

What is Sentiment Analysis?

- Sentiment analysis is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes.
- Also known as Opinion Mining, Opinion Extraction, Sentiment Mining, Subjectivity Analysis, Emotion Analysis.

Sentiment Analysis~NLP

- ❖ Sentiment analysis is a **highly restricted NLP problem** because the system **does not need to fully understand** the semantics of each document but only needs to understand some aspects of it, i.e., positive or negative sentiments and their target entities or topics.

Sentiment Analysis Applications

- ❖ No longer be necessary to **conduct surveys, opinion polls, and focus groups** in order to gather public opinions.
- ❖ **Customer feedback** collected from emails and call centers or results from surveys conducted by the organizations.

Different Levels of Analysis

- ❖ Document level
- ❖ Sentence level
- ❖ Entity and Aspect level

Document Level

- ❖ The task at this level is to classify whether a **whole opinion document** expresses a positive or negative sentiment.
- ❖ *For example*, given a product review, the system determines whether the review expresses an overall positive or negative opinion about the product.

Sentence Level

- ❖ Determines whether **each sentence** expressed a positive, negative, or neutral opinion (Neutral usually means no opinion).
- ❖ Example, “*I love OnePlus 7T.*” - positive sentiment and “*I didn’t like the packaging of device.*” - negative sentiment.

Entity and Aspect Level

- ❖ It is based on the idea that an opinion consists of a **sentiment (positive or negative)** and a **target (of opinion)**.
- ❖ For Example, "*The iPhone's call quality is good, but its battery life is short.*"
 - **Aspect:** call quality and battery life, **Entity:** iPhone.
 - The sentiment on iPhone's call quality is positive, but the sentiment on its battery life is negative.
 - The call quality and battery life of iPhone are the opinion targets.

Sentiment Lexicon

- ❖ The most important indicators of sentiments are **sentiment words**,
 - *Example*, good, wonderful, and amazing are positive sentiment words and bad, poor, and terrible are negative sentiment words.
- ❖ Apart from individual words, there are also **phrases and idioms**,
 - *Example*, cost someone an arm and a leg.
- ❖ A list of such words and phrases is called a **sentiment lexicon**.
- ❖ Sentiment lexicon is **necessary but not sufficient** for sentiment analysis.

Sentiment Lexicon- Issue 1

- ❖ A +ive or -ive sentiment word may have **opposite orientations** in different application domains.
- ❖ Example, “**suck**” usually indicates negative sentiment, but

“This camera sucks.”

“This vacuum cleaner really sucks.”

Sentiment Lexicon- Issue 2

- ❖ A sentence containing sentiment words may not express any sentiment.
- ❖ Question (interrogative) sentences and conditional sentences are two important types,
- ❖ Example,

“Can you tell me which Sony camera is good ?”

“If I can find a good camera in the shop, I will buy it.”

Sentiment Lexicon- Issue 3

- ❖ **Sarcastic sentences** with or without sentiment words are hard to deal with.
- ❖ Sarcasms are **not so common** in consumer reviews about products and services, but are very common in political discussions, which make political opinions hard to deal with.
- ❖ **Example,**

"What a great car! It stopped working in two days."

Sentiment Lexicon- Issue 4

- ❖ Many sentences without sentiment words can also imply opinions.
- ❖ Many of these sentences are actually objective sentences that are used to express some factual information.
- ❖ Example,

"This washer uses a lot of water."

expresses a negative sentiment about the washer since it uses a lot of resource (water)

"After sleeping on the mattress for two days, a valley has formed in the middle."

expresses a negative opinion about the mattress.

Problem Definition

- ❖ We use the following review about a Canon camera to introduce the problem.

Posted by: John Smith

Date: September 10, 2011

"(1) I bought a Canon G12 camera six months ago. (2) I simply love it. (3) The picture quality is amazing. (4) The battery life is also long. (5) However, my wife thinks it is too heavy for her."

- ❖ From this review, we notice a few important points.

Analysis of Review..

The review has a number of opinions, both positive and negative, about Canon G12 camera.

