Sentiment analysis

OPINION MINING PROBLEM

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

Why user opinions necessary?

Customer reviews

4.3 out of 5

3,704 global ratings

5 star	72%
4 star	12%
3 star	4%
2 star	2%
1 star	10%

By feature

Touch Screen	4.8
Face recognition	4.5
Battery life	4.2
Fingerprint reader	4.2
Value for money	4.0
Giftable	3.7

Apple iPhone 13, 128GB, Starlight - Unlocked (Renewed)

perfect phone

Reviewed in the United States on August 18, 2023

Service Provider: T-Mobile | Color: Red | Size: 128GB | **Verified Purchase**

i never do reviews but i am so happy with my purchase. i bought an excellent condition iphone and it's perfect! would totally recommend anyone getting one!

Cuts out when I talk

Reviewed in the United States on October 3, 2023

Service Provider: Unlocked | Color: Red | Size: 128GB | **Verified Purchase**

I guess this is what you get when you don't buy new. Cuts out when you talk on the phone. I I've tried updating it and resetting it, and every time I make a phone call it cuts in and out. The phone is pretty much useless.

Why user opinions necessary?

- Potential customers want to know the opinions of prior users before they use a service or purchase a product.
- Businesses find public or consumer opinions about their products and services.

- Opinions are key influencers of human behaviors.
  - Our beliefs and choices are considerably conditioned on how others see and evaluate the world.

The opinion mining problem

- The customers' opinions and evaluations is analyzed toward the entities of interest and their attributes.
- Keywords: opinion mining, sentiment analysis

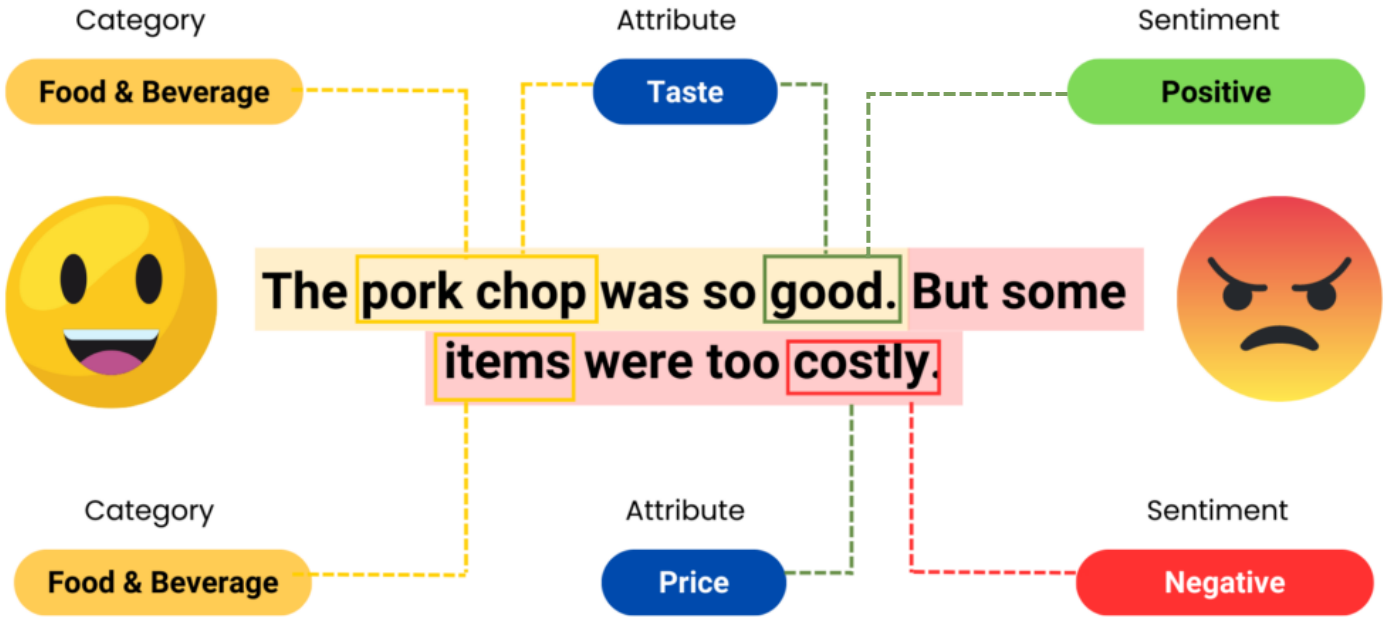


Image credit: Google Image

User opinions

- Product reviews are publicly available on the Web.

