Sentiment Analysis with R

Methods Fair - UoM

Dr Giulia Grisot

giulia.grisot@manchester.ac.uk

Lecturer in Digital Humanities, The University of Manchester

May 15, 2025

Today's workshop - prerequisites

- Basic understanding of R programming
- Familiarity with text mining and natural language processing concepts
- RStudio installed on your computer
- Access to internet

What is sentiment analysis?

Sentiment Analysis is a **subfield of Natural Language Processing** (NLP) that involves **determining the emotional/sentiment value of a text** .

What is sentiment analysis?

Sentiment analysis can be used to:

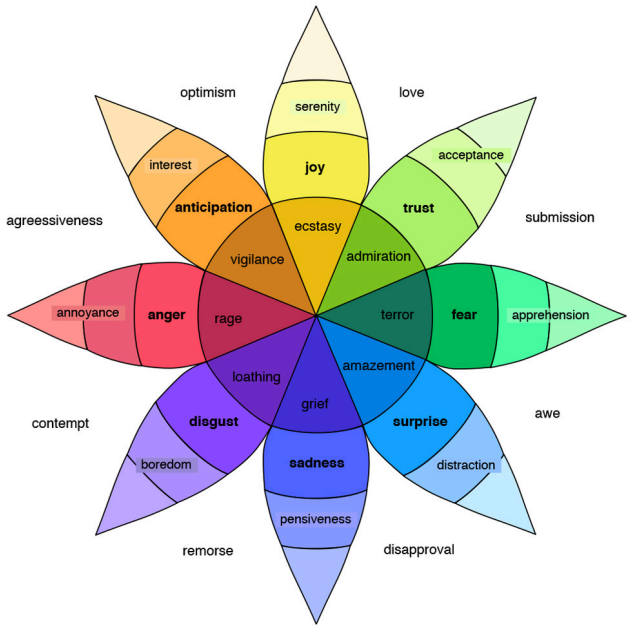
- **understand human behaviour** in digital spaces
- **analyse texts** from books, social media, blogs, websites, and other digital platforms
- analyse **opinions** and cultural representations
- study the **evolution of language** and sentiment over time

Sentiment analysis is, in short, **the process of extracting the emotional value of a text**.

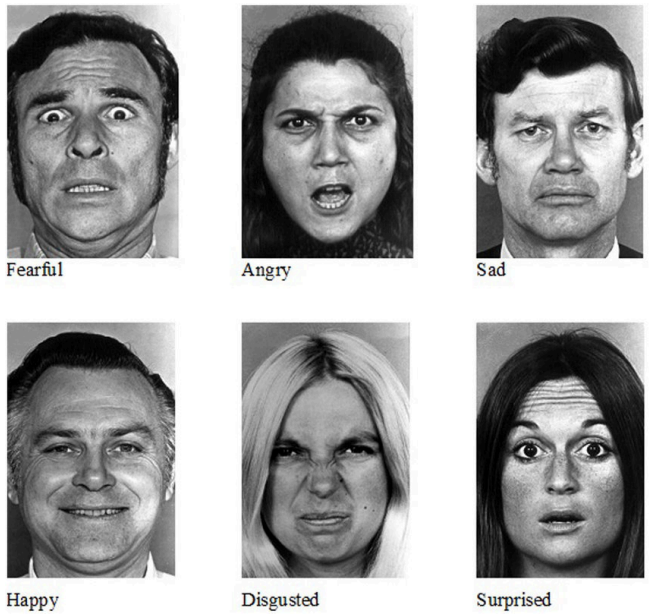
not to be confused with the emotions the reader may feel when reading a text

In its most simple form, it accounts for either **positive** or **negative** sentiment. However, more complex models can account for a wider range of **emotions** , such as joy, anger, sadness, and fear.

