CS224C: NLP for CSS

Sentiment and Emotion

Diyi Yang

Stanford CS

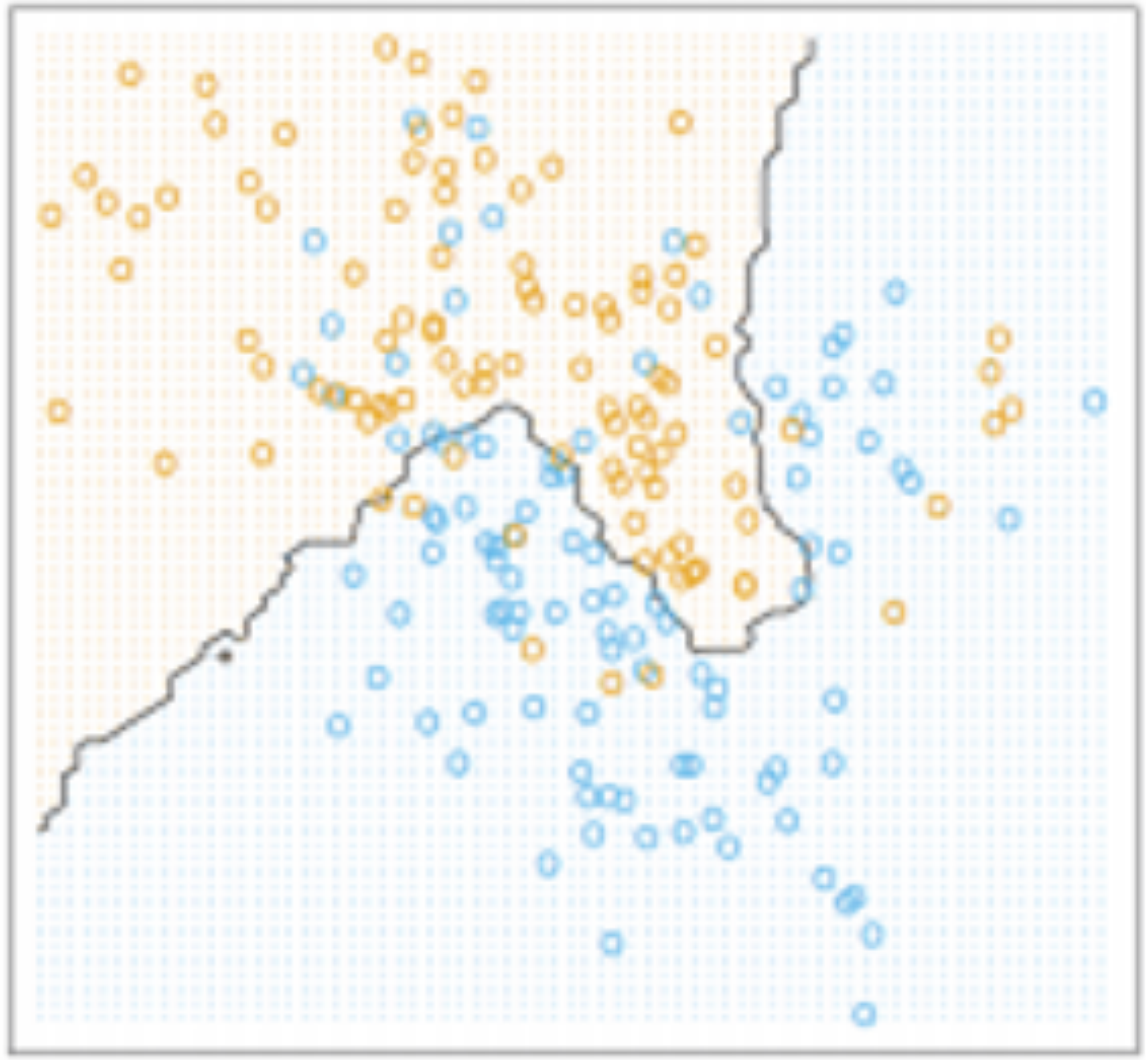
Announcements

Resources on How to Do Research on Canvas

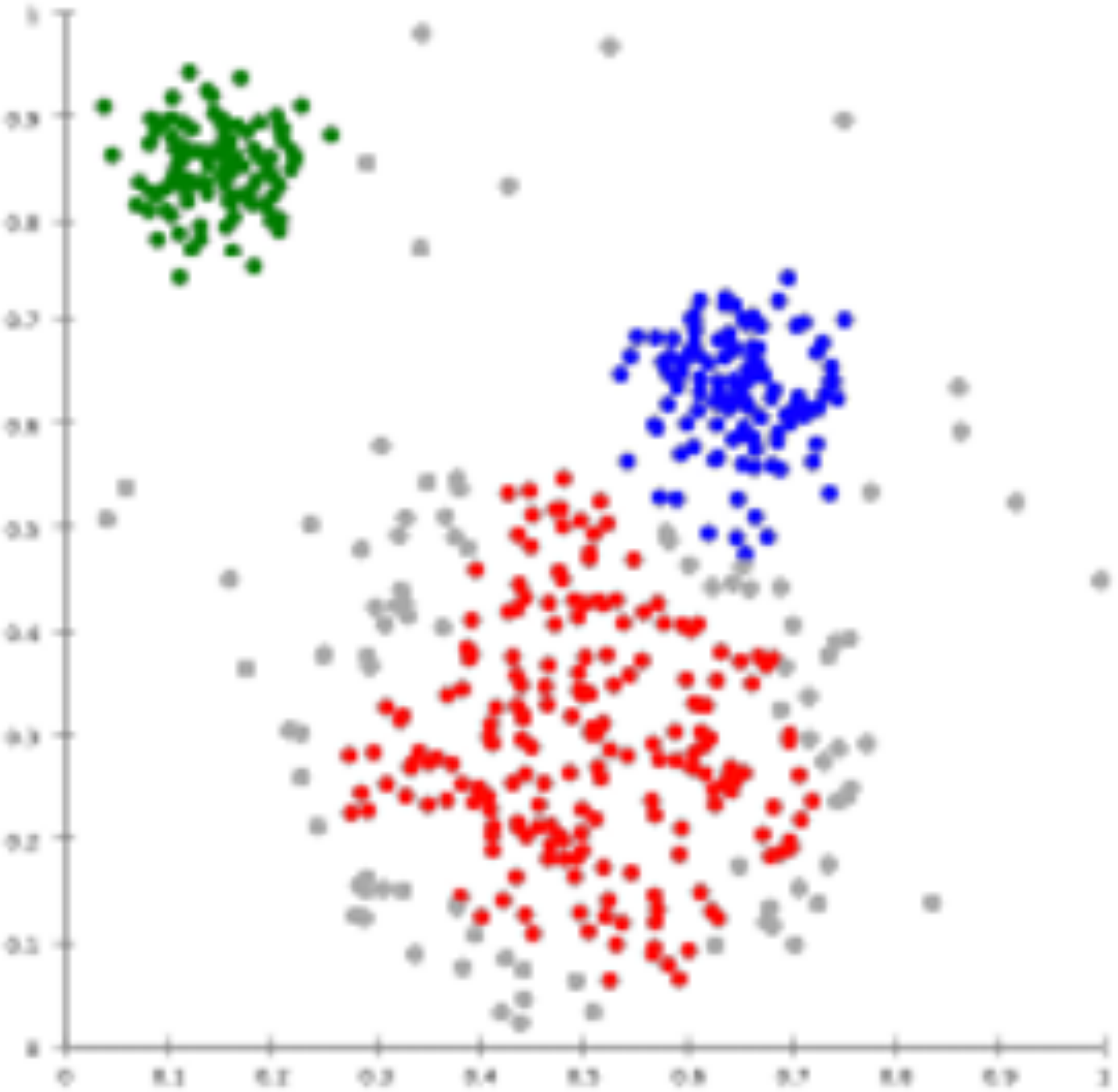
Reading responses for next Tuesday (submit as “replies” to the reading thread)

Homework 1 is out!