Polls, surveys, and focused groups to gather public opinions may no longer be necessary.

- Distilling the information contained in opinion sites remains a formidable task.
  - Opinionated text in long forum postings and blogs is not always easily deciphered.
  - The proliferation of diverse sites

User opinions

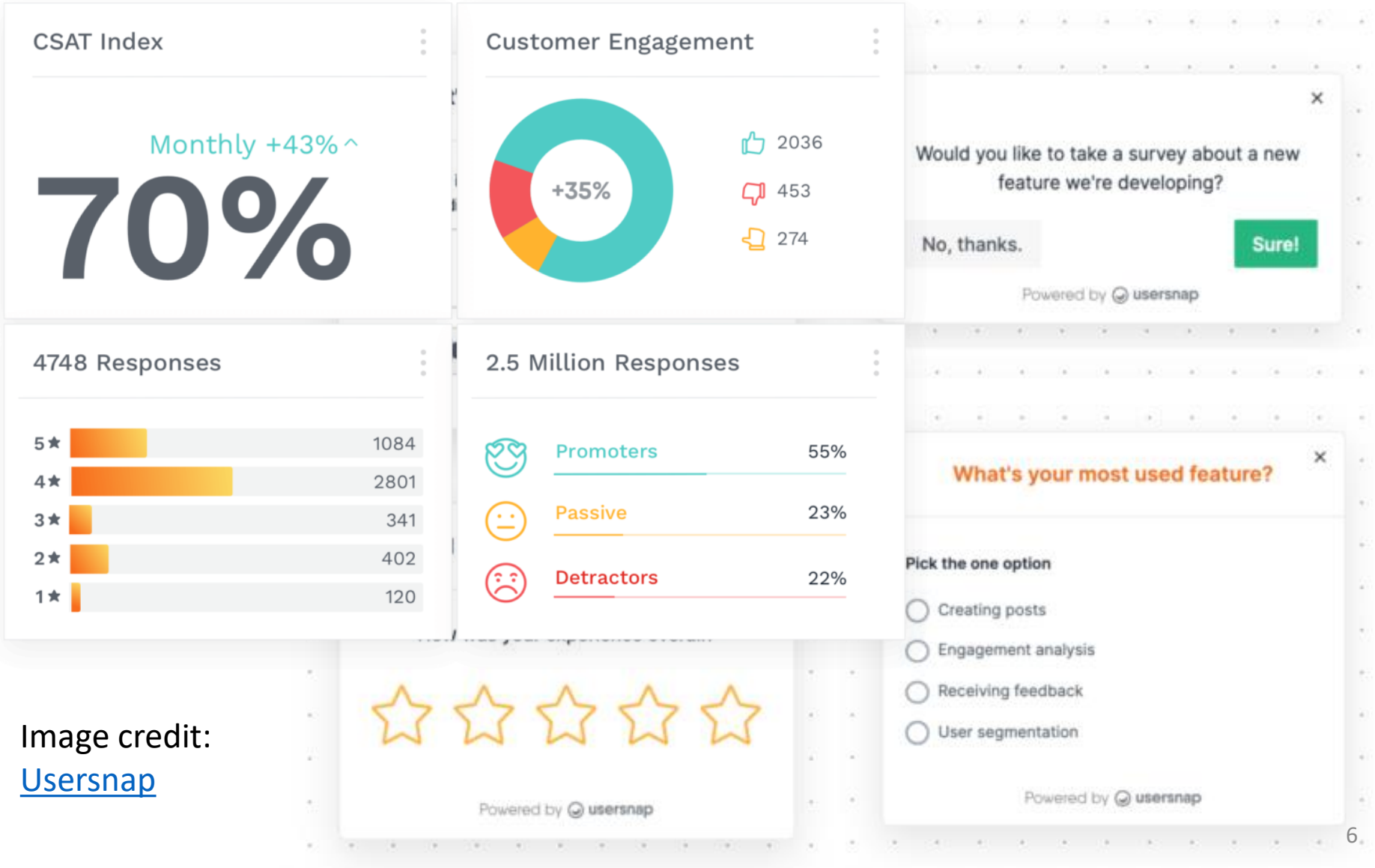
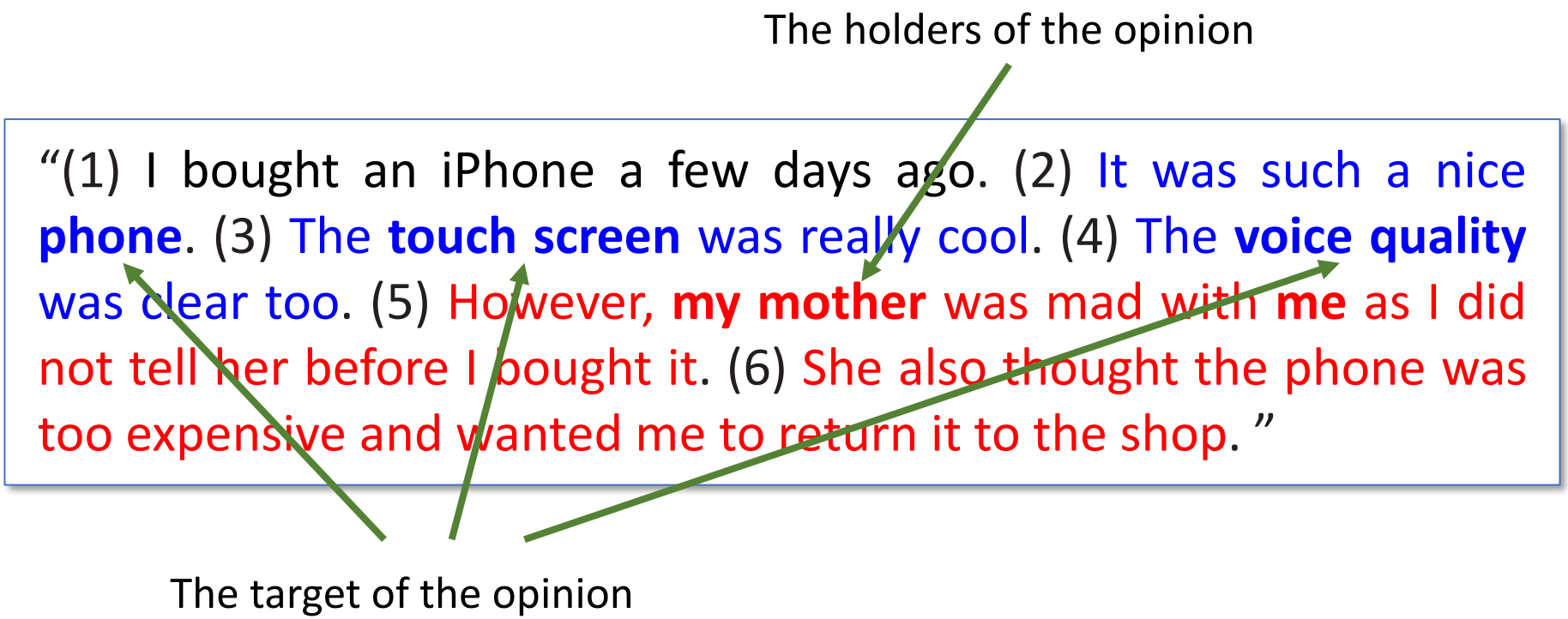


Image credit:
[Usersnap]

Automated opinion mining

- Objective: Identify relevant sites and accurately summarize the information and opinions contained in them
- Humans have difficulty in producing consistent results when the data to be processed is large.
  - *Mental and physical limitations* over time
  - *Considerable biases:* people often pay greater attention to opinions that are consistent with their own preferences
- Automated processes offer better objective opinion analysis.