- ❖ Sentence (2) - a positive opinion about the Canon camera as a whole.
- ❖ Sentence (3) - a positive opinion about its picture equality.
- ❖ Sentence (4) - a positive opinion about its battery life.
- ❖ Sentence (5) - a negative opinion about the weight of the camera.

Analysis of Review(Definition of Opinion as Tuple)..

Definition: An opinion consists of two key components:

$$(g, s)$$

where

g - target (Any entity about which an opinion has been expressed)

s - sentiment, (strength/intensity or type of the sentiment.)

Example, the target of the opinion in sentence (2) - *Canon G12*.

Analysis of Review(Issues with this Definition)..

- The holder of the opinions in sentences (2), (3), and (4) is the **author** of the review (“ John Smith”), but for sentence (5), it is the wife of the author.
- The **date of the review** is September 10, 2011. This date is important in practice because one often wants to know how opinions change with time and opinion trends.

Analysis of Review(Opinion as a Quadruple)..

Definition: An opinion is a quadruple,

$$(g, s, h, t)$$

where,

g - opinion (or sentiment) target,

s - sentiment about the target,

h - opinion holder

t - time when the opinion was expressed.

Analysis of Review(Improvement)..

- ❖ **Observation:** In sentence (3), the **opinion target** is actually “picture quality of Canon G12,” but the sentence mentioned only “picture quality.”
- ❖ **Solution:** The target can often be decomposed and described in a structured manner with multiple levels, which greatly facilitate both mining of opinions and later use of the mined opinion results.
- ❖ For example, “picture quality of Canon G12” ~ (Canon-G12, picture-quality).

Analysis of Review(Entity)..

Definition: An entity e is a product, service, topic, issue, person, organization, or event. It is described with a pair,

$$e: (T, W)$$

where,

T - a hierarchy of parts, sub-parts, and so on,

W - a set of attributes of e .

Example, For a Canon G12. It has a set of attributes, e.g., picture quality, size, and weight, and a set of parts, e.g., lens, viewfinder, and battery.

Analysis of Review(Improvement)..

- ❖ **Observation:** Sentence (2) expresses a positive opinion about the **entity** Canon G12 camera **as a whole**.
Sentence (3) expresses a positive opinion on the attribute of picture quality of the camera. Clearly, one can also express opinions about parts or components of the camera.
- ❖ **Solution:** We simplify the hierarchy to two levels and use the term aspects to denote both parts and attributes. The root node is still the entity itself, but the second level (also the leaf level) nodes are different aspects of the entity.

Analysis of Review(Quintuple Definition of Opinion)..

Definition: An opinion is a quintuple,

$$(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$$

where

e_i is the name of an entity,

a_{ij} is an aspect of e_i ,

s_{ijkl} is the sentiment on aspect a_{ij} of entity e_i

h_k is the opinion holder,

t_l is the time when the opinion is expressed by h_k .

Quintuple Definition of Opinion(Important Points)

- ❖ We purposely use subscripts. The opinion s_{ijkl} must be given by opinion holder h_k about aspect a_{ij} of entity e_i at time t_l .
- ❖ The opinion defined here is just one type of opinion, called regular opinion. Another type is comparative opinion, which needs a different definition.
- ❖ This definition provides a framework to transform unstructured text to structured data.
- ❖ The definition covers most but not all possible facets of the semantic meaning of an opinion, which can be arbitrarily complex. For example, in situation "*The viewfinder and the lens are too close.*" which expresses an opinion but does not cover the context of the opinion.

Objective of Sentiment Analysis

Given an opinion document d , discover all opinion quintuples $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$ in d .

Problem in Extraction of Quintuple

- For example, Motorola may be written as Mot, Moto, and Motorola. We need to recognize that they all refer to the same entity.

Definition: An **entity category** represents a unique entity, while an **entity expression** is an actual word or phrase that appears in the text indicating an entity category.

Problem in Extraction of Quintuple

- For example, picture, image, and photo are the same aspect for cameras. We thus need to extract aspect expressions and categorize them.