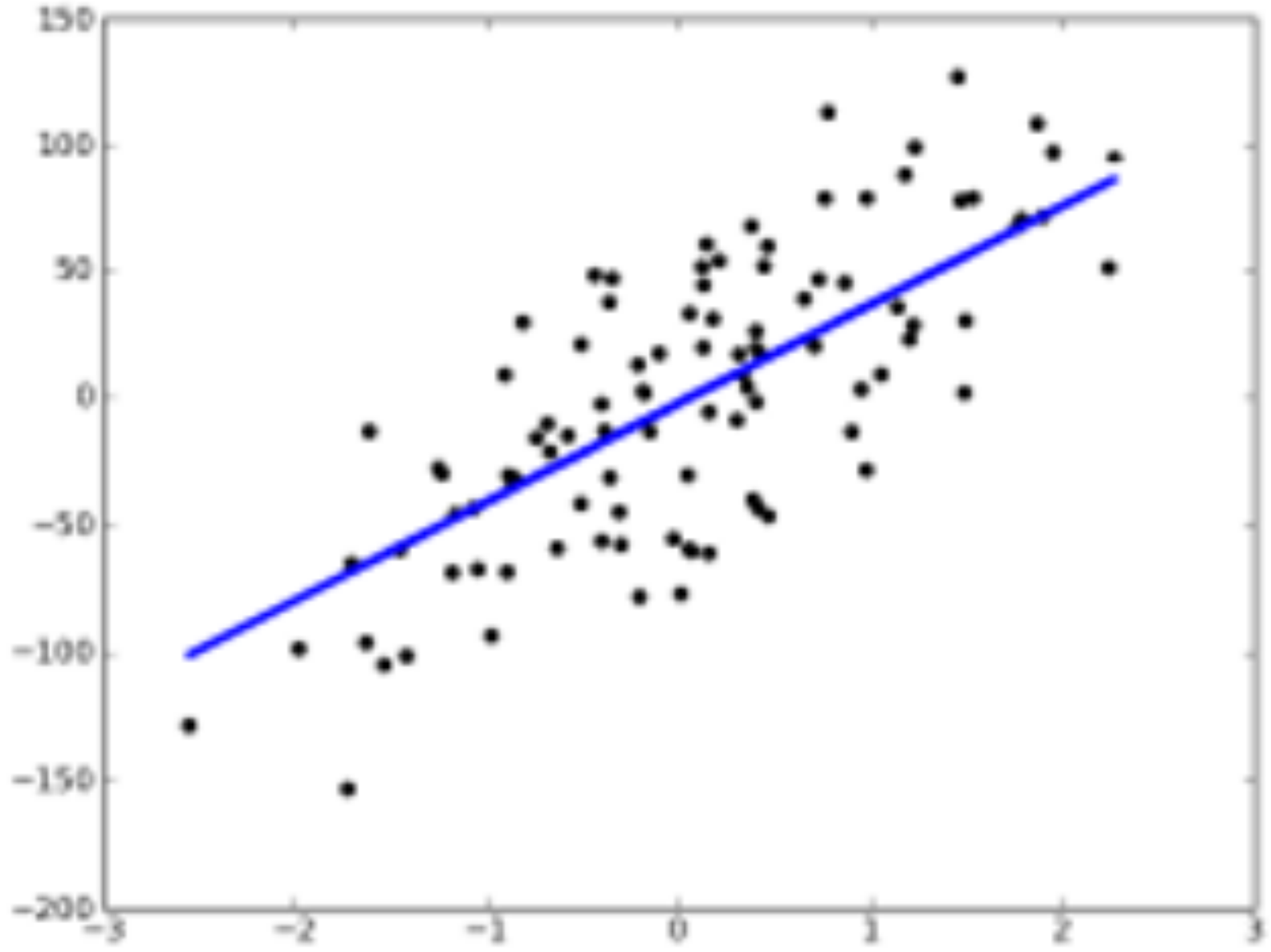
Common Methods



Classification



Clustering



Regression

Let's find different groups of people in support groups

Imagine this is on an online social support community …

1. **Why** is this a clustering task?
2. **What** is "group" of people?
3. How can we get the **ground truth**?
4. How **many** groups?
5. What **features** should we use?
6. How can we **evaluate** it?

Let’s find different groups of people in support groups

Imagine this is on an online social support community …

We need to come up with **a lot of features**

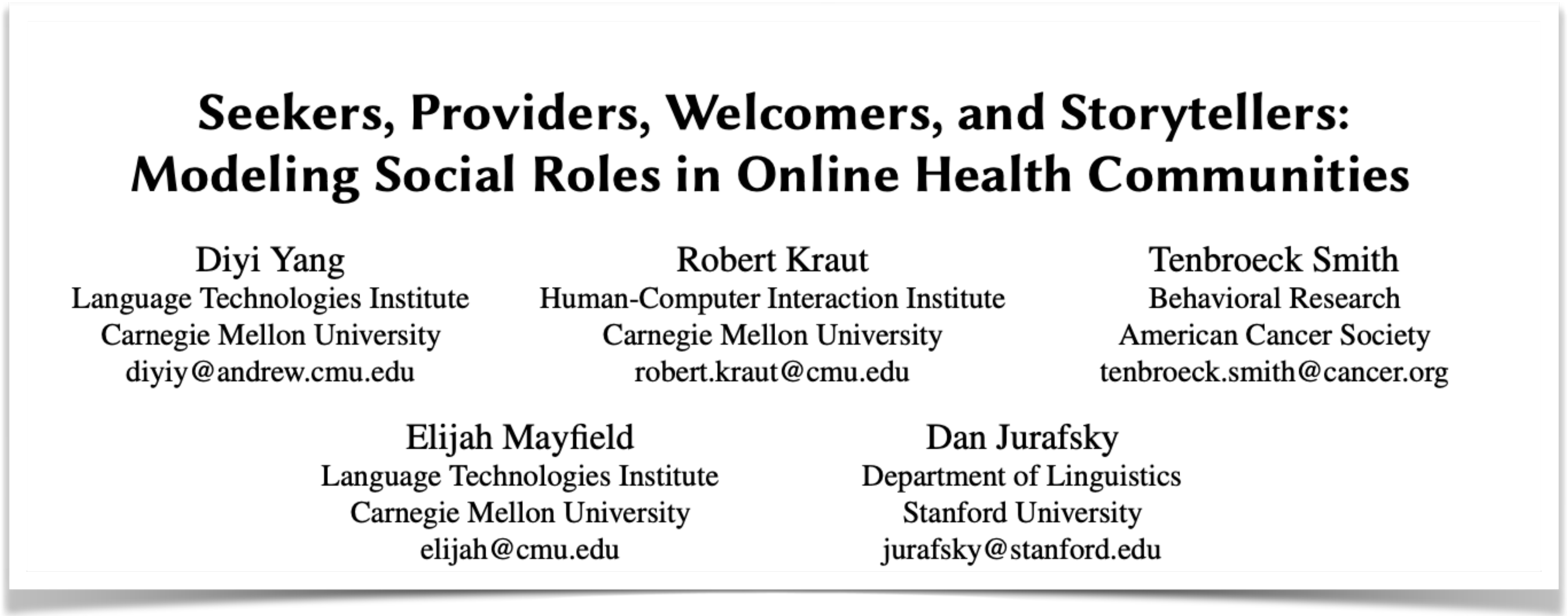
Agent: members on CSN …

Interaction: medical/treatment topics, emotions …

Expectation: report to moderators …

Context: private vs. public discussion …

Goal: social support …


Seekers, Providers, Welcomers, and Storytellers: Modeling Social Roles in Online Health Communities

Diyi Yang
Language Technologies Institute
Carnegie Mellon University
diyiy@andrew.cmu.edu

Robert Kraut
Human-Computer Interaction Institute
Carnegie Mellon University
robert.kraut@cmu.edu

Tenbroeck Smith
Behavioral Research
American Cancer Society
tenbroeck.smith@cancer.org

Elijah Mayfield
Language Technologies Institute
Carnegie Mellon University
elijah@cmu.edu

Dan Jurafsky
Department of Linguistics
Stanford University
jurafsky@stanford.edu

Let's find different groups of people in support groups

Agent: members on CSN …
Interaction: medical/treatment topics, emotions …
Expectation: report to moderators …
Context: private vs. public discussion …
Goal: social support

The Facet of Goal: Social Support

Since you are a triple positive they can put you on hormones and the chance of recurrence is low. Listen to your chemo nurse ...

Informational Support

It gives me faith that you can have cancer and live a full life. Sorry to hear that. God bless you. Please stay strong!

Emotional Support

Let’s find different groups of people in support groups

Intuition: *a user is a mixture of different social roles*

Let's find different groups of people in support groups

Emotional Support Provider	Private Support Provider
Newcomer Welcomer	All-round Expert
Informational Support	
Story Sharer	
Informational Support	
Private Communica	

Role Name	Prevalence (%)	Typical Behaviors Listed in Importance
Emotional Support Provider	33.3	Provide emotional support Provide empathy Participate in a large number of cancer-specific forums
Welcomer	15.9	Frequently talking to newcomers Provide encouragement Higher number of replies
Informational Support Provider	13.3	Provide informational support Higher usage of words related to symptoms and treatment
Story Sharer	10.2	Higher level of self-disclose Seek emotional support Initialize higher number of threads

Let's find different groups of people in support groups

Work with 6 moderators on CSN to assess the derived roles

*" It seems very **comprehensive** and there are so many different examples, so I feel like it is **covered very well** with your different roles and labels. "*

The identified roles were comprehensive

Is it a classification/regression/clustering problem?

I want to predict a star value {1,2,3,4,5} for a product review

I want to find all of the texts that have allusions to Paradise Lost

I want to predict the stock price

I want to tell which team will win

I want to associate photographs of cats with animals in a taxonomic hierarchy

I want to reconstruct an evolutionary tree for languages

Computational Social Science in the Age of Big Data

danah boyd and Kate Crawford (2012), “Critical Questions for Big Data,” Information, Communication and Society

1 “Big data” changes the definition of knowledge

How do computational methods/quantitative analysis pragmatically affect epistemology?

Restricted to what data is available (twitter, data that’s digitized, google books, etc.). How do we counter this in experimental designs?

Establishes alternative norms for what “research” looks like

2 Claims to objectivity and accuracy are misleading

Data collection, selection process is subjective, reflecting belief in what matters.

Model design is likewise subjective

- model choice (classification vs. clustering etc.)
- representation of data
- feature selection

Claims need to match the sampling bias of the data

3 Bigger data is not always better data

Uncertainty about its source or selection mechanism [Twitter, Google books]

Appropriateness for question under examination

How did the data you have get there?

Are there other ways to solicit the data you need?

Remember **the value of small data**: interview and qualitative studies

4 Taken out of context, big data loses its meaning

A representation (through features) is a necessary approximation; what are the consequences of that approximation?

Example: quantitative measures of “tie strength” and its interpretation

5 Just because it is accessible does not make it ethical

Anonymization practices for sensitive data (even if born public)

Accountability both to research practice and to subjects of analysis

6 Limited access to big data creates new digital divides

Inequalities in access to data and the production of knowledge

Privileging of skills required to produce knowledge

Sentiment and Affect

Overview

- Emotion
- Subjectivity
- LIWC
- Empath
- Semi-supervised and supervised approaches to infer affect

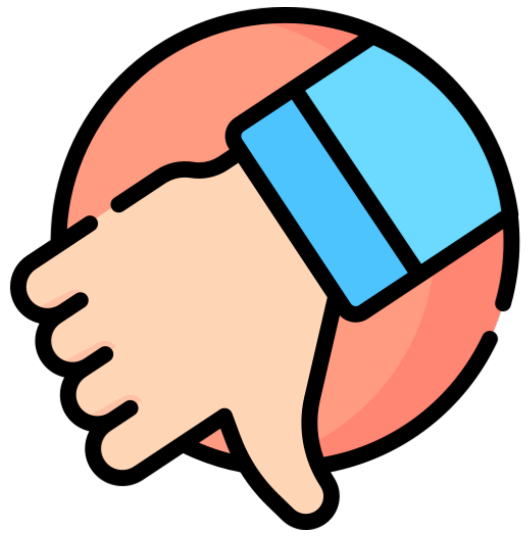
Some slides are adapted based on *Lexicons for Sentiment, Affect, and Connotation* from Speech and Language Processing (3rd ed. draft) Dan Jurafsky and James H. Martin (https://web.stanford.edu/~jurafsky/slp3/)

Lexicon

- A (usually hand-built) list of words that correspond to some meaning or class
- Possibly with numeric values
- Commonly used as simple classifiers, or as features to complex classifiers

Why Lexicons for Sentiment and Affect

- Easy to use
- Interpretable
- Fast to calculate



- Fail to consider negation or word order
- Can't deal with context

Scherer's typology of affective states

Emotion: brief organically synchronized evaluation of a major event

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

Mood: diffuse non-caused low-intensity long-duration change in subjective feeling

cheerful, gloomy, irritable, listless, depressed, buoyant

Interpersonal stance: affective stance toward another person in a specific interaction

distant, cold, warm, supportive, contemptuous

Attitudes: enduring, affectively colored beliefs, dispositions towards objects or persons

liking, loving, hating, valuing, desiring

Personality traits: stable personality dispositions and typical behavior tendencies

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): the pleasantness of the stimulus