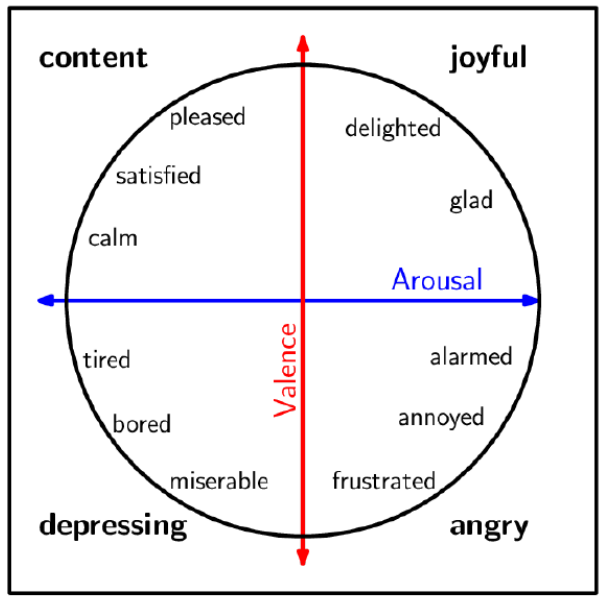
Emotions/Sentiment Models



Plutchik wheel (Plutchik 1980)



Ekman's six basic emotions (Ekman 1992)



Russell's circumplex model (Russell 1980)

So, How Does Sentiment Analysis Work?

Lexicon-based SA

Lexicon-based approaches use predefined word lists (lexicons) with assigned sentiment scores.

Example:

- “This book is fantastic ! It tells wonderful and beautiful things…”
- “In those days, sadness shadowed my heart”

Lexicon-based SA

Lexicon-based approaches use predefined word lists (lexicons) with assigned sentiment scores.

Example:

- “This book is fantastic ! It tells wonderful and beautiful things…”
- “In those days, sadness shadowed my heart”

Strengths and limitations

- Strengths: Simple, interpretable, no training required.
- Limitations: Struggles with negation, sarcasm, and context.

nrc_lex

Filter

	word	sentiment
1	abacus	trust
2	abandon	fear
3	abandon	negative
4	abandon	sadness
5	abandoned	anger
6	abandoned	fear
7	abandoned	negative
8	abandoned	sadness
9	abandonment	anger
10	abandonment	fear
11	abandonment	negative
12	abandonment	sadness
13	abandonment	surprise
14	abba	positive
15	abbot	trust
16	abduction	fear
17	abduction	negative
18	abduction	sadness
19	abduction	surprise
20	aberrant	negative
21	aberration	disgust
22	aberration	negative
23	abhor	anger
24	abhor	disgust
25	abhor	fear
26	abhor	negative
27	abhorrent	anger
28	abhorrent	disgust
29	abhorrent	fear

Screenshot

jockers_lex

Filter

	word	jockers_polarity
1	abandon	-0.75
2	abandoned	-0.50
3	abandoner	-0.25
4	abandonment	-0.25
5	abandons	-1.00
6	abducted	-1.00
7	abduction	-0.50
8	abductions	-1.00
9	aberrant	-0.60
10	aberration	-0.80
11	abhor	-0.50
12	abhorred	-1.00
13	abhorrent	-0.50
14	abhors	-1.00
15	abilities	0.60
16	ability	0.50
17	abject	-1.00
18	ablaze	-0.25
19	abnormal	-0.50
20	aboard	0.25
21	abolish	-0.50
22	abominable	-0.50
23	abominably	-1.00
24	abominate	-1.00
25	abomination	-0.50
26	abort	-0.50
27	aborted	-0.80
28	abortion	-0.80
29	abortive	-1.00

afinn_lex

Filter

	word	value
1	abandon	-2
2	abandoned	-2
3	abandons	-2
4	abducted	-2
5	abduction	-2
6	abductions	-2
7	abhor	-3
8	abhorred	-3
9	abhorrent	-3
10	abhors	-3
11	abilities	2
12	ability	2
13	aboard	1
14	absentee	-1
15	absentees	-1
16	absolve	2
17	absolved	2
18	absolves	2
19	absolving	2
20	absorbed	1
21	abuse	-3
22	abused	-3
23	abuses	-3
24	abusive	-3
25	accept	1
26	accepted	1
27	accepting	1
28	accepts	1
29	accident	-2