Definition: An **aspect category** of an entity represents a unique aspect of the entity, while an **aspect expression** is an actual word or phrase that appears in the text indicating an aspect category.

Aspect Expressions

Aspect expressions are usually nouns and noun phrases but can also be verbs, verb phrases, adjectives, and adverbs.

- ❖ **Explicit Aspect Expression:** Aspect expressions that are nouns and noun phrases
 - Example, "*picture quality*" in "*The picture quality of this camera is great*"
- ❖ **Implicit Aspect Expression:** Aspect expressions that are not nouns or noun phrases
 - Example, "*expensive*" is an implicit aspect expression in "*This camera is expensive.*"

Summarization

Task 1 (entity extraction and categorization): Extract all entity expressions in D, and categorize or group synonymous entity expressions into entity clusters (or categories). Each entity expression cluster indicates a unique entity e_i .

Task 2 (aspect extraction and categorization): Extract all aspect expressions of the entities, and categorize these aspect expressions into clusters. Each aspect expression cluster of entity e_i represents a unique aspect a_{ij} .

Task 3 (opinion holder extraction and categorization): Extract opinion holders for opinions from text or structured data and categorize them. The task is analogous to the above two tasks.

Summerization(Cont.)..

Task 4 (time extraction and standardization): Extract the times when opinions are given and standardize different time formats. The task is also analogous to the above tasks.

Task 5 (aspect sentiment classification): Determine whether an opinion on an aspect a_{ij} is positive, negative or neutral, or assign a numeric sentiment rating to the aspect.

Task 6 (opinion quintuple generation): Produce all opinion quintuples $(e_j, a_{ij}, s_{ijkl}, h_k, t_l)$ expressed in document d based on the results of the above tasks.

Example..

Example 4: Posted by: big John

Date: Sept. 15, 2011

(1) I bought a Samsung camera and my friends brought a Canon camera yesterday. (2) In the past week, we both used the cameras a lot. (3) The photos from my Samy are not that great, and the battery life is short too. (4) My friend was very happy with his camera and loves its picture quality. (5) I want a camera that can take good photos. (6) I am going to return it tomorrow.

Solution(1)..

1. **Task 1** should extract the entity expressions, “Samsung,” “Samy,” and “Canon,” and group “Samsung” and “Samy” together as they represent the same entity.
2. **Task 2** should extract aspect expressions “picture,” “photo,” and “battery life,” and group “picture” and “photo” together as for cameras they are synonyms.
3. **Task 3** should find the holder of the opinions in sentence (3) to be big John (the blog author) and the holder of the opinions in sentence (4) to be big

Solution(2)..

- Task 4 should also find the time when the blog was posted is Sept-15-2011.
- Task 5 should find that sentence (3) gives a negative opinion to the picture quality of the Samsung camera and also a negative opinion to its battery life. Sentence (4) gives a positive opinion to the Canon camera as a whole and also to its picture quality. Sentence (5) seemingly expresses a positive opinion, but it does not.

Solution(3)..

(Samsung, picture_quality, negative, big John, Sept-15-2011)

(Samsung, battery_life, negative, big John, Sept-15-2011)

(Canon, GENERAL, positive, big John's_friend, Sept-15-2011)

(Canon, picture_quality, positive, big John's_friend, Sept-15-2011)

Types of Opinions - Regular & Comparative Opinions

- ❖ **Regular Opinions:** A sentiment only on an particular entity or an aspect of the entity,
 - For Example, "*Coke tastes very good.*" which expresses a positive sentiment on the aspect taste of Coke.
- ❖ **Comparative Opinions:** A comparative opinion compares multiple entities based on some of their shared aspects,
 - For Example, "*Coke tastes better than Pepsi.*" which compares Coke and Pepsi based on their tastes (an aspect) and expresses a preference for Coke.

Type of Regular Opinions- Regular & Indirect Opinion

- ❖ **Direct opinion:** A direct opinion refers to an opinion expressed directly on an entity or an entity aspect. Example, "*The picture quality is great.*"
- ❖ **Indirect opinion:** An indirect opinion is an opinion that is expressed indirectly on an entity or aspect of an entity based on its effects on some other entities.
 - Example, the sentence "*After injection of the drug, my joints felt worse*" describes an undesirable effect of the drug on "*my joints*", which indirectly gives a negative opinion or sentiment to the drug.

