Sentiment Analysis with R

Manika Lamba
RLadies Gaborone
08th October 2022

sentiment analysis

Negative   Neutral   Positive

https://textmining-infopros.github.io
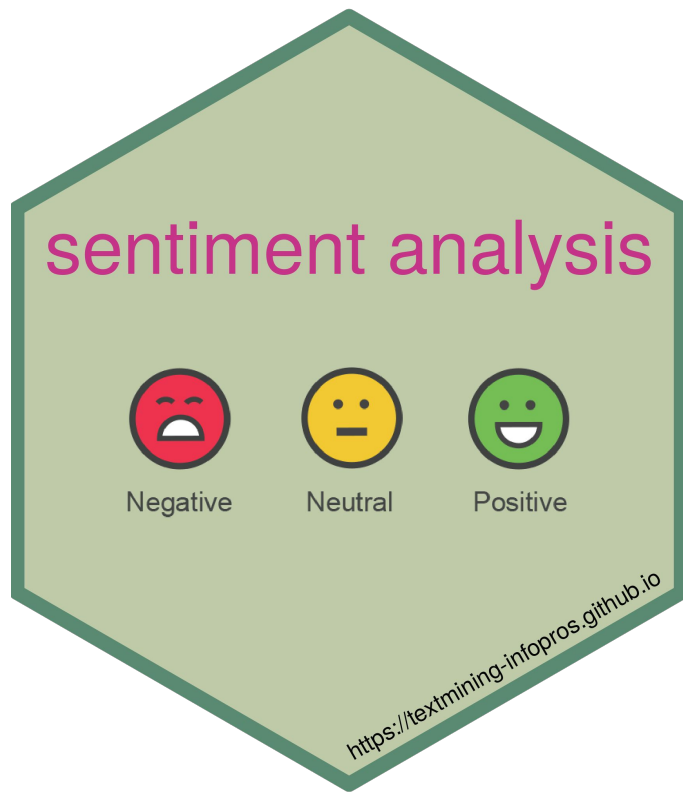
SENTIMENT ANALYSIS USING R

WITH MANIKA LAMBA, AUTHOR OF 'TEXT MINING FOR INFORMATION PROFESSIONALS'

10 AM CAT 08 OCTOBER
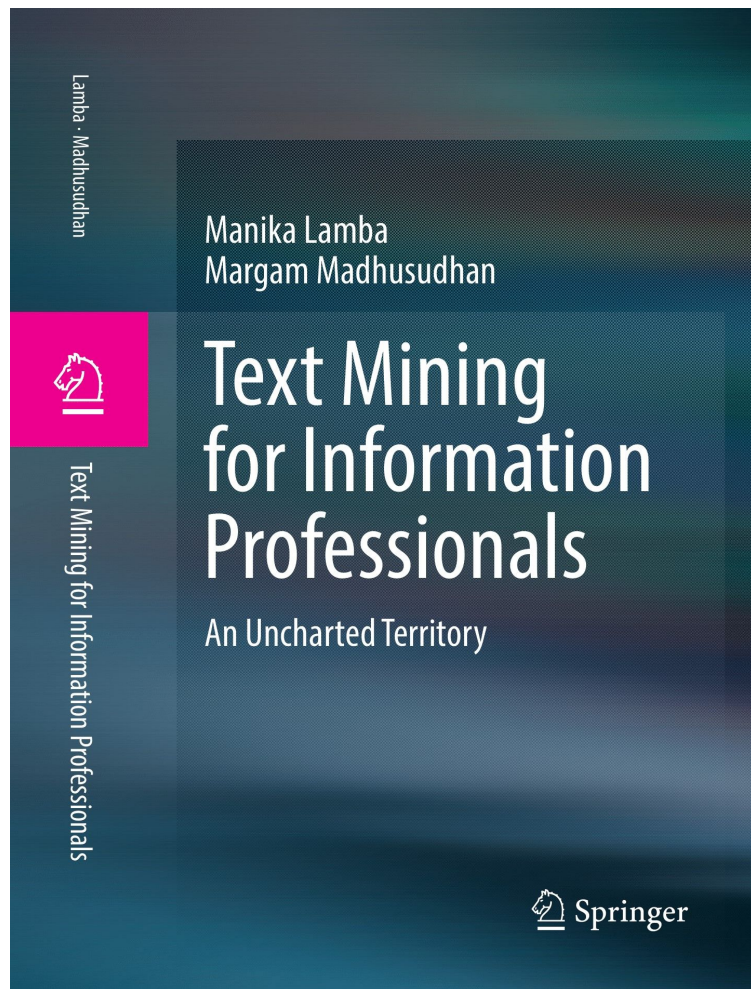
WWW.MEETUP.COM/RLADIES-GABORONE/EVENTS/286102816/

CONTACT US AT: GABORONE@RLADIES.ORG

https://manika-lamba.github.io   @lamba_manika

2

Manika Lamba
Margam Madhusudhan

Text Mining for Information Professionals

An Uncharted Territory

Springer

ToC

Amazon

Publisher Website

Author Website

GitHub
(https://github.com/textmining-infopros/)

3

Learning Goal

By the end of this talk you will:

- Understand what sentiment analysis is and how it works

- Read text from a dataset & tokenize it

- Use a sentiment lexicon to analyze the sentiment

- Visualize the sentiments and emotions

What is Sentiment Analysis?

