Sentiment Analysis

Ramaseshan Ramachandran

DEFINITION

The process of computing and classifying opinions from an unstructured text

CLASSIFICATION OF EMOTION

- Emotion: Most of living react in the same way to something happening outside or inside them - anger, sadness, joy, fear, shame, pride, elation, desperation
- Mood Definition: A diffuse affect state characterized by a change in subjective feeling, often without apparent cause
  - *Intensity*: Low intensity but relatively long duration
  - *Examples*: Cheerful, gloomy, irritable, listless, depressed, buoyant
- Interpersonal Stance: Affective stance taken toward another person in a specific interaction
  - *Examples* : Distant, cold, warm, supportive, contemptuous
- Attitudes: Relatively enduring, affectively colored beliefs, preferences, and predispositions
  - *Components*: Beliefs, preferences, and predispositions
  - *Objects*: Objects or persons

OTHER NAMES

- Opinion mining
- Evaluation Analysis
- Appraisal Assessment
- Attitude mining
- Emotion extraction
- Subjectivity analysis
- Aspect extraction
- Affect extraction
- Review mining
- . . .

CORE CONCEPTS

- Sentiment reflects emotional tone of text
- Polarity: positive, negative, or neutral
- Emotion detection for joy, anger, sadness
- Aspect-based sentiment: specific components focus
- Analyzed at document, sentence, phrase level

FORMAL DEFINITION

- An opinion consists of a target and sentiment.
- Target (g): Entity or aspect opinion is about.
- Sentiment (s): Positive, negative, neutral, or numeric score.
- Example: 1-5 stars or sentiment polarity.
- Polarity types: Positive, negative, or neutral sentiment.
- Targets can include entire entities or specific aspects.
- Example: Target in "Canon G12" is camera.
- Example: Target in "Picture quality" is image aspect.

TARGET DOMAIN EXAMPLES

Sentiments on the attributes of products

- Example: Mobile phone case - design, fitment, price, durability, protection, shell type (poly carbonate, plastic), color, weight, etc.

Services

- Banks, restaurants, sports centers, fitness centers, repair shops, etc.

Individuals

- Example: Fitness instructors/trainers, teachers,

Public Issues

- Political, non-political, governance, etc.

Social media

- Monitoring social media for issues, products, trends, etc

Events

- Music events, workshops, Topics

. . .

SAMPLES

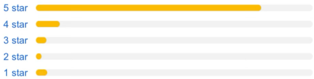
Product details

Smartphone · Single SIM · iOS · 5G · Wireless Charging · Dual Lens · With OLED Display · Facial Recognition · 2778 x 1284 · Water Resistant

The camera system receives its most significant boost yet, with next-generation technology that captures far more information, while the super-slick Pro Motion display ensures a smoother experience when surfing, gaming, or simply checking social media.

Reviews

4.6
★★★★★
11,884 reviews



Long battery life (478) · Easy to use (249) · Heavy (215) · Performs well (190) · Attractive (150) · Easy to set up (110) · Large display (110) · Quality display (86)

★★★★★ 7 December, 2021

THE BOTTOM LINE Apple's iPhone 13 Pro Max is the ultimate mobile content creation machine, with the best camera and longest battery life of any iPhone. The ultimate phone for photo and video creators! The iPhone 13 Pro Max (starting at $1,099) is the ultimate professional content creator's phone. It combines Apple's excellent camera algorithms and software support with true two-day battery life for a massive phone that's always ready to realize your dreams. While the standard iPhone 13 (starting at $799) seems to be the best choice for most people, with a terrific balance of size, power, battery life, and price, the iPhone 13 Pro Max is a terrific alternative for heavy users and artists, with its killer cameras and beautiful buttress of a battery. The iPhone 13 Pro ... More

chantal.v · Review provided by influenster.com

All reviews

Customer reviews

4.2 out of 5

2,071 global ratings

5 star	57%
4 star	23%
3 star	10%
2 star	4%
1 star	6%

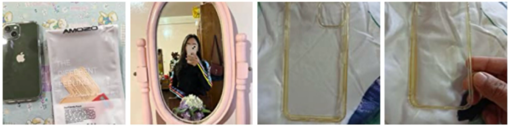
How are ratings calculated?

By feature

Value for money	4.2
Sturdiness	4.1
Durability	4.1
Sheerness	4.0

See less

Reviews with images



See all customer images

Read reviews that mention

value for money · yellow within · good quality · within a week · fits perfectly · within few days · started turning · perfect fit · fit for iphone · back cover · overall good · sturdy back · attracts a lot

Top reviews

Top reviews from India

Amazon Customer

Not sure about the Durability, But so far, so good!