Type of Opinions- Explicit and Implicit Opinions

- ❖ **Explicit opinion:** An explicit opinion is a subjective statement that gives a regular or comparative opinion.
 - Example, "*Coke tastes great,*" and "*Coke tastes better than Pepsi.*"
- ❖ **Implicit (or implied) opinion:** An implicit opinion is an objective statement that implies a regular or comparative opinion.
 - Example, "*I bought the mattress a week ago, and a valley has formed,*" and "*The battery life of Nokia phones is longer than Samsung phones.*"

Explicit opinions are easier to detect and to classify than implicit opinions.

Subjectivity Classification

- ❖ **Definition:** The task of determining whether a sentence is subjective or objective.
- ❖ An *objective sentence* presents some factual information about the world, while a *subjective sentence* expresses some personal feelings, views, or beliefs.
 - Example, objective sentence is "*iPhone is an Apple product.*" and subjective sentence is "*I like iPhone.*"

Subjectivity and Opinionated

- ❖ Subjectivity totally unrelated to being opinionated. By opinionated, we mean that a document or sentence expresses or implies a positive or negative sentiment.
- ❖ A subjective sentence may not express any sentiment.
 - Example, "*I think that he went home*" is a subjective sentence, but does not express any sentiment.
- ❖ Objective sentences can imply opinions or sentiments due to desirable and undesirable facts.
 - Example, "The earphone broke in two days." and "I brought the mattress a week ago and a valley has form."

Emotions

- ❖ **Definition:** Emotions are our subjective feelings and thoughts.
- ❖ The strength of a sentiment or opinion is typically linked to the intensity of certain emotions, e.g., joy and anger.
- ❖ Opinions that we study in sentiment analysis are mostly evaluations.
- ❖ According to consumer behavior research, evaluations can be broadly categorized into two types: *rational evaluations* and *emotional evaluations*.

Rational Evaluation & Emotional Evaluation

- ❖ **Definition (Rational Evaluation):** Such evaluations are from rational reasoning, tangible beliefs, and utilitarian attitudes.
 - Example, "The voice of this phone is clear," "This car is worth the price,"
- ❖ **Definition(Emotional Evaluation):** Such evaluations are from non-tangible and emotional responses to entities which go deep into people's state of mind.
 - Example, "I love iPhone," "I am so angry with their service people"

Document Sentiment Classification

Given an opinion document d evaluating an entity, determine the overall sentiment s of the opinion holder about the entity, i.e., determine s expressed on aspect GENERAL in the quintuple

$(_, \text{GENERAL}, s, _, _)$,

where the entity e , opinion holder h , and time of opinion t are assumed known or irrelevant (do not care).

Assumption: Sentiment classification or regression assumes that the opinion document d (e.g., a product review) expresses opinions on a single entity e and contains opinions from a single opinion holder h .

Sentiment Classification Using Supervised Learning

- ❖ Like other supervised machine learning applications, the key for sentiment classification is the engineering of a set of effective features.
- ❖ Training and testing data used are normally product reviews.
- ❖ Some of the example features are the following:
 - Terms and their frequency
 - Part of speech.
 - Sentiment words and phrases
 - Sentiment shifters

Terms and their Frequency

- ❖ These features are individual words (unigram) and their n-grams with associated frequency counts.
- ❖ In some cases, word positions may also be considered.
- ❖ The **TF-IDF Weighting Scheme** (*tf-idf is a weighting scheme that assigns each term in a document a weight based on its term frequency (tf) and inverse document frequency (idf). The terms with higher weight scores are considered to be more important.*) from information retrieval may be applied too.

Part of Speech

- ❖ Words of different parts of speech (POS) may be treated differently.
- ❖ For example, it was shown that adjectives are important indicators of opinions.
- ❖ However, one can also use all POS tags and their n-grams as features.

