

*Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*

---

# Sentiment Analysis Essentials

---

Bing Liu  
Department of Computer Science  
University Of Illinois at Chicago  
liub@cs.uic.edu

# Introduction

- Sentiment analysis (SA) or opinion mining
  - computational study of opinion, sentiment, appraisal, evaluation, and emotion.
- Why is it important?
  - Businesses and organizations
    - Benchmark products and services; marketing intelligence
  - Individuals
    - Make decisions to purchase products or services
    - Find public opinions about political candidates and issues
  - Rise of social media → opinion data

# SA: A fascinating problem!

## ■ Popular research

- popular research topic in NLP and data mining
  - very rich problem, many inter-related sub-problems
- Spread from CS to management & social sciences

## ■ Unlimited applications

- > 300 companies in USA alone.

## ■ Intellectually challenging

- CEO 1: *“Anyone who claims >70% accuracy is lying”*
- CEO 2: *“Our system is as bad as everyone else’s”*

---

# Roadmap

- Sentiment analysis problem
- Document and sentence sentiment classification
- Aspect-based sentiment analysis
  - Aspect extraction
  - Aspect sentiment classification
- Summary

# Sentiment analysis: core problem

## ■ It consists of two abstractions

### (1) Opinion definition. What is an opinion?

- Can we provide a structured definition?
  - If we cannot structure a problem, we probably do not understand the problem.

### (2) Opinion summarization. why?

- Opinions are subjective. An opinion from a single person (unless a VIP) is often not sufficient for action.
- We need opinions from many people, and thus the need for opinion summarization.

# Two main types of opinions

(Jindal and Liu 2006; Liu, 2010)

- **Regular opinions:** Sentiment/opinion expressions on some target entities
  - **Direct opinions:**
    - “The **touch screen** is really cool.”
  - **Indirect opinions:**
    - “After taking the drug, my pain has gone.”
- **Comparative opinions:** Comparisons of more than one entity.
  - E.g., “iPhone is better than Blackberry.”

# (I): Definition of an opinion

- **Id:** **Abc123** on **5-1-2008** -- “I bought an *iPhone* yesterday. It is such a nice *phone*. The *touch screen* is really *cool*. The *voice quality* is *clear* too. It is much *better* than my *Blackberry*. However, *my mom* was *mad* with *me* as I didn’t tell her before I bought the *phone*. She thought the *phone* was too *expensive*”
- **Definition:** An *opinion* is a quadruple (Liu, 2012),  
(*target*, *sentiment*, *holder*, *time*)
- This definition is concise, but not easy to use.
  - Target can be complex, e.g., “I bought an iPhone. The *voice quality* is amazing.”
    - *Target* = *voice quality*? (not quite)

# A more practical definition

(Hu and Liu 2004; Liu, 2010, 2012)

- An *opinion* is a quintuple

(*entity*, *aspect*, *sentiment*, *holder*, *time*)

where

- *entity*: target entity (or object).
- *aspect*: aspect (or feature) of the entity.
- *sentiment*: +, -, or neu, a rating, or an emotion.
- *holder*: opinion holder.
- *time*: time when the opinion was expressed.

- *Aspect-based sentiment analysis*



# Our example blog in quintuples

- **Id: Abc123 on 5-1-2008** *“I bought an **iPhone** a few days ago. It is such a nice **phone**. The **touch screen** is really cool. The **voice quality** is clear too. It is much better than my old **Blackberry**, which was a terrible **phone** and so **difficult to type** with its **tiny keys**. However, **my mom** was mad with me as I did not tell her before I bought the **phone**. She also thought the phone was too **expensive**, ...”*
- **In quintuples**
  - (iPhone, GENERAL, +, Abc123, 5-1-2008)
  - (iPhone, touch\_screen, +, Abc123, 5-1-2008)
  - ....
- We will discuss comparative opinions later.