Arousal (strong, weak): the intensity of emotion provoked by the stimulus

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

Plutchick's wheel of emotion

8 basic emotions

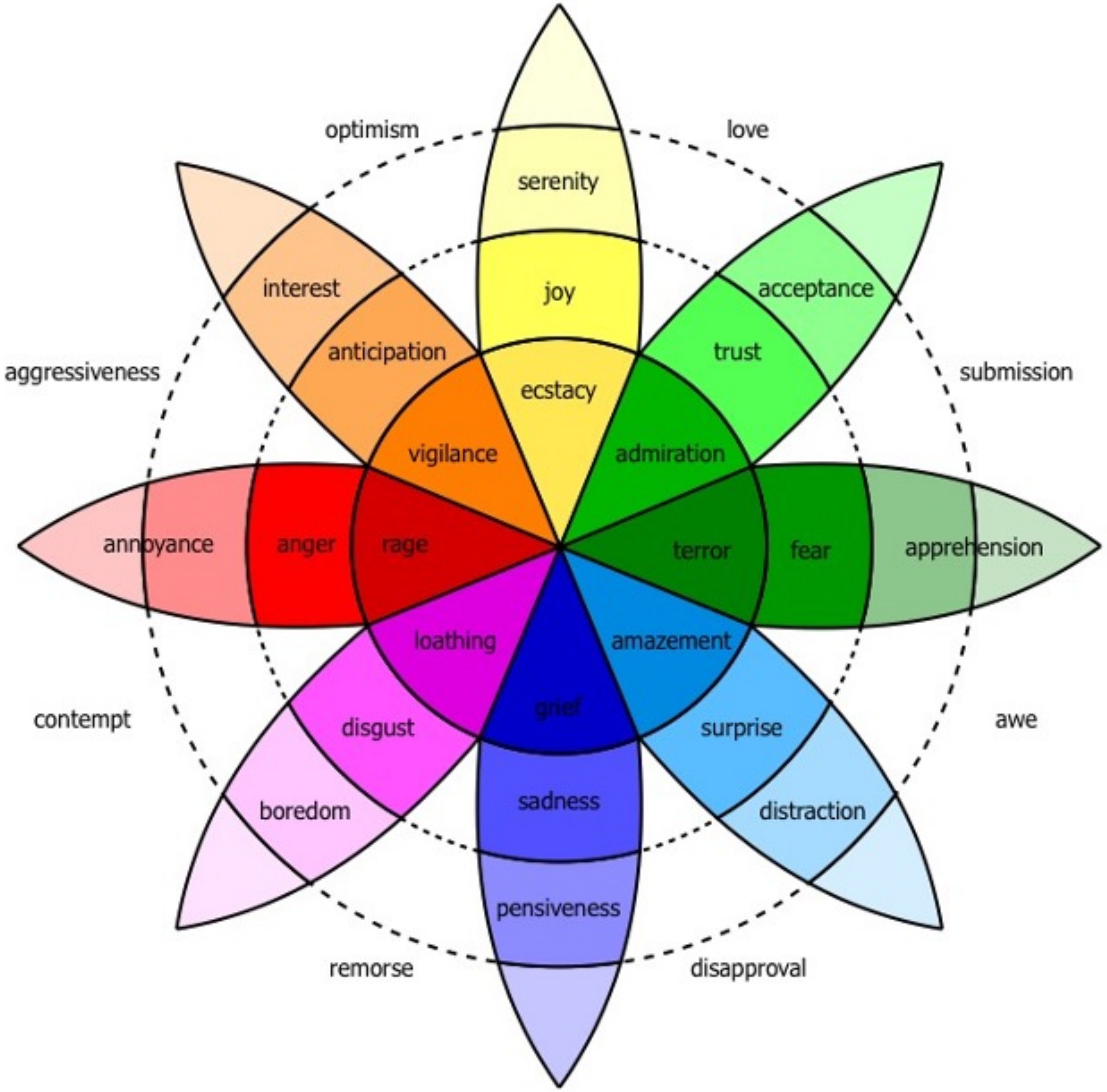
In four opposing pairs:

joy–sadness

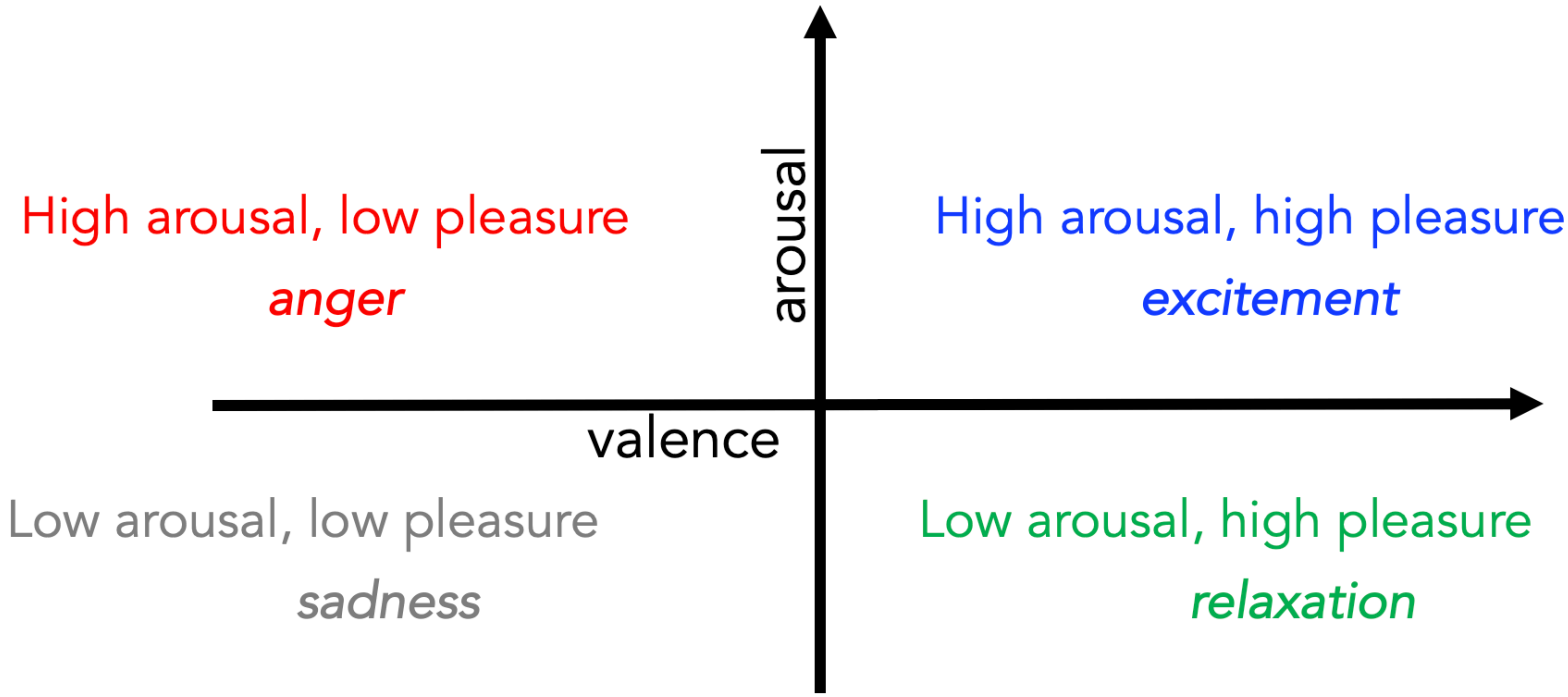
anger–fear

trust–disgust

anticipation–surprise



Valence/Arousal Dimensions



Some Sentiment Lexicons

The General Inquirer

Positive (1915 words), and Negative (2291 words)