Opinion mining: An example



- What we want to mine or extract from the above review?
  - Positive/negative opinions and emotions
  - The holder and target of the opinion

Basic concepts: Entity

- An entity represents target object that has been evaluated.
  - E.g., a product, service, person, event, organization, or topic.
- It is associated with a pair, $\boldsymbol{e}: (\boldsymbol{T}, \boldsymbol{W})$.
  - T is a hierarchy of components (and sub-components, etc.)
  - W is a set of attributes of $\boldsymbol{e}$.
  - Each component or sub-component also has its own set of attributes

Basic concepts: Entity

I do not like iPhone.

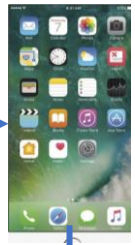
voice quality, size, weight,…

The voice quality of iPhone is clear.

part-of

battery life, battery size,…

Battery run down too fast.



resolution, brightness,…

…

An opinion can be expressed on any node of the hierarchy and any attribute of the node.

Basic concepts: Entity

- Entity name: the name of an entity given by the user.
- Entity expression: an actual word/phrase that has appeared in text indicating an entity

MOTOROLA

"Hello, Moto."

Basic concepts: Aspect

- The aspects of an entity e are components and attributes of e.
- Aspect name: the name of an aspect given by the user
- Aspect expression: an actual word/phrase that has appeared in text indicating an aspect

Name: voice quality

Expression: sound, voice, voice quality

Aspect expression

- Aspect expressions are usually *nouns and noun phrases*.
  - They can also be verbs, verb phrases, adjectives, and adverbs.
- Explicit aspect expression: noun or noun phrase
  - E.g., “sound” in “The sound of this phone is clear.”
- Implicit aspect expression: other types
  - Many of them are adjectives and adverbs to imply specific aspects.
    - E.g., “large” in “This phone is too large.” implies *size.*
  - They can be phrases of quite complex structures.
    - E.g., “fit in pockets” in “This phone will not easily fit in pockets.” implies *size* and/or *shape*.

Basic concepts: Opinion holder

- An opinion holder (or opinion source) is the organization or person that expresses the opinion.

- Opinion holders in news articles are more reliable thanks to their professed identities.
  - Those in product reviews and blogs can concealed themselves as anonymous users.

Basic concepts: Types of opinions

Regular opinions

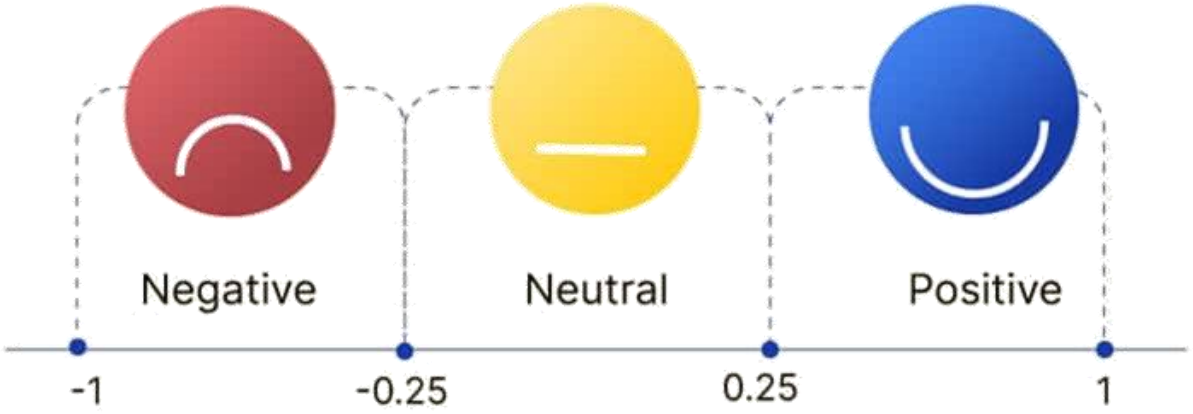
- An attitude, emotion, or appraisal about an entity or an aspect of the entity from an opinion holder.
  - E.g., iPhone 15 Pro Max is very expensive.

Comparative opinion

- The similarities between two or more entities, and/or
- A preference of the opinion holder based on some of the shared aspects of the entities.
  - E.g., Galaxy S23 Ultra is cheaper than iPhone 15 Pro Max.