- Sentiment analysis (also referred to as subjectivity analysis or opinion mining or emotion artificial intelligence) is a NLP technique that identifies important patterns of information and features from a large text corpus

- It analyzes *thought, attitude, views, opinions, beliefs, comments, requests, questions, and preferences* expressed by an author based on emotion rather than a reason in the form of text towards entities like *services, issues, individuals, products, events, topics, organizations, and their attributes*

- It finds the author's **overall emotion** for a text where text can be *blog posts, product reviews, online forums, speech, database sources, social media data, and documents*

What is Sentiment Analysis? (Continue)

It usually consists of three elements depending on the context:

1. Opinions or emotions: An opinion is also referred to as polarity, whereas emotions can be qualitative such as sad, joy, angry, surprise, disgust, or happy or quantitative such as rating a movie on a scale of one to ten.

2. Subject: It refers to the subject of the discussion where one opinion can discuss more than one aspect of the same subject, for instance, the camera of the phone is great, but the battery life is disappointing.

3. Opinion holder: It refers to the author/person who expresses the opinion.

What is Sentiment Analysis? (Continue)

Texts are thus categorized as:

- **subjective** if they reflect opinion
- **objective** if they express a fact

> **SUBJECTIVITY**

- **positive** if they present a state of satisfaction, bliss, or happiness
- **negative** if they present a state of dejection, disappointment, or sorrow
- **neutral** if they present a state that is neither negative nor positive

POLARITY

Polarity is measured on a scale of **−1 to +1**, where −1 means very negative, 0 means neutral, and +1 means very positive. On the other hand, **subjectivity** is measured on a scale of **0−1**, where 0 means very objective, and 1 means very subjective.

Applications

- Business
- Customer Feedback
- Brand Monitoring
- Reputation Management
- Customer Support
- Product Analysis
- Market Research
- Competitive Research
- Voice of Employee
- Voice of Customer

- Financial Marketing
- Social Media Monitoring

Pros

- Classifying data at a large scale based on their polarity
- Real-time analysis
- Comparatively few categories/attributes like polarity compared to text categorization
- Having non-dependent categories/attributes
- Having a relation between topic, domain, and user as opposed to text categorization
- Simple and efficient

Cons

- Complexity in determining the real meaning of the expressions expressed by the opinion holder
- Sarcasm, irony, and implication are common and hard to decipher
- Different words might have different polarity or subjectivity in different contexts
- Same sentence or phrases might have different meaning in different domains
- Sentiments can be negated in different ways and it is difficult to identify such negations
- Dependent on order
- Content-dependent opinion words cannot be processed
- Can result in under- or over-analyzed sentiments if the used dictionary is too sparse or exhaustive

Procedure

1. Pre-processing of data. It may include (i) **removal of replies, mentions, URLs, hashtags, retweets**, (ii) **correction of spelling errors** using a specific dictionary such as Hunspell dictionary, and (iii) **replacing of abbreviations and shorthand notions** using a specific dictionary such as SMS (Short Message Service) dictionary for social media data in addition to other pre-processing

2. Feature extraction to extract the aspects from the processed data. This will be used to **compute the polarity and subjectivity.** This may include determining the **n-gram with their frequency, part-of-speech tagging, identifying phrases and idioms, positioning of terms, negation, or syntactic patterns like collocations**

Procedure (Continue)

3. Performing sentiment analysis using an appropriate algorithm or Open-source tool. The tools calculate the compound sentiment score for each article/post, giving an overall sentiment score

4. Visualizing the sentiments

Available R Packages

Some of the popular packages in R to perform sentiment analysis include:

- syuzhet
- SentimentAnalysis
- mscstexta4r
- Sentimentr
- Quanteda
- tidytext

Data

PART-I

A CSV file containing 5000 book reviews web-scrapped from Amazon in 2018

PART-II

Speeches given by the President of the United States to a joint session of congress every year in 228 .txt files

Research Questions

PART-I

RQ1: How the narrative and emotions are structured across the book reviews?

PART-II

RQ1: How sentiment has changed over time, from 1989 to 2017?

RQ2: Whether different presidents tend to have more negative or more positive sentiment?