents the association strength (the regression coefficient), while the color (ranging from cold to hot) represents the relative frequency of the word/phrase (from low to high).



Stance and Connotations

Combining Sentiment Lexicons

With Sematic Roles or Dialog Social Roles

Stance

- *Another kind of affective meaning is what Scherer (2000) calls interpersonal stance, the 'affective stance taken toward another person in a specific interaction interpersonal stance coloring the interpersonal exchange'.*

- *Extracting this kind of meaning means automatically labeling participants for whether they are friendly, supportive, distant.*

For example Ranganath et al. (2013) studied a corpus of speed-dates, in which participants went on a series of 4-minute romantic dates, wearing microphones.

Each participant labeled each other for how flirtatious, friendly, awkward, or assertive they were.

Ranganath et al. (2013) then used a combination of lexicons and other features to detect these interpersonal stances from text.

Connotation Frames

- *The lexicon sentiment define a word as a point in affective space.*
- *A connotation frame, by contrast, is lexicon that incorporates a richer kind of grammatical structure, by combining affective lexicons with the frame semantic lexicons .*

The basic insight of connotation frame lexicons is that **a predicate** like a verb expresses connotations about the verb's arguments (Rashkin et al. 2016, Rashkin et al. 2017)

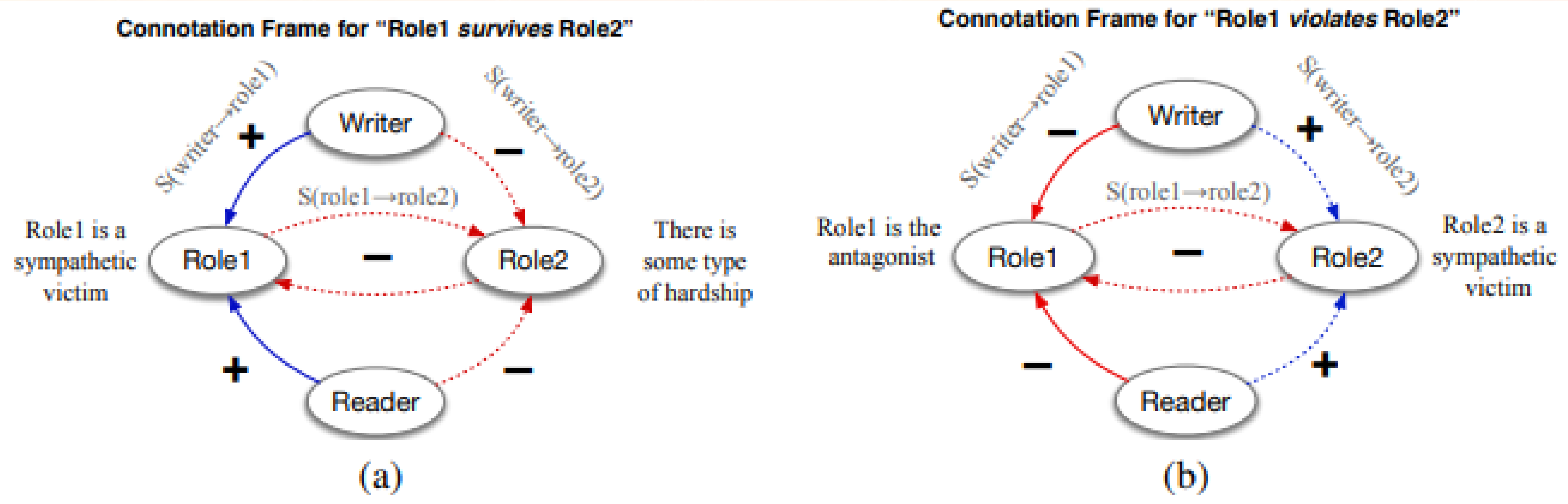
Example 1: *Country A* ***violated*** *the sovereignty of Country B.*

- *By using the verb violate, the author is expressing their sympathies with Country B, portraying Country B as a victim, and expressing antagonism toward the agent Country A*

Example 2: *the teenager ...* ***survived*** *the Boston Marathon bombing"*

- *using the verb survive, the author is expressing that the bombing is a negative experience, and the subject of the sentence the teenager, is a sympathetic character*

aspects of connotation are inherent in the meaning of the verbs



Connotation frames for **survive** and **violate**.

(a) For **survive**, the writer and reader have positive sentiment toward Role1, the subject, and negative sentiment toward Role2, the direct object.

(b) For **violate**, the writer and reader have positive sentiment instead toward Role2, the direct object.

Affect Computing - Conclusions

- *Many kinds of affective states can be distinguished, including emotions, moods, attitudes (which include sentiment), interpersonal stance, and personality.*
  - *Emotion can be represented by fixed atomic units often called basic emotions, or as points in space defined by dimensions like valence and arousal.*
  - *Words have connotational aspects related to these affective states, and this connotational aspect of word meaning can be represented in lexicons.*
  - *Affective lexicons can be built by hand, using crowd sourcing to label the affective content of each word. Lexicons can also be learned in a fully supervised manner, when a convenient training signal can be found in the world, such as ratings assigned by users on a review site.*
  - *Words can be assigned weights in a lexicon by using various functions of word counts in training texts, and ratio metrics like log odds ratio informative Dirichlet prior.*
  - *Personality is often represented as a point in 5-dimensional space*