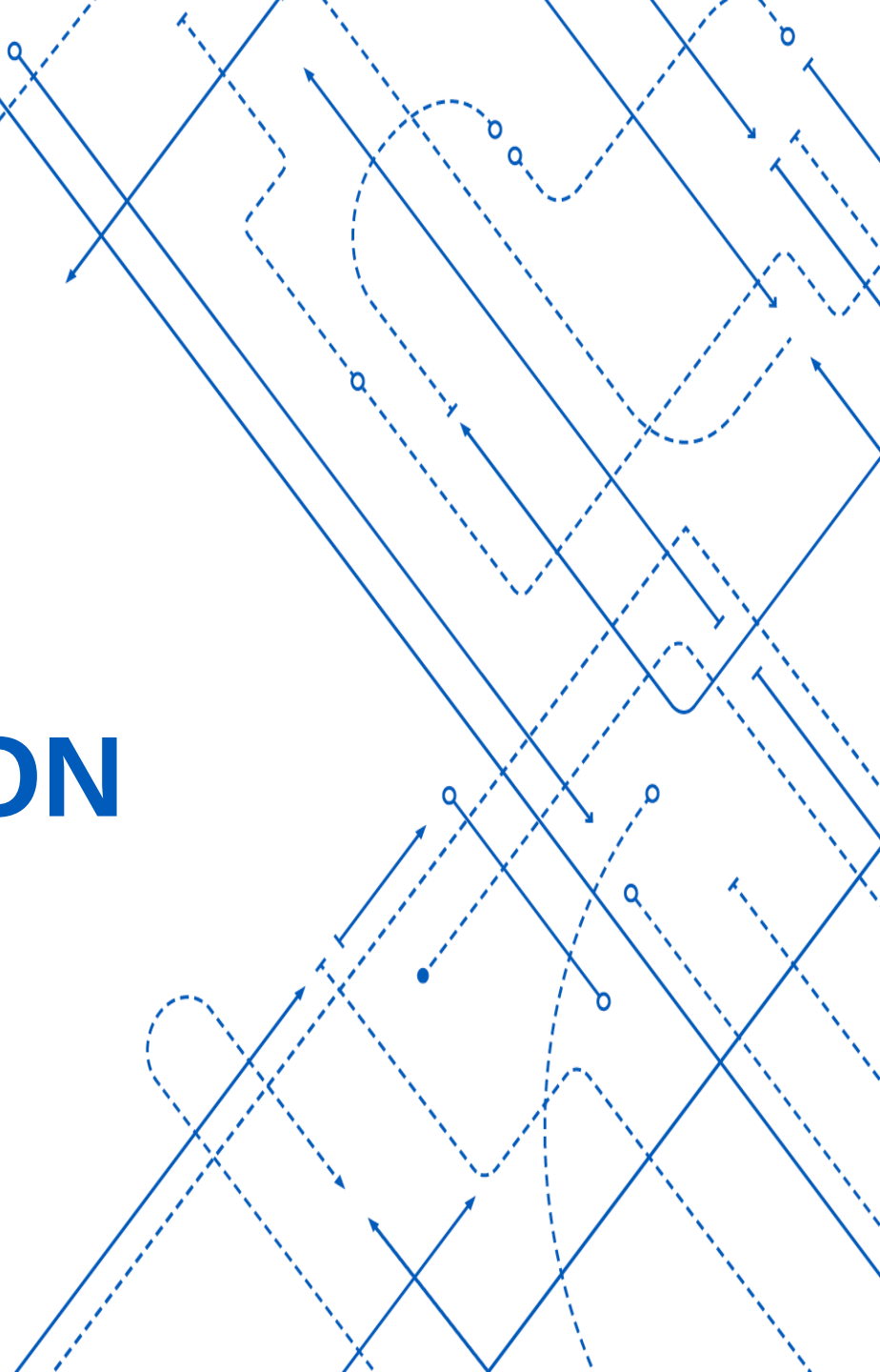
Sentiment analysis

SENTIMENT CLASSIFICATION

Nguyen Ngoc Thao
nnthao@fit.hcmus.edu.vn



Content outline

- Sentiment analysis: A categorization
- Document sentiment classification
- Sentence subjectivity and sentiment classification
- Aspect-based opinion mining

Sentiment analysis: A categorization

Tasks in sentiment analysis

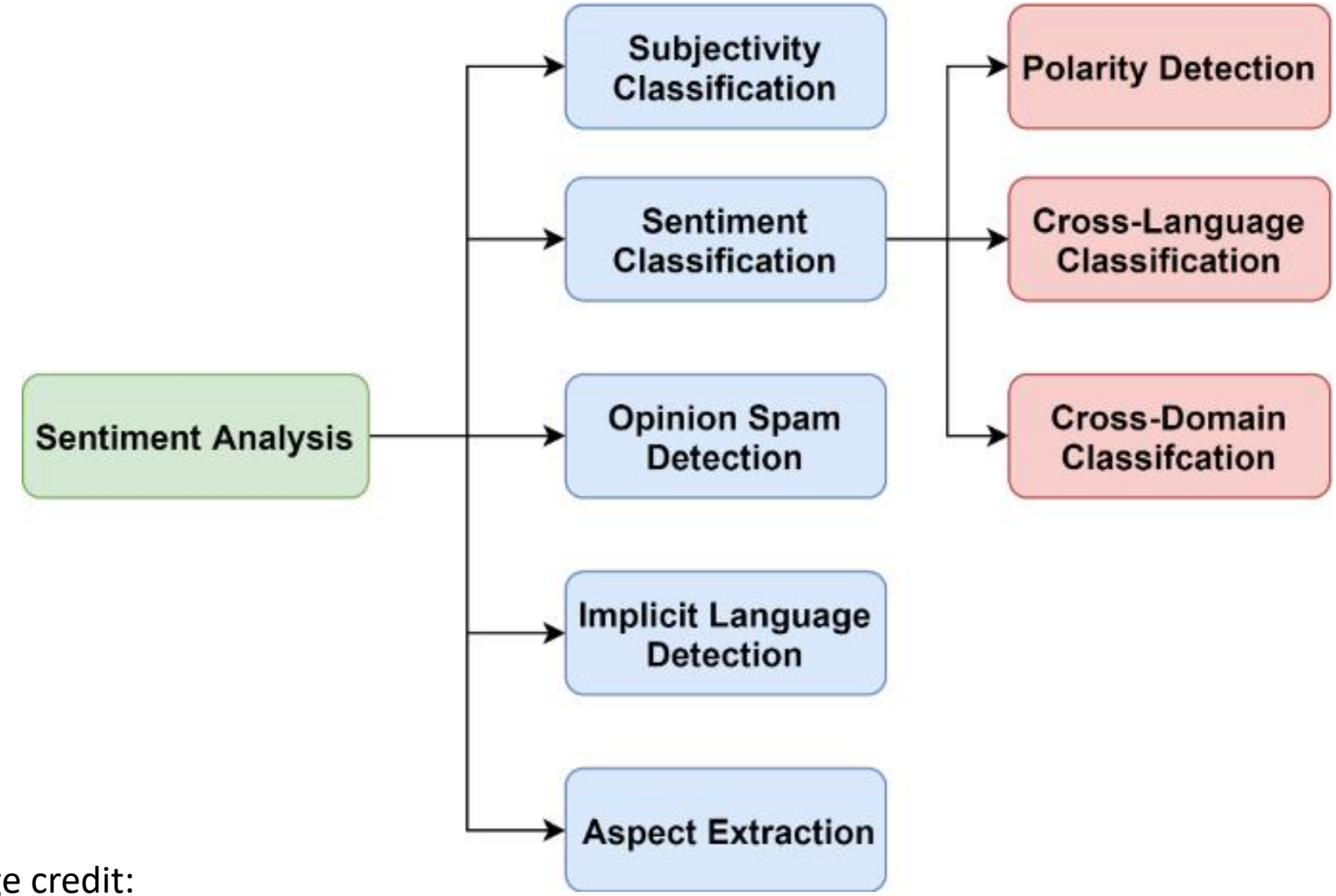


Image credit:
[Artificial Intelligence Review]

Sentiment classification levels

- Sentiment analysis has been investigated on several levels.