Reviewed in India on 20 October 2022

Customers say

Customers like the condition and appearance of the shoes. They mention that it's value for money, the shoes are sturdy and premium. Some say that the durability is questionable and that the sole starts damaging within 2 uses. Opinions are mixed on fit, comfort, and quality.

AI-generated from the text of customer reviews

Appearance | Condition | Quality | Comfort | Fit | Performance | Durability | Sole

Customer reviews

4.1 out of 5

419 global ratings

5 star	52%
4 star	24%
3 star	14%
2 star	3%
1 star	7%

How are ratings calculated?

To calculate the overall star rating and percentage breakdown by star, we don't use a simple average. Instead, our system considers things like how recent a review is and if the reviewer bought the item on Amazon. It also analyses reviews to verify trustworthiness.

Customers say

Customers like the appearance and weight of the artificial plant. They mention it looks real and good for decoration. However, some customers are not happy that the pot is made of plastic. They are mixed on quality and size.

AI-generated from the text of customer reviews

Appearance | Weight | Quality | Size | Material | Packaging

Reviews with images

See all photo

SAMPLE SENTENCES

- When you don’t want to spend a whole lot of cash but want a great deal....
- This is the shop to buy from
- It is a very cute case
- Cannot argue with the price or appearance
- The jewels do fall off
- It is a beautiful phone case but it is also hard to remove
- Arrived broken and very flimsy
- Fits perfectly but needs a little attention at the installation
- The design and color combination makes the case simple yet elegant, and not too bold and flashy Don’t believe that these screen protectors have glue in them

- Sentiment analysis can be treated as a classification task.
- Assigns input text into predefined sentiment categories.

  Examples of sentiment categories
  - Positive
  - Negative
  - Neutral
- Classification output can also be represented on a numerical scale.

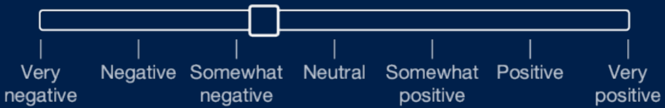


Figure: The labeling interface[1]

WHY USE A SCALE?

- Provides finer granularity than simple positive, negative, or neutral labels.
- Enables better differentiation between sentiments of similar types.
- Allows for more nuanced analysis, e.g., distinguishing between slightly positive (+5) and highly positive (+20).
- Facilitates advanced tasks like regression-based sentiment prediction.

APPROACHES TO SENTIMENT ANALYSIS

- Rule-based: lexicons and manual rules
- Machine learning: Naive Bayes, SVM, others
- Deep learning: CNNs, RNNs, Transformers
- Trade-offs: simplicity vs. computational complexity

- Context impacts interpretation of sentiment
- Sarcasm and irony are difficult to detect
- Sentiment varies across different domains
- Handling ambiguous and mixed sentiments is tricky
- Multilingual and multimodal sentiment needs innovation

APPLICATIONS

- Analyze customer reviews for feedback insights
- Monitor political opinions and public reactions
- Assess sentiment in healthcare and treatments
- Predict audience reactions to entertainment content
- Monitor brand sentiment on social media

- TextBlob: easy sentiment classification library
- VADER: social media sentiment lexicon-based tool
- NLTK: preprocessing and sentiment classification
- Commercial APIs: IBM Watson, AWS Comprehend

TAXONOMY

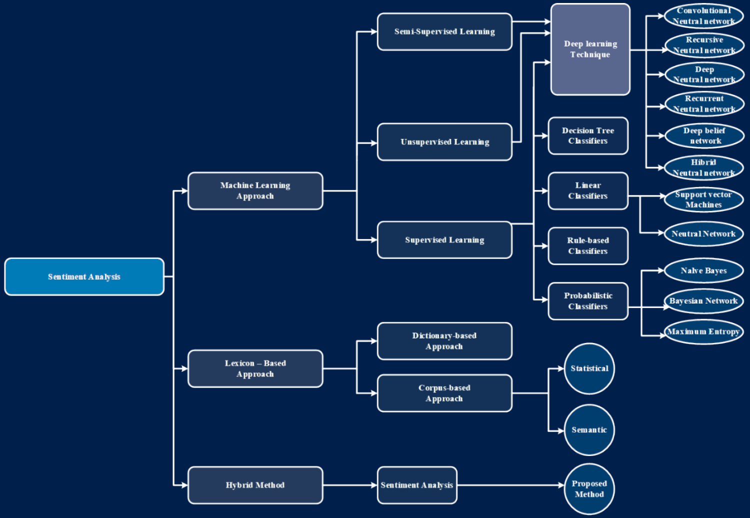


Figure: Taxonomy of sentiment analysis techniques[2].

