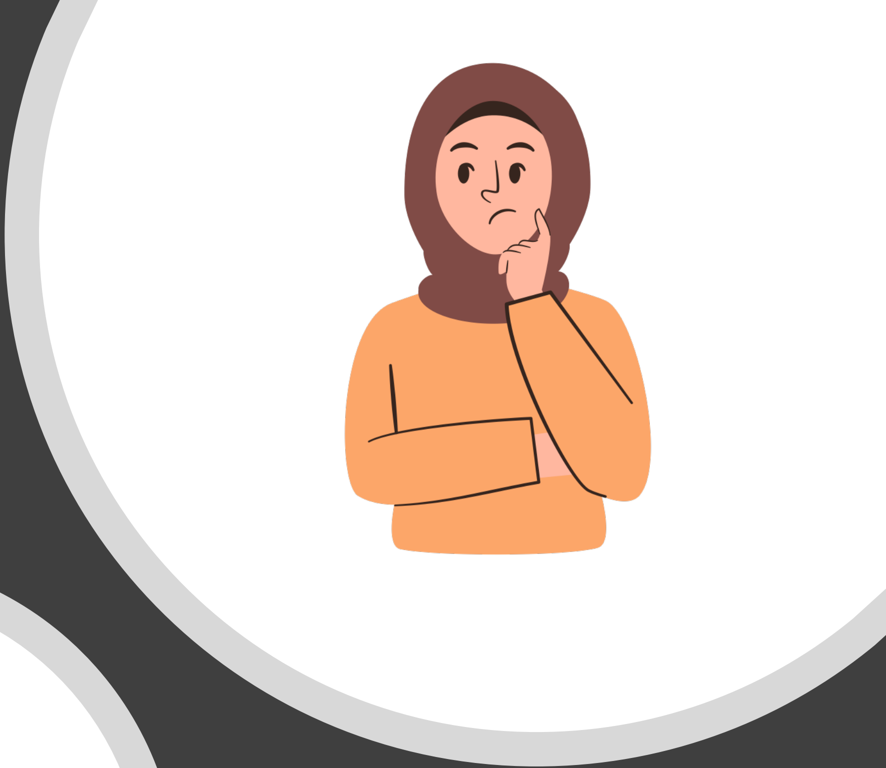
Lexicons for Sentiment, Affect, and Connotation

Natalie Parde

UIC CS 421

Language is tricky for many reasons!

- Ambiguity
- Abstractness
- Tone
- Polarity
- Subjectivity



To understand meaning and intent, we often have to read between the lines.

- Difficult for people, and extremely challenging for machines!
- One approach:
  - **Affective analysis**

Affective Analysis

- The automated analysis of the **emotions**, **moods**, **attitudes**, **interpersonal stance**, or **personality** that is conveyed or evoked by a language sample
- Closely intertwined with social science research
  - Determining views towards a specific topic
  - Assessing public opinion
  - Interpreting intent

When can these different forms of analysis be useful?

- **Attitudes** help us figure out what people like or dislike
  - Useful for processing and interpreting reviews
  - Useful for measuring public sentiment
- **Emotions** and **moods** help us measure engagement or frustration, among other factors
  - Useful for studying how people interact with automated systems
  - Useful for psycholinguistic tasks

When can these different forms of analysis be useful?

- **Interpersonal stance** can help us understand perspectives and characteristics of multi-party interaction
  - Useful for determining views with respect to specific topics
  - Useful for summarizing conversations (is the interaction friendly or awkward?)
- **Personality** can help us customize interactive agents
  - Useful for matching user expectations
  - Useful for optimizing user experience

Automated Affect Recognition

- Classification of text into predetermined affective categories
- Commonly performed using supervised learning
- Useful features:
  - N-grams
  - Features derived from **affective lexica**

What is an affective lexicon?

- Known list of words corresponding to different affective dimensions
  - Optionally including scores indicating the closeness of their association with a given dimension
- This information makes it easier to predict affect for an entire instance
  - If the input contains many positive words, it is likely a positive input!

What are some common forms of affective analysis?

SENTIMENT ANALYSIS

EMOTION RECOGNITION

CONNOTATION FRAMING

Emotion Recognition

- Emotion can be defined in numerous ways
- In some frameworks, emotion is an **atomic unit**
  - IsHappy = TRUE
- In other frameworks, emotion is a **point along a multi-dimensional continuum**
  - Happiness = 0.78