Basic concepts: Opinion orientation

- Opinion orientations are categorization of user opinions.
- They can be discrete or interval values.
  - {positive, negative, neutral} vs. a scale of 1-5 stars.
- Keywords: sentiment/semantic orientation, or polarity.



- Neutral is often interpreted as no opinion in practice.

Opinions and Emotions

- Emotions are subjective feelings and thoughts, which may have different intensities.
- The strengths of opinions are related to the intensities of certain emotions, e.g., joy and anger.

Image credit: Medium

A formal definition of Opinion

- An opinion is a quintuple, $\langle e_i, a_{ij}, oo_{ijkl}, h_k, t_l \rangle$.
  - oo_{ijkl} is the orientation of the opinion about aspect a_{ij} of entity e_i, given by the option holder h_k at the time t_l.
  - When an opinion is on the entity itself as a while, the aspect is set to the default value GENERAL.
- Opinion quintuples give an excellent source of information for generating both qualitative and quantitative summaries.

Opinion quintuple: An example

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.

- (Motorola, voice_quality, negative, bigXyz, Nov-4-2010)
- (Motorola, camera, positive, bigXyz, Nov-4-2010)
- (Nokia, GENERAL, positive, bigXyz's girlfriend, Nov-4-2010)
- (Nokia, voice_quality, positive, bigXyz's girlfriend, Nov-4-2010)

Opinion definition: Important remarks

- The five pieces of information in the quintuple must conform to one another.
- These components are vital but not mandatory in all cases.
  - It can be problematic when one is missing, e.g., “The picture quality is great,” $\rightarrow$ whose picture quality is that?
  - Summarizing opinions from a population does not require knowing every opinion holder.
  - Extra information may be needed in some applications, e.g., sex and age of customers in marketing
- The quintuple as a nested relation rather than a flat relation.

Opinion definition: Important remarks

- Opinions include direct opinions and indirect opinions.
- Direct opinions are expressed directly on entities or aspects.
  - E.g., “The voice quality of this phone is great.”
- Indirect opinions are expressed based on their effects on some other entities.
  - E.g., “After taking this drug, my hand felt better” → a desirable effect of the drug on “my hand” → indirect positive opinion to the drug.

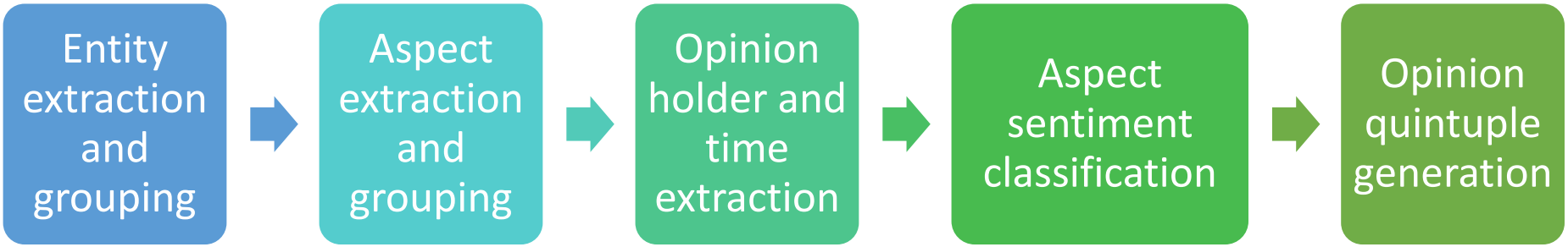
- This definition provides a basis for transforming unstructured text into structured data, which is later stored in RDBMS.

Aspect-based opinion mining

- Model of entity: An entity e_i is represented by itself as a whole and a finite set of aspects, $A_i = \{a_{i1}, a_{i2}, .., a_{in}\}$.
  - The entity itself can be expressed with any one of a finite set of entity expressions, $OE_i = \{oe_{i1}, oe_{i2}, .., oe_{in}\}$.
  - Each aspect $a_{ij} \in A_i$ can be depicted by any one of a finite set of aspect expressions, $AE_{ij} = \{ae_{ij1}, ae_{ij2}, .., ae_{ijm}\}$.
- Model of opinionated document: An opinionated document d contains opinions on a set of entities $\{e_1, e_2, .., e_r\}$ from a set of opinion holders $\{h_1, h_2, .., h_p\}$.
  - The opinions on each entity e_i are expressed on the entity itself and a subset A_{id} of its aspects.