MPQA Subjectivity Cues Lexicon

6885 words on strong/weak subjectivity

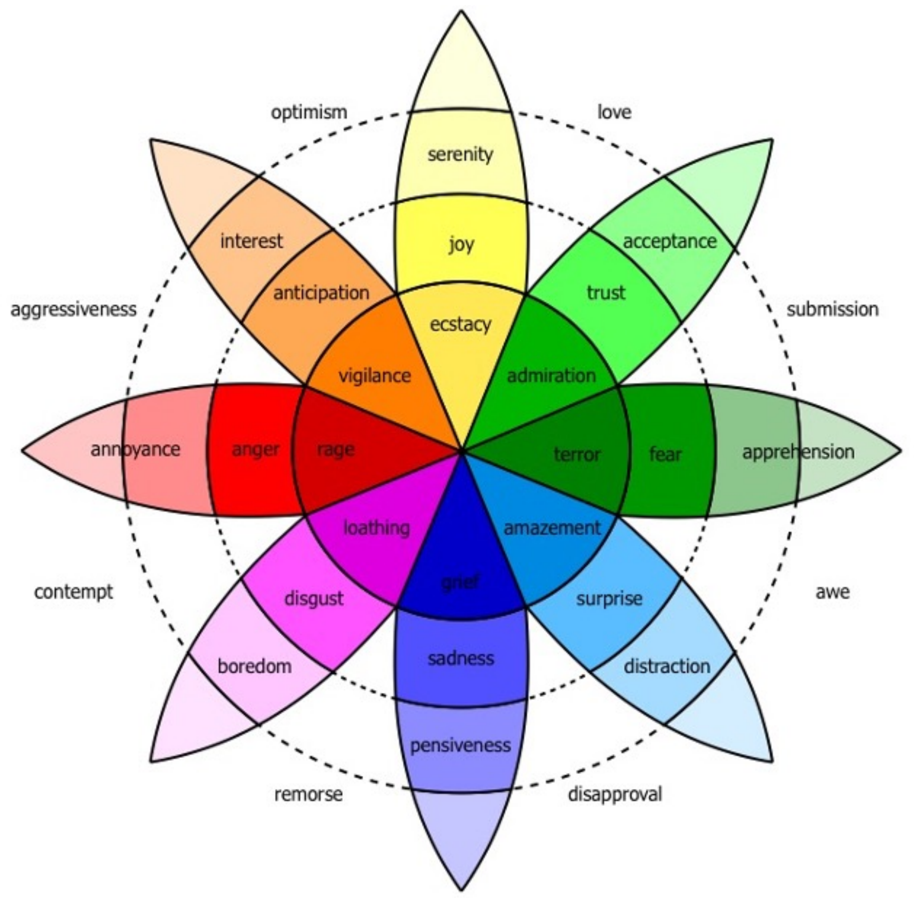
Is a subjective word positive or negative?

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

Words with consistent sentiment across lexicons

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

NRC Emotion Lexicon



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

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

Another Widely Used Lexicon: LIWC

LIWC: Linguistic Inquiry and Word Count

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

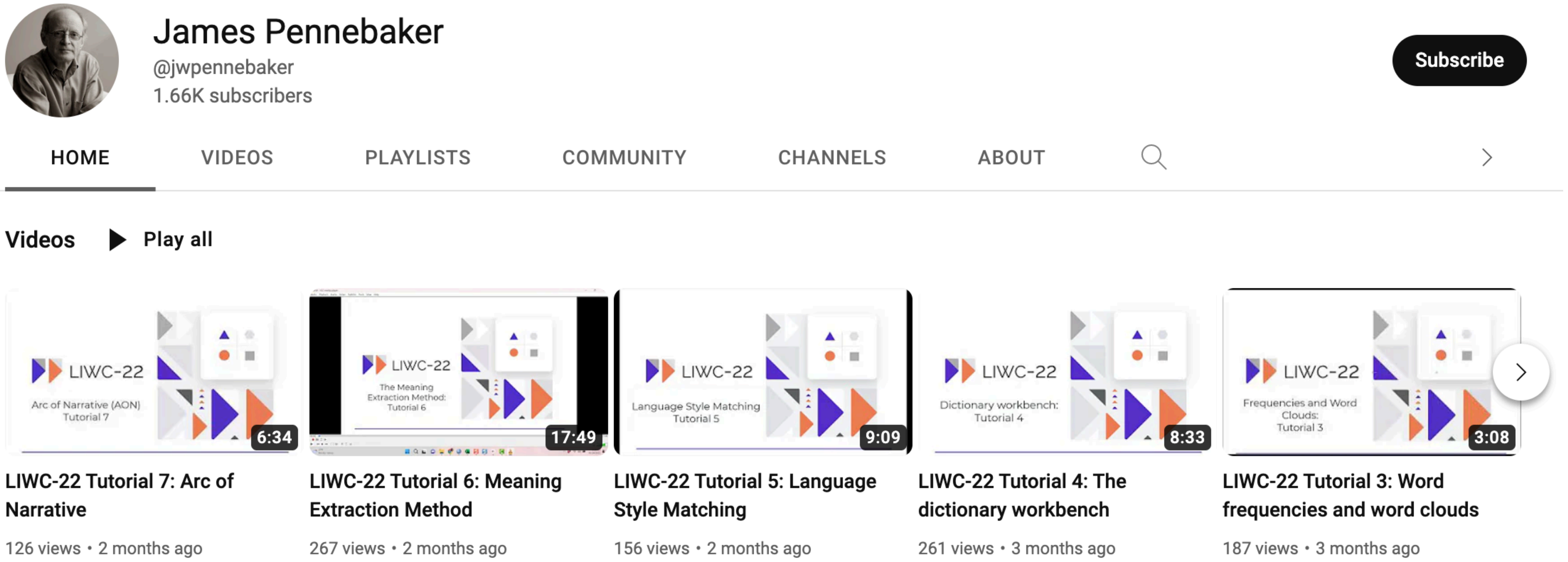
http://www.liwc.net/

2300 words

>70 classes

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



Concreteness versus Abstractness

Definition:

The degree to which the concept denoted by a word refers to a perceptible entity.