Common Emotion Frameworks

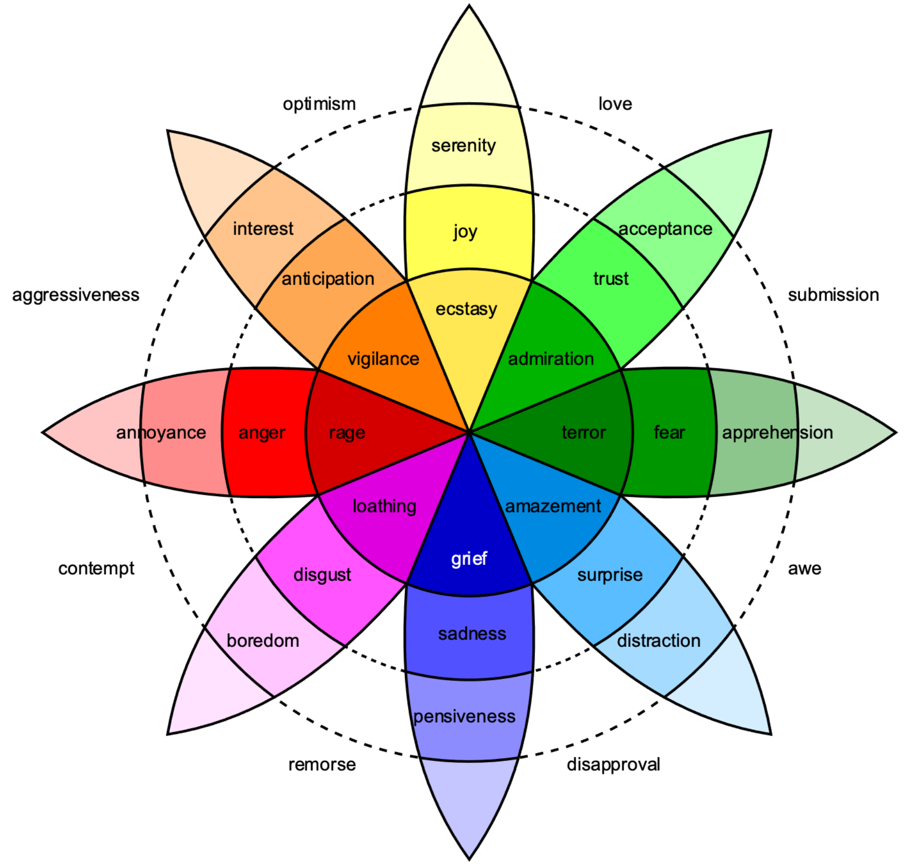
- Ekman’s six basic emotions
- Plutchik’s wheel of emotion

Ekman’s Basic Emotions

- Paper:
  - Ekman, P. (1999). Basic emotions. In T. Dalgleish & M. J. Power (Eds.), Handbook of cognition and emotion (pp. 45–60). John Wiley & Sons Ltd. https://doi.org/10.1002/0470013494.ch3
- Six basic emotions:
  - Happiness
  - Sadness
  - Anger
  - Fear
  - Disgust
  - Surprise
- Emotions are distinct from one another
- Generally known to be present across cultures

Plutchik Wheel of Emotion

- Paper:
  - Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In *Theories of emotion* (pp. 3-33). Academic press. https://doi.org/10.1016/B978-0-12-558701-3.50007-7
- Situates eight basic emotions in a wheel
- Emotions located opposite one another also oppose one another semantically
- Stronger emotions derived from the basic emotions are located at more internal locations
- Weaker emotions derived from the basic emotions are located at more external locations



Atomic vs. Continuous Emotions

- Ekman and Plutchik both define emotion as an atomic unit
- Emotion along a continuum is often represented using a set of common dimensions
  - **Valence:** Pleasantness (e.g., positive or negative)
  - **Arousal:** Intensity of emotion provoked (e.g., strong or weak)
  - (Sometimes) **Dominance:** Degree of control exerted (e.g., active or passive)
- Sentiment is sometimes viewed as a measure of valence

Sentiment and Affect Lexicons

- A wide range of resources for sentiment and affect recognition are available for public use!
- Can be highly useful for performing automated sentiment or affect analysis

General Inquirer

- Classic resource created during the 1960s
  - Stone, P. J., In Kirsch, J., & Cambridge Computer Associates. (1966). *The general inquirer: A computer approach to content analysis*. Cambridge, Mass: M.I.T. Press.
- 1915 positive words
- 2291 negative words
- Additional words associated with other categories
- Link:
  - http://www.webuse.umd.edu:9090/

MPQA Subjectivity Lexicon

- Collection of positive and negative words from existing lexicons
  - 2718 positive words
  - 4912 negative words
- Additional subjective words learned via bootstrapping, with manually-provided sentiment and subjectivity levels
  - Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 347-354).
- Link:
  - https://mpqa.cs.pitt.edu/lexicons/subj_lexicon/

Opinion Lexicon

- Positive and negative words collected from product reviews via bootstrapping
  - Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 168-177).
- 2006 positive words
- 4783 negative words
- Link:
  - https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html