bing_lex

Filter

	word	sentiment
1	2-faces	negative
2	abnormal	negative
3	abolish	negative
4	abominable	negative
5	abominably	negative
6	abominate	negative
7	abomination	negative
8	abort	negative
9	aborted	negative
10	aborts	negative
11	abound	positive
12	abounds	positive
13	abrade	negative
14	abrasive	negative
15	abrupt	negative
16	abruptly	negative
17	abscond	negative
18	absence	negative
19	absent-minded	negative
20	absentee	negative
21	absurd	negative
22	absurdity	negative
23	absurdly	negative
24	absurdness	negative
25	abundance	positive
26	abundant	positive
27	abuse	negative
28	abused	negative
29	abuses	negative

Other SA methods

- **Rule-based SA**: Uses hand-crafted rules to identify sentiment.
- **Machine Learning-based SA**: Uses labeled datasets to train models for sentiment classification.
- **Deep Learning-based SA**: Uses neural networks (e.g., RNNs, LSTMs, Transformers) to model sentiment.
- **Aspect-based SA**: Extracts sentiment related to specific aspects of a text.

Multiple SA Applications

- **SA for the study of historical texts**: Sprugnoli et al., *Towards sentiment analysis for historical texts* (2016).
- **SA and cognitive studies**: Jacobs et al., *What's in the brain that ink may character* (2017).
- **SA for political discourse**: Thakur, *Sentiment analysis of the public discourse on Twitter about COVID-19* (2023).
- **SA in literary studies**: Reagan et al., *The emotional arcs of stories are dominated by six basic shapes* (2016).

SA Critical Aspects

Subjectivity: SA may struggle with nuanced or ambiguous language

Ambiguity and Polysemy: Words with multiple meanings can lead to misinterpretations.

Irony and Sarcasm: Algorithms may fail to detect sarcasm or irony.

Context Sensitivity: Sentiment analysis can fail to account for broader context.

Data Bias: Models trained on biased data may not generalize well.

Domain Specificity: Models trained on general datasets may need to be fine-tuned for domain-specific accuracy.

Temporal Dynamics: Sentiment can evolve over time, and static models may miss these changes.

References

Ekman, Paul. 1992. “An Argument for Basic Emotions.” *Cognition and Emotion* 6 (3-4): 169–200. https://doi.org/10.1080/02699939208411068.

Jacobs, Arthur M, and Annette Kinder. 2017. “‘The Brain Is the Prisoner of Thought’: A Machine-Learning Assisted Quantitative Narrative Analysis of Literary Metaphors for Use in Neurocognitive Poetics.” *Metaphor and Symbol* 32 (3): 139–60. https://doi.org/10.1080/10926488.2017.1338015.

Plutchik, Robert. 1980. “A General Psychoevolutionary Theory of Emotion.” In *Theories of Emotion*, 3–33. Elsevier. https://doi.org/10.1016/B978-0-12-558701-3.50007-7.

Reagan, Andrew J, Lewis Mitchell, Dilan Kiley, Christopher M Danforth, and Peter Sheridan Dodds. 2016. “The Emotional Arcs of Stories Are Dominated by Six Basic Shapes.” *EPJ Data Science* 5 (1): 31. https://doi.org/10.1140/epjds/s13688-016-0093-1.

Russell, James A. 1980. “A Circumplex Model of Affect.” *Journal of Personality and Social Psychology* 39 (6): 1161–78. https://doi.org/10.1037/h0077714.

Sprugnoli, Rachele, Sara Tonelli, Alessandro Marchetti, and Giovanni Moretti. 2016. “Towards Sentiment Analysis for Historical Texts.” *Digital Scholarship in the Humanities* 31 (4): 762–72. https://doi.org/10.1093/llc/fqv027.

Thakur, Nirmalya. 2023. “Sentiment Analysis and Text Analysis of the Public Discourse on Twitter about COVID-19 and MPox.” *Big Data and Cognitive Computing* 7 (2): 116. https://doi.org/10.3390/bdcc7020116.