TAG	DESCRIPTION	TAG	DESCRIPTION
CC	Coordinating conjunction	PRP\$	Possessive pronoun
CD	Cardinal number	RB	Adverb
DT	Determiner	RBR	Adverb, comparative
EX	Existential <i>there</i>	RBS	Adverb, superlative
FW	Foreign word	RP	Particle
IN	Preposition or subordinating conjunction	SYM	Symbol
JJ	Adjective	TO	<i>to</i>
JJR	Adjective, comparative	UH	Interjection
JJS	Adjective, superlative	VB	Verb, base form
LS	List item marker	VBD	Verb, past tense
MD	Modal	VBG	Verb, gerund or present participle
NN	Noun, singular or mass	VBN	Verb, past participle
NNS	Noun, plural	VBP	Verb, non-3rd person singular present
NNP	Proper noun, singular	VBZ	Verb, 3rd person singular present
NNPS	Proper noun, plural	WDT	Wh-determiner
PDT	Predeterminer	WP	Wh-pronoun
PRP\$	—	WDT	—

Sentiment words and Phrases

- ❖ Sentiment words are words in a language that are used to express positive or negative sentiments.
- ❖ For example, good, wonderful, and amazing are positive sentiment words, and bad, poor, and terrible are negative sentiment words.
- ❖ Apart from individual words, there are also sentiment phrases and idioms, e.g., cost someone an arm and a leg.

Sentiment Shifter

- ❖ These are expressions that are used to change the sentiment orientations, e.g., from positive to negative or vice versa.
- ❖ Negation words are the most important class of sentiment shifters. Example, the sentence “I don’t like this camera ” is negative.
- ❖ Suchshifters also need to be handled with care because not all occurrences of such words mean sentiment changes. For example, “not” in “not only . . . but also” does not change sentiment orientation.

Sentiment Analysis Using Unsupervised Learning

- ❖ Classification based on some fixed syntactic patterns that are likely to be used to express opinions.
- ❖ The syntactic patterns are composed based on part-of-speech (POS) tags. The algorithm given in Turney (2002) consists of three steps.

Step -1

- ❖ Two consecutive words are extracted if their POS tags conform to any of the patterns. See Table. The nouns or verbs act as the contexts because in different contexts a JJ, RB, RBR and RBS word may express different sentiments.
- ❖ As an example, in the sentence “This piano produces beautiful sounds ”, “beautiful sounds ” is extracted as it satisfies the first pattern.

Note: The adjective (JJ) “unpredictable” may have a negative sentiment in a car review as in “unpredictable steering,” but it could have a positive sentiment in a movie review as in “unpredictable plot.”

TABLE 3.2: Patterns of POS tags for extracting two-word phrases

	FIRST WORD	SECOND WORD	THIRD WORD (NOT EXTRACTED)
1	JJ	NN or NNS	anything
2	RB, RBR, or RBS	JJ	not NN nor NNS
3	JJ	JJ	not NN nor NNS
4	NN or NNS	JJ	not NN nor NNS
5	RB, RBR, or RBS	VB, VBD, VBN, or VBG	Anything

Step 2

- ❖ It estimates the sentiment orientation (SO) of the extracted phrases using the pointwise mutual information (PMI) measure:

$$PMI(term_1, term_2) = \log_2 \left(\frac{\Pr(term_1 \wedge term_2)}{\Pr(term_1)\Pr(term_2)} \right).$$

$\Pr(term_1 \wedge term_2)$ is the actual co-occurrence probability of term 1 and term 2 .

$\Pr(term_1)\Pr(term_2)$ is the co-occurrence probability of the two terms

- ❖ PMI measures the degree of statistical dependence between two terms.

Step 2(Cont..)

- ❖ The sentiment orientation (SO) of a phrase is computed based on its association with the positive reference word “excellent” and the negative reference word “poor ”:

$$SO(\text{phrase}) = PMI(\text{phrase}, \text{"excellent"}) - PMI(\text{phrase}, \text{"poor"})$$

Step 3

Given a review, the algorithm computes the average SO of all phrases in the review and classifies the review as positive if the average SO is positive and negative otherwise.