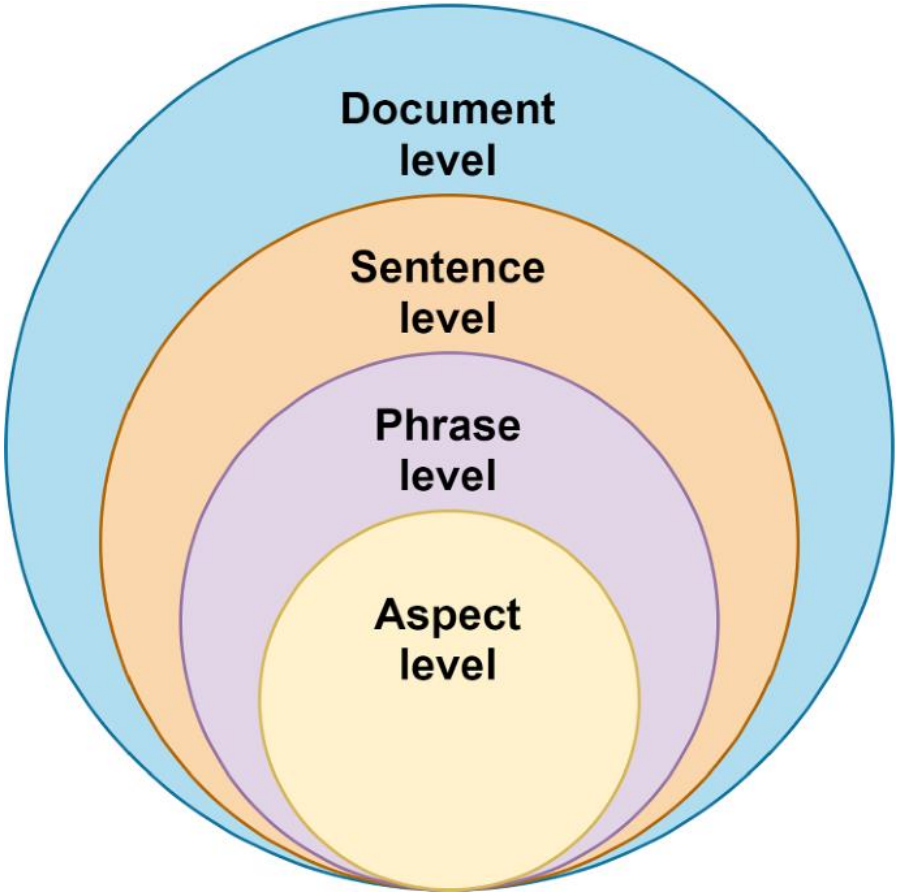


Image credit:
Artificial Intelligence Review

Sentiment classification approaches

- Researchers are continuously