COMM 605

Week 9: Dictionary/Lexicon-based Sentiment Analysis

Ryan Y. Wang, Ph.D.

ryanwang@rit.mail.edu

School of Communication

October 23, 2023

What is sentiment analysis?

Sentiment analysis measures the polarity or tonality of texts by identifying and assessing expressions people use to evaluate or appraise persons, entities (e.g., products, services) or events (**Pang and Lee 2008**; **Soroka 2014**).

Sentiment on social media: people use social media to express their feelings, attitude, views, arguments, opinions on wide range of topic:

- Business: Users reviews, customers attitude and trends, sales performance
- Health: Depression, public health
- Politics: election prediction, polarization, public opinion or media tone
- Engagement in Advertising/Strategic communication: “if it bleeds, it leads.”

Measurement of sentiment analysis

Level of analysis:

- word-level, sentence-level, document-level, aspect-level, concept-level

Measurement:

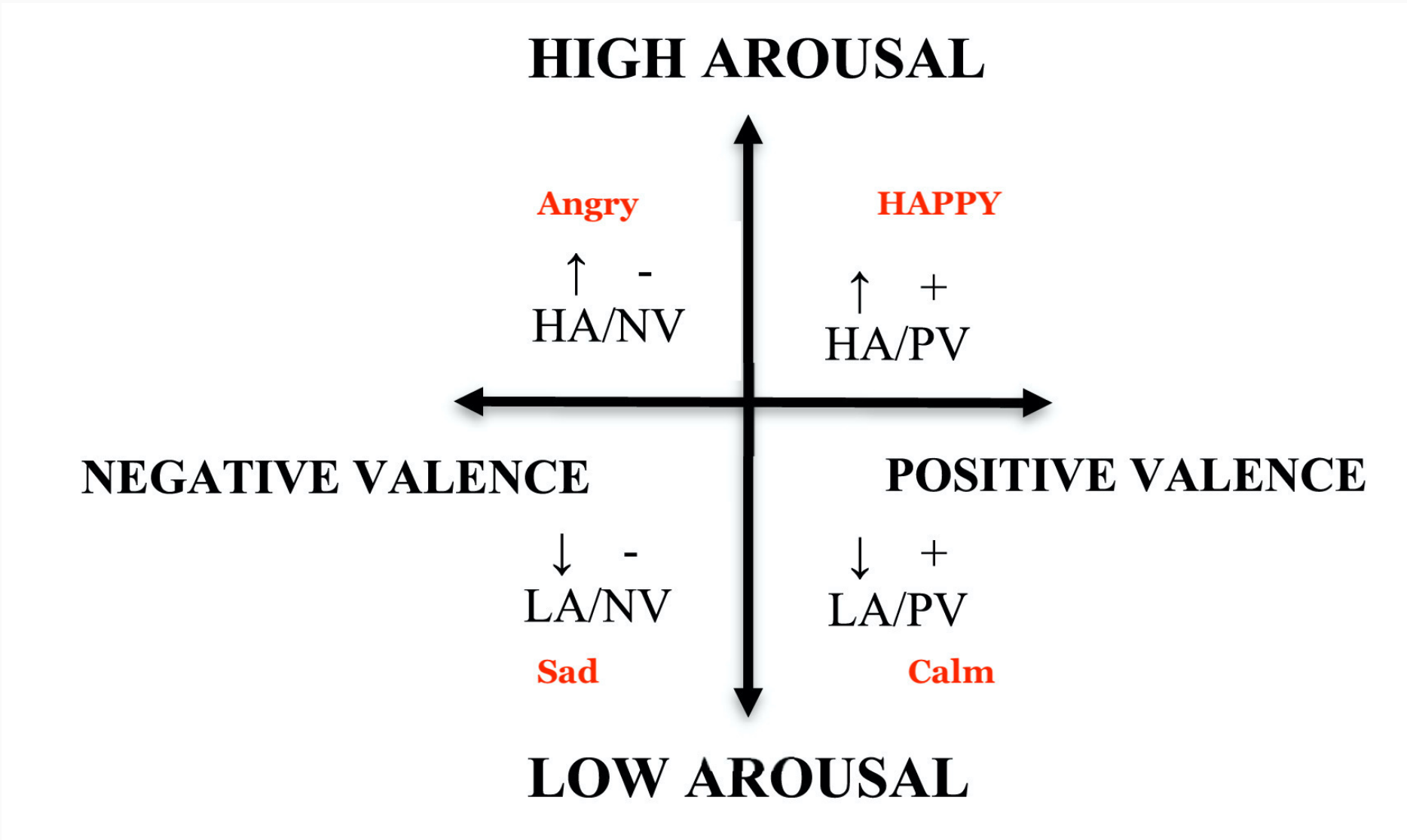
- Polarity-based setniment:a polarity is normally a categorical attribute: positive or negative.
- Valence-based sentiment: a valence is in the form of a continuous value indicating a degree of sentiment.

What about emotions?

Emotions: complex states of feeling lead to a change in thoughts, actions, behavior, and personality.

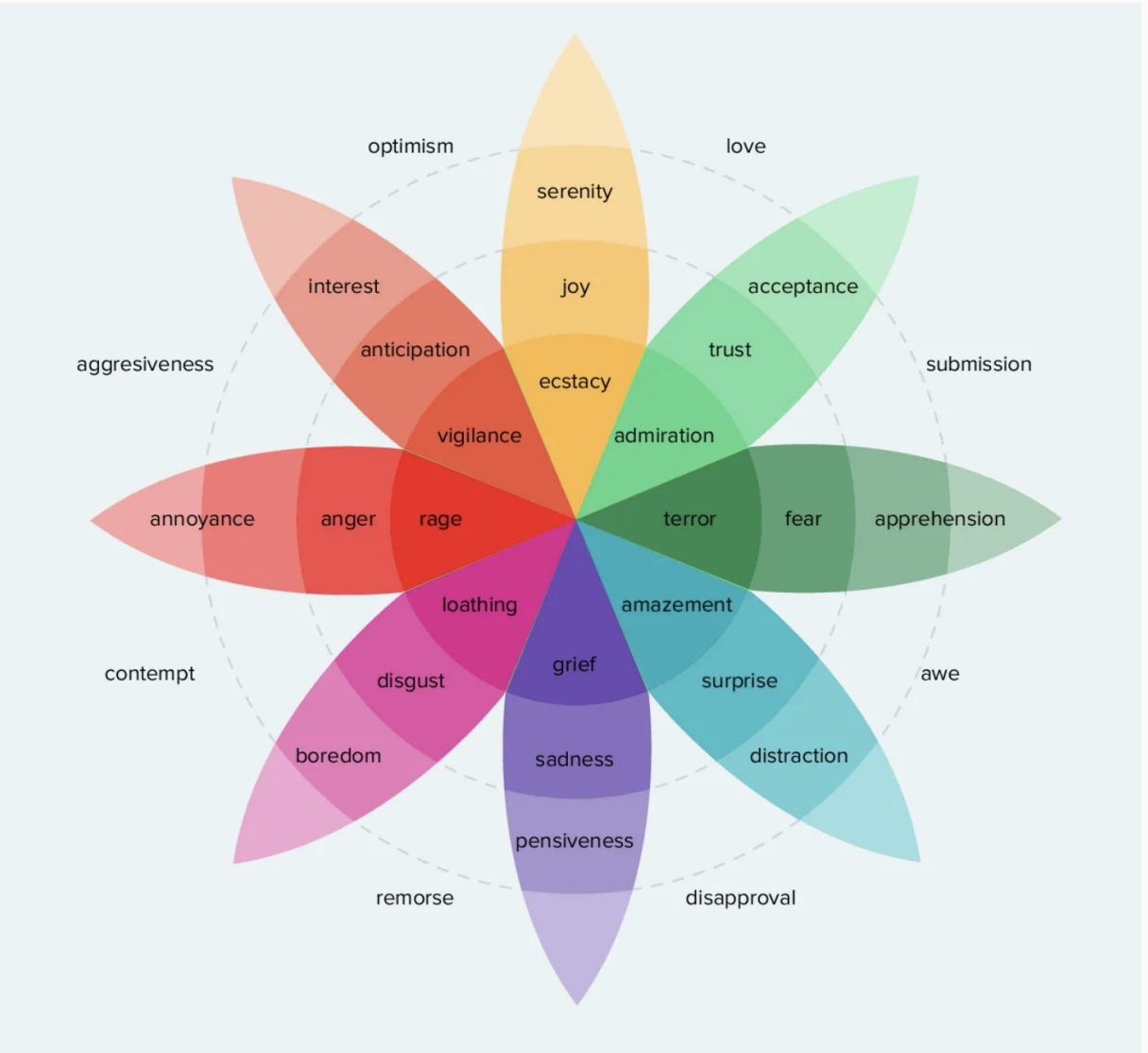
- Dimensional emotion model (**Russel, 1980**)
- Discrete emotion model (**Plutchik, 2003**)

Dimensional emotion model



Valence-arousal Plane

Discrete emotion model



Wheel of Emotions

Methods in sentiment analysis

- Dictionary-based (this tutorial): **Linguistic Inquiry and Word Count(LIWC)**
  - It makes use of a predefined list of words, where each word is associated with a specific sentiment.
  - A basic assumption of using a dictionary is that it contains the most important words required for rating a text.
- Supervise machine learning and Deep learning (pre-trained model, large language model)
  - It relies on supervised classification approaches based on training data, where sentiment detection is framed as a binary (i.e., positive/negative/neutral).

Questions: are they able to produce the same sentiment rating of texts as a human coder? - van Atteveldt (2021):

- The best performance is still attained with trained human or crowd coding
- None of the used dictionaries come close to acceptable levels of validity
- Machine learning, especially deep learning, substantially outperforms dictionary-based methods but falls short of human performance

Empirical examples

- Dubovi and Tabk (2021): public engagement with science on YouTube
  - Detect emotion via lexicon-based sentiment analysis based on `syuzhet` package
  - Statistical analysis on cognitive engagement (argumentative expression) with emotional and behavioral engagement
- Harris and Krishnan (2023): Suicide Prevention Public Service Announcements and User Comments on YouTube
  - (Human) Content analysis on 72 videos: gain-loss frame; argument-narrative format
  - (Computational) Textual analysis on users' comments
    - Sentiment analysis: negative and positive
    - Latent semantic analysis: help seeking or avoidant

Thank you!