Lecture 04:
Sentiment Analysis

OVERVIEW

1. What is sentiment analysis?
2. Motivations for sentiment analysis
3. Types of sentiment analysis
4. How sentiment analysis works
5. Challenges
6. Application

WHAT IS SENTIMENT ANALYSIS?

- Sentiment analysis refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials.

- Technique to analyzing textual data to determine whether it is positive, negative, or neutral (or other predefined class)

Also known as….

- Opinion mining
- Sentiment mining
- Subjective detection

WHAT IS SENTIMENT ANALYSIS?

- Identify the orientation of opinion in a piece of text

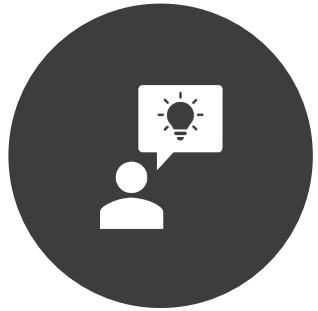
- Can be generalized to a wider set of emotions

MOTIVATION

KNOWING SENTIMENT IS A NATURAL ABILITY FOR HUMAN BEING.

CAN A COMPUTER BE TRAINED TO DO IT?

SENTIMENT ANALYSIS AIMS AT GETTING SENTIMENT-RELATED KNOWLEDGE ESPECIALLY FROM THE HUGE AMOUNT OF TEXT FROM THE INTERNET.



SENTIMENT ANALYSIS IS A MACHINE LEARNING TECHNIQUE THAT USES NLP TO IDENTIFY POSITIVE AND NEGATIVE SENTIMENT IN TEXT.

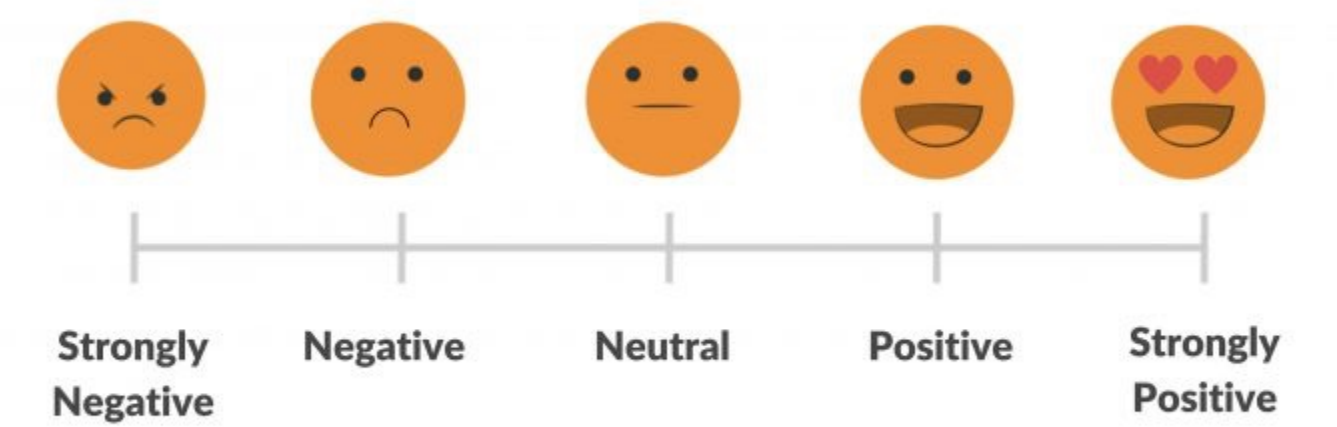
MOTIVATION

- Companies and other brands can use this NLP technique to detect
  - Brand monitoring
  - Social Media Monitoring (user sentiments from feedback),
  - Market research,
  - understand customers' needs
  - Customer services

- For example, sentiment analysis can help you analyse 10,000+ reviews related to your product. You can use the insights to determine if the customers are happy with your product and customer service.

TYPES OF SENTIMENT ANALYSIS

1. **Fine-grained sentiment analysis:** looks at text polarity to get insights on sentiments expressed



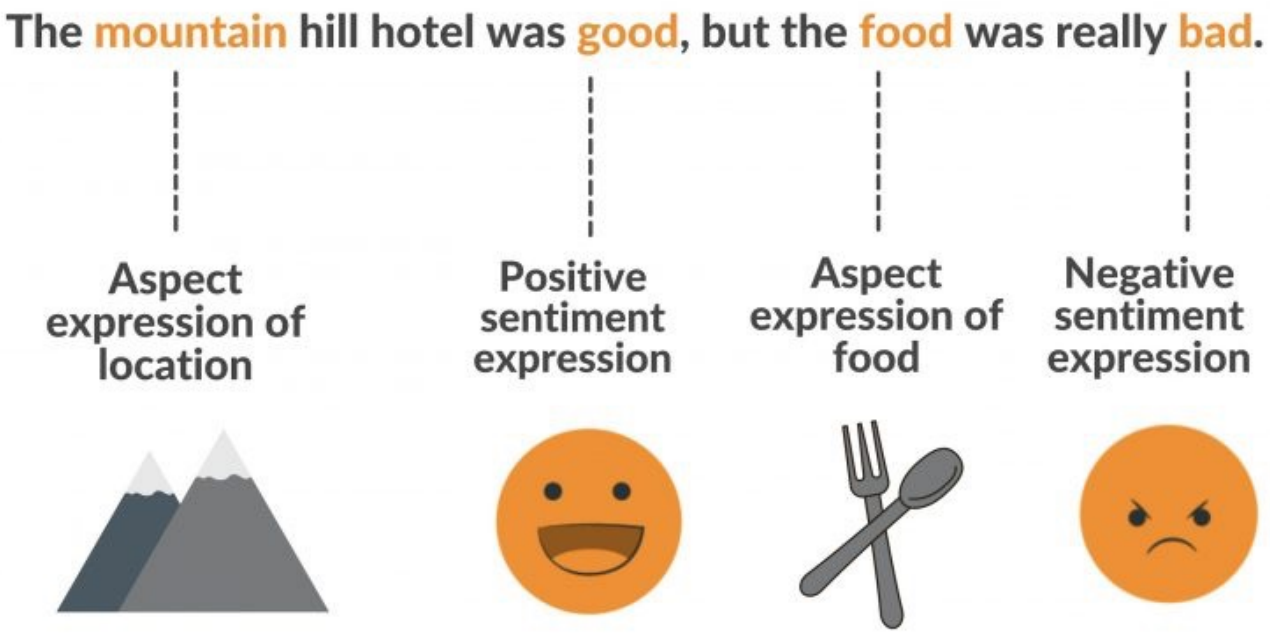
TYPES OF SENTIMENT ANALYSIS

2. Emotion detection: focus on detecting emotions such as happiness, frustration, anger, sadness

- Example:
- **Negative:** Your customer support is soo **bad**, I **regret** buying your product.
- **Positive:** This is **bad**-ass, I **regret** not trying it earlier.

TYPES OF SENTIMENT ANALYSIS

- **3. Aspect-based Sentiment Analysis:** when businesses want to know aspect of their products that are often discussed



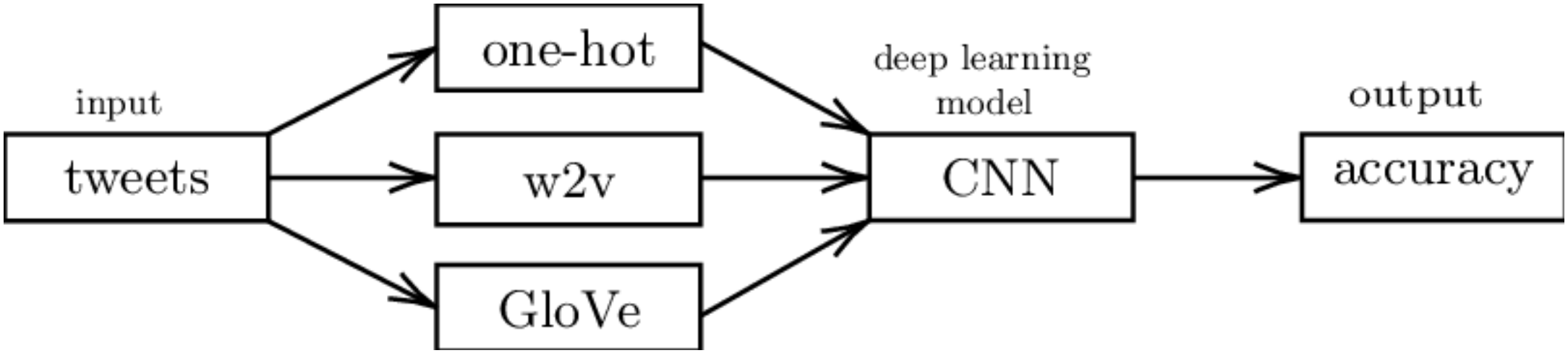
HOW SENTIMENT ANALYSIS WORKS

Sentiment analysis uses natural language processing to interpret human language and machine learning to identify the emotions expressed in textual data.

Three main approach to develop a sentiment analysis model (depends on the volume of data and target accuracy).

- **Rule-Based:** performs sentiment analysis based on predefined rules
- **Automatic:** leverage machine learning techniques to learn sentiments from data.
- **Hybrid:** combines both rule-based and automatic analysis approaches.

