Social Media Analytics: Introduction and Sentiment Analysis

COMP61332: Text Mining
Week 5
Riza Batista-Navarro

Social Media Analytics (SMA)

Social media: a channel for people to express their **thoughts**, **beliefs** and **opinions**

- highly subjective
- individually, cannot be taken as facts
- but can provide insight if information is aggregated

Social Media Analytics

> computational analysis of information made available in **online data-sharing platforms**

What is SMA applied on?

Historically: Amazon product reviews

More recently:

- **opinions** on products or services delivered by key industry players
- **perceptions** of social-economic issues, political discussions (e.g., elections)
- **emotions** and **appeals** in the wake of urgent crises (e.g., natural disasters, disease outbreaks, terror attacks)

Sources of data

Social media or networking sites

Twitter

Facebook

Weibo

Instagram

Snapchat

Sources of data

User-generated content: data contributed by online users

multimedia-sharing platforms

review sites

online fora

Types of data

Structured data

- user profile information
  - age
  - gender
  - location
- interactions
  - number of posts
  - number of replies (retweets, likes)
  - number of followers

Types of data

Unstructured data: the actual text

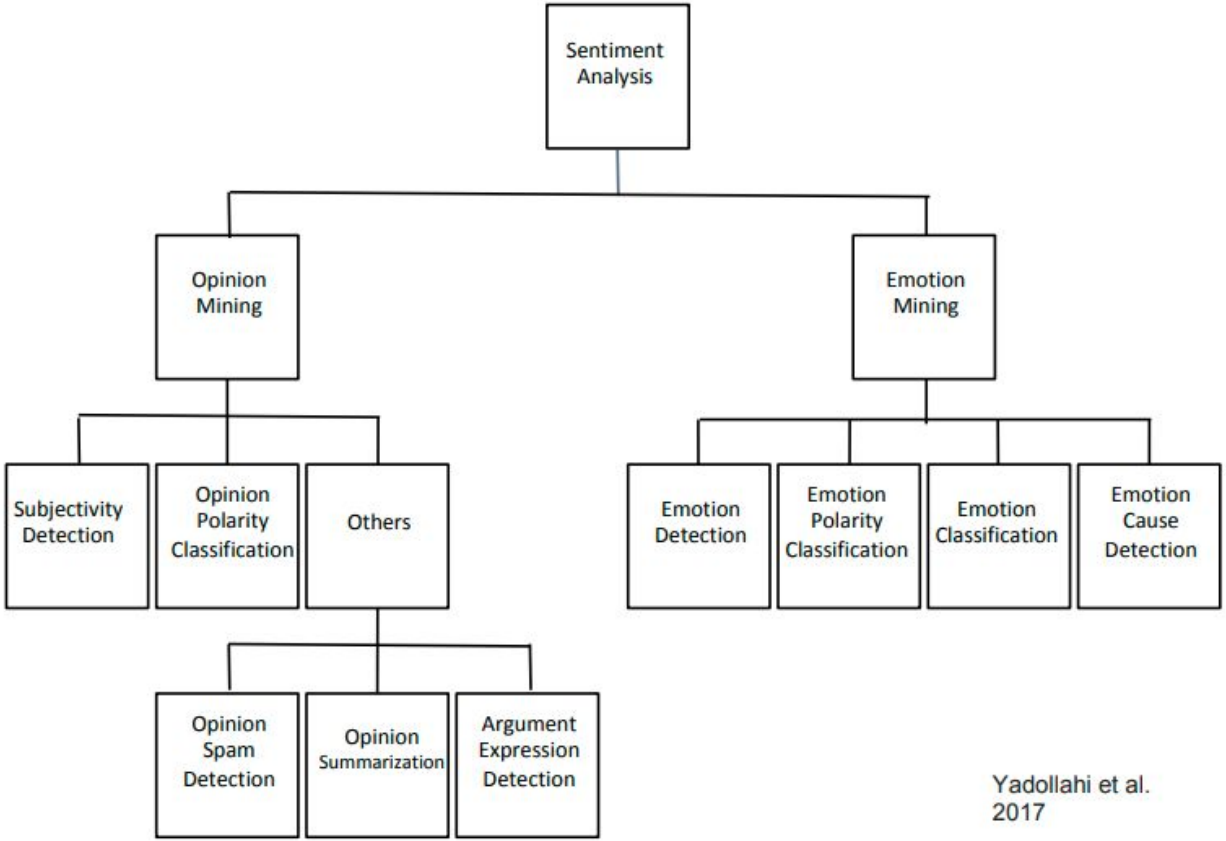
- can be short, often are incomplete sentences
- noisy
  - intentional misspellings (*"very tiiiiired", "helloooo!"*)
  - acronyms (*"AFAIK", "BFF"*)
  - slang (*"woke"*)
  - code-mixing: combining different languages
  - URLs, hashtags, user mentions, emojis

Typical Text Mining Tasks

Sentiment Analysis/Emotion Detection

Topic Modelling

Sentiment Analysis



Yadollahi et al. 2017

Sentiment Analysis

Given a piece of text, does it contain any sentiment?

If yes, is it positive or negative?

Possible labels: **neutral** (no sentiment), **positive**, **negative**

Examples:

> *Today's lecture covers sentiment analysis.* [neutral]
>
> *It's fun and easy to do sentiment analysis!* [positive]
>
> *I don't like reading all of the negative Tweets!* [negative]

Sentiment Analysis

Cast as a **text classification** task

Approaches:

- Lexicon/dictionary-based + rules
- Machine learning-based

Sentiment Analysis using Dictionaries

Calculating a sentiment score by matching words

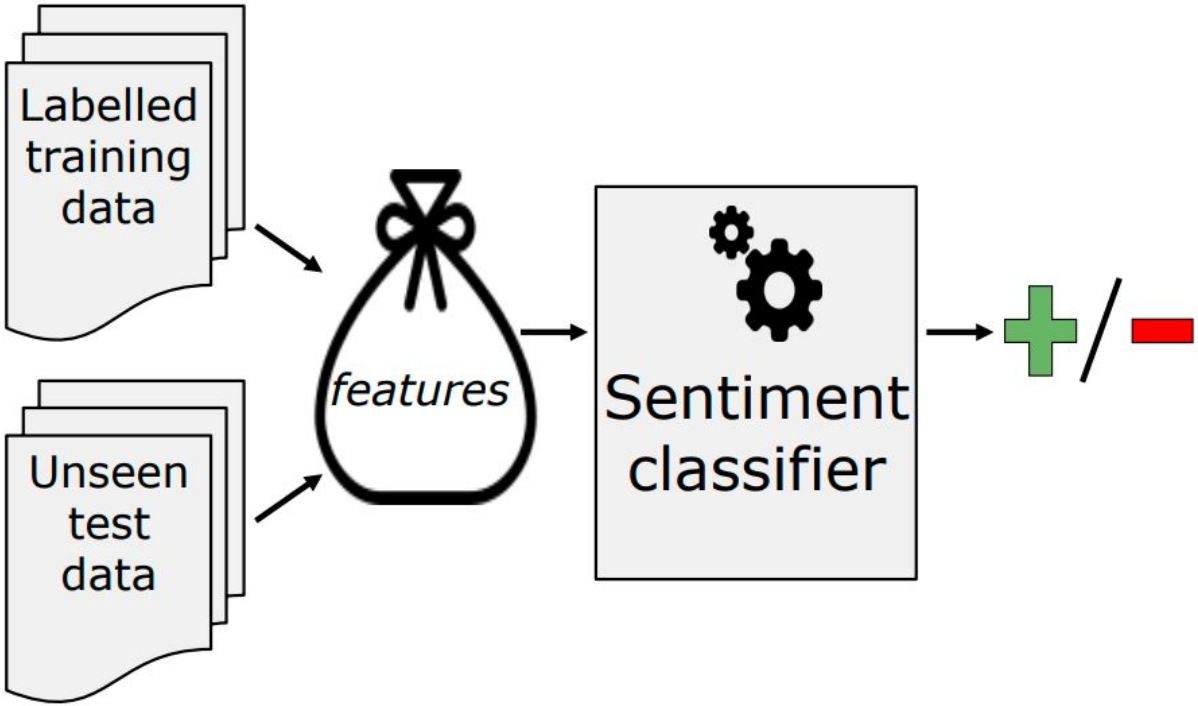
+1 +1

It is **fun** and **easy** to do sentiment analysis!

-1 -1

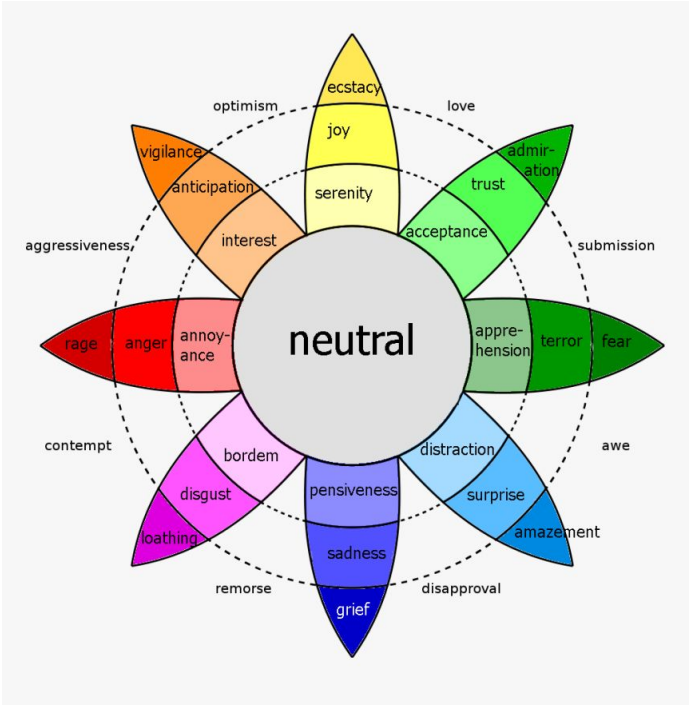
I **don't like** reading all of the **negative** Tweets!

Sentiment Analysis using Machine Learning



Emotion Detection

Bears similarities to Sentiment Analysis but **finer-grained**; cast as a **multi-class classification** task



Plutchik's wheel