Lexicon:

37,058 English words and 2,896 two-word expressions

Rating from 1 (abstract) to 5 (concrete)

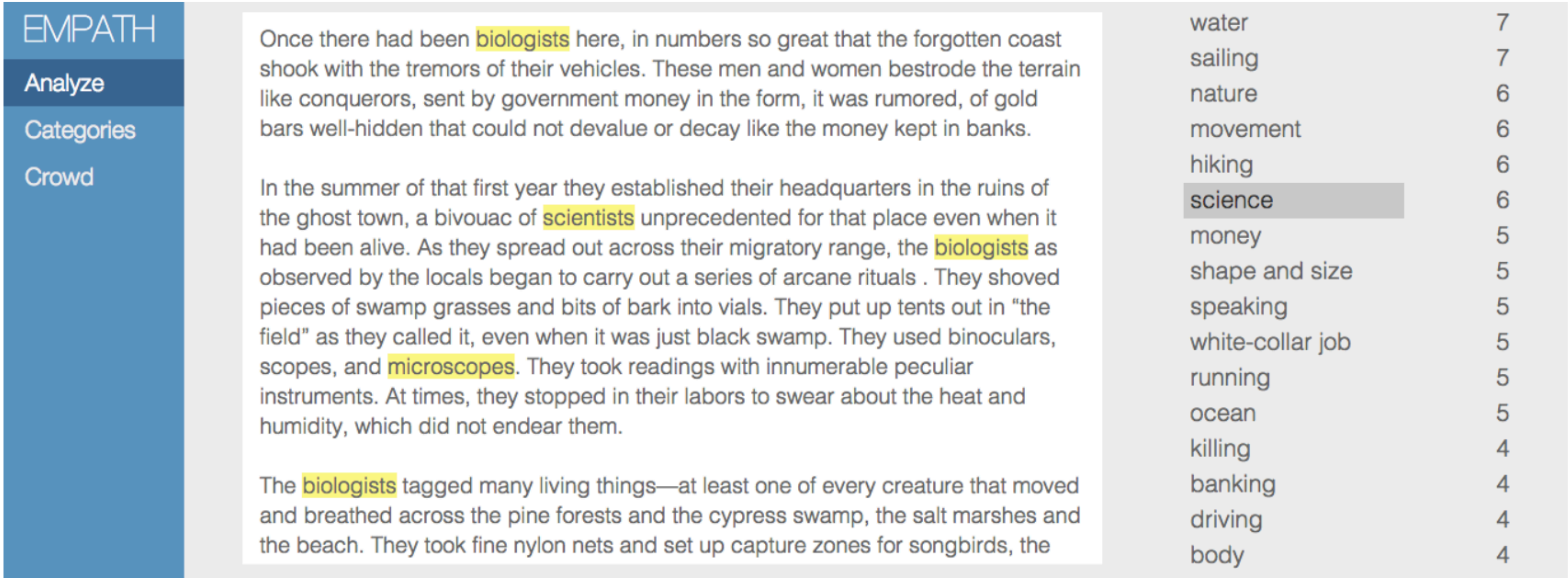
Brysbaert, M., Warriner, A. B., and Kuperman, V. (2014) Concreteness ratings for 40 thousand generally known English word lemmas Behavior Research Methods.

Concreteness versus Abstractness

Some example ratings from the final dataset of 40,000 words and phrases

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

Empath



Fast, Ethan, Binbin Chen, and Michael S. Bernstein. "Empath: Understanding topic signals in large-scale text." In Proceedings of the 2016 CHI conference on human factors in computing systems, pp. 4647-4657. 2016.

Empath

Generate categories from seed words using word embeddings

Broad set of 200 built-in categories:

```
Technology = {iPad, android, …}
Violence = {bleed, punch, …}
Government = {embassy, democrat, …}
```

When someone is punching the printer in the computer lab because of a paper jam.

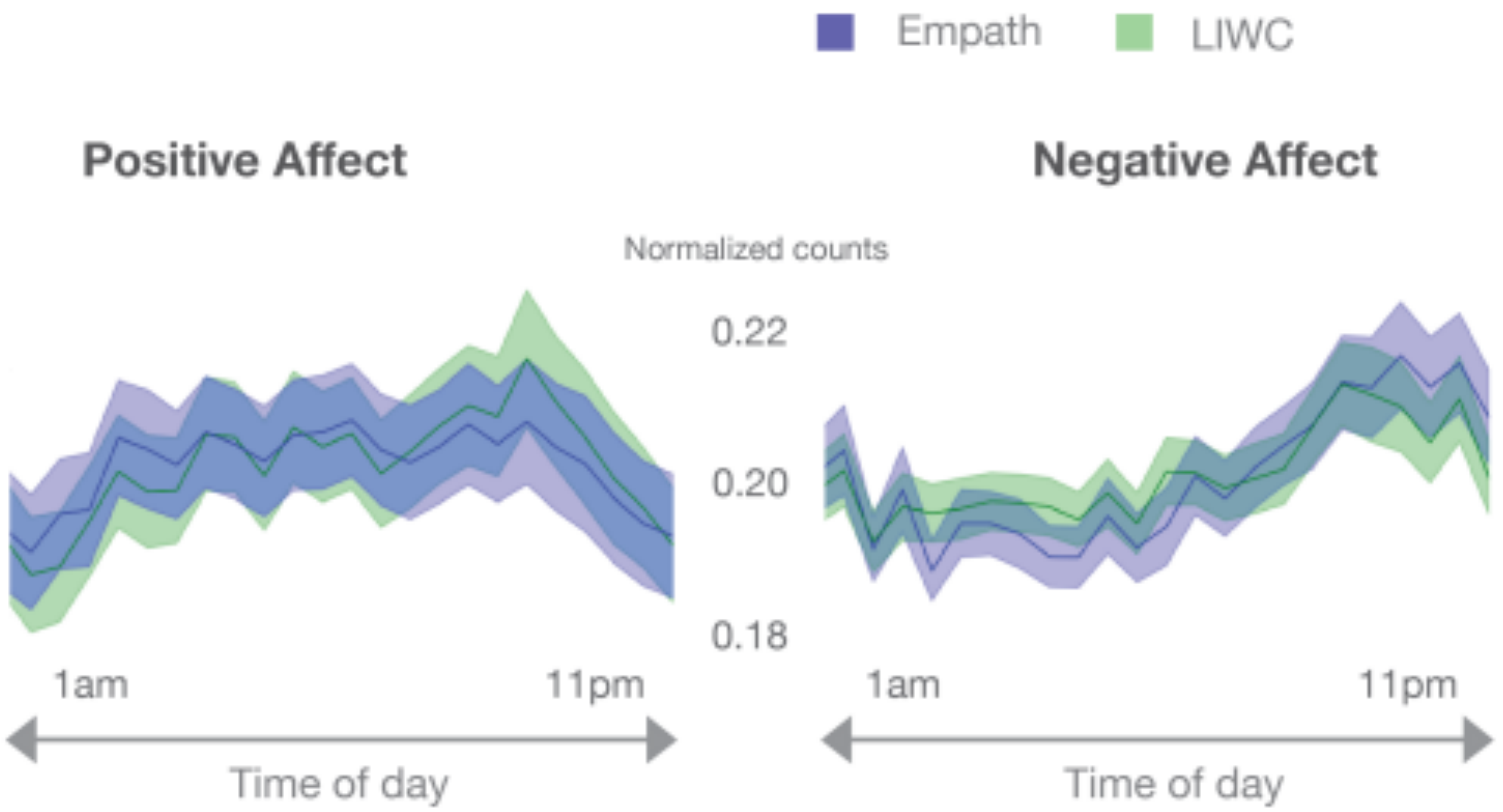
I'm scared to learn cause I'm scared of truth.