NRC Valence, Arousal, and Dominance Lexicon

- 20,000 words labeled with valence, arousal, and dominance scores
  - Mohammad, S. (2018a). Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 174-184).
- Link:
  - https://saifmohammad.com/WebPages/nrc-vad.html

NRC Word-Emotion Association Lexicon

- Approximately 14,000 words labeled for the eight basic emotions from Plutchik's wheel of emotions
  - Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3), 436-465.
- Link:
  - https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

NRC Emotion Intensity Lexicon

- Approximately 10,000 words labeled with continuous scores for the eight basic emotions from Plutchik's wheel of emotions
  - Mohammad, S. (2018b). Word Affect Intensities. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.
- Link:
  - https://www.saifmohammad.com/WebPages/AffectIntensity.htm

Linguistic Inquiry and Word Count

- Approximately 2300 words across 73 lexical resources associated with different psychological tasks
  - Pennebaker, J. W., Booth, R. J., and Francis, M. E. (2007). *Linguistic Inquiry and Word Count: LIWC 2007*. Austin,TX.
- Link:
  - https://www.liwc.app/
- Actively maintained and updated (most recent version is from 2022)
- Not free!

Brysbaert Concreteness Lexicon

- Approximately 40,000 words labeled with continuous concreteness labels ranging from 1-5
  - Brysbaert, M., Warriner, A.B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46, 904-911.
- Link:
  - http://crr.ugent.be/archives/1330

Personality and Stance

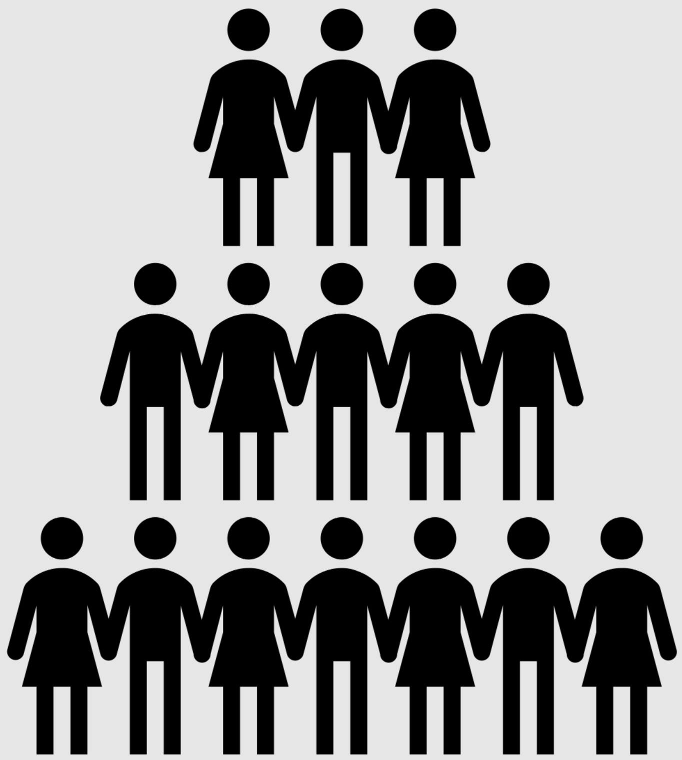
- Two other popular forms of affective analysis:
  - Personality detection
  - Stance detection
- Personality detection focuses on recognizing and classifying predefined aspects of a user's personal character
- Stance detection focuses on recognizing a user's opinion towards a specific topic

Personality Dimensions

- Most work in NLP makes use of the “Big Five” **personality dimensions**
  - Extroversion vs. Introversion
  - Emotional Stability vs. Neuroticism
  - Agreeableness vs. Disagreeableness
  - Conscientiousness vs. Unconscientiousness
  - Openness to Experience
- Several corpora exist containing language samples annotated with these personality dimensions
- Paper:
  - Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual review of psychology*, 41(1), 417-440.

Affective Stance

- One’s position towards another (or towards a topic) during an interaction
  - Friendly
  - Distant
  - Supportive
  - Unsupportive
- Corpora also exist containing conversational exchanges labeled with each party’s affective stance



How to build a lexicon?

- Expert labels
  - Very reliable 🙂
  - Costly and time-consuming 😕
- Crowdsourced labels
  - Less reliable 😕
  - Inexpensive and quick 🙂