TRADITIONAL APPROACHES

- Lexicon based Approach
- Naive Bayes Classifier
- Support Vector Machines (SVM)
- Effective in high-dimensional spaces for binary or multi-class classification
- Logistic Regression
- Decision Trees and Random Forests
- K-Nearest Neighbors (KNN)

FUTURE DIRECTIONS

- Combining text, voice, and video sentiment
- Real-time sentiment analysis for live events
- Summarizing sentiment across vast datasets
- Enhancing model explainability and transparency

Lexicon based Approach

- A **lexicon** is a collection of words or vocabulary items in a language.
- It includes information about meanings, forms, and usage.
- Different domains (linguistics, NLP) may define lexicons slightly differently.

KEY ASPECTS OF A LEXICON

- **Linguistics**:
  - Entire set of words and phrases in a language or a person's mental vocabulary.
  - Includes part of speech, pronunciation, and definitions.
- **Computational Lexicon**:
  - Structured resource for NLP containing words and their linguistic properties.
  - May include synonyms, antonyms, semantic relationships, and grammar rules.

LEXICON VS. DICTIONARY

- Both list words and meanings, but a lexicon is broader in scope.
- Lexicons focus on linguistic and semantic relationships.
- Dictionaries often provide formal definitions and examples.

- **General Purpose**:
  - WordNet: Organizes words into synsets and provides relationships like hypernyms.
- **Domain-Specific**:
  - Medical or legal lexicons tailored to specialized fields.
- **Multilingual Lexicons**:
  - BabelNet: Bridges words across languages.

APPLICATIONS OF A LEXICON

- **Natural Language Processing**:
  - Tokenization and word segmentation.
  - Named Entity Recognition (NER).
  - Sentiment analysis using sentiment lexicons.
- **Language Learning**:
  - Provides learners with vocabulary and usage examples.
- **Search Engines and Information Retrieval**:
  - Expands user queries with synonyms and related terms.

CONCLUSION

- A lexicon is essential for understanding and processing language.
- It serves as the backbone of both human cognition and computational systems.
- Plays a critical role in NLP, linguistics, and language learning.

WHAT IS LEMMATIZATION?

- Lemmatization is the process of reducing words to their base or dictionary form, known as the **lemma**.
- It considers the context and performs morphological analysis to ensure the resulting word is meaningful.

EXAMPLES OF LEMMATIZATION

Word	**Lemma**
running	run
studies	study
children	child
better	good

HOW LEMMATIZATION WORKS

1. **Morphological Analysis**: Analyzes the structure of the word.
2. **POS Tagging**: Identifies the word's part of speech to determine the correct lemma.
3. **Dictionary Lookup**: Ensures the lemma is an actual word.

- **Python Libraries**:
  - **NLTK**: Provides the `WordNetLemmatizer`.
  - **spaCy**: Efficient lemmatization in its pipeline.
  - **Stanford CoreNLP**: Robust lemmatization tool.

APPLICATIONS OF LEMMATIZATION

- Preprocessing text for machine learning and NLP tasks.
- Enhancing search engine accuracy by normalizing queries.
- Useful in sentiment analysis, topic modeling, and text classification.

Lemmatize all sentences

Let L represent the sentiment lexicon pairs, the word and its polarity

$Positives = 0,\ Negatives = 0$

Test Sentence $S = \{w_1, w_2, w_3 ... w_n\}$

For w in S

If w found in L and $w == positive$, then

$Positives = Positives$+1

Else

$Negatives = Negatives + 1$

If $Positives > Negatives$ then

Return $positive$

Else

Return $negative$

NAIVE BAYES ALGORITHM

- A generative model for binary classification
- Probabilistic principles for predictions are used

  Key principles
  - Class-conditional independence
  - Bayes' theorem
- Simple and effective for many classification tasks.

- Naive Bayes assumes the following generative process:
  1. A biased coin determines the class label $y \in \{0, 1\}$ with $P(y = 1) = p$.
  2. d independent coins generate binary features $\mathrm{x} = [x_1, x_2, \ldots, x_d]$:

$$P(x_i = 1|y) = P_y(i), \quad P(x_i = 0|y) = 1 - P_y(i).$$

Class-conditional independence

- Given the class label y, all features $x_1, x_2, \ldots, x_d$ are independent.
- Reduces the number of parameters from 2^d to $2d$.