Spending days off class to online chat with course support. Help me Adeep!

Now Russia is the nation going thru the kicking out the ruling party drama.

social media	**war**	**violence**	**technology**	**fear**	**pain**	**hipster**	**contempt**
facebook	attack	hurt	ipad	horror	hurt	vintage	disdain
instagram	battlefield	break	internet	paralyze	pounding	trendy	mockery
notification	soldier	bleed	download	dread	sobbing	fashion	grudging
selfie	troop	broken	wireless	scared	gasp	designer	haughty
account	army	scar	computer	tremor	torment	artsy	caustic
timeline	enemy	hurting	email	despair	groan	1950s	censure
follower	civilian	injury	virus	panic	stung	edgy	sneer

Empath correlates with LIWC well



Lexicon based computing for sentiment/affect

Ratio of words in a sentence belonging to a category

The number of words in a given sentence belonging to a category k

$$r_k = \frac{\mathbf{n}_k}{\mathbf{n}}$$

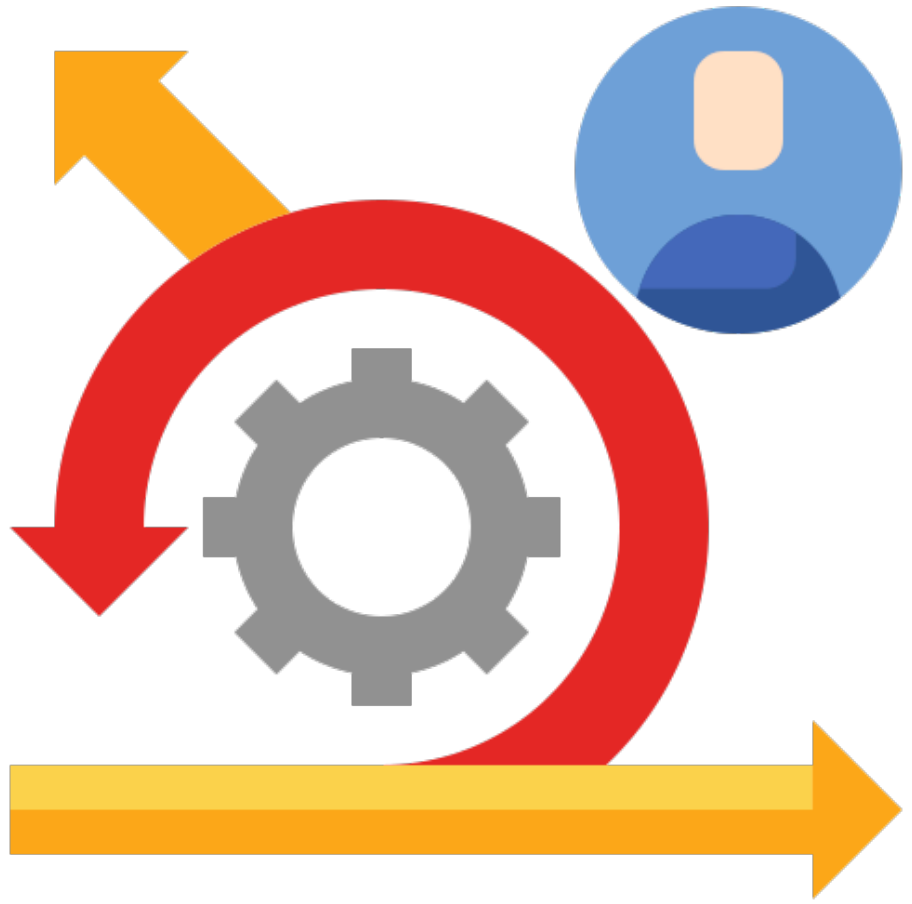
The total number of words in a given sentence

So far, only lexicon based approaches …

Supervised approaches exist

Or building lexicons via human annotation

Or semi-supervised induction



Semantic Axis Methods

Start with seed words like good or bad for the two poles

For each word to be added to lexicon

1. Compute a word representation
2. Use this to measure its distance from the poles
3. Assign it to the pole it is closer to

Initial Seeds for Different Domains

- Start with a single large seed lexicon and rely on the induction algorithm to fine-tune it to the domain
- Choose different seed words for different genres:

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

Computing word representation

Can just use off-the-shelf static embeddings

word2vec, GloVe, etc.

Or compute on a corpus

Or fine-tune pre-trained embeddings to a corpus

Representing each pole

Start with embeddings for seed words:

$$S^+ = \{E(w_1^+), E(w_2^+), ..., E(w_n^+)\}$$
$$S^- = \{E(w_1^-), E(w_2^-), ..., E(w_m^-)\}$$

Pole centroids are:

$$\mathbf{V}^+ = \frac{1}{n}\sum_1^n E(w_i^+)$$
$$\mathbf{V}^- = \frac{1}{m}\sum_1^m E(w_i^-)$$

Semantic axis is:

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

Word score is cosine with axis

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

Supervised Learning of Word Sentiment

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

Log odds ratio

Log likelihood ratio: does "horrible" occur more % in corpus i or j?

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

Log odds ratio

Log odds ratio: does "horrible" have a higher odds in corpus i or j?