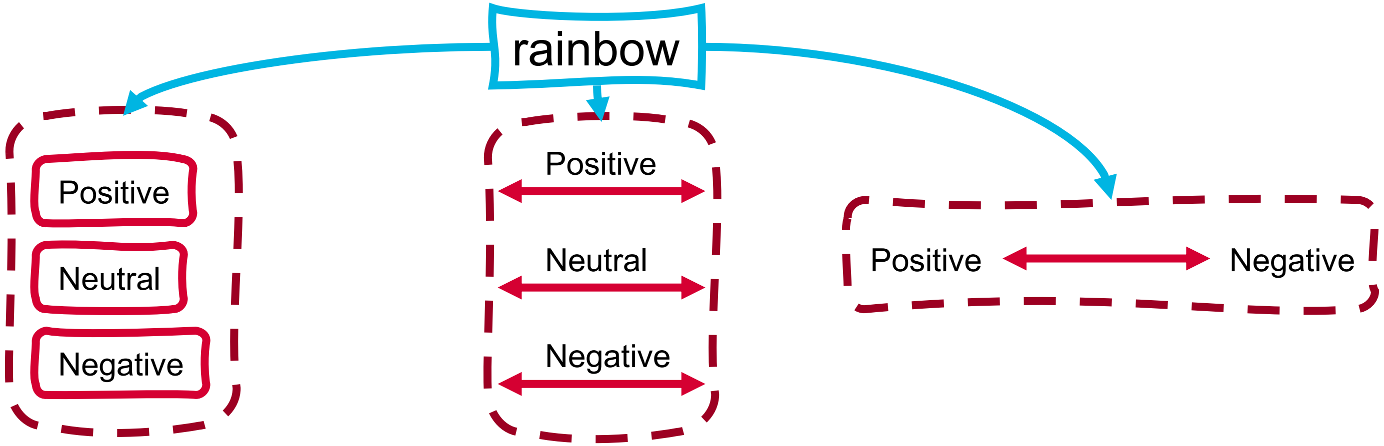
Crowdsourcing Resources

- Amazon Mechanical Turk:
  - https://www.mturk.com
- Appen:
  - https://appen.com
- Your own online survey

Annotation Schemata

- What kinds of labels will be permissible for your annotators?
- How will they know which labels to select?

Positive: A word that evokes a happy or content emotion.
Examples: *love*, *great*, *happy*
Neutral: A word that does not particularly evoke any emotion.
Examples: *pencil*, *refrigerator*, *khaki*
Negative: A word that evokes a sad or angry emotion.
Examples: *violence*, *evil*, *upset*



Adjudication

- Third-party adjudicator
- Majority label
- Average label

Semi-Supervised Induction of Affect Lexicons

- **Semi-supervised label induction:** The process of labeling new, unlabeled instances based on their similarity to instances in a small, labeled seed set
- Two main families:
  - **Axis-based** induction
  - **Graph-based** induction

Axis-Based Lexicon Induction

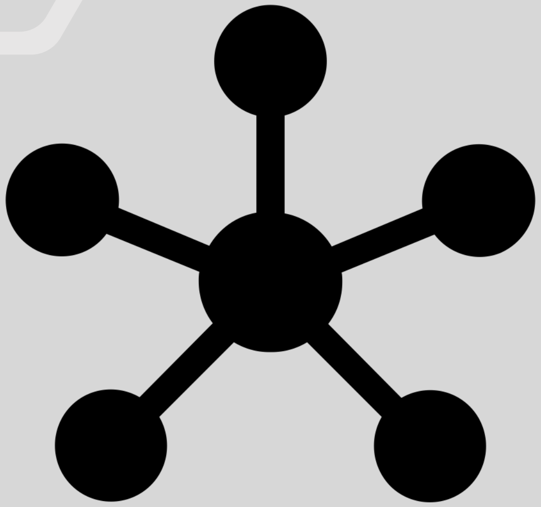
- Given a seed set, how similar is the instance to positive instances and how different is it from negative instances?
- Seed sets can be determined by hand in two different ways:
  1. Start with a single seed lexicon, and fine-tune it to the domain using induction techniques
  2. Choose different seed words to represent different domains
- Both of these methods allow us to handle varying affect associated with seed words in different contexts

Axis-Based Lexicon Induction

- Once we've determined our seed words:
  - Compute an embedding for each seed word
  - Find the centroid of the embeddings for positive words, and the centroid of the embeddings for negative words
    - $\mathbf{V}^{+} = \frac{1}{n}\sum_{n}^{n} E(w_i^{+})$
  - Compute the axis by subtracting one centroid from another
    - $\mathbf{V}_{\boldsymbol{axis}} = \mathbf{V}^{+} - \mathbf{V}^{-}$
  - Compute the similarity between a given word embedding and the axis
    - $\text{score}(w) = \cos(E(w), \mathbf{V}_{axis}) = \frac{E(w) \cdot \mathbf{V}_{axis}}{\|E(w)\| \|\mathbf{V}_{axis}\|}$
  - Higher similarities indicate closer alignment with the positive class

As an alternative….

