SENTIMENT ANALYSIS

Kristina Lerman
USC Information Sciences Institute
DSCI 552 – Spring 2021
March 10, 2021

Why text analysis

- Volume of text data is growing exponentially, necessitating methods for automatically organizing, understanding, searching and summarizing them
  - Uncover hidden topical patterns in collections.
  - Annotate documents according to topics.
  - Using annotations to organize, summarize and search.
- Application: Use text data to learn about people
  - Feelings, Moods → Sentiment analysis, affective computing
  - Conceptual biases
  - …

What and why

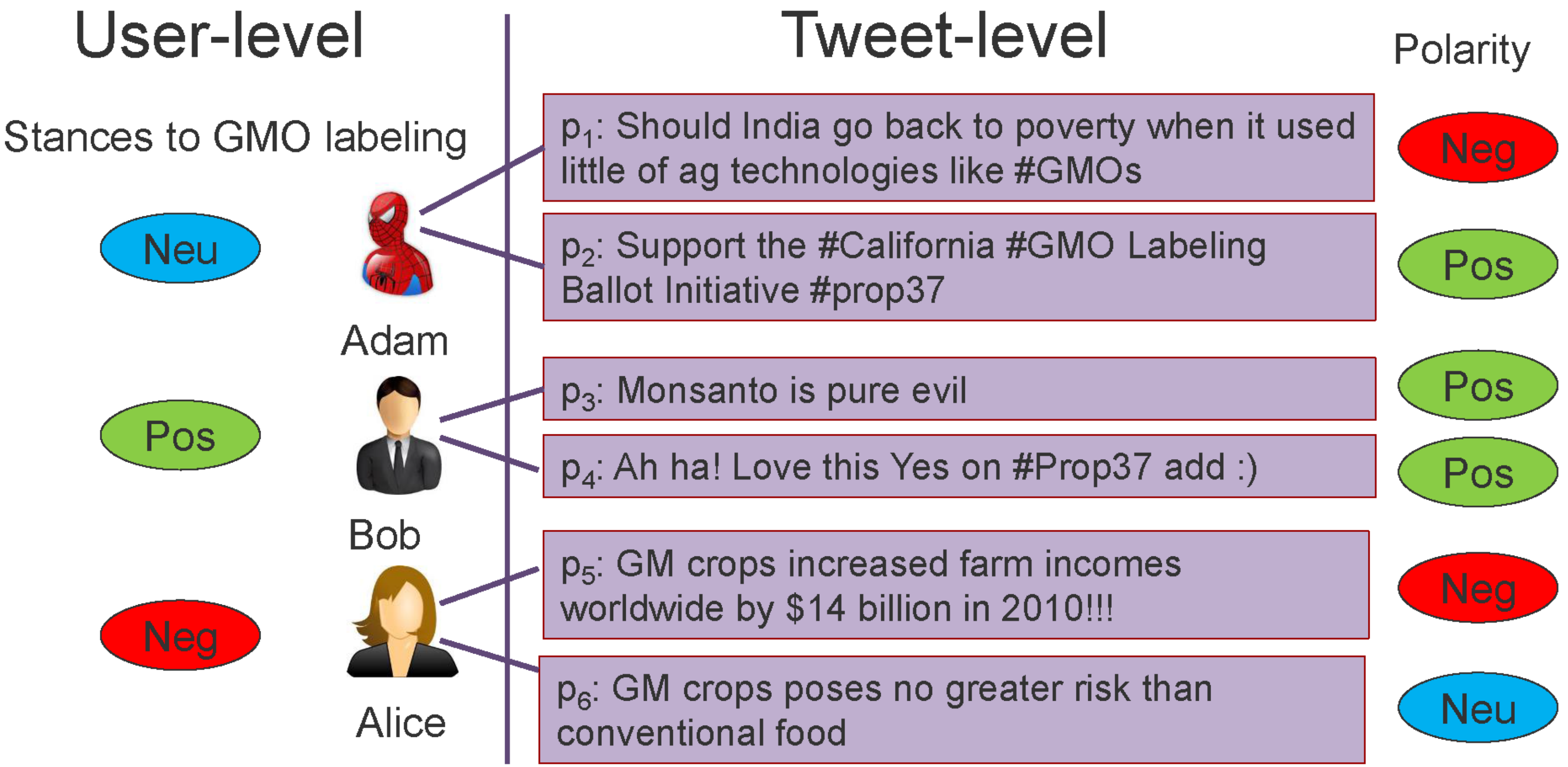
- What is sentiment analysis
- Tools
- Applications
- Fairness aspects

How do people feel about movies?

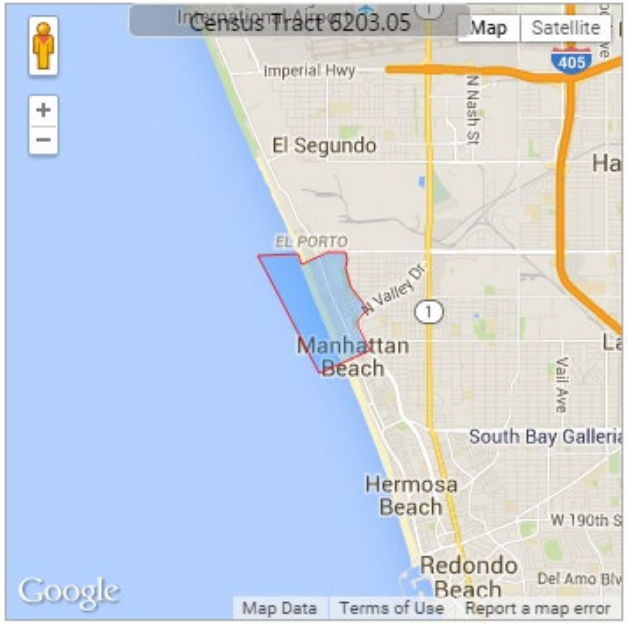
Estimate this…

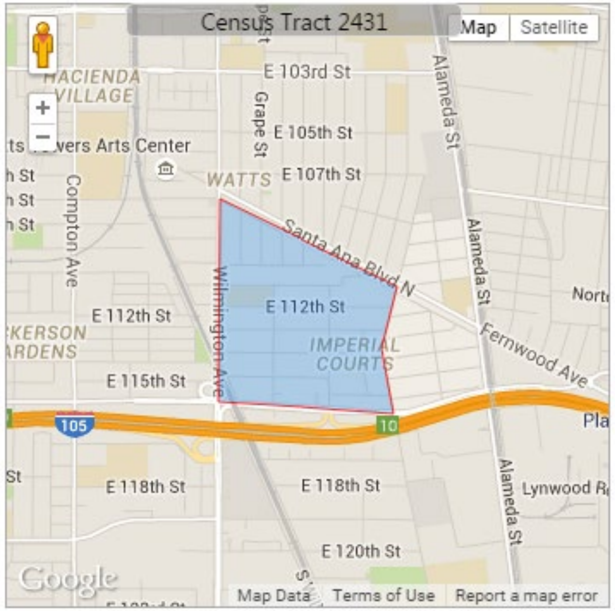
…using only this!

How do people feel about issues?



Where are people happier?





Other applications of sentiment analysis

- Help consumers and brands understand the opinions being expressed about
  - Events
    - Court decisions, protests, acts of congress
  - Products
    - Movies, consumer electronics
  - People
    - Political candidates, Dictators
  - Locations
    - Restaurants, hotels, vacation destinations

Mood and emotion

- Emotions are physiological in origin
  - Influenced by levels of neurotransmitters, hormones, …
  - Moods are emotional feelings lasting for days
- Emotions also depend on external factors
  - Daily routing, work, commuting, eating, …
  - Products used by a person
- Two dimensions of emotion
  - Positive affect
    - Enthusiasm, delight, activeness, alertness, happiness, …
  - Negative affect
    - Distress, fear, anger, guilt, disgust, sadness, …
- Can we accurately measure emotions from text?