# Subjectivity

- **Subjective sentences** express personal feelings, views, emotions, or beliefs. **Objective sentences** present factual information (Wiebe 2000).
  - A subjective sentence may not contain a +/- opinion, e.g., “*I think he went home*”
- **Most opinionated sentences are subjective, but objective sentences can imply opinions too** (Liu, 2010)
  - “The machine stopped working in the second day”
  - “We brought the mattress yesterday, and a body impression has formed.”
  - “After taking the drug, there is no more pain”

# Emotion

- No agreed set of basic emotions among theorists.
- Based on Parrott (2001), we have 6 basic emotions: love, joy, surprise, anger, sadness, and fear.
- Emotions and opinions are not equivalent.
- **Definition (Emotion):** It is a quintuple,  
*(entity, aspect, emotion\_type, feeler, time)*
- E.g., “*I am so mad with the hotel manager because he refused to refund my booking fee*”
  - **Entity:** hotel, **Aspect:** manager, **emotion\_type:** anger  
**feeler:** I, **time:** unknown

## (II): Opinion summary (Hu and Liu 2004)

- With a lot of opinions, a summary is necessary.
  - Not traditional text summary: from long to short.
  - Text summarization: defined operationally based on algorithms that perform the task
- Opinion summary (OS) can be defined precisely,
  - not dependent on how summary is generated.
- Opinion summary needs to be quantitative
  - 60% positive is very different from 90% positive.
- Main form of OS: *Aspect-based opinion summary*

# Aspect-based opinion summary<sup>1</sup>

(Hu & Liu, 2004)

“I bought an *iPhone* a few days ago. It is such a nice *phone*. The *touch screen* is really cool. The *voice quality* is clear too. It is much better than my old *Blackberry*, which was a terrible *phone* and so *difficult to type* with its *tiny keys*. However, *my mother* was mad with me as I did not tell her before I bought the *phone*. She also thought the phone was too *expensive*, ...”

1. Originally called *feature-based opinion mining and summarization*

## Feature Based Summary of iPhone:

### Feature1: **Touch screen**

Positive: 212

- The *touch screen* was really cool.
- The *touch screen* was so easy to use and can do amazing things.

...

Negative: 6

- The *screen* is easily scratched.
- I have a lot of difficulty in removing finger marks from the *touch screen*.

...

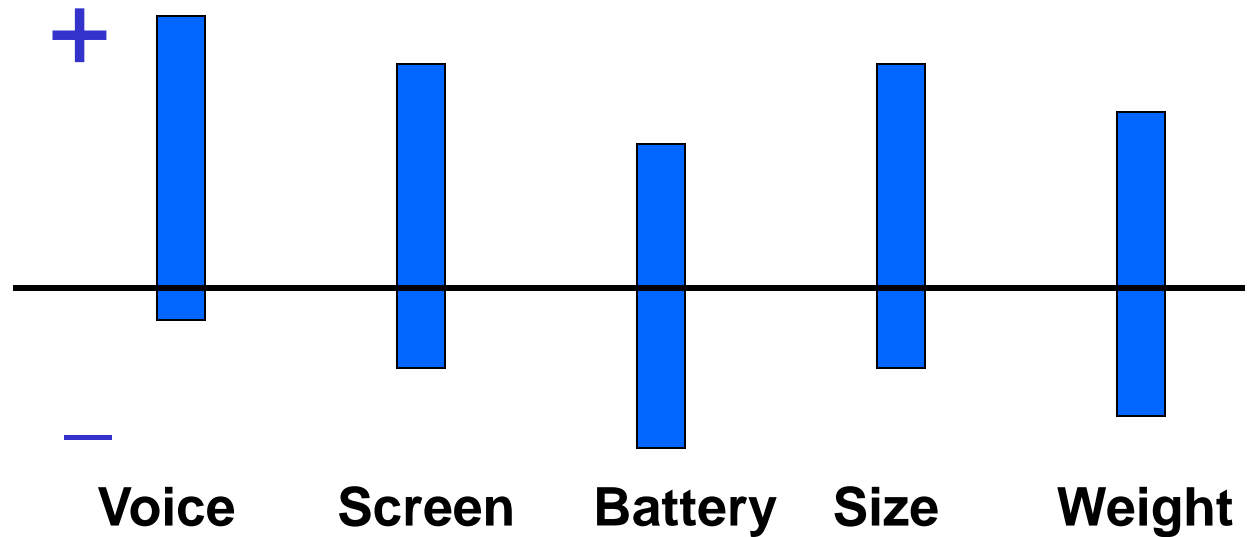
### Feature2: **voice quality**

...