trying to figure out ways for better accuracy and lower computational cost.

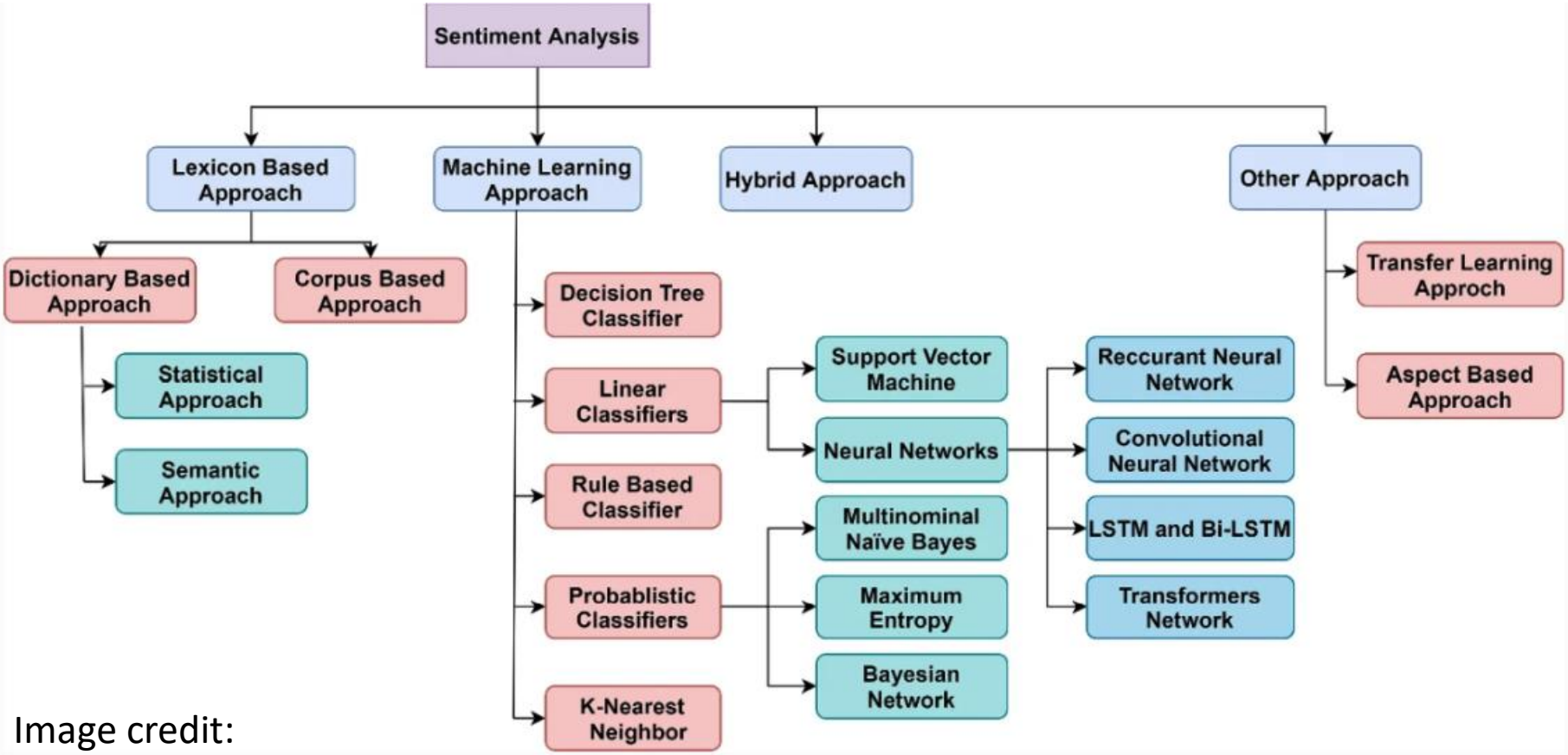
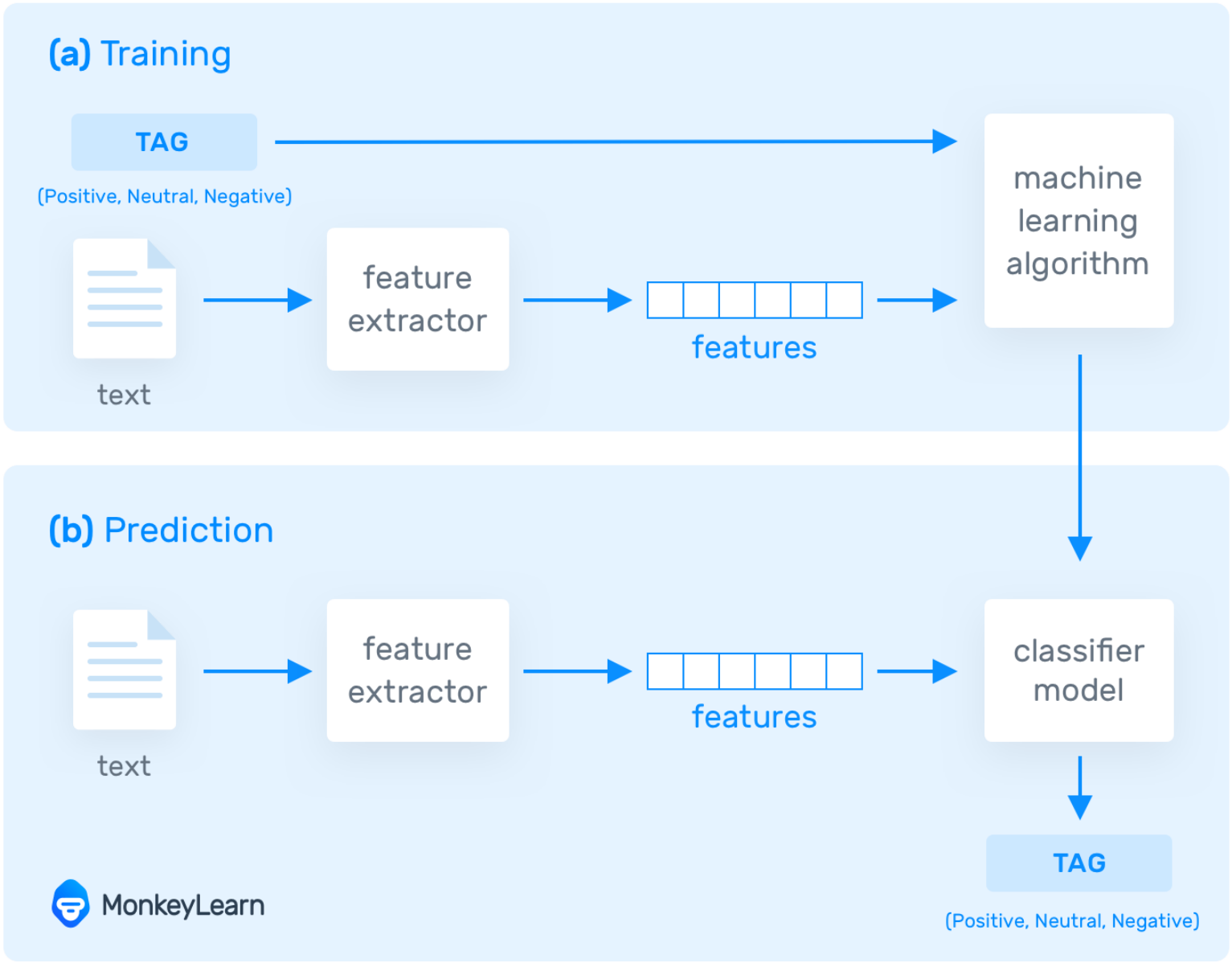


Image credit:
Artificial Intelligence Review

How Does Sentiment Analysis Work?



Applications of sentiment analysis

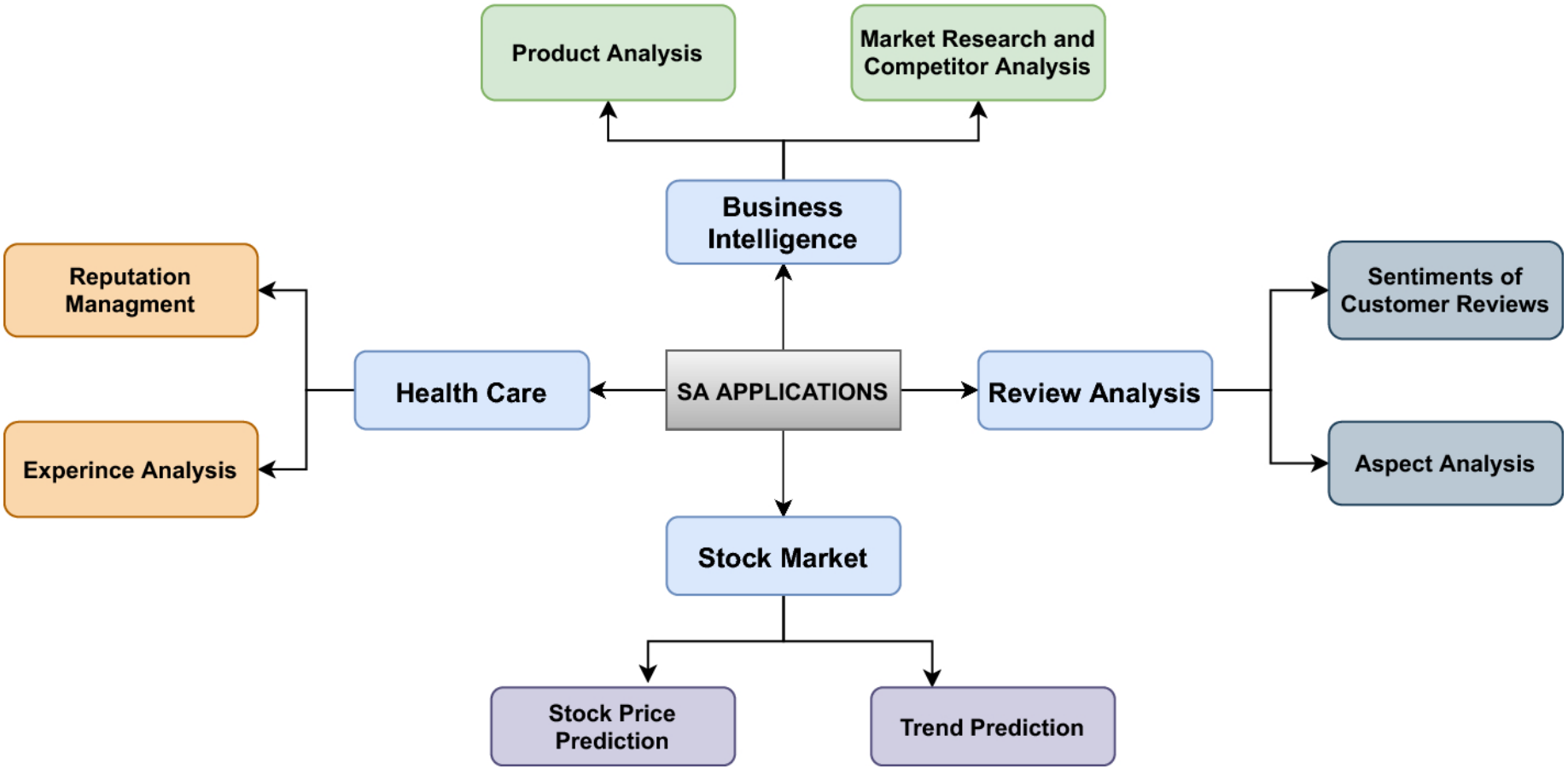


Image credit:
[Artificial Intelligence Review]

Document-level sentiment classification

Problem definition

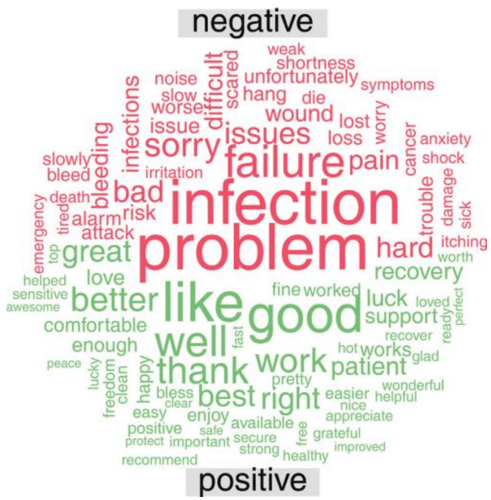
- The whole document is considered as the basic unit and a single polarity is given to this document.
- Problem definition: Given an opinionated document d about an entity e, determine the opinion orientation oo on e.
  - Determine oo on aspect GENERAL, while e, h, and t are assumed known or irrelevant.
- Assumption: The document d conveys opinions on a single entity e, which is provided by an opinion holder h.

Supervised learning approaches

- A supervised learning problem with three classes: positive, negative, and neutral.
- Training and testing data used are mostly product reviews.
  - E.g., 4-5 stars: positive, 1-2 stars: negative, and 3 stars: neutral.
- Any supervised learning methods can be applied.
  - E.g., naïve Bayesian classification, support vector machines, etc.

Feature engineering

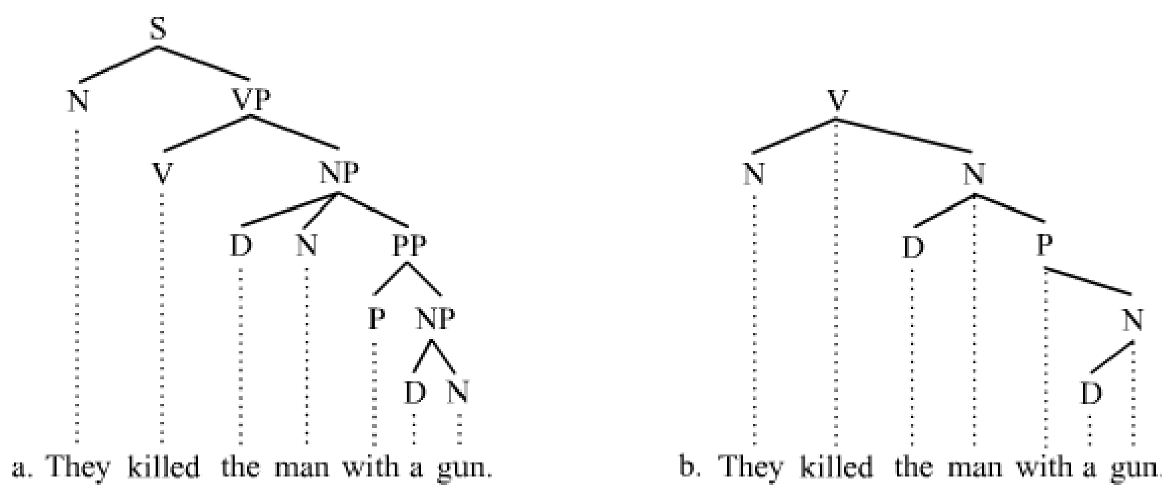
- Terms and their frequency
- Opinion words and phrases



- Part of speech
- Rules of opinions

1.	POSITIVE	::=	P
2.		|	PO
3.		|	orientation shifter N
4		|	orientation shifter NE
5.	NEGATIVE	::=	N
6.		|	NE
7.		|	orientation shifter P
8.		|	orientation shifter PO

- Syntactic dependency



Phrase structure grammar **Dependency grammar**

Unsupervised learning approaches

- Extract phrases with adjectives or adverbs using POS tagging
  - E.g., extract “beautiful pictures” from “This camera produces beautiful pictures”.
- The opinion orientation (SO) of a phrase is computed as follows

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

  - The pointwise mutual information (PMI) bet ween $term_1$ and $term_2$ is defined as

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

- For a review, compute the average $\overline{SO}$ of all phrases in the review and classify the review as recommended if $\overline{SO} > 0$.

Limitations

- Document-level sentiment classification gives no details on what people liked and/or disliked.
- It is not easily applicable to non-reviews, e.g., forum and blog postings, which evaluate multiple entities.
- Some documents may not evaluate products but still contain a few opinion sentences.

Sentence-level sentiment classification

Sentence subjectivity

- An objective sentence presents some factual information about the world.
- A subjective sentence expresses some personal feelings, views, or beliefs.

Posted by: bigXyz on Nov-4-2010: (1) I bought a Motorola phone and my girlfriend bought a Nokia phone yesterday. (2) We called each other when we got home. (3) The voice of my Moto phone was unclear, but the camera was good. (4) My girlfriend was quite happy with her phone, and its sound quality. (5) I want a phone with good voice quality. (6) So I probably will not keep it.

Objective sentences are marked in green color. Subjective sentences are in blue.

Sentence subjectivity

- A subjective sentence may not contain an opinion.
  - E.g., “I want a phone with good voice quality.” is subjective, but it does not express any opinion.
- An objective sentence can imply some opinion.
  - E.g., “The earphone broke in two days.” is objective while implying a negative opinion.

Sentence-level sentiment classification

- Each sentence's polarity will be determined independently.
- Techniques for document-level classification are applicable to individual sentences.
- Problem Definition: Given a sentence s, two sub-tasks are:
  1. Subjectivity classification. Decide whether the sentence s is subjective or objective.
  2. Sentence-level sentiment classification. If s is subjective, identify which opinion it conveys.
- Assumption: The sentence expresses a single opinion from a single opinion holder.

Limitations

- A compound and complex sentence may reveal more than one opinion.
  - E.g., "The picture quality of this camera is amazing and so is the battery life, but the viewfinder is too small for such a great camera,"
- Furthermore, the opinions can be embedded in phrases.
  - E.g., "Apple is doing very well in this terrible economy." $\rightarrow$ the opinion on "Apple" is clearly positive but on "economy" it is negative.
- Not all subjective sentences have opinions. Many objective sentences can also imply opinions $\rightarrow$ mine both types.

Aspect-level sentiment classification

Opinions are complicated

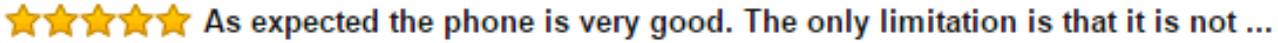
- A positive opinion about an entity does not mean that the author has positive opinions on all aspects of the entity.

Nothing wrong with it

By Jason on April 23, 2017

Size: 128 GB | Color: Black | Verified Purchase

It's a good phone but the price is a little high



As expected the phone is very good. The only limitation is that it is not ...

By Amazon Customer on October 30, 2016

Size: 128 GB | Color: Gold | Verified Purchase

Thanks. As expected the phone is very good. The only limitation is that it is not possible charging and listening to music at the same time.

- Document-level and sentence-level classification could not resolve the above issue.

Problem definition

- Assign a polarity to every aspect in the sentence, and then aggregate a sentiment for the whole sentence
- Problem definition: Given an opinionated document d, discover every quintuple $\langle e_i, a_{ij}, oo_{ijkl}, h_k, t_l \rangle$ in d.
  - A complete process of five tasks needs to be performed.
- We only focus on the two core tasks: Aspect extraction and Aspect sentiment classification.

Aspect sentiment classification

- **Objective:** Decide the opinion orientation expressed on each aspect in a sentence
- The sentence-level and clause-level methods can be useful.
  - Apply the method to each sentence/clause containing some aspects $\rightarrow$ these aspects will take the defined opinion orientation.
- These methods have difficulty dealing with mixed opinions in a sentence and opinions that need phrase level analysis.
  - E.g., “Apple is doing very well in this terrible economy.”
- Text in blogs and forum is informal with grammatical errors.

Lexicon-based approach

- The opinion lexicon is a list of opinion words and phrases.
- Lexicon-based approach uses an opinion lexicon to decide the opinion orientation in a sentence.
- It includes the following steps

1. Mark opinion words and phrases
2. Handle opinion shifters
3. Handle but-clauses
4. Aggregating opinions

Mark opinion words and phrases

- Given a sentence that contains one or more aspects.
- Mark all opinion words and phrases in the sentence.
  - E.g., “The *picture quality* of this camera is not **great**, but the *battery life* is **long**.”
- Each positive word is assigned the opinion score of +1 and each negative word with a score of -1.
  - E.g., “The *picture quality* of this camera is not **great**[+1], but the *battery life* is **long**[?]**.**”
  - “great” is a positive opinion word, but “long” is context-dependent.

Handle opinion shifters

- Opinion shifters are words/phrases that can shift opinion orientations.
- Negation words: not, never, none, nobody, nowhere, neither, and cannot.
  - E.g., “The *picture quality* of this camera is **not great**[-1], but the battery life is long”.
- Modal auxiliary verbs: would, should, could, might, must, and ought
  - E.g., “The brake could be improved.”
- Some presuppositional items: barely, hardly, fail, omit, neglect, etc.
  - E.g., “It works” vs, “It hardly works.”, or “This camera fails to impress me.”
- Additionally, sarcasm changes orientation too,.
  - E.g., “What a great car, it failed to start the first day.”
- Not every appearance of a shifter changes the opinion orientation: “not only … but also.”

Handle but-clauses

- The opinion orientation before and after but are opposite to each other if the opinion on one side cannot be determined.
  - E.g., "The *picture quality* of this camera is **not great**[-1], but the *battery life* is **long**[+1]"
- Similar phrases: "with the exception of," "except that," and "except for", etc.
- Some structures with "but" don't show an opinion change.
  - E.g., "not only … but also…", or "Car-x is great, but Car-y is better."

Aggregating the opinions

- Given sentence *s* with a set of aspects $\{a_1, a_2, .., a_m\}$ and a set of opinion words $\{ow_1, ow_2, .., ow_n\}$ with opinion scores
- The opinion orientation for each aspect a_i in *s* is as follows:

$$\boldsymbol{score(a_i, s) = \sum_{ow_j \in s} \frac{ow_j.oo}{dist(ow_j, a_i)}}$$

  - ow_j is an opinion word/phrase in s, $dist(ow_j, a_i)$ is the distance between aspect a_i and opinion word ow_j in s, and $ow_j.oo$ is the opinion score of ow_j.
- The opinion is considered as positive if $score > 0$, negative if $score < 0$, and neutral otherwise.

Aspect extraction from user reviews

- Format 1 – Pros, cons, and the detailed review: The reviewer describes some brief pros and cons separately and write a detailed/full review.

My SLR is on the shelf

by camerafun4. Aug 09 '04

Pros: Great photos, easy to use, very small
Cons: Battery usage; included memory is stingy.

I had never used a digital camera prior to purchasing this Canon A70. I have always used a SLR … **Read the full review**

- Pros and Cons are often brief with short phrases or sentence segments.
- Many information extraction techniques can be applied.
  - E.g., Conditional Random Fields (CRF), Hidden Markov Models (HMM), and mining sequential rules [72]