Main ideas

- Text messages (tweets, blog posts) use distinctive words to convey emotions
  - Identify features (words, linguistic features) that are highly indicative of emotions
- Train classifier to recognize emotion in text
  - Supervised machine learning
    - Need labeled data to train classifier
    - Features are noisy. How to filter them to improve classifier performance?
    - What classifier to use?
  - Automatically classify the emotion of a new text message using only the features of the message

Sentiment of Twitter posts

["Twitter as a Corpus for Sentiment Analysis and Opinion Mining" by Pak & Paroubek]

- Main idea
  - People widely use microblogging platforms (e.g., Twitter) to express opinions. Understanding opinions would be useful for marketing and social sciences
  - But, it is challenging to extract sentiment from microblog posts, because they are very short (e.g., 140 characters)
- Contributions
  - Automatically collect training data from Twitter
  - Use linguistic features to automatically recognize the sentiment of posts
    - Positive, negative, objective

Twitter sentiment

- Twitter posts often express opinions
  - Which posts express positive sentiment? Negative sentiment?

funkeybrewster: @redeyechicago I think Obama's visit might've sealed the victory for Chicago. Hopefully the games mean good things for the city.	+
vcurve: I like how Google celebrates little things like this: Google.co.jp honors Confucius Birthday — Japan Probe	+
mattfellows: Hai world. I hate faulty hardware on remote systems where politics prevents you from moving software to less faulty systems.	−
brroooklyn: I love the sound my iPod makes when I shake to shuffle it. Boo bee boo	+
MeganWilloughby: Such a Disney buff. Just found out about the new Alice in Wonderland movie. Official trailer: http://bit.ly/131Js0 I love the Cheshire Cat.	+

- Posts are short: few words to go by to recognize an opinion

Sentiment classification

- Train classifier to recognize positive and negative sentiment
- But, need lots of training data containing posts expressing positive and negative opinions, as well as objective posts not expressing an opinion

Training data collection – **ground truth**

- Query Twitter for posts containing
  - Happy emoticons… :-), :), =), :D, … → subjective post, expressing positive affect
  - Sad emoticons… :-(, :(, =(, ;(, … → subjective post, expressing negative affect
  - Links to news articles → objective posts

Do linguistic features help?

- Linguistic analysis of words
  - Part-of-speech (POS) **tag**: noun, adjective, personal pronoun, verb, …

n adj

- Study the distribution of POS tags in each data set
  - Compare the prevalence of a tag T across two data sets

$$\left[P_{1,2} = \frac{N_1 - N_2}{N_1 + N_2} \right]$$

ex. $N_1 = \oplus$
$N_2 = \ominus$

  - Where N_1 = number of occurrences of tag T in set 1 (e.g., positive posts)
  - And N_2 = number of occurrences of tag T in set 2 (e.g., negative posts)

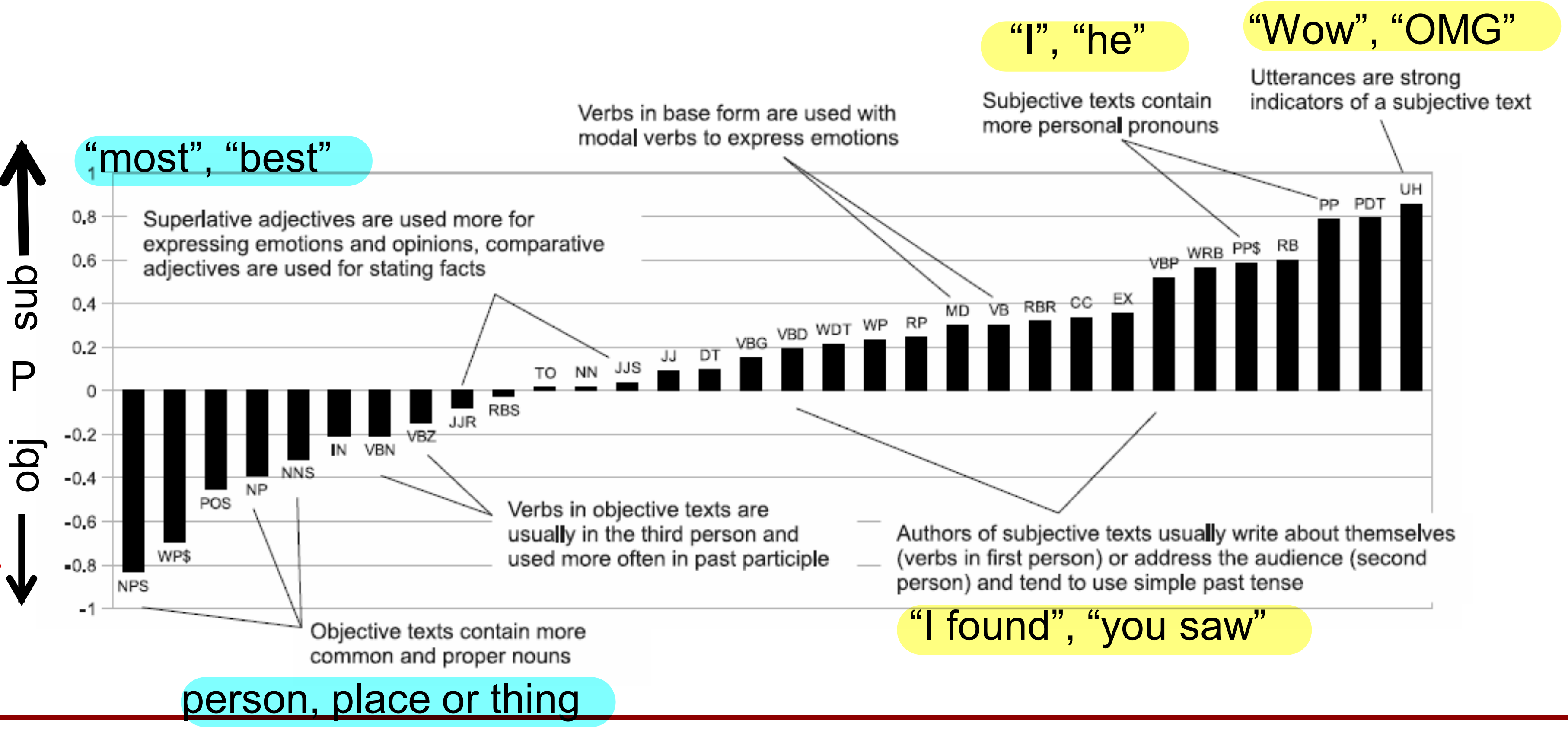
Subjective vs objective posts

主观 客观

- Relative prevalence of POS tags across subjective posts (positive or negative) and objective posts