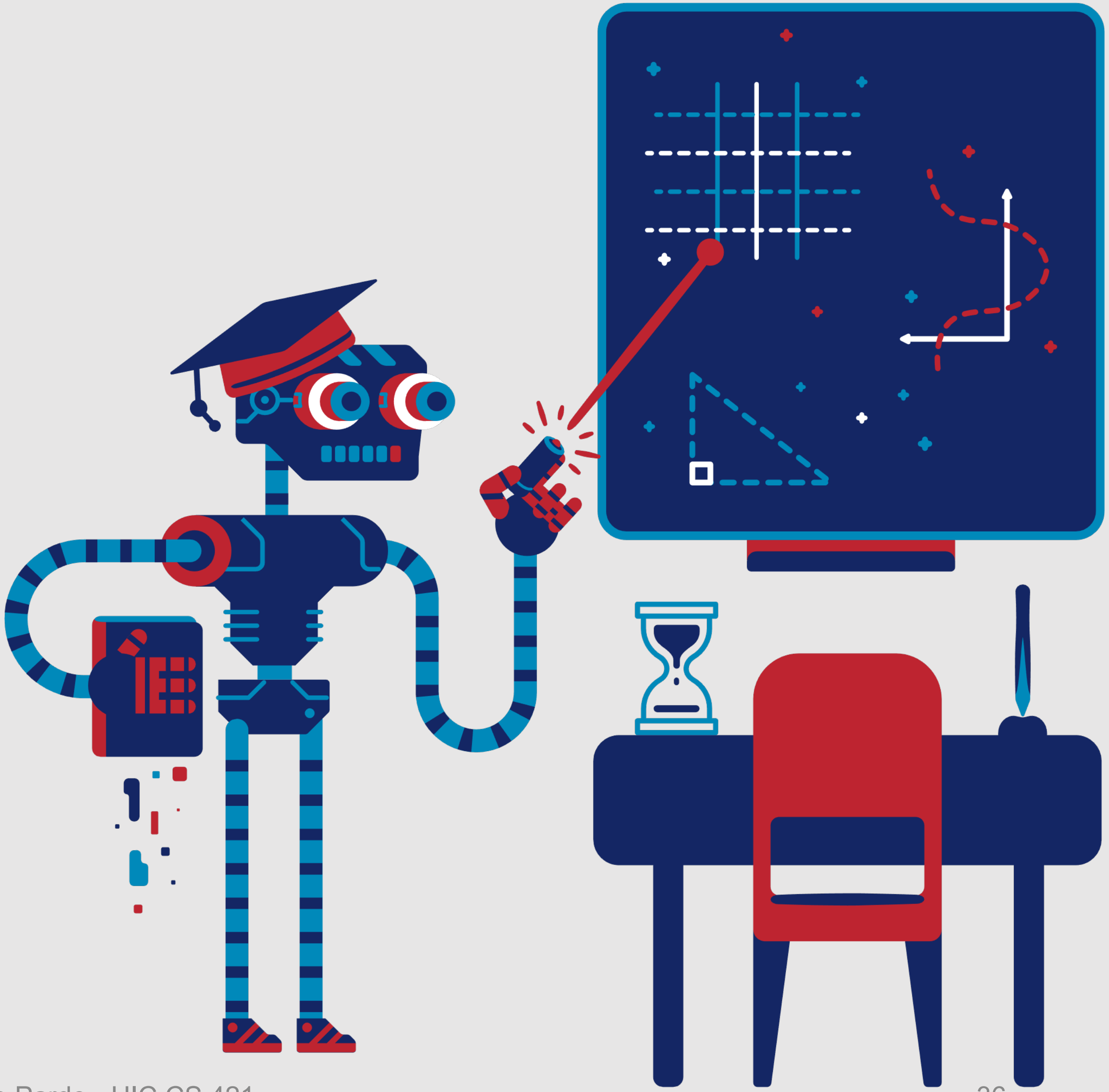
- Graph-based induction techniques allow us to define lexicons by propagating sentiment labels on graphs