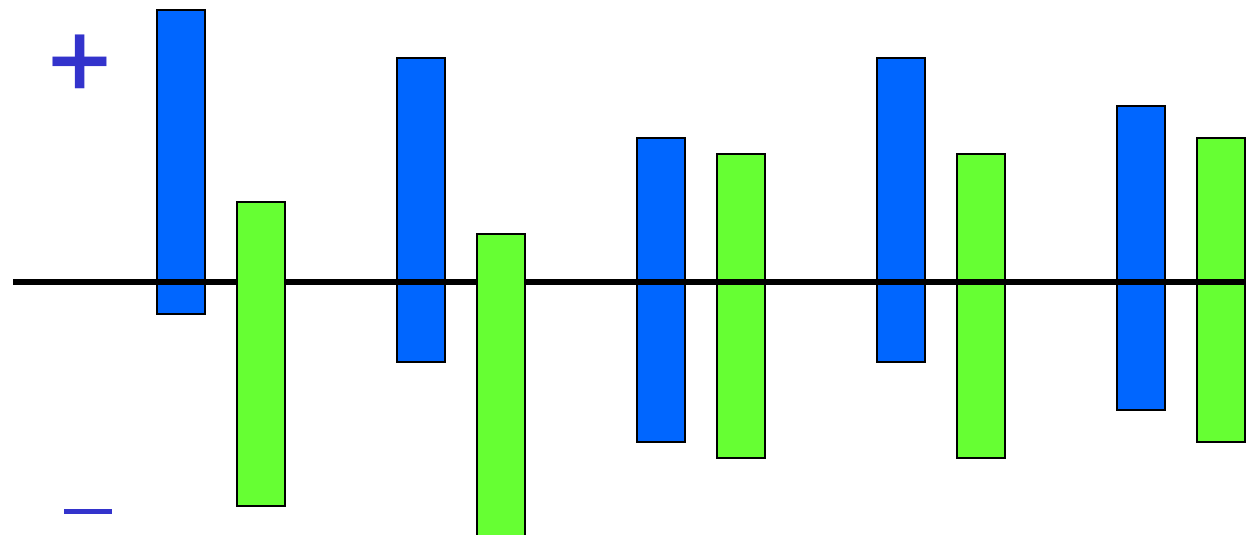
Note: We omit opinion holders

# Opinion Observer (Liu et al. 2005)

- Summary of reviews of **Cell Phone 1**



- Comparison of reviews of **Cell Phone 1** and **Cell Phone 2**



# Aspect-based opinion summary

bing HP printer

ALL RESULTS

Shopping

POPULAR FEATURES

- all
- Affordability
- Speed
- Print Quality
- Reliability
- Ease Of Use
- Brand
- Installation
- Size
- Compatibility

SHOPPING

HP LaserJet 1020 - printer

from \$119

★★★★★

The HP LaserJet 1020 is a high-quality printer.

user reviews

product

Google products sony camera Search Products

**Sony Cyber-shot DSC-W370 14.1 MP Digital Camera (Silver)**

[Overview](#) - [Online stores](#) - [Nearby stores](#) - [Reviews](#) - [Technical specifications](#) - [Similar items](#) - [Accessories](#)

 **\$140 online, \$170 nearby**

★★★★☆ 159 reviews +1 0

**Reviews**

Summary - Based on 159 reviews

1 2 3 stars 4 stars 5 stars

**What people are saying**

<a href="#">pictures</a>	■■■■■	"We use the product to take quickly photos."
<a href="#">features</a>	■■■■■	"Impressive panoramic feature."
<a href="#">zoom/lens</a>	■■■■■	"It also record better and focus better on sunny days."
<a href="#">design</a>	■■■■■	"It has the slightest grip but it's sufficient."
<a href="#">video</a>	■■■■■	"Video zoom is choppy."
<a href="#">battery life</a>	■■■■■	"Even better, the battery lasts long."
<a href="#">screen</a>	■■■■■	"I Love the Sony's 3" screen which I really wanted."