PARAMETERIZATION

- The model parameters are:
  - p: Probability of $y = 1$.
  - $P_1(i)$: Probability of $x_i = 1$ given $y = 1$.
  - $P_0(i)$: Probability of $x_i = 1$ given $y = 0$.
- Total parameters: $2d + 1$ (one for p, d for each class).

- Using Maximum Likelihood Estimation (MLE):
  - **MLE for p:**

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

  - **MLE for $P_y(i)$:**

$$\hat{P}_y(i) = \frac{\sum_{j=1}^{n} \mathbb{I}[x_{j,i} = 1 \wedge y_j = y]}{\sum_{j=1}^{n} \mathbb{I}[y_j = y]}$$

- $\mathbb{I}[\cdot]$: Indicator function (1 if condition true, 0 otherwise).

- Goal: Predict class label $\hat{y}_{\text{test}}$ for test point $\mathbf{x}_{\text{test}}$.
- **Bayes' theorem:**

$$P(y|\mathbf{x}_{\text{test}}) = \frac{P(\mathbf{x}_{\text{test}}|y)P(y)}{P(\mathbf{x}_{\text{test}})}$$

- Decision rule simplifies to:

$$\hat{y}_{\text{test}} = \begin{cases} 1, & \text{if } P(\mathbf{x}_{\text{test}}|y=1)P(y=1) > P(\mathbf{x}_{\text{test}}|y=0)P(y=0), \\ 0, & \text{otherwise.} \end{cases}$$

- Use class-conditional independence assumption:

$$P(\mathrm{x}_{\text{test}}|y) = \prod_{i=1}^{d} P_y(i)^{x_i}(1 - P_y(i))^{1-x_i}$$

- Final prediction rule:

$$\hat{y}_{\text{test}} = \begin{cases} 1, & \text{if } \prod_{i=1}^{d} \hat{P}_1(i)^{x_i}(1 - \hat{P}_1(i))^{1-x_i} \cdot \hat{p} > \prod_{i=1}^{d} \hat{P}_0(i)^{x_i}(1 - \hat{P}_0(i))^{1-x_i} \cdot (1 - \hat{p}), \\ 0, & \text{otherwise.} \end{cases}$$

Advantages:

- Simple and computationally efficient.
- Scalable to high-dimensional data.
- Performs well despite strong independence assumptions.

Limitations:

- Independence assumption often unrealistic.
- Can struggle with imbalanced datasets.

- Naive Bayes is a fundamental algorithm combining generative modeling with efficient parameter estimation.
- Its simplicity and effectiveness make it widely used in classification tasks.
- While independence assumptions may not hold, it often performs well in real-world scenarios.

Questions?

Naive Bayes Algorithm

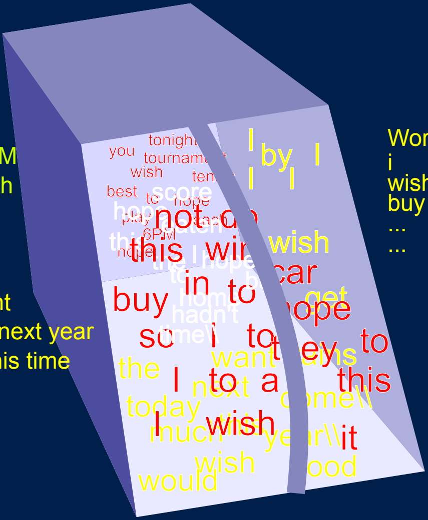
FIND THE SENDER OF THE EMAIL

Assume that Ram and Raj exchanged the following emails

Ram	Raj
I wish you the best	I hope to play tennis tonight
I hope to reach home by 6PM	I hope to win this tournament
I wish to go home early	I hope to buy this car in the next year
I do not want to buy this	I wish to get a good score this time
I hope it rains today	I wish they would come

Who would have sent this email "I wish you would come"

I wish you the best
I hope to reach home by 6PM
I wish I hadn't eaten so much
I do not want to buy this
I hope it rains today
I hope to play tennis tonight
I hope to win this tournament
I hope to buy this car in the next year
I wish to get a good score this time
I wish they would come



Word Frequency
i
wish
buy
...
...