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

Log odds ratio with a prior

Now with prior

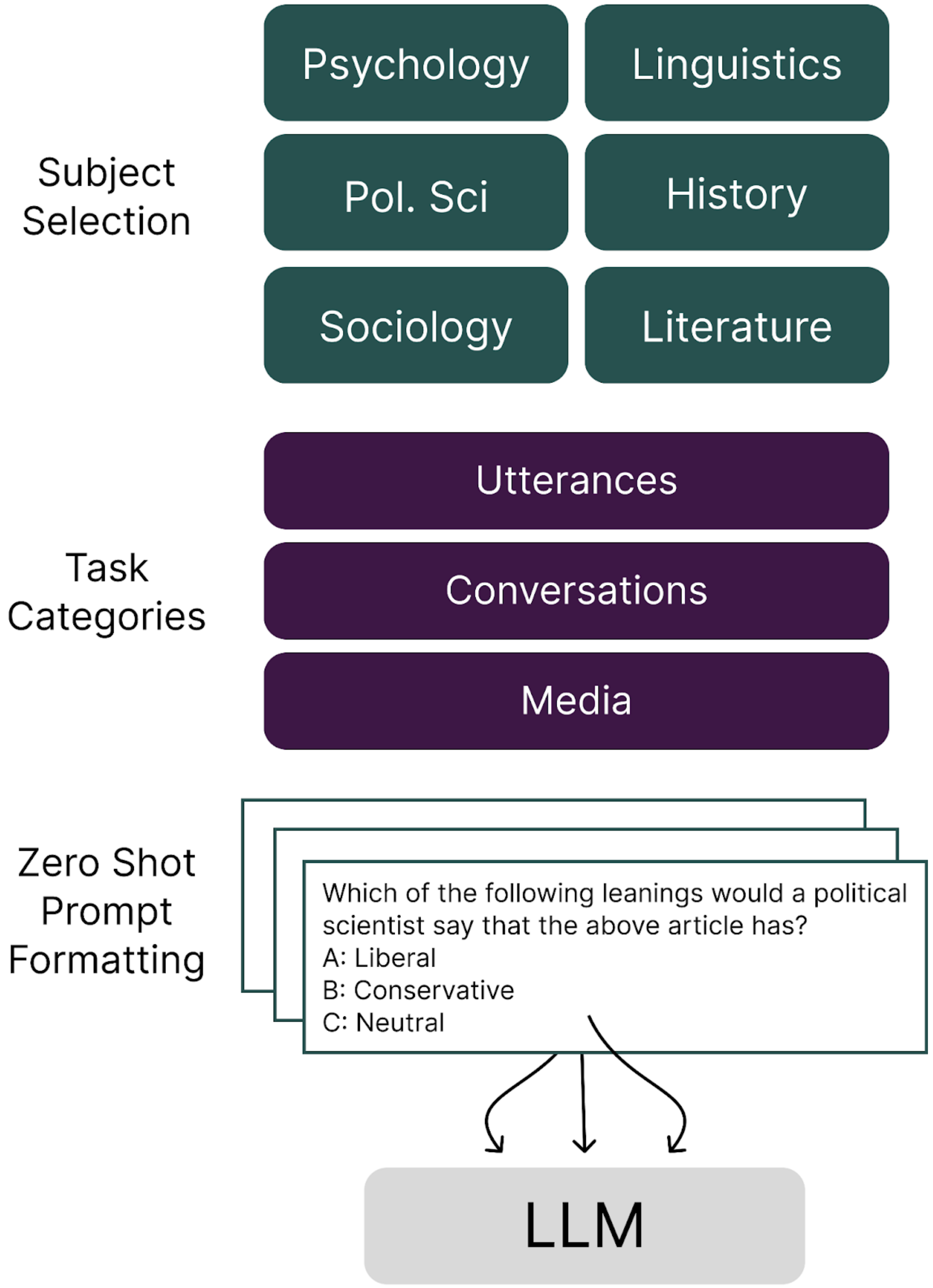
$$\delta_w^{(i-j)} = \log\left(\frac{f_w^i + \alpha_w}{n^i + \alpha_0 - (f_w^i + \alpha_w)}\right) - \log\left(\frac{f_w^j + \alpha_w}{n^j + \alpha_0 - (f_w^j + \alpha_w)}\right)$$

n^i= size of corpus *i*, n^j = size of corpus *j*, f_w^i= count of word *w* in corpus i, f_w^j= count of word *w* in corpus *j*, α_0 is the size of the background corpus, and α_w = count of word *w* in the background corpus.)

Top 50 words associated with bad (= 1-star) reviews

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*

Jurafsky, D., V. Chahuneau, B. R. Routledge, and N. A. Smith. 2014. Narrative framing of consumer sentiment in online restaurant reviews. First Monday, 19(4).



Using LLMs for Emotion Understanding

Ziems, Caleb, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, and Diyi Yang. "Can large language models transform computational social science?." Computational Linguistics (2024): 1-55.

Using LLMs for Emotion Understanding

Model / Data	Baselines		FLAN-T5					FLAN	text-001				text-002	text-003	Chat	
	Rand	Finetune	Small	Base	Large	XL	XXL	UL2	Ada	Babb.	Curie	Dav.	Davinci	Davinci	GPT3.5	GPT4
Utterance Level Tasks																
Dialect	3.3	3.0	0.2	4.5	23.4	24.8	30.3	32.9	0.5	0.5	1.2	9.1	17.1	14.7	11.7	23.2
Emotion	16.7	71.6	19.8	63.8	69.7	65.7	66.2	70.8	6.4	4.9	6.6	19.7	36.8	44.0	47.1	50.6
Figurative	25.0	99.2	16.6	23.2	18.0	32.2	53.2	62.3	10.0	15.2	10.0	19.4	45.6	57.8	48.6	17.5
Humor	49.5	73.1	51.8	37.1	54.9	56.9	29.9	56.8	38.7	33.3	34.7	29.2	29.7	33.0	43.3	61.3
Ideology	33.3	64.8	18.6	23.7	43.0	47.6	53.1	46.4	39.7	25.1	25.2	23.1	46.0	46.8	43.1	60.0
Impl. Hate	16.7	62.5	7.4	14.4	7.2	32.3	29.6	32.0	7.1	7.8	4.9	9.2	18.4	19.2	16.3	3.7
Misinfo	50.0	81.6	33.3	53.2	64.8	68.7	69.6	77.4	45.8	36.2	41.5	42.3	70.2	73.7	55.0	26.9
Persuasion	14.3	52.0	3.6	10.4	37.5	32.1	45.7	43.5	3.6	5.3	4.7	11.3	21.6	17.5	23.3	56.4
Sem. Chng.	50.0	62.3	33.5	41.0	56.9	52.0	36.3	41.6	32.8	38.9	41.3	35.7	41.9	37.4	44.2	21.2
Stance	33.3	36.1	25.2	36.6	42.2	43.2	49.1	48.1	18.1	17.7	17.2	35.6	46.4	41.3	48.0	76.0
Conversation Level Tasks																
Discourse	14.3	49.6	4.2	21.5	33.6	37.8	50.6	39.6	6.6	9.6	4.3	11.4	35.1	36.4	35.4	16.7
Empathy	33.3	71.6	16.7	16.7	22.1	21.2	35.9	34.7	24.5	17.6	27.6	16.8	16.9	17.4	22.6	6.4
Persuasion	50.0	33.3	9.2	11.0	11.3	8.4	41.8	43.1	6.9	6.7	6.7	33.3	33.3	53.9	51.7	28.6
Politeness	33.3	75.8	22.4	42.4	44.7	57.2	51.9	53.4	16.7	17.1	33.9	22.1	33.1	39.4	51.1	59.7
Power	49.5	72.7	46.6	48.0	40.8	55.6	52.6	56.9	43.1	39.8	37.5	36.9	39.2	51.9	56.5	42.0
Toxicity	50.0	64.6	43.8	40.4	42.5	43.4	34.0	48.2	41.4	34.2	33.4	34.8	41.8	46.9	31.2	55.4
Document Level Tasks																
Event Arg.	22.3	65.1	–	–	–	–	–	–	–	–	8.6	8.6	21.6	22.9	22.3	23.0
Event Det.	0.4	75.8	9.8	7.0	1.0	10.9	41.8	50.6	29.8	47.3	47.4	44.4	48.8	52.4	51.3	14.8
Ideology	33.3	85.1	24.0	19.2	28.3	29.0	42.4	38.8	22.1	26.8	18.9	21.5	42.8	43.4	44.7	51.5
Tropes	36.9	-	1.7	8.4	13.7	14.6	19.0	28.6	7.7	12.8	16.7	15.2	16.3	26.6	36.9	44.9

Summary

- ✓ Emotion can be represented by fixed atomic units often called basic emotions, or as points in space defined by dimensions like valence and arousal.
- ✓ Affective lexicons can be built by hand, using crowd sourcing to label the affective content of each word.
- ✓ Lexicons can be built with semi-supervised, bootstrapping from seed words using similarity metrics like embedding cosine.
- ✓ Lexicons can 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, and ratio metrics like log odds ratio informative Dirichlet prior