Aspect-based opinion mining

- **Objective:** Given a collection of opinionated documents D, discover all opinion quintuples $\langle e_i, a_{ij}, oo_{ijkl}, h_k, t_l \rangle$ in D.
- We need to perform the following tasks to fulfil the objective.



- None of the above tasks is fully solved problem.

Opinion mining: Applications

- Aspect-based opinion summary: a common form of summary to capture the views from a large number of opinion holders.

Cellular phone 1:

Aspect: **GENERAL**

Positive:	125	<individual review sentences>
Negative:	7	<individual review sentences>

Aspect: **Voice quality**

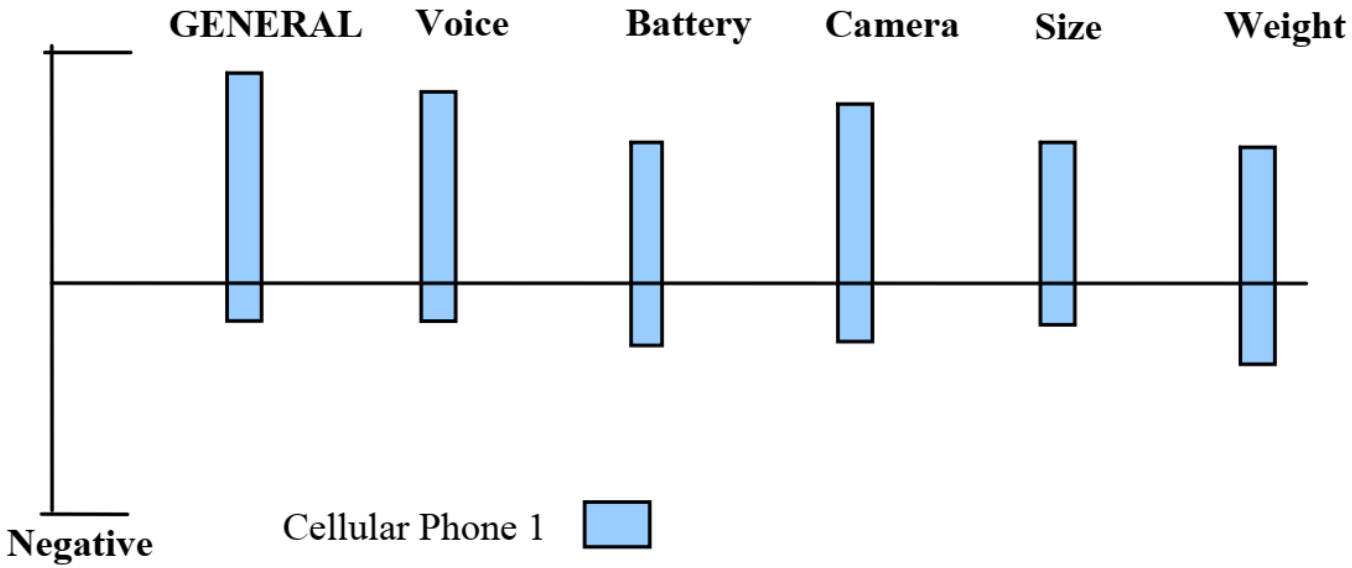
Positive:	120	<individual review sentences>
Negative:	8	<individual review sentences>

