ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
CAMPUS DI FORLÌ

91258 / B0385
Natural Language Processing

Lesson 4. Rule-based Sentiment Analysis (+ Naïve Bayes)

Alberto Barrón-Cedeño
a.barron@unibo.it

08/10/2025

Previously

- Pre-processing (e.g., tokenisation, stemming, stopwording)
- BoW representation
- Dot product

Table of Contents

Sentiment Analysis (with VADER)

Sentiment Analysis

It does not refer to actual sentiment (e.g., love or hate)[1]
It is about positive and negative perceptions (plus neutral)

a

This monitor is definitely a good value. Does it have superb color and contrast? No. Does it boast the best refresh rate on the market? No. But if you're tight on money, this thing looks and preforms great for the money. It has a Matte screen which does a great job at eliminating glare. The chassis it's enclosed within is absolutely stunning.
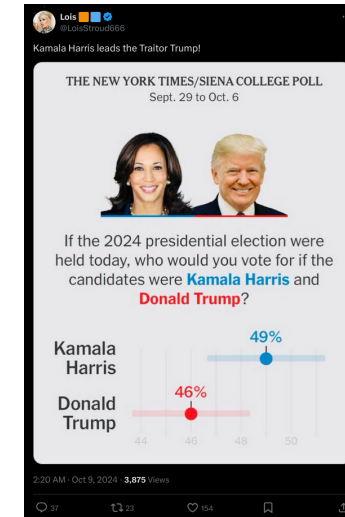
POSITIVE

His [ssa] didnt concede until July 12, 2016. Because he was throwing a tantrum. I can't say this enough: [kcuF] Bernie Sanders.

NEGATIVE

From (Lane et al., 2019, p. 62–65)

[1] That's emotion analysis; e.g., ▤ Fernicola et al. (2020); ▤ Zhang et al. (2022)

Sentiment Analysis

https://x.com/LoisStroud666/status/1843808652802801745

Valence Aware Dictionary for sEntiment Reasoning
(Hutto and Gilbert, 2014)[2]

- It has a lexicon packed with tokens and their associated "sentiment" score
- It counts all tokens belonging to each category: [pos, neu, neg]
  . . . and combine them to determine the sentiment

  </> Let us see it working

[2] https://ojs.aaai.org/index.php/ICWSM/article/view/14550
https://github.com/cjhutto/vaderSentiment

Into ML

Machine Learning

"[. . .] an umbrella term for **solving problems** for which development of algorithms by human programmers would be cost-prohibitive"

"[. . .] the problems are solved by helping machines **"discover" their "own" algorithms**, without needing to be explicitly told what to do by any human-developed algorithms."

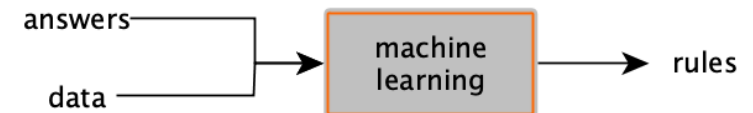`https://en.wikipedia.org/wiki/Machine_learning`

Machine Learning
A change of paradigm

From hand-crafted rules

To training

Diagrams borrowed from L. Moroney's Introduction to TensorFlow for Artificial Intelligence, Machine Learning, and Deep Learning

---

Supervised vs Unsupervised

Supervised The algorithms build a mathematical model of a set of data including. . .
- inputs
- desired outputs

Unsupervised The algorithms take a set of data that contains. . .
- only inputs

. . . and find structure in the data

`https://en.wikipedia.org/wiki/Machine_learning`

Naïve Bayes

Naïve Bayes

1. Introduced in the IR community by Maron (1961)
2. First machine learning approach
3. It is a supervised model
4. It applies Bayes' theorem with strong (naïve) independence assumptions between the features:
   - they are independent
   - they contribute "the same"

Naïve Bayes
A conditional probability model

Given an instance represented by a vector

$$\mathbf{x} = (x_1, \ldots, x_n) \tag{1}$$

representing n independent features $x_1, x_2, x_3, \ldots, x_{n-2}, x_{n-1}, x_n$

n could be $|V|$ (the size of the vocabulary)[3]

The model assigns to instance $\mathbf{x}$ the probability

$$p(C_k \mid \mathbf{x}) = p(C_k \mid x_1, \ldots, x_n) \tag{2}$$

for each of the k possible outcomes C_k

where $C_k = \{c_1, \ldots, c_k\}$

From https://en.wikipedia.org/wiki/Naive_Bayes_classifier

[3]$|\mathbf{x}|$ is the cardinality of $\mathbf{x}$

Naïve Bayes'

The rest of Naïve Bayes' is covered in session 5

References

Fernicola, F., S. Zhang, F. Garcea, P. Bonora, and A. Barrón-Cedeño
2020. Ariemozione: Identifying emotions in opera verses. In *Italian Conference on Computational Linguistics*.

Hutto, C. and E. Gilbert
2014. VADER:A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*, Ann Arbor, MI.

Lane, H., C. Howard, and H. Hapkem
2019. *Natural Language Processing in Action*. Shelter Island, NY: Manning Publication Co.

Maron, M.
1961. Automatic indexing: An experimental inquiry. *Journal of the ACM*, 8:404–417.

Zhang, S., F. Fernicola, F. Garcea, P. Bonora, and A. Barrón-Cedeño
2022. AriEmozione 2.0: Identifying Emotions in Opera Verses and Arias. *IJCoL*, 8(2).