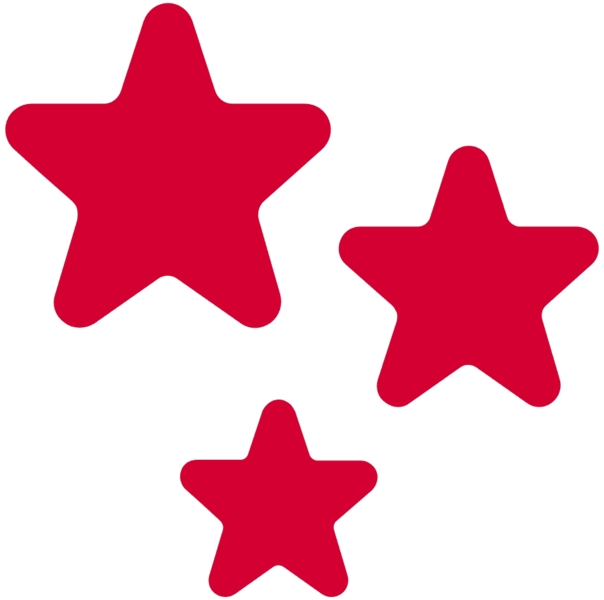


Graph-Based Lexicon Induction

- Given a graph that connects words with their nearest neighbors, how likely is it that a random walk from a positive word ends on the given word?
  - Define a graph that connects each word to its *k* nearest neighbors, with edges weighted by word similarity
  - Identify words in the graph belonging to a labeled seed set
  - Starting at a word from the seed set, perform an edge-weighted random walk
  - Assign an unlabeled word's score based on the probability of landing on it during a random walk from a positive seed and a random walk from a negative seed
    - $\text{score}^{+}(w_i) = \frac{\text{score}^{+}(w_i)}{\text{score}^{+}(w_i)+\text{score}^{-}(w_i)}$
  - Repeat multiple times using bootstrapping, and assign confidence to word scores based on their standard deviation across multiple runs

How can we use supervised machine learning to predict a word's sentiment?





Lots of supervision signals (potential labels) exist in real-world data.

- One example: Review scores
  - 1-5 stars
  - Rating from 1-10
  - Often associated with free-form review text
- We can use these scores and associated text to learn polarity distributions for words

Normalized Word Likelihood

- **Document-level sentiment classifier** → any statistical or neural methods we've learned about so far!
- **Word-level sentiment classifier** → also consider simple probabilistic measures
- **Normalized word likelihood**
  - $P(w|c) = \frac{\text{count}(w,c)}{\sum_{w \in C} \text{count}(w,c)}$

Normalized word likelihood helps us find a word's sentiment distribution across classes.

How likely the word is to be associated with one star, two stars, and so on!

We can then visualize this distribution using a **Potts diagram**

Potts Diagrams

- Mechanism for visualizing word sentiment
  - Sentiment class vs. normalized word likelihood
- Characteristic patterns:
  - **"J" shape:** Strongly positive word
  - **Reverse "J" shape:** Strongly negative word
  - **"Hump" shape:** Weakly positive or negative word
- Patterns may also correspond to different types of word classes
  - Emphatic and attenuating adverbs

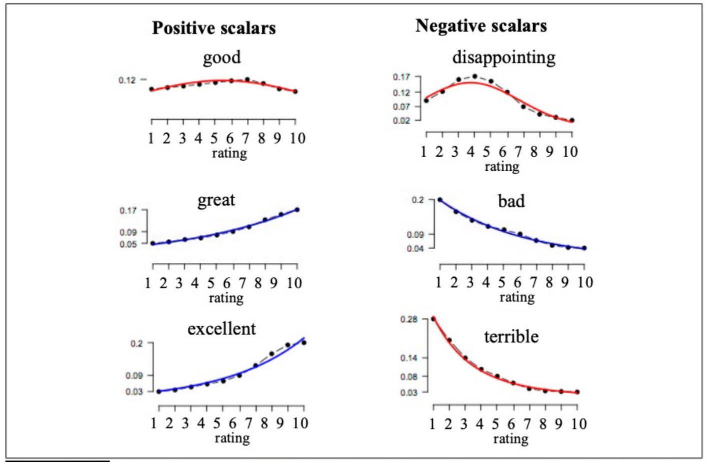


Figure 21.10 Potts diagrams (Potts, 2011) for positive and negative scalar adjectives, showing the J-shape and reverse J-shape for strongly positive and negative adjectives, and the hump-shape for more weakly polarized adjectives.

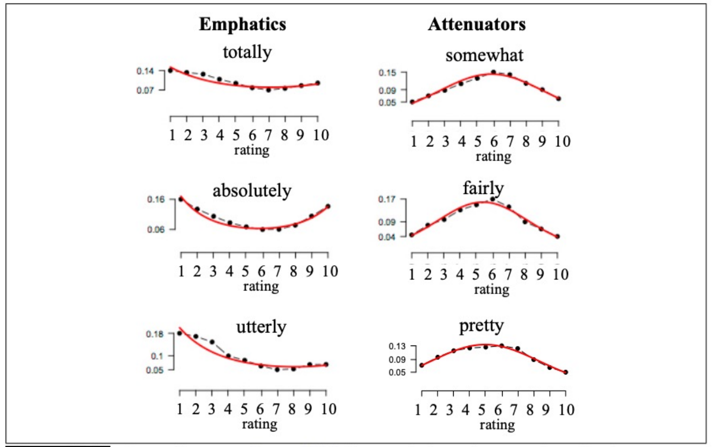


Figure 21.11 Potts diagrams (Potts, 2011) for emphatic and attenuating adverbs.

Sometimes we want to figure out which words are most closely associated with a class rather than just plotting frequency distributions.

- Certain words will have high frequencies across many classes, but these are less helpful for understanding class membership
- Certain words will have low frequencies across all classes, but still might be more informative if they are more clearly associated with a specific class

Log Odds Ratio with Informative Dirichlet Prior

- Is computing raw difference in frequency sufficient for identifying whether a word is very positive or very negative?
  - Not necessarily!
  - **Highly frequent words may have large raw differences even with small relative differences**, and highly infrequent words may have small raw differences even with large relative differences
- More sophisticated approach: **Log odds ratio with an informative Dirichlet prior**

Log Odds Ratio

- Originally proposed for measuring partisan speech used by US politicians
  - Some words are likelier to be used by Republicans, and other words are likelier to be used by Democrats
- Generalizes to any other problem domain for which lexical trends are anticipated to be different
- Key goal: Find words that are statistically overrepresented in one category of text compared to another

Log Odds Ratio

- **Probability of word *w* existing in corpus *i*:**
  - $P^i(w) = \frac{f_w^i}{n^i}$
- **Log odds ratio:**
  - $\mathrm{lor}(w) = \log \frac{P^i(w)}{1-P^i(w)} - \log \frac{P^j(w)}{1-P^j(w)}$
  - $\mathrm{lor}(w) = \log \frac{f_w^i}{n^i - f_w^i} - \log \frac{f_w^j}{n^j - f_w^j}$

Dirichlet Intuition

- Use a large background corpus to get a prior estimate of our expected frequency for each word *w*
- To do so:
  - Add the counts from that corpus to our numerator and denominator
  - This basically shrinks the counts toward that prior (how big are the differences, *given what we would expect from a large background corpus*?)

Prior-Modified Log Odds Ratio

- Modifying the previous equation with an informative Dirichlet prior:
  - $\delta_w^{(i-j)} = \log \frac{f_w^i + \alpha_w}{n^i + \alpha_0 - (f_w^i + \alpha_w)} - \log \frac{f_w^j + \alpha_w}{n^j + \alpha_0 - (f_w^j + \alpha_w)}$

Size of corpus

Count of *w* in corpus

Size of background corpus

Count of w in background corpus

Log Odds Ratio with Informative Dirichlet Prior

- Estimate of variance for the modified log odds ratio:
  - $\sigma^2\left(\hat{\delta}_w^{(i-j)}\right) \approx \frac{1}{f_w^i + \alpha_w} + \frac{1}{f_w^j + \alpha_w}$
- Final statistic for a word is then the z-score of its modified log odds ratio:
  - $\frac{\hat{\delta}_w^{(i-j)}}{\sqrt{\sigma^2\left(\hat{\delta}_w^{(i-j)}\right)}}$

Key Modifications to Log Odds Ratio

- This method thus modifies the commonly used log odds ratio by:
  - Using z-scores of the log odds ratio
    - Controls for the amount of variance in a word's frequency
  - Using counts from a background corpus to provide a prior count for words
    - Helps leverage additional useful information

This ultimately gives us a useful tool for analysis!

Words associated with one-star restaurant reviews

worst, rude, terrible, horrible, bad, awful, disgusting, bland, tasteless, gross, mediocre, overpriced, worse, poor

Words associated with five-star restaurant reviews

Great, best, love, delicious, amazing, favorite, perfect, excellent , awesome, friendly, fantastic, fresh, wonderful, incredible, sweet, yum

What is affect recognition?

- Affect recognition: The task of automatically determining how a given input makes would be characterized, based on some specified range of categories
  - Happy vs. sad
  - Extroverted vs. introverted
  - Friendly vs. distant

Affect Recognition

- Typically framed as a supervised learning task
- Large datasets:
  - N-gram features
- Very large datasets:
  - N-gram features, pruned based on frequency or **pointwise mutual information (PMI)**
    - $\text{PMI}(x;y) = \log \frac{p(x,y)}{p(x)p(y)}$

Features from External Lexicons

- **Indicator Function:**
  - $f_L(x) = \begin{cases} 1 \; if \; \exists w : w \in L \; \& \; w \in x \\ 0 \quad \text{otherwise} \end{cases}$
- **Count-Based Function:**
  - $f_L(x) = \sum_{w \in x} \text{count}_L(w)$
- **Weighted Count-Based Function:**
  - $f_L(x) = \sum_{w \in x} \theta_w^L \text{count}_L(w)$

Lexicon-based features can shed new light on interesting social science problems!

- Does one's use of positive language correlate with one's level of extroversion?
- Is more concrete language likely to evoke more neutral emotions?
- Is there a relationship between the number of "difficult" words and the overall subjectivity of an input?

What if we don’t have labeled training data to build a supervised model for sentiment or affect recognition?

Use sentiment resources (such as those described earlier!) to perform sentiment analysis directly

Using Lexicons for Sentiment Recognition

1. Assign a positive label to instances that contain more positive than negative words from the lexicon

2. Assign a negative label to instances that contain more negative than positive words

3. Assign a neutral label in the event of a tie

More formally….

- Define a threshold λ indicating the minimum percentage of positive or negative words needed for a positive or negative classification
- Select a sentiment class as follows:
  - $f^+ = \sum_{w \in x} \theta_w^+ \text{count}_+(w)$
  - $f^- = \sum_{w \in x} \theta_w^- \text{count}_-(w)$
  - $\text{sentiment} = \begin{cases} + \text{ if } \frac{f^+}{f^-} > \lambda \\ - \text{ if } \frac{f^-}{f^+} > \lambda \\ 0 \text{ otherwise} \end{cases}$

Entity-Centric Affect

- Sometimes we don't need (or want) to recognize the affect of an entire input
  - Scope may be too broad!
- We can also learn to predict the affect of a single entity within the input

Methods for Entity-Centric Affect Recognition

- One way to do this: Leverage both affect lexica and contextual word embeddings
- First, extract a contextual embedding for each instance of a word
- Then, average those embeddings
- Repeat this process for all words
- Learn to map from averaged word embeddings to lexicon-based scores corresponding to affective dimensions

How do we make predictions using this technique?

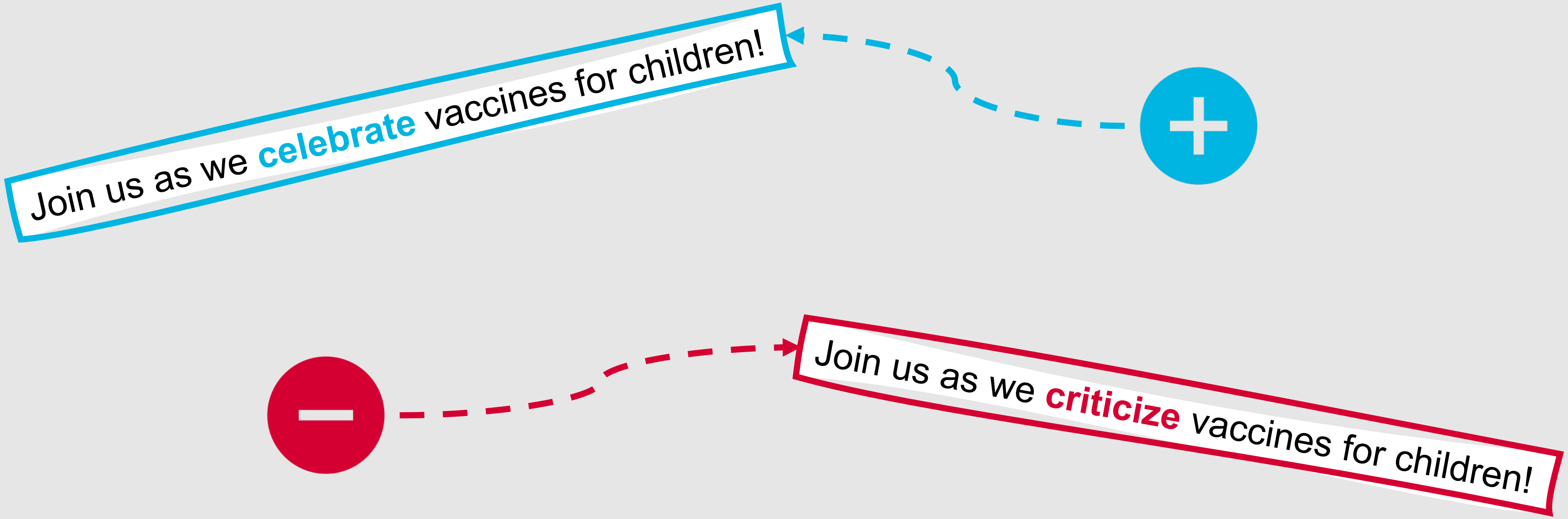
- When a new entity is encountered without a known affective score:
  - Create a new average embedding based on all instances of that entity in context
  - Use that embedding to predict scores associated with the different affective dimensions

Connotation Frames

- So far, we've mainly focused on representing affect using n-dimensional affective space
- We can also represent affective meaning using **connotation frames**

Connotation Frames

- Indicate affective properties commonly associated with words
  - Similar to how verb frames indicate selectional preferences



Connotation Frames

- Can be:
  - Manually constructed
  - Learned automatically
- Downloadable collection of connotation frames:
  - https://hrashkin.github.io/connframe.html

Summary: Lexicons for Sentiment, Affect, and Connotation

- Lexicons can help us distinguish many kinds of **affective states**
- Emotion can be represented using fixed **atomic units** or **dimensions** in a continuous space
- **Affective lexicons** can be built by hand, in a semi-supervised manner, or using fully supervised methods
- Words can be assigned weights in a lexicon based on frequency measures and ratio metrics like **log odds ratio with an informative Dirichlet prior**
- **Connotation frames** express richer affective relationships, similar to those seen with semantic frames