Aspect: **Battery**

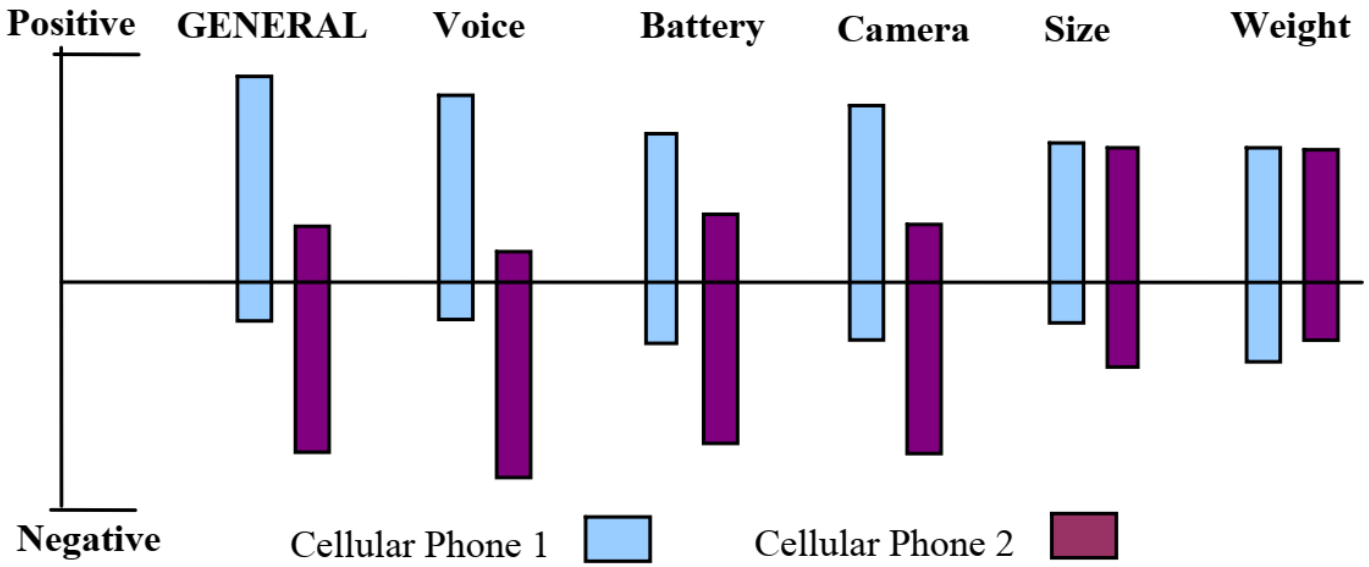
Positive:	80	<individual review sentences>
Negative:	12	<individual review sentences>

...

An aspect-based opinion summary of cellular phone 1 on different aspects (e.g., the phone as a whole, voice quality, battery, etc.).



Visualization of aspect-based summary of opinions on a cellular phone



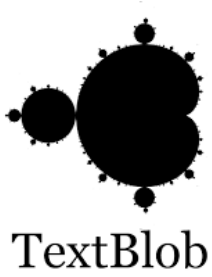
Visual opinion comparison of two cellular phones

Sentiment analysis tools

- Factors to pick the right sentiment analysis tool

Accuracy | Efficiency | Customization | Integration

- Python packages and libraries: read more in here



VADER
(GitHub)

spaCy

flair

PyTorch

1. The opinion quintuple

- Consider the following review about the Nokia mobile phone.

"This phone is good with a huge array of features built into it. I purchased the phone last week and have been using till then. I haven't had any problem till now. The design is sleek, and the color screen has good resolution. It is very light weight and has a good signal strength. However, the main problem that I think is the sound quality. It is not as good as the Samsung phones that I have used earlier. When talking the voice is not very clear. But I would definitely recommend this phone. Go for it ..."

- Identify four opinions mentioned in the review.
- For each opinion, represent it as a tuple of aspect A, opinion orientation OO, and cite the phrases used to detect OO.

2. The opinion quintuple

- Consider the following review about the ABC restaurant.

"My wife took me here on my birthday for breakfast. It took me one hour to get a table on a non-weekend day, and the waitress was slow in service. The selection on the menu was great and so were the restaurant's view. However, I still think that the food is overpriced, and the portions are incredibly small. I tried shrimp salad, salmon sashimi, pasta and French wine. Meanwhile, my wife had Caesar salad, beefsteak and Chilean wine. The salmon sashimi is the worst of all my dishes. It was not fresh at all."

- Identify four opinions mentioned in the review.
- For each opinion, represent it as a tuple of aspect A, opinion orientation OO, and cite the phrases used to detect OO.

List of references

- Bing Liu. 2007. *Web Data Mining-Exploring Hyperlinks, Contents, and Usage Data.* Springer Series on Data-Centric Systems and Applications. **Chapter 11.1**.


THE END