## user reviews

view: positive comments (44)

speed ■■■■■ 96%

The quality is as good as any laserjet printer I've used and the speed is fast.  
Love Reading [www.amazon.com](http://www.amazon.com) 3/17/2006 [more...](#)

Quick and fast transaction.  
Arthur L. Taylor [www.amazon.com](http://www.amazon.com) 2/5/2008 [more...](#)

It's small and fast and very reliable.  
Muffinhead's mom [www.amazon.com](http://www.amazon.com) 1/9/2007 [more...](#)

# Roadmap

- Sentiment analysis problem
- Document and sentence sentiment classification
- Aspect-based sentiment analysis
  - Aspect extraction
  - Aspect sentiment classification
- Summary



# Document sentiment classification

- **Classify a review based on its overall sentiment.**
  - **Classes:** Positive, Negative (and possibly neutral)
- **Is this review positive or negative?**
  - **John - 5/1/2008 --** *“I bought an iPhone yesterday. It is such a nice phone. The touch screen is really cool. The voice is clear too. It is much better than my old Blackberry. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive, ...”*
- **Solution methods: classification**
  - Supervised (Pang et al. 2002; Xia & Zong 2011, Li et al, 2012 ...)
  - Unsupervised - based on patterns (Turney, 2002; ...)

# Sentence sentiment classification

- **Classify each sentence based on its sentiment.**
  - ❑ (1) **Classes:** Subjective, Objective.
  - ❑ (2) **Classes:** Positive, Negative, Neutral
  - ❑ Ignore mixed: *“Apple is going well in this bad economy.”*
- **Is this sentence positive, negative, or neutral?**
  - ❑ *“Trying out Chrome because Firefox keeps crashing.”*
- **Solution: (un)supervised text classification**
  - ❑ (1) Subjectivity classification (Wiebe *et al.*, 1999; Socher *et al.* 2013...)
  - ❑ (2) Sentiment classification (Yu & Hatzivassiloglou, 2003; ...)

# Roadmap

- Sentiment analysis problem
- Document and sentence sentiment classification
- **Aspect-based sentiment analysis**
  - **Aspect extraction**
  - **Aspect sentiment classification**
- Summary

# We need to go further

- Sentiment classification at both the document and sentence (or clause) levels are useful, but
  - They do not identify the targets of opinions, i.e.,
    - Entities and their aspects
    - Without knowing targets, opinions are of limited use.
- We need to go to the entity and aspect level.
  - *Aspect-based opinion mining and summarization* (Hu and Liu 2004).
  - We thus need the full opinion definition.  
(*entity*, *aspect*, *sentiment*, *holder*, *time*)

# Aspect extraction

- **Goal:** Given an opinion corpus, extract all aspects
- Main approaches:
  - ❑ (1) Finding frequent nouns and noun phrases
  - ❑ (2) Exploiting opinion and target relations
  - ❑ (3) Supervised learning
  - ❑ (4) Topic modeling
  - ❑ Many others (Meng and Wang, 2009; Moghaddam, and Ester 2010; Liu, Xu, and Zhao, 2013, 2014; Zhou, Wan, and Xiao, 2013; Xu et al, 2013; 2014; Zhao, Qin & Liu 2014 ...)

# (1) Frequent nouns and noun phrases

- Nouns (NN) that are frequently mentioned are likely to be true **aspects** (Hu and Liu 2004).
- Improved (Hu and Liu, 2004) by removing some frequent noun phrases that may not be aspects (Popescu and Etzioni, 2005)
  - It identifies part-of relationship
    - Each noun **phrase** is given a pointwise mutual information score between the phrase and **part discriminators** associated with the product class, e.g., a scanner class.
    - E.g., “of scanner”, “scanner has”, etc, which are used to find parts of scanners by searching on the Web:

## (2) Exploiting opinion & target relation

- **Key idea:** opinions have targets, i.e., opinion terms are used to modify aspects and entities.
  - “The pictures are absolutely amazing.”
  - “This is an amazing software.”
- The syntactic relation is approximated with the **nearest** noun phrases to the opinion word in (Hu and Liu 2004).
- The idea was generalized to
  - **syntactic dependency** in (Zhuang et al 2006)
  - **double propagation** in (Qiu et al 2009).

# Extract aspects using DP (Qiu et al. 2009; 2011)

- *Double propagation (DP)*
  - Based on the definition earlier, **an opinion should have a target**, entity or aspect.
- Use dependency of opinions & aspects to extract both aspects & opinion words.
  - **Knowing one helps find the other.**
  - E.g., “*The **rooms** are **spacious***”
- It extracts both aspects and opinