Sentiment analysis

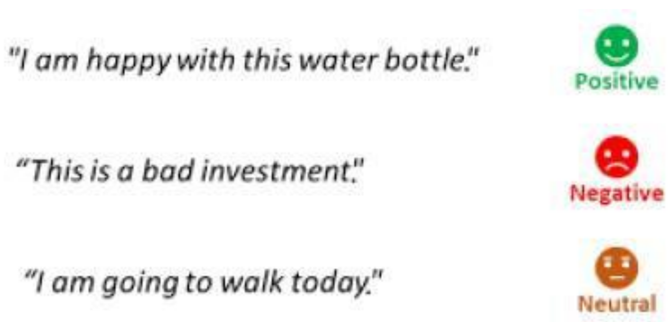
Prof. Luca Cagliero
Dipartimento di Automatica e Informatica
Politecnico di Torino

Outline

- Problem statement
- Levels and applications
- Subtasks
- The Sentiment Analysis pipeline
- State-of-the-art models

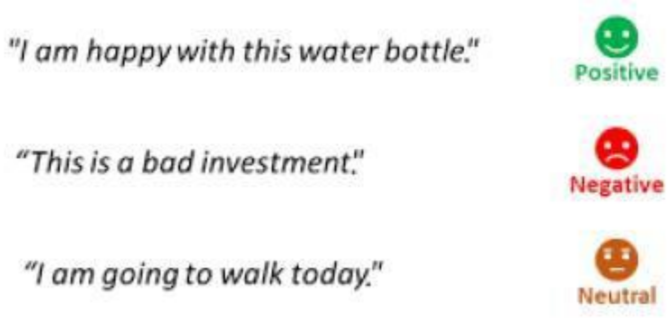
Sentiment analysis

- Extract writer's feeling, opinion, emotions, likes/dislikes
  - Also known as **opinion mining**
- Identify the opinion/human behavior of a person from plain text
- It can rely either on traditional NLP rules or on Machine Learning



Sentiment analysis

- Use cases
  - Hotel review analysis
  - News trading
  - Hate speech detection
  - Advertisement placing
  - . . .



Example of sentiment analyzer: VADER

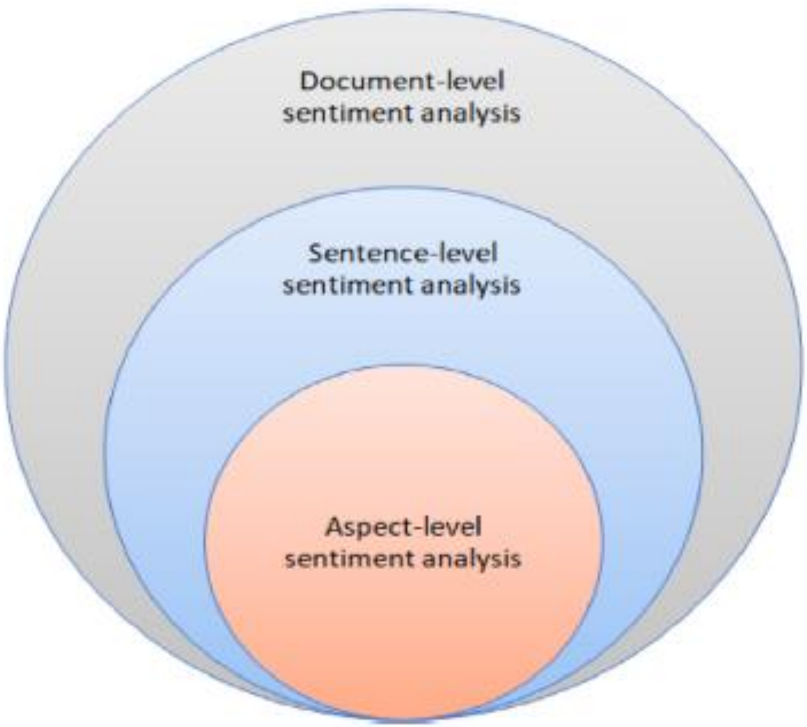
- Input
  sentiment_analyzer_scores("The phone is super cool.")
- Output
  The phone is super cool --- {'neg': 0.0, 'neu':0.326, 'pos': 0.674, 'compound': 0.7351}

https://github.com/cjhutto/vaderSentiment

Sentiment analysis: a special case of text categorization

- Assign a predefined label to a document or a text snippet
  - Also known as text classification
- Supervised process
  - given a set of labeled documents/snippets, learn how to automatically label new, unlabeled ones
- If multiple labels are allowed
  - multi-label text classification
- It can rely either on traditional NLP rules or on Machine Learning

Levels and applications of sentiment analysis



A comprehensive survey on sentiment analysis: Approaches, challenges and trends. Marouane Birjali, Mohammed Kasri, Abderrahim Beni-Hssane. Knowledge-Based Systems. 2021

Document-level sentiment analysis

- Classify whether the whole document expresses a negative or positive sentiment or opinion
- Works best when the document is written by one person
- Unsuitable for documents that evaluate/compare multiple entities
- Common approaches
  - Estimate the document sentiment polarity as a combination of the sentence-level scores
  - Embed document-level content into vector representations and measure the distance from labeled documents

Sentence-level sentiment analysis

- Classify a sentence that expresses
  - A factual information
  - A subjective expressing view
  - An opinion
- Can be based on entity recognition
- It does not find precisely what people like or dislike

Levels and applications of sentiment analysis

- Aspect-level
  - Fine-grained analysis
  - Find sentiments with respect to specific implicit aspects or entities
- Example

The camera of iPhone X is awesome but the battery does not last long

Entity Camera -> Positive sentiment

Entity Battery -> Negative sentiment

Aspect-level sentiment analysis: example of application

- Hotel review analysis
  - Aspects
    - *Location*
    - *Food*
    - *WiFi*
    - *Service*
    - *Room*
    - ...

Image taken from https://www.trustyou.com/wp-content/uploads/2015/08/trivago.png

Sentiment analysis: applications

- Business Intelligence
  - Make product/service improvements
  - Study the customers' feedback
  - Plan new marketing strategies
  - Forecast market movements
- Recommender systems
  - Suggest relevant items to users
    - E.g., movies, music, products to buy
- Government intelligence
  - Identify opinions on government policies
  - Monitor publics' mood in real time
- Healthcare
  - Shape healthcare services
  - Allow healthcare actors to obtain information about diseases, epidemics, treatments, patients' mood, etc.