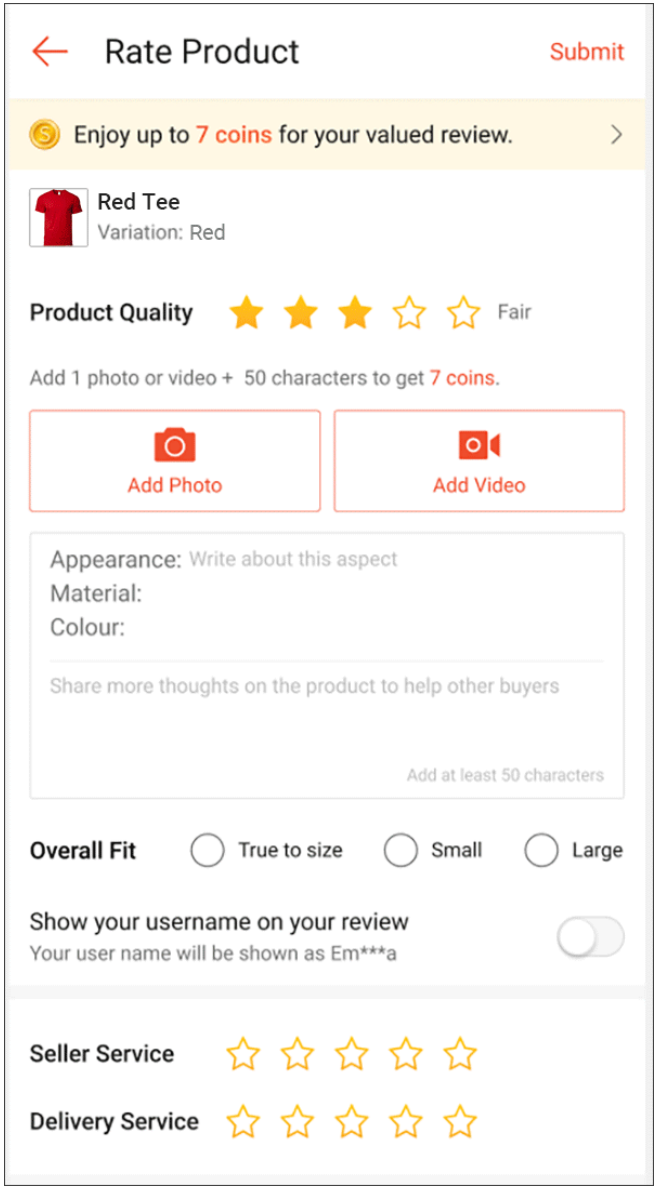
Aspect extraction from user reviews

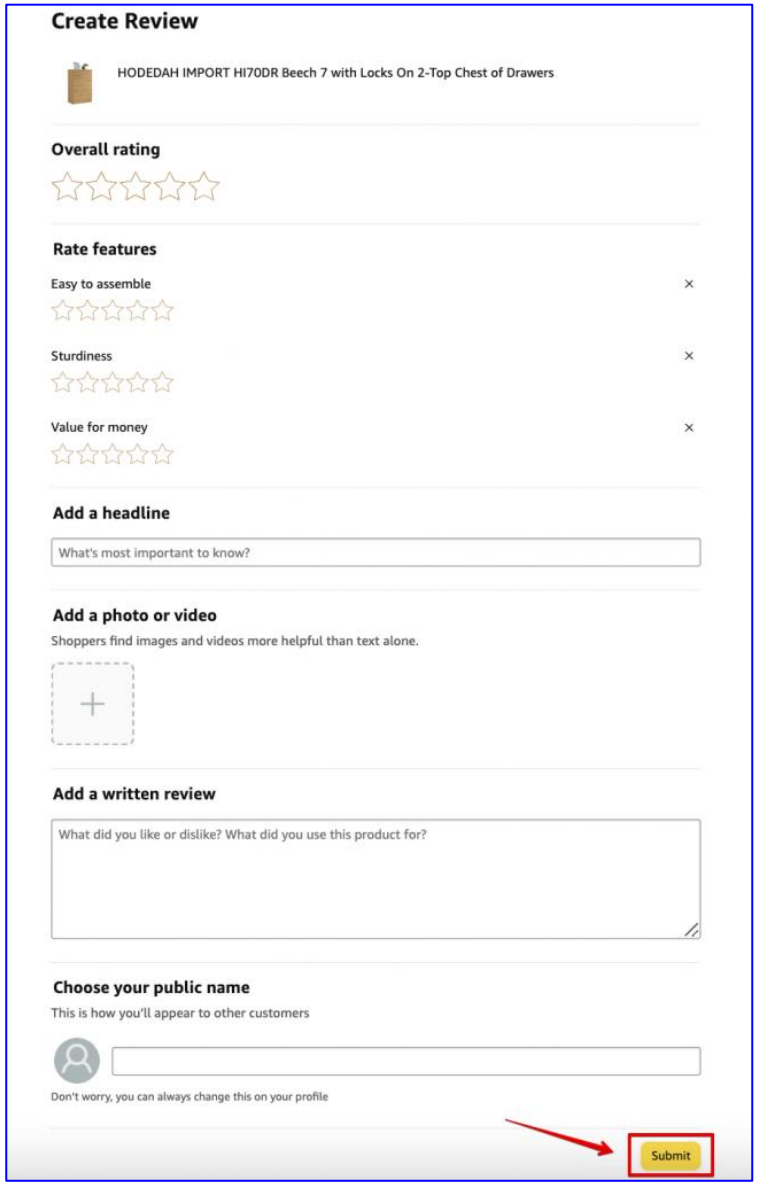
- Format 2 – Free format: The reviewer can write freely, i.e., no separation of pros and cons.

> **GREAT Camera.**, Jun 3, 2004
>
> Reviewer: **jprice174** from Atlanta, Ga.
>
> I did a lot of research last year before I bought this camera... It kinda hurt to leave behind my beloved nikon 35mm SLR, but I was going to Italy, and I needed something smaller, and digital.
>
> The pictures coming out of this camera are amazing. The 'auto' feature takes great pictures most of the time. And with digital, you're not wasting film if the picture doesn't come out. …

- The extraction is more general, performing on complete sentences.
- Previous algorithms are ineffective since these sentences are more complex with lots of noise.

User reviews: An example





List of references

- Bing Liu. 2007. *Web Data Mining-Exploring Hyperlinks, Contents, and Usage Data.* Springer Series on Data-Centric Systems and Applications. **Chapter 11.2-5**.
- Wankhade, M., Rao, A.C.S. & Kulkarni, C. A survey on sentiment analysis methods, applications, and challenges. Artif Intell Rev 55, 5731–5780 (2022). https://doi.org/10.1007/s10462-022-10144-1