Who would have sent this email "I wish you would come" This question can be answered by using Bayes theorem

Text Analytics

Let us consider two random variables $X\, and\, Y$. Then Joint probability, $P(X = x, Y = y)$, refers to the probability that the variable X takes the value x and the variable Y takes the value y. The conditional probability $P(Y = y|X = x)$ refers to the probability that the variable Y takes the value y given the observation the variable X takes the value x

$$P(X, Y) = P(Y|X) \times P(X) = P(X|Y) \times P(Y) \tag{1}$$

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \tag{2}$$

- Map Bayes theorem using the statistical properties of the data
- Let **X** $=$ and Y represent the random variables, where **X** is a set of attributes or is a attribute variable and Y represent a class.
- The relationship between **X** and Y can be found using the conditional probability $P(Y|\mathbf{X})$
- The conditional probability $P(Y|\mathbf{X})$ is known as posterior probability of Y
- $P(Y)$ is known as the prior probability
- In the classification problem, it is important to learn the parameters $P(Y|\mathbf{X})$. Given the attributes of the email (TF,TF-IDF), find the class to which the email belongs - in this case the person who sent it.

The parameters are obtained from the training data - the corpus of emails written by Ram and Raj. During the training process, we will learn $P(Y|\mathbf{X})$ for every word in the corpus

SUPERVISED CLASSIFICATION

- Set of input parameters/attributes $\mathbf{X} = X_1, X_2, \ldots, X_m$and a fixed set of classes $Y = y_1, y_2, \ldots, y_n$
- Every element of the training set, $D = d_1, d_2, \ldots, d_n$ is manually assigned a class

  $(d_1, y_1), (d_2, y_2), (d_3, y_1), \ldots$
- Goal is to learn the classifier, so that it can map a new document $\hat{d}$ to any of the classes, $y \in Y$
- Bayes classifier would assign a probability based on the observation to the new document to aid the class selection
- The probability score for each class is computed as given by the equation

  $$P(Y|\mathbf{X}) = \frac{P(\mathbf{X}|Y)P(Y)}{P(\mathbf{X})}$$
- The class will be found using $\arg\max_{y \in Y} P(Y|\mathbf{X})$

ESTIMATING THE CONDITIONAL PROBABILITY $P(X_i|Y)$

$$
\begin{aligned}
\hat{y} &= \arg\max_{y \in Y} P(Y|\mathbf{X}) && (3)\\
&= \arg\max_{y \in Y} P(\mathbf{X}|y)P(y) && (4)\\
&= \arg\max_{y \in Y} P(y)P(X_1, X_2, X_m|Y) && (5)\\
&= \arg\max_{y \in Y} P(y)P(X_1|y) \times P(X_2|y) \times \ldots P(X_m|y) && (6)\\
&= \arg\max_{y \in Y} P(y) \prod_{i=1}^{m} P(X_i|y) && (7)
\end{aligned}
$$

1. Prior probability - $P(y) = \dfrac{Count(y)}{Count(Y)}$
2. Learn $P(X_1|y) = \dfrac{Count(X_1, y)}{Count(Y)}$

Word	Frequency

Text Analytics

Assume that Ram and Raj share emails exchanged emails using the words given in the table. A new email arrives with just three words - ***motivate, profit and product***. Find the sender using the historical information given in the table

Historical Information

Ram	Raj
motivate(0.24)	motivate(0.05)
profit(0.3)	profit(0.35)
product(0.26)	product(0.35)
leadership(0.08)	leadership(0.15)
operations(0.12)	operations(0.10)

Who would have used these words (motivate, profit and product) in the email ?
Is it possible to apply this technique to identify the sentiments of a movie review with two classes **Good** and **bad**?

Assume the following likelihoods for each word being part of a positive or negative review, and equal prior probabilities for each class - positive and negative ($P(positive) = 0.5$ and $P(negative) = 0.5$)

word	positive	negative
I	0.09	0.16
love	0.07	0.06
to	0.05	0.07
fill	0.29	0.06
credit	0.04	0.15
card	0.08	0.11
application	0.06	0.04

What class Naive Bayes classifier would assign to the sentence "I do not like to fill in the application form?"

[1] Richard Socher et al. "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank". In: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Ed. by David Yarowsky et al. Seattle, Washington, USA: Association for Computational Linguistics, Oct. 2013, pp. 1631–1642. URL: https://aclanthology.org/D13-1170/.

[2] Nhan Cach Dang, María N. Moreno-García, and Fernando De la Prieta. "Sentiment Analysis Based on Deep Learning: A Comparative Study". In: *Electronics* 9.3 (Mar. 2020), p. 483. ISSN: 2079-9292. DOI: 10.3390/electronics9030483. URL: http://dx.doi.org/10.3390/electronics9030483.