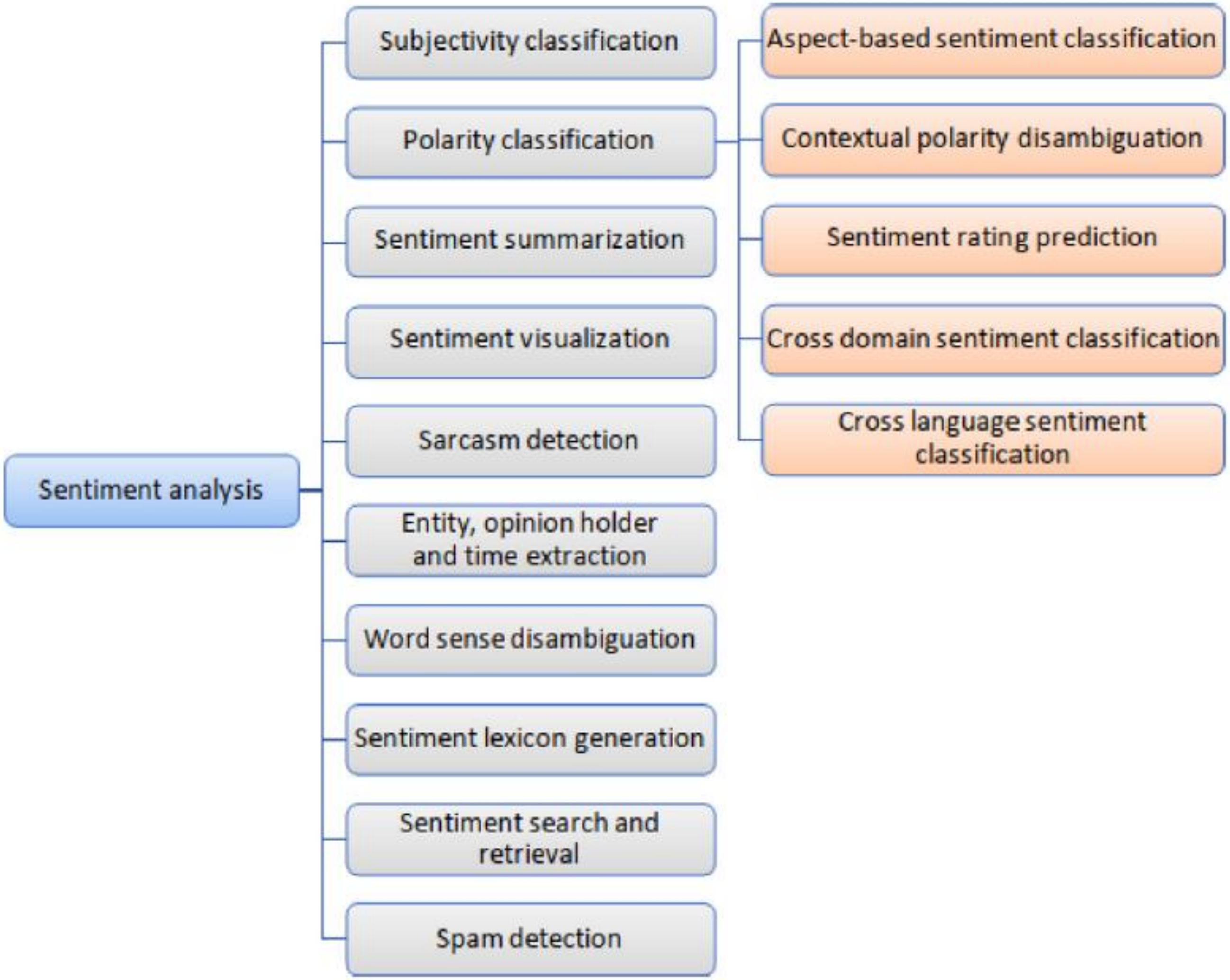
Data sources for Sentiment Analysis

- Social media
  - Analysis of User-Generated Content
  - Individual and collective life and behavior
  - Web- and mobile-based Internet Applications
- Review websites
  - Where people can post opinions or reviews about a particular entity
  - Often integrated into e-commerce platforms
- Forums
  - Posting text messages to share thoughts and ideas
- Interviews
  - Transcripted record of completed oral interviews
  - Either in real time or from audio or video recordings

Data connectors for Sentiment Analysis

- Application Programming Interfaces
  - Provide access to textual content using HTTP-based protocols
  - Based on search query
    - Keyword-based, location-based, etc.
- Open datasets
  - Data collected by academic institutions
  - Useful for Deep Learning model training
- Web scraping
  - Automatic process of extracting data from websites that hold valuable data
  - Main challenges: quality, noise, privacy
- Crowdsourcing
  - Outsource data creation or annotation
  - Useful for creating large data volumes in short time
  - E.g., Amazon Mechanical Turk (https://www.mturk.com/)

Sentiment analysis subtasks



A comprehensive survey on sentiment analysis: Approaches, challenges and trends. Marouane Birjali, Mohammed Kasri, Abderrahim Beni-Hssane. Knowledge-Based Systems. 2021

Data preprocessing for Sentiment Analysis

- Tokenization
  - Necessary for sentence- and aspect-level sentiment analysis
- Stopword removal
  - Unnecessary when embedding models are applied
- Text cleaning
  - Abbreviations/repetitions/punctuation can be relevant to sentiment detection
    - E.g., *Ooookkkkkkk!!!!!!*

Feature extraction for Sentiment Analysis

- N-gram presence and frequency
  - Commonly used in Information Retrieval
  - Based on weighting scheme
    - E.g., Tf-idf, BM25
- POS tagging
  - Adjectives and adverbs are very important indicators of opinions

Feature extraction for Sentiment Analysis

- Opinion words/phrases
  - Expressions that are commonly associated with positive or negative sentiment
    - E.g., *Worderful, awful*
  - They can be domain-dependent
    - E.g., *capital increase* in a financial news
- Negations
  - Also called *opinion shifters* or *valence shifters*
  - Words that may shift/change the opinion orientation and reverse the sentiment polarity
    - E.g., *not, never, nobody, nowhere, neither*
  - Important notice
    - These words could be accidentally removed by stopword elimination