SENTIMENT ANALYSIS



HOW SENTIMENT ANALYSIS WORKS – HYBRID

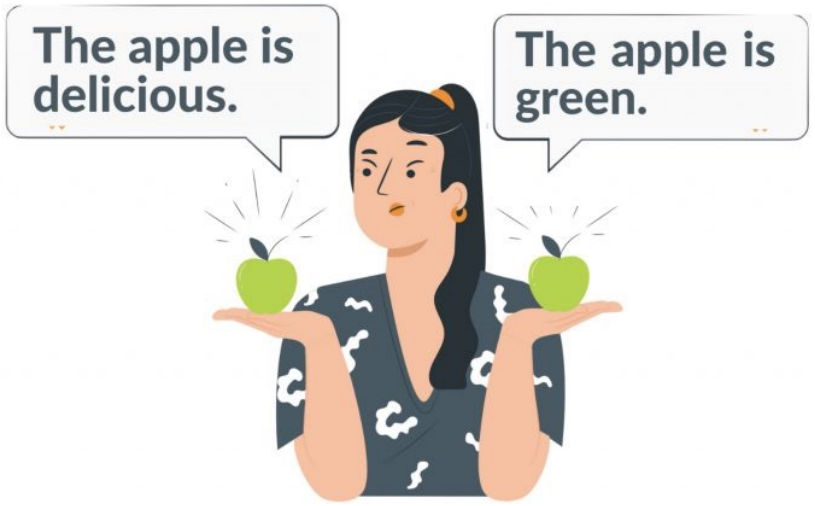
Bring together elements of rule-based and automatic systems.

- Start with automatic
- Use rule based to fine tune output
- Vice versa

A major benefit of these methods is that they usually give more precise results.

CHALLENGES - SUBJECTIVITY

- Sentiment analysis is one of the most difficult task in NLP.
- **Text Subjectivity and Tone:** textual data is usually of two types:
  - Subjective: often contains sentiment
  - Objective: does not often contain sentiment



CHALLENGES

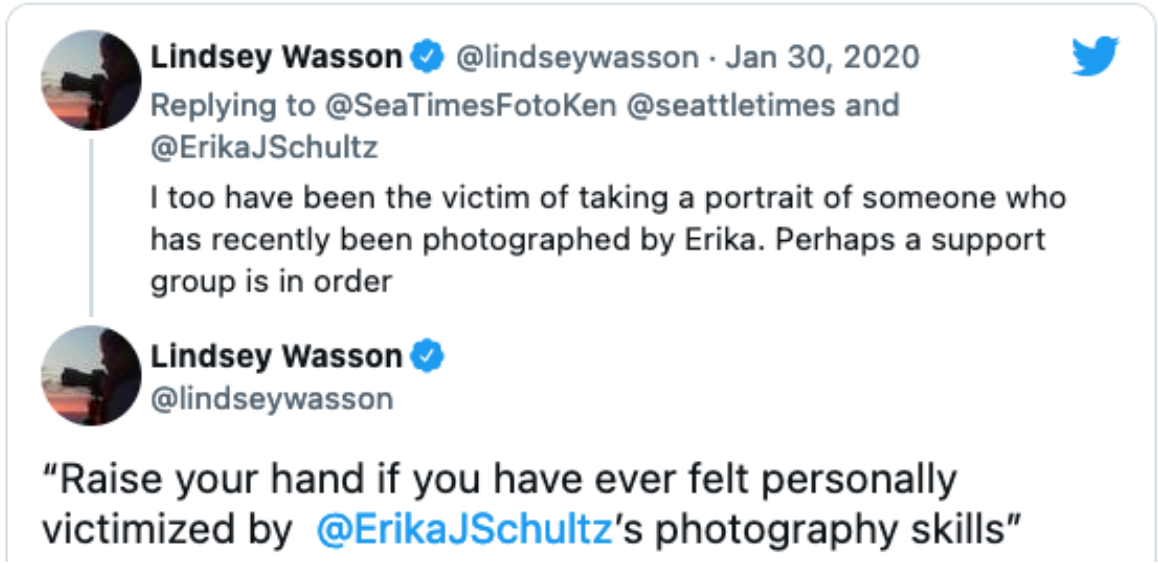
- **Text Polarity and Context**
  - Analysing text with context increases accuracy
  - Computers learn to analyse context only when they are specifically included

- Question:
  - *What is it that you like about the game?*
    - Positive context in the question
  - *What is it that you don't like about the game?*
    - Negative context

CHALLENGES

- **Irony and Sarcasm in Text**
  - Very difficult sentiments for computers to detect
  - People communicate negative sentiments through positive words

What's the sentiment of this tweet?



Based on the context the author is praising the skills

CHALLENGES

Opinion orientation

- change according to valence shifters (e.g., negation: *not* etc.)
- ***But*** clauses ("the pictures are good, but the battery life ...")
- Dictionary-based: Use semantic relations (e.g., synonyms, antonyms)

Corpus-based orientation

- learn from labelled examples
- Disadvantage: need these (expensive!)
- Advantage: domain dependence

CHALLENGES

- **Text Comparison**
  - One main challenge in general NLP is to capture comparison in text

CHALLENGES

- **Emojis in Text**
  - Emojis play crucial rule in expressing sentiments
  - Increasing popularity in the use of emojis to communicate
  - Emojis are replacing slangs
  - The need for character and word level consideration to account for emoji

SUMMARY

- Generally modeled as a classification or regression task
- Features:
  - Negation is important
  - Using all words works well for some task
  - Finding subsets of words may help in some task
    - Hand-built polarity lexicons
    - Use seeds and semi-supervised learning to induce lexicon