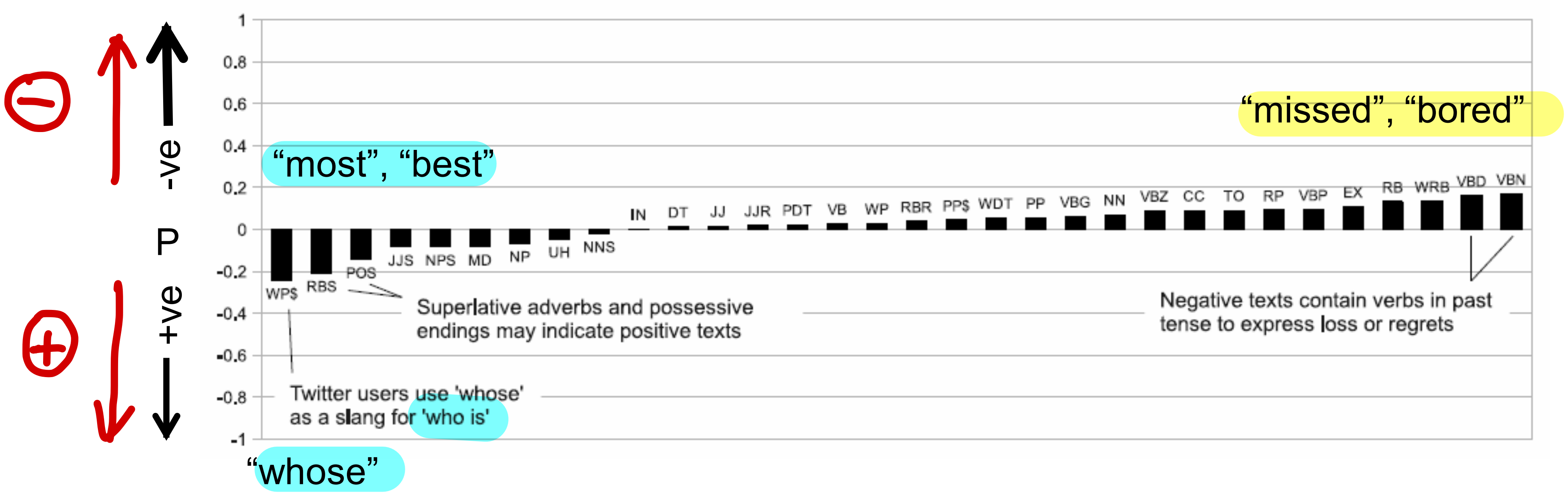
主 sub

客 obj

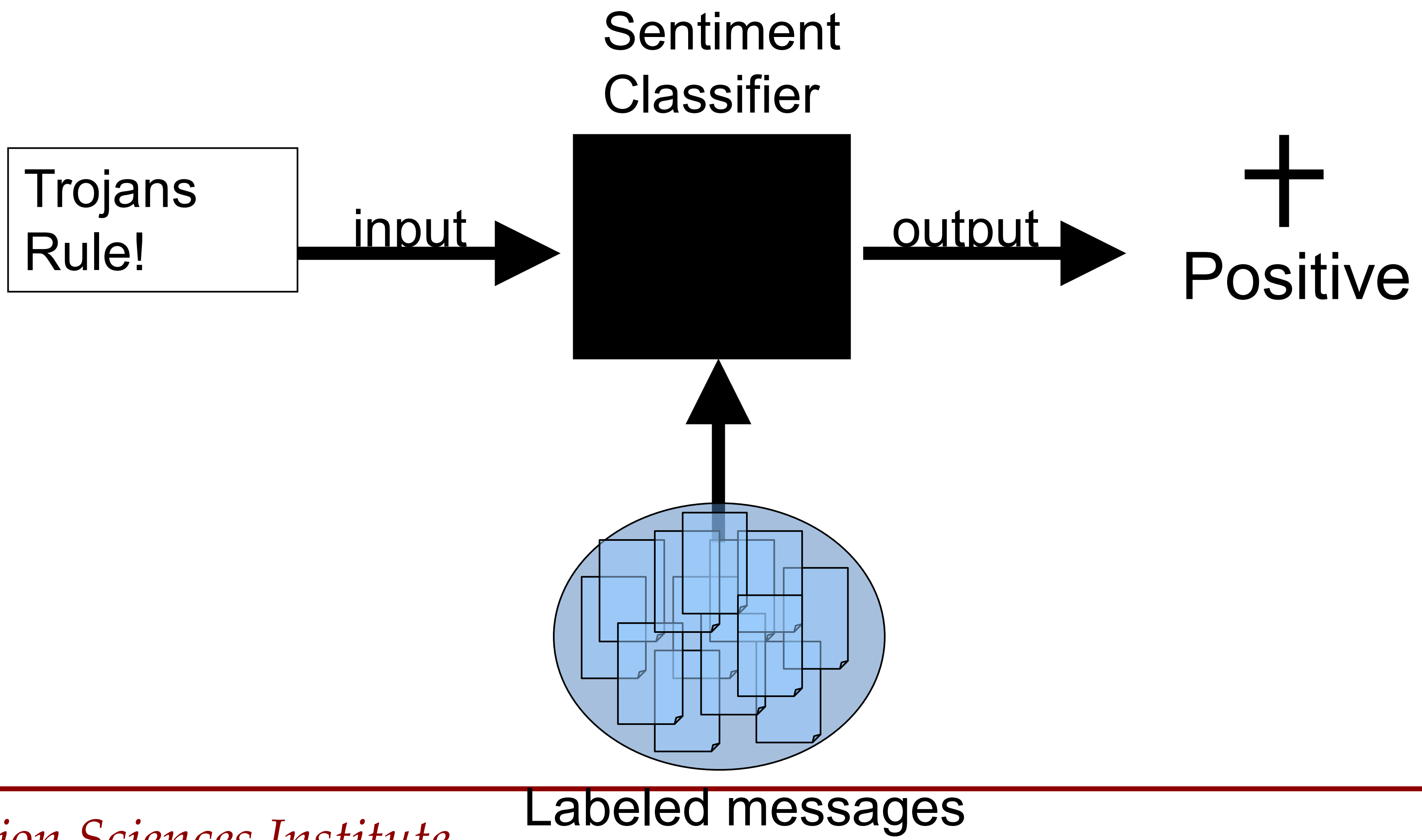


Negative vs Positive

- Relative prevalence of POS tags across negative and positive posts
- Prevalence has less discriminative power than for objective vs subjective posts



Supervised Machine Learning



Given a message M, what is its sentiment s?

- Classifier 1: Naïve Bayes with words only

$$P(s|M) = \frac{P(s)P(M|s)}{P(M)} \sim P(M|s) \begin{matrix} + \\ - \end{matrix}$$

  – Assume that n-grams *g* in a message *M* are conditionally independent, given sentiment *s*

$$P(s|M) \sim \prod_{g \in M} P(g|s)$$

P(+|M) P(g|+)
P(-|M) P(g|-)

- Classifier 2: Naïve Bayes with words and linguistic tags
  – Assume tags *t* are conditionally independent given *s*

$$P(s|M) \sim P(G|s)P(T|S) \sim \prod_{g \in M} P(g|s) \prod_{t \in M} P(t|s)$$

- Calculate P(+|M), P(-|M), P(objective|M)

P(+|M) P(-|M) P(objective|M)

$P(+|\text{trojan}) = \frac{6}{16}$ $P(-|\text{trojan}) = \frac{5}{16}$

Unigram (1-gram)	**Positive Message Count**	**Negative Message Count**	**Objective Message Count**
trojans	6	5	5
rule	22	6	25
great	40	1	2
home	10	10	10
bad	2	30	2
news	3	7	44
Total count	5000	5000	5000

Example of calculating *P*(*s*|*M*)

P(+|"trojans rule") = P(+)* product of probabilities P(unigrams|+)

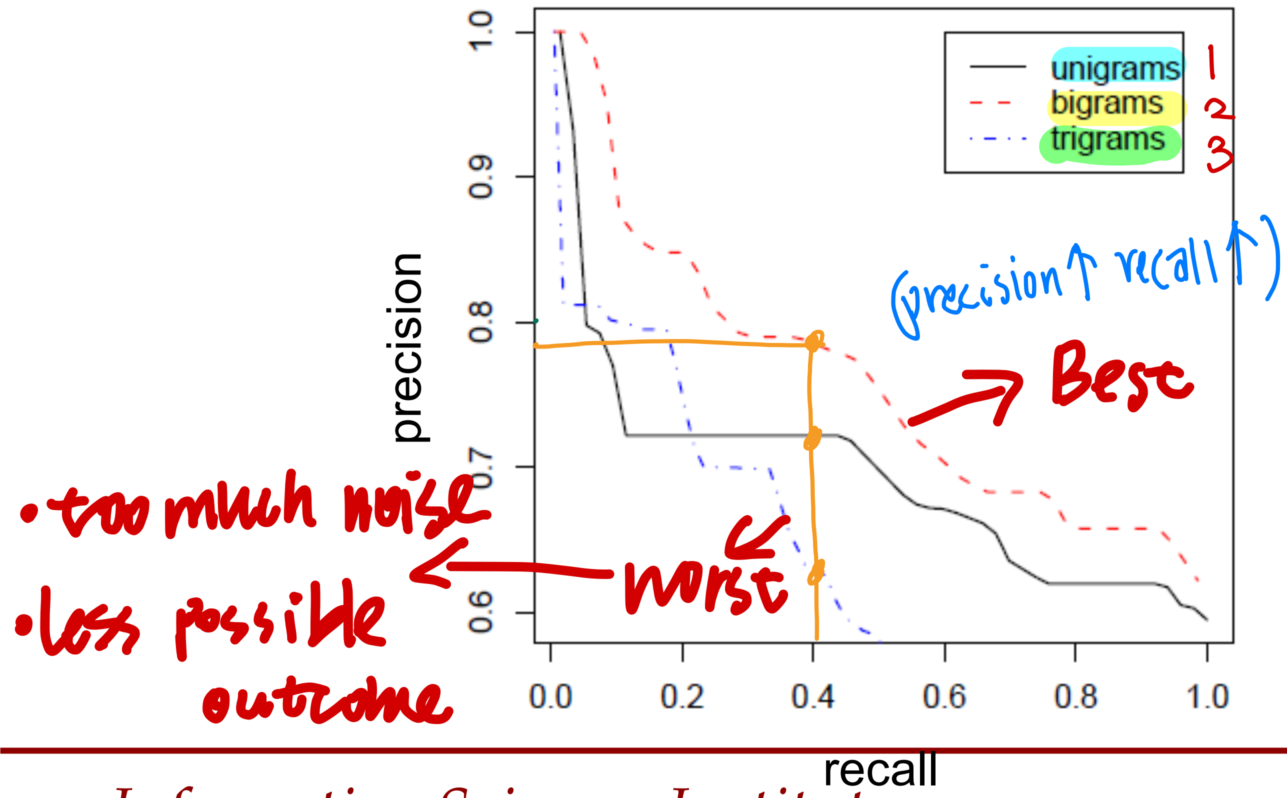
= P(+) * P("trojans"|+) * P("rule"|+)

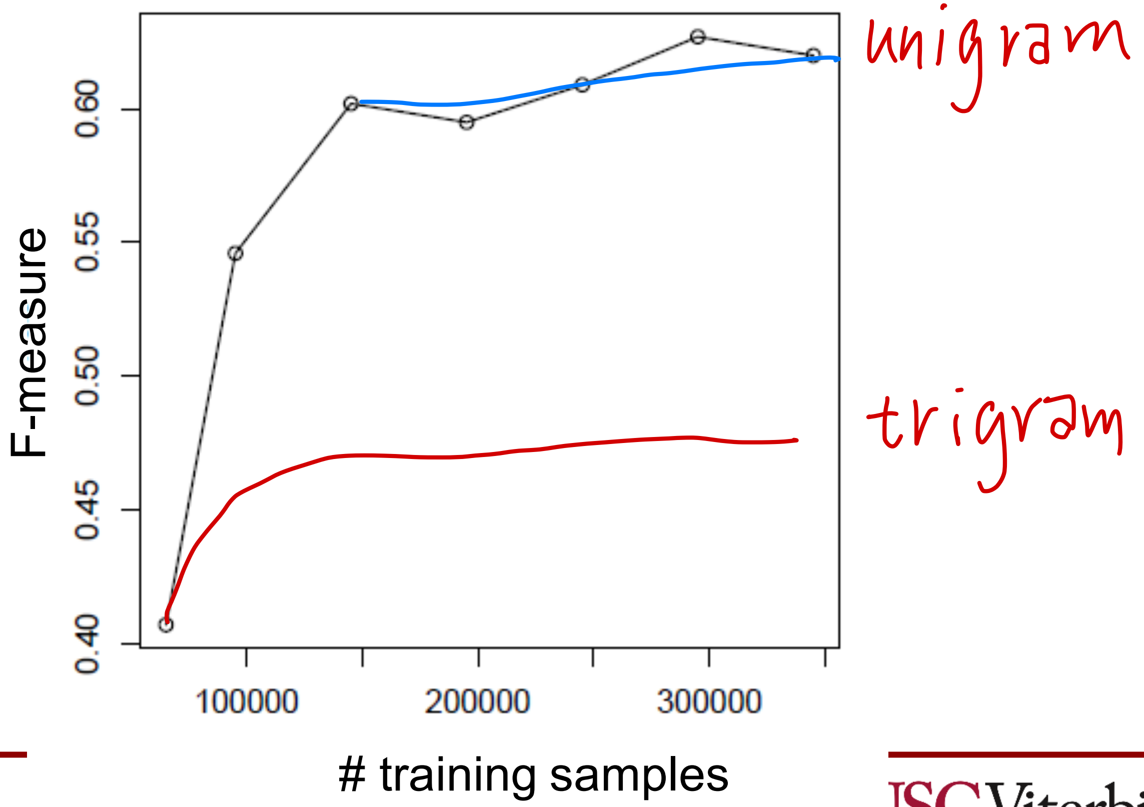
$$=\underset{\frac{1}{3}}{0.333} * 6/\underset{16}{\cancel{5000}} * 22/\underset{25}{\cancel{5000}}$$

- Similarly for P(-|"trojans rule") and P(obj|"trojans rule")
    - Sentiment with the largest probability wins.

Results

- Classify the sentiment of 200 messages.
- Ground truth: messages were manually annotated for their sentiment





Summary

- Authors of Twitter messages use linguistic features to describe emotions (positive or negative sentiment messages) or state facts (objective messages)
  - Some part-of-speech tags may be strong indicators of emotional text
- Use examples of positive, negative, and objective messages collected from Twitter to train a classifier
  - Recognize sentiment of a new message based on its words and POS tags

Sentiment analysis tools

- LIWC: dictionary approach
  - Words annotated with over 80 psychological categories, including positive and negative sentiment
- WKB lexicon → general tool
  - 14,000 English words annotated with emotional valence (positivity/negativity) and arousal (strength of emotion).
  - Also exists in Spanish
- SentiStrength
  - Positive and negative sentiment
- Vader
  - Valence Aware Dictionary for sEntiment Reasoning

Social media

① Measuring emotions: Warriner's (WKB) lexicon

- Sentiment lexicon
- Three dimensions of emotion
  - Valence
  - Arousal
  - Dominance

3维

- Collected emotional ratings for 14,000 words, the majority of the well-known English content words
  - Participants recruited through Amazon Mechanical Turk
  - shown lists of 350 words, including 10 calibrator words
  - asked to rate each word along 3 dimensions using a 9 pt scale

[Wariner et el., (2013) "Norms of valence, arousal, and dominance for 13,915 English lemmas"]

How do you feel while reading each word?

- Valence
  - 9: Happy, pleased, satisfied, contented, hopeful
  - 1: Unhappy, annoyed, melancholic, despaired, or bored
- Arousal
  - 9: Excited, stimulated, excited, frenzied, jittery, wide-awake
  - 1: Calm, relaxed, sluggish, dull, sleepy, or unaroused
- Dominance (in-control / out-control)
  - 9: In control, influential, important, dominant, autonomous
  - 1: Controlled, influenced, cared-for, submissive, or guided
- 5: Neutral, neither happy nor sad [not excited nor at all calm; neither in control nor controlled]

ex

	Valence		Arousal		Dominance	
Lowest	pedophile	1.26	grain	1.60	dementia	1.68
	rapist	1.30	dull	1.67	Alzheimer’s	2.00
	AIDS	1.33	calm	1.67	lobotomy	2.00
	leukemia	1.47	librarian	1.75	earthquake	2.14
	molester	1.48	soothing	1.91	uncontrollable	2.18
	murder	1.48	scene	1.95	rapist	2.21
Highest	excited	8.11	motherfucker	7.33	rejoice	7.68
	sunshine	8.14	erection	7.37	successful	7.71
	relaxing	8.19	terrorism	7.42	smile	7.72
	lovable	8.26	lover	7.45	completion	7.73
	fantastic	8.36	rampage	7.57	self	7.74
	happiness	8.48	insanity	7.79	incredible	7.74

(2) Measuring emotions: LIWC

- Linguistic Inquiry and Word Count
- James W. Pennabaker, U. Texas @ Austin
  - "Virtually no one in psychology has realized that low-level words can give clues to large-scale behaviors"
  - Book: *The Secret Life of Pronouns* (2011)
- 4,500 words and word stems
  - Each in one or more psychological categories
    - "cried" in sadness, negative emotion, overall affect, verb, past tense verb.
    - More than 80 categories, including positive, negative affect

LIWC Category	Examples	No. of Words
Positive Emotion	Love, nice, good, great	406
Negative Emotion	Hurt, ugly, sad, bad, worse	499

Testing LIWC Online

http://liwc.net/liwcresearch07.php

Gender of author: Male *Age of author:* 30

Type or paste the text you want analysed into the box below and then hit the submit button.

We identified individual-level diurnal and seasonal mood rhythms in cultures across the globe, using data from millions of public Twitter messages. We found that individuals awaken in a good mood that deteriorates as the day progresses—which is consistent with the effects of sleep and circadian rhythm—and that seasonal change in baseline positive affect varies with change in daylength. People are happier on weekends, but the morning peak in positive

Submit text for online LIWC analysis

LIWC Results

Details of Writer: 30 year old Male
Date/Time: 7 February 2012, 4:56 pm

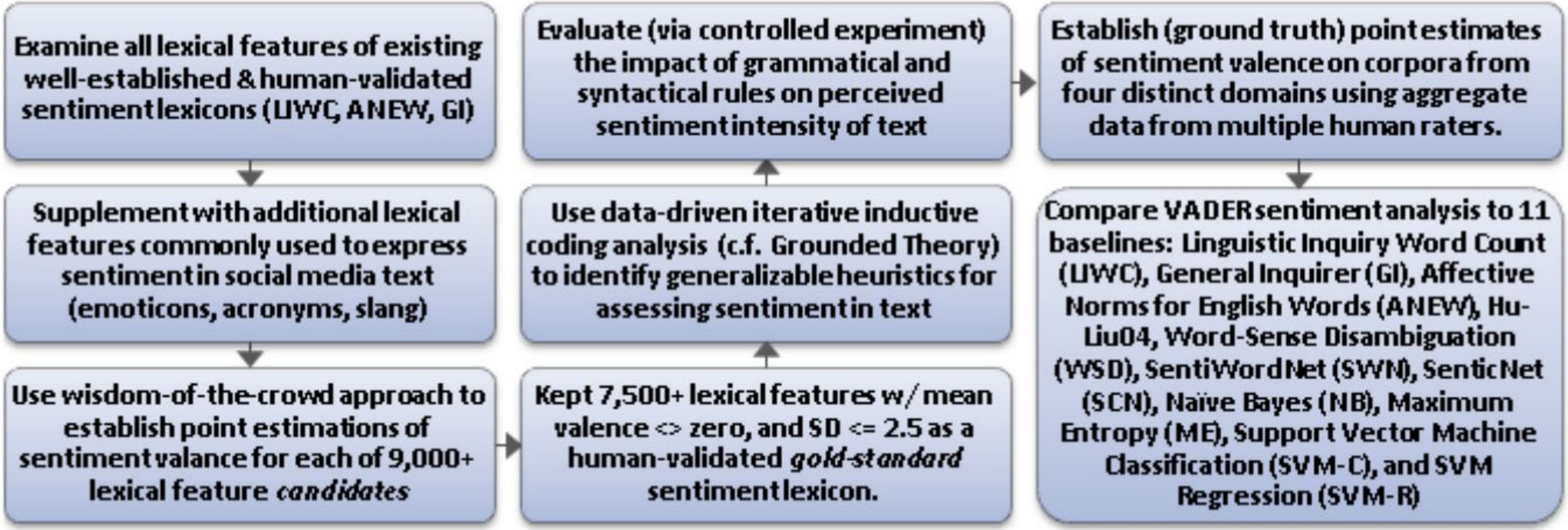
LIWC Dimension	***Your Data***	***Personal Texts***	***Formal Texts***
Self-references (I, me, my)	2.33	11.4	4.2
Social words	9.30	9.5	8.0
Positive emotions	4.65	2.7	2.6
Negative emotions	0.00	2.6	1.6
Overall cognitive words	6.98	7.8	5.4
Articles (a, an, the)	5.81	5.0	7.2
Big words (> 6 letters)	31.40	13.1	19.6

The text you submitted was 86 words in length.

3 VADER (emotion)

- gold-standard sentiment lexicon that is especially attuned to microblog-like contexts.
- Rules for grammatical and syntactical conventions that humans use when expressing or emphasizing sentiment intensity
- Performs well on social media posts (better than LIWC)

VADER



Examine all lexical features of existing well-established & human-validated sentiment lexicons (LIWC, ANEW, GI)
Supplement with additional lexical features commonly used to express sentiment in social media text (emoticons, acronyms, slang)
Use wisdom-of-the-crowd approach to establish point estimations of sentiment valance for each of 9,000+ lexical feature *candidates*
Kept 7,500+ lexical features w/ mean valence <> zero, and SD <= 2.5 as a human-validated *gold-standard* sentiment lexicon.
Use data-driven iterative inductive coding analysis (c.f. Grounded Theory) to identify generalizable heuristics for assessing sentiment in text
Evaluate (via controlled experiment) the impact of grammatical and syntactical rules on perceived sentiment intensity of text
Establish (ground truth) point estimates of sentiment valence on corpora from four distinct domains using aggregate data from multiple human raters.
Compare VADER sentiment analysis to 11 baselines: Linguistic Inquiry Word Count (LIWC), General Inquirer (GI), Affective Norms for English Words (ANEW), Hu-Liu04, Word-Sense Disambiguation (WSD), SentiWordNet (SWN), SenticNet (SCN), Naïve Bayes (NB), Maximum Entropy (ME), Support Vector Machine Classification (SVM-C), and SVM Regression (SVM-R)

APPLICATIONS

Global mood patterns

"Diurnal and seasonal moods vary with work, sleep and daylength across diverse cultures" by Golder and Macy

- Can automated sentiment analysis be applied to social media data to provide a global picture of human mood?
- Do moods have a time scale: diurnal, seasonal?

Corpus of Twitter tweets

→ data available

- Up to 400 public messages from each user
- 2.4 million individuals worldwide
- 509 million messages between 2/08-1/10
- 84 identified countries
- English only
- Date, Time, and country latitude

FB: data ✗ available

reddit: less use than FB / twitter

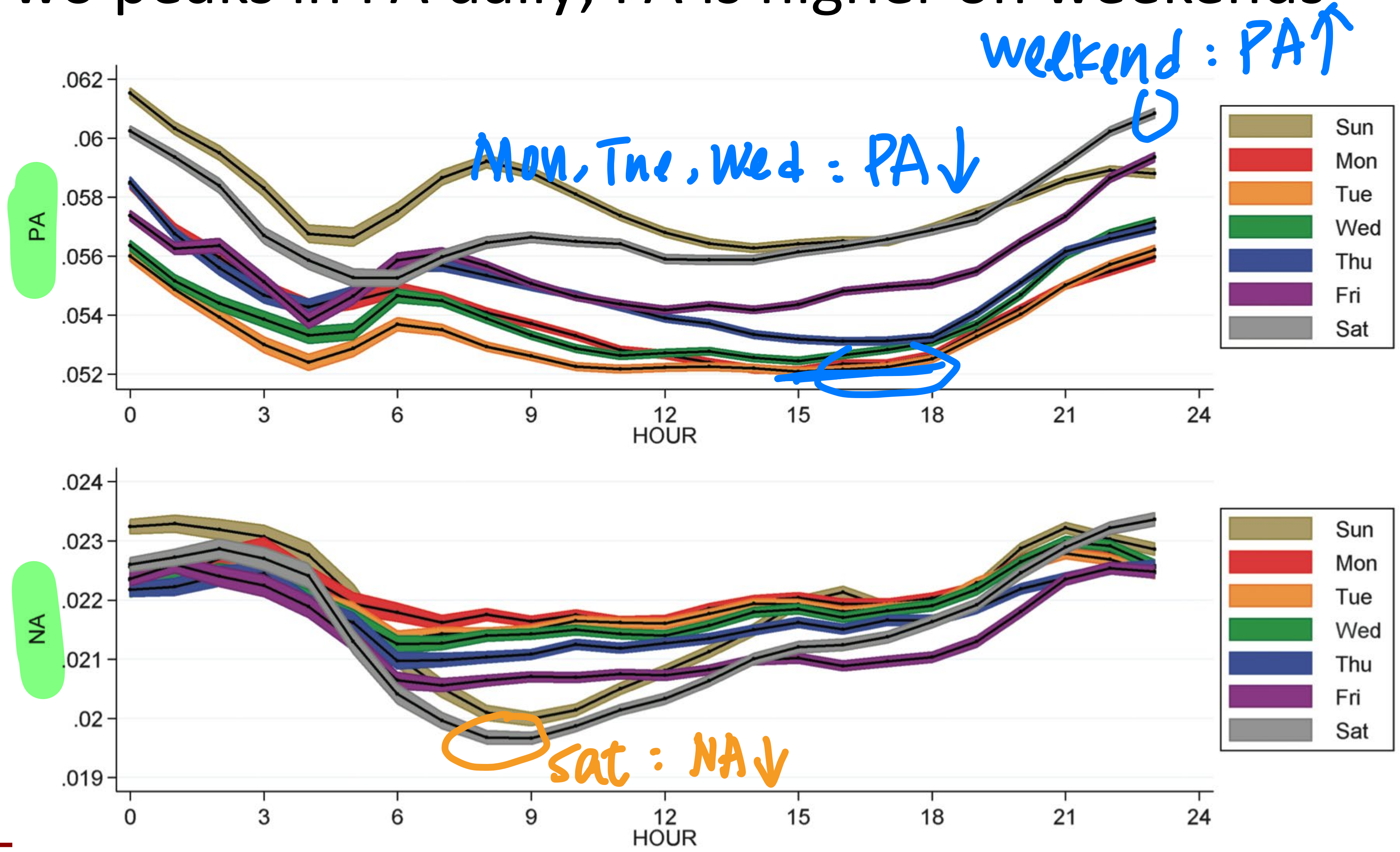
Methodology

- Examined within-individual Positive Affect (PA) and Negative Affect (NA) independently
  - E.g., fraction of PA words appearing in an individual's messages every hour

$$PA_u(h) = \frac{\|PAWORDS_u(h)\|}{\|WORDS_u(h)\|}$$

  - To eliminate between-individual variation, subtract the mean: $PA_u^*(h) = PA_u(h) - \langle PA_u(h) \rangle$
- Additional analysis on 4 English-speaking regions: Africa, India, UK/Aus, US/Can

Two peaks in PA daily; PA is higher on weekends



Informa **Fig. 1.** Hourly changes in individual affect broken down by day of the week (top, PA; bottom, NA). Each series shows mean affect (black lines) and 95% confidence interval (colored regions).

Mood governed by diurnal cycles, not culture

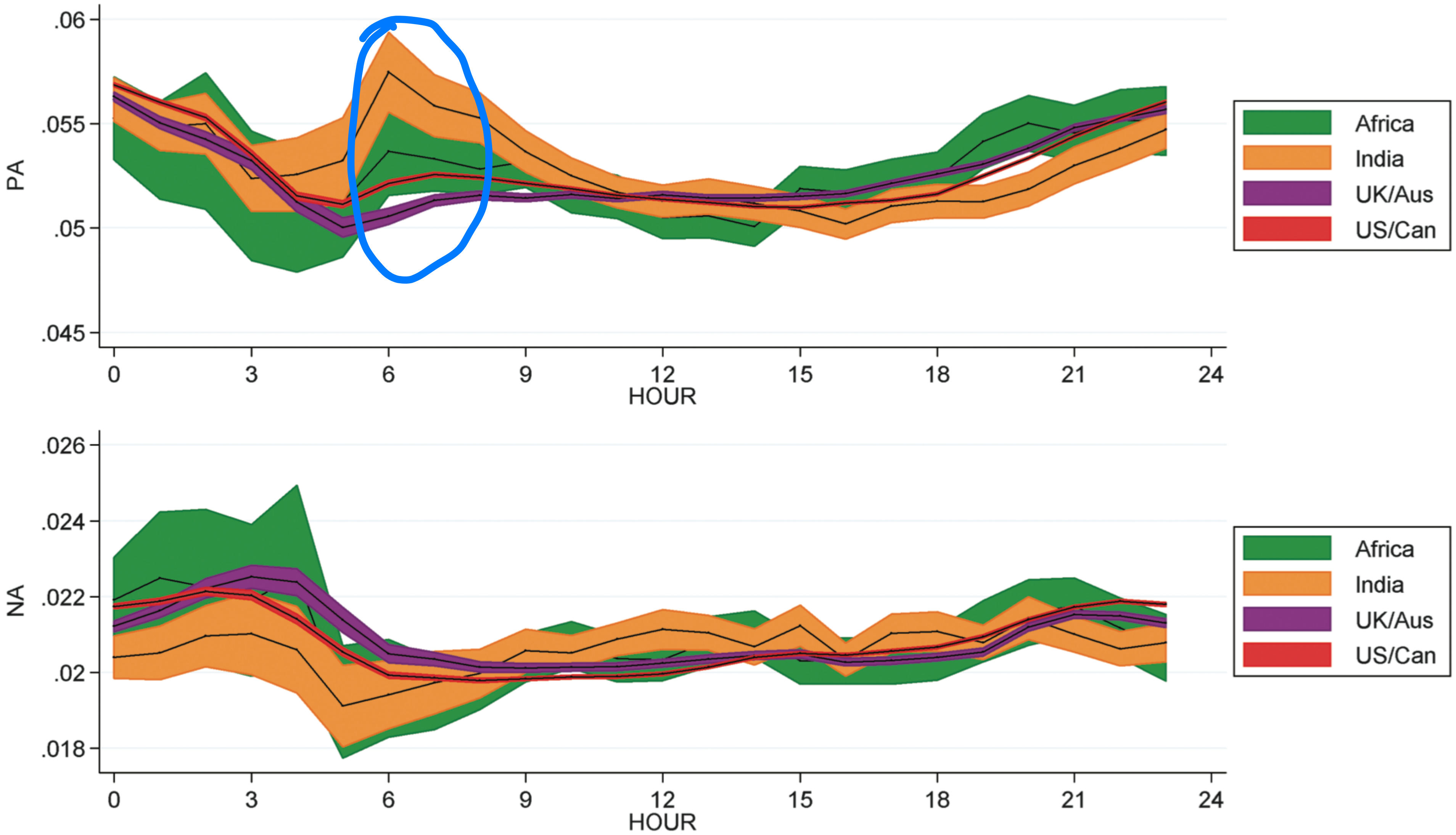


Fig. 2. Hourly changes in individual affect in four English-speaking regions. Each series shows mean affect (black lines) and 95% confidence interval (colored regions).

PA is higher when days are growing longer

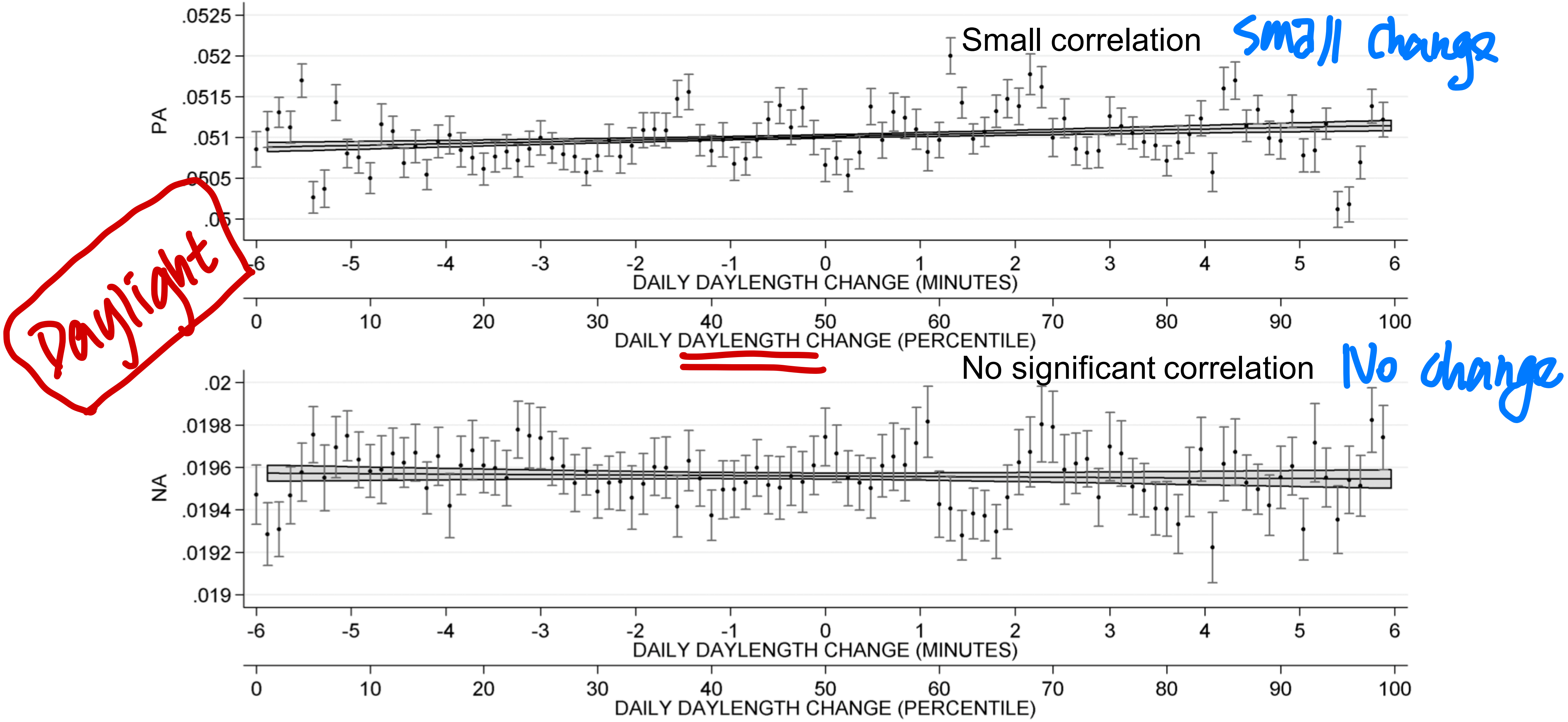


Fig. 3. Line of best fit through the 14.3 million person-month observations (affect by minutes gained or lost per day). For visual reference, 100 aggregate observations binned by percentiles are superimposed.

Summary

- Confirm findings from psychological studies
  - Psychology studies are small scale, on homogeneous population vs Twitter study on large, heterogeneous population
  - Mood changes are associated with diurnal (sleep-wake) cycles
    - PA highest in the morning and before midnight
    - PA highest on weekends
    - Universal and independent of culture
  - Seasonal mood changes
    - PA decreases as days grow shorter → "winter blues"
- Possible to do psychology through text analysis of social media data

FAIRNESS

Creating a racist sentiment classifier

http://blog.conceptnet.io/posts/2017/how-to-make-a-racist-ai-without-really-trying/

- Use pre-trained **word embeddings** to represent the meanings of words
  - GLOVE
- Acquire **ground truth data** for use in training and testing the algorithm
  - 6800 positive and negative words (maybe bias)
  - http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar
- **Train a classifier**, using gradient descent, to recognize other positive and negative words based on their word embeddings
  - 95% accuracy on words not in training data
  - Generalizes to words not in the ground truth data
- Compute **sentiment scores** for sentences of text using this classifier

sentiment	
fidget	-9.931679
interrupt	-9.634706
staunchly	1.466919
imaginary	-2.989215
taxing	0.468522
world-famous	6.908561
low-cost	9.237223
disapointment	-8.737182
totalitarian	-10.851580
bellicose	-8.328674
freezes	-8.456981
sin	-7.839670
fragile	-4.018289
fooled	-4.309344
undecided	-2.816172
handily	2.339609
demonizes	-2.102152
easygoing	8.747150
unpopular	-7.887475

-10 neg

+10 pos

Let’s analyze the sentiment of some texts

Restaurants

text_to_sentiment("Let's go get Italian food“) → 2.043

text_to_sentiment("Let's go get Chinese food") → 1.409

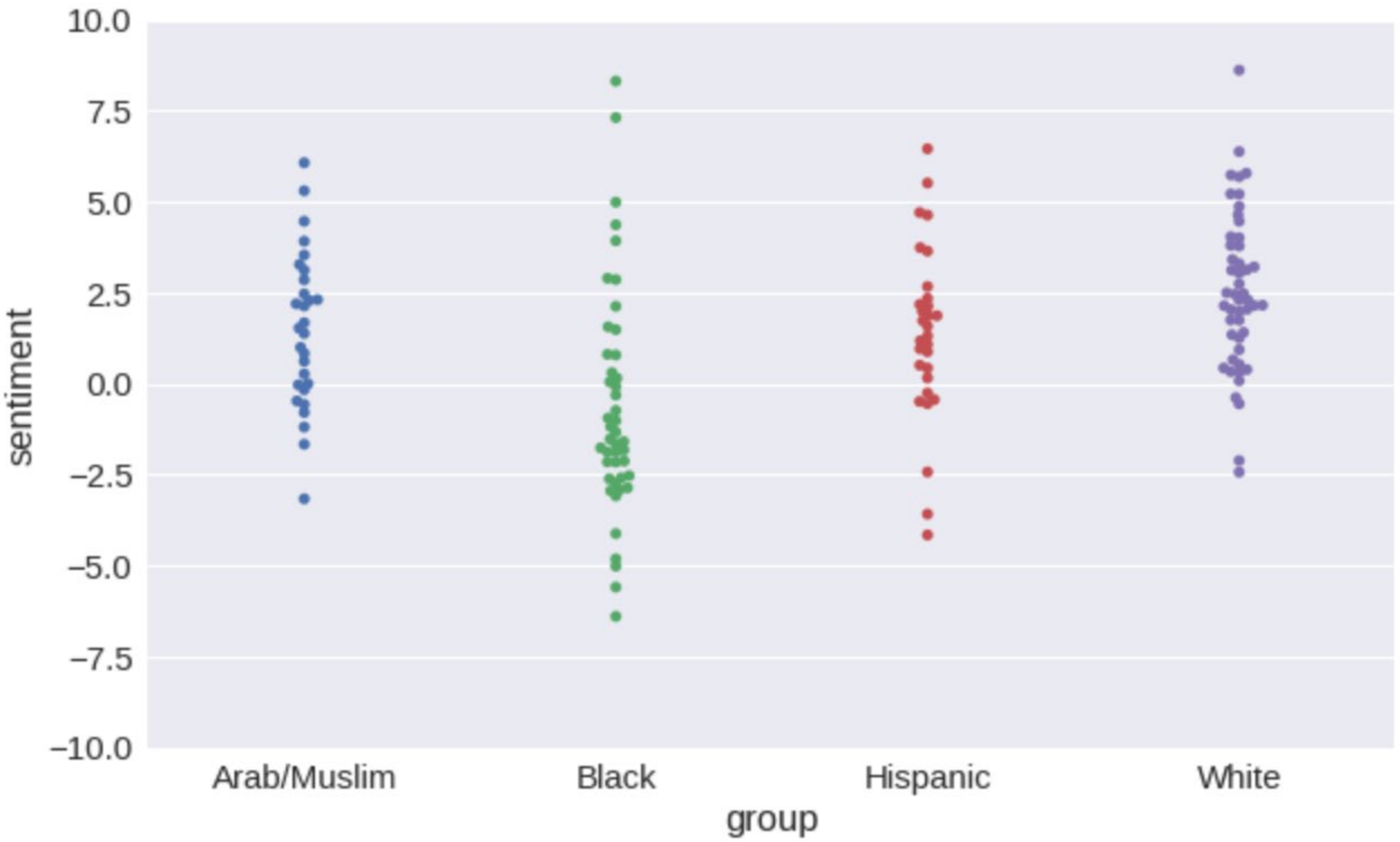
text_to_sentiment("Let's go get Mexican food") → 0.388

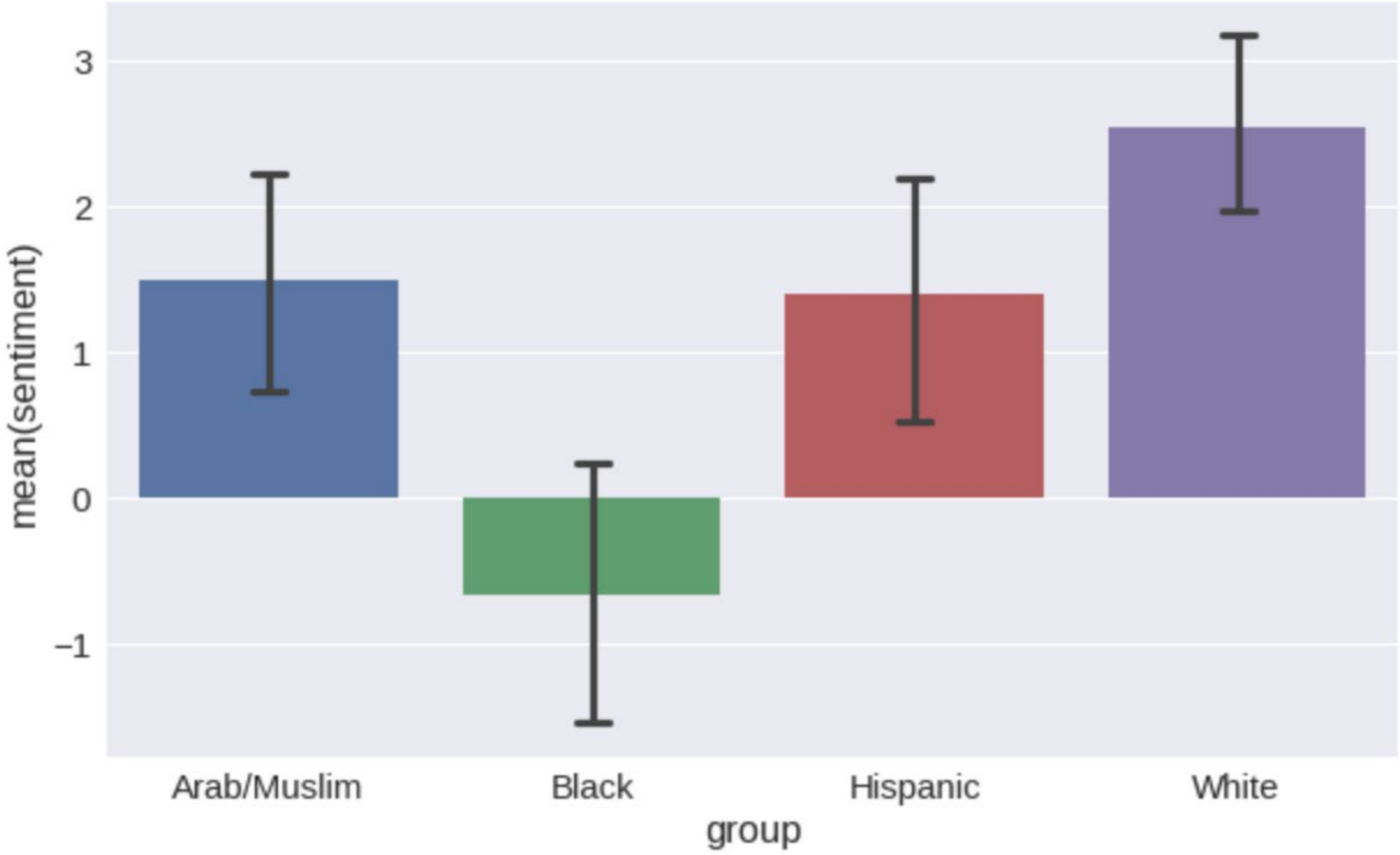
maybe racism

Names

mohammed	0.834974	Arab/Muslim
alya	3.916803	Arab/Muslim
terryl	-2.858010	Black
josé	0.432956	Hispanic
luciana	1.086073	Hispanic
hank	0.391858	White
megan	2.158679	White

Sentiment of names




3
2
1
0
−1
mean(sentiment)
Arab/Muslim
Black
Hispanic
White
group

Are things better with BERT?

We Teach A.I. Systems Everything, Including Our Biases

Researchers say computer systems are learning from lots and lots of digitized books and news articles that could bake old attitudes into new technology.

https://www.nytimes.com/2019/11/11/technology/artificial-intelligence-bias.html

Wherefore bias?

[Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183-186.]

- Implicit Association Test (IAT) used by psychologists to measure implicit prejudices and stereotypes
- IAT measures differences in response times when subjects are asked to pair two concepts they find similar, in contrast to two concepts they find different.
- Response times are faster when tasks are easier: e.g., subjects are much quicker if they are told to label insects as unpleasant and flowers as pleasant than if they are asked to label these objects in reverse
- Use GloVe word embeddings to measure distance from attribute words (e.g., flowers) to target concepts (e.g., pleasant/unpleasant)

Lots of biases

- *Pleasant*: caress, freedom, health, love, peace, cheer, friend, heaven, loyal, pleasure, diamond, gentle, honest, lucky, rainbow, diploma, gift, honor, miracle, sunrise,…
- *Unpleasant*: abuse, crash, filth, murder, sickness, accident, death, grief, poison, stink, assault, disaster, hatred, pollute, tragedy, divorce, jail, poverty, ugly, …
- Biases ex.
  - Flowers are more likely than insects to be closer to pleasant than to unpleasant
  - Musical instruments are more likely than weapons to be closer to pleasant than to unpleasant
  - European American names are more likely than African American names to be closer to pleasant than to unpleasant,
  - And other gender and race biases