Feature selection for Sentiment Analysis

- Distinguish between
  - Relevant features
  - Irrelevant features (to be removed)
  - Redundant features (to be removed)
- Methods
  - Lexicon-based
    - Human-generated features
      - Collect term with a strong sentiment
      - Find synonyms etc.
    - Accurate but expensive!
    - E.g., WordNet
  - Statistical
    - Automatic feature analysis
    - Efficient but prone to errors

Statistical feature selection for Sentiment Analysis

- Filter approach
  - Based on statistical tests/measures applied to the training data
    - Chi Square test, Mutual Information, Information Gain
  - No Machine Learning
- Wrapper approach
  - Evaluate a subset of features based on the resulting performance of the applied ML algorithm
    - E.g., SVM classification
  - Computationally intensive

Statistical feature selection for Sentiment Analysis

- Embedded approach
  - Incorporate feature selection into the ML algorithm execution
  - Model dependent
    - E.g., decision trees
- Hybrid approach
  - A mix of the above

Machine Learning approaches for Sentiment Analysis

- Supervised
  - Require labeled data
    - Linear, probabilistic, rule-based, decision trees, etc.
- Unsupervised
  - Applied in the absence of labeled data
    - E.g., partition-based clustering, hierarchical clustering
- Semi-supervised
  - A small set of initial labeled data is used to guide the feature learning procedure
  - Commonly associated with self-training
    - Step 1: the classifier is trained using a small amount of labeled data
    - Step 2: the trained classifier is used to classify the unlabeled data and transform them in the newly labeled data
    - Step 3: repeat Step 1 using all the labeled data (both new and old)

Machine Learning approaches for Sentiment Analysis

- Generative
  - It assumes that data in different polarity classes follow different distributions
  - Estimate the distribution parameters
    - Requires at least one labeled data per class
- Co-training
  - It assumes that data can be represented by two independent views
    - Each view has information about each data
  - Two separate classifiers are trained to teach each other based on the shared information between them
  - The process stops when all unlabeled data have been used or a specific number of iterations has been reached

Machine Learning approaches for Sentiment Analysis

- Self-training
  - It assumes that data in different polarity classes follow different distributions
  - Estimate the distribution parameters
    - Requires at least one labeled data per class
- Graph-based
  - Model data as a graph
    - Vertices are instances
      - E.g., sentences
    - Edges indicate similarities between instances
- Multi-view learning
  - Treat the problem by considering multiple viewpoints
  - Train a classifier per view
  - The overall result is an agreement between multiple views
    - As in ensemble methods

Machine Learning approaches for Sentiment Analysis

- Reinforcement learning
  - Based on a trial-and-error approach
  - An agent is rewarded in the next time step based on the evaluation at the previous stage
  - Foster interactions between the agent and the environment
  - Learn repetitive action-effect combinations
    - Similar to human beings

Machine Learning approaches for Sentiment Analysis

- Transfer learning
  - Also called *cross-domain sentiment analysis*
  - Apply the knowledge already learnt to address one task to another task
    - E.g., from movie reviews to book reviews

- Multimodal learning
  - Include modalities other than text
    - E.g., audio, video, images
  - Require information fusion
    - E.g., in a unified embedding space

Machine Learning approaches for Sentiment Analysis

- Deep Learning
  - Learn latent feature representation in both supervised and unsupervised manner
  - Require ad hoc computational resources

Deep Learning solutions

- Deep Learning
  - Learn latent feature representation in both supervised and unsupervised manner
  - Require ad hoc computational resources

Attention-based approaches to aspect-based sentiment analysis

- Adopt attention-based Neural Networks to implicitly aspects with opinion words
- Key idea
  - Opinion words usually do not appear too far from the aspect

Attention-based approaches to aspect-based sentiment analysis

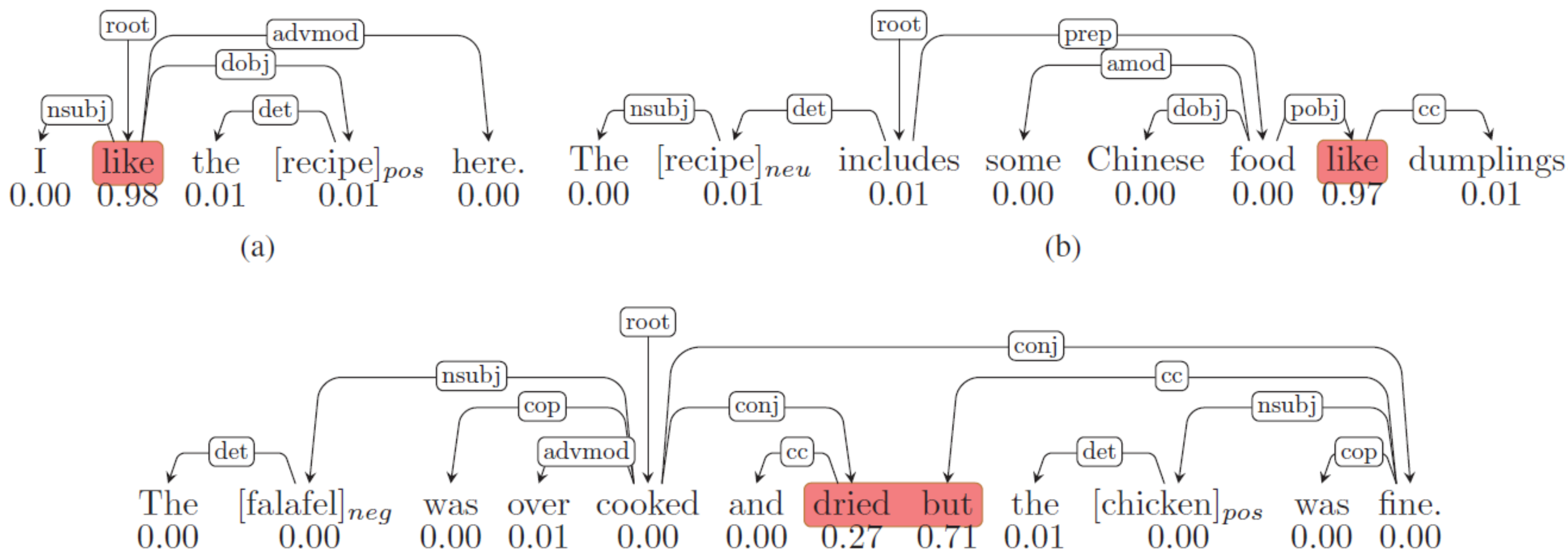
- Explainable approach
  - Identify the aspects
    - E.g., using Named Entity Recognition
  - Analyze the attention score map to detect the words that mainly influence that aspect

Legend: Negative Neutral Positive

True Label	Predicted Label	Attribution Label	Attribution Score	Word Importance
1	POSITIVE (1.00)	POSITIVE	1.96	[CLS] i like you , i love you [SEP]

Relational Graph Attention Network for Aspect-based Sentiment Analysis. Kai Wang, Weizhou Shen, Yunyi Yang, Xiaojun Quan, Rui Wang. ACL 2020: 3229-3238

Combine dependency parsing and attention-based neural network learning



Relational Graph Attention Network for Aspect-based Sentiment Analysis. Kai Wang, Weizhou Shen, Yunyi Yang, Xiaojun Quan, Rui Wang. ACL 2020: 3229-3238

Additional reading on R-GAT

- Relational Graph Attention Network for Aspect-based Sentiment Analysis. Kai Wang, Weizhou Shen, Yunyi Yang, Xiaojun Quan, Rui Wang. ACL 2020: 3229-3238
- Download and read the paper: https://aclanthology.org/2020.acl-main.295/

Additional reading on Hate Speech Detection

- Hate speech detection using static BERT embeddings. Gaurav Rajput and Narinder Singh punn and Sanjay Kumar Sonbhadra and Sonali Agarwal. 2021
- Download and read the paper: https://arxiv.org/pdf/2106.15537.pdf

Acknowledgements and copyright license

- Acknowledgements
  - I would like to thank Dr. Moreno La Quatra, who collaborated to the writing and revision of the teaching content
- Affiliation
  - The author and his staff are currently members of the Database and Data Mining Group at Dipartimento di Automatica e Informatica (Politecnico di Torino) and of the SmartData interdepartmental centre
    - https://dbdmg.polito.it
    - https://smartdata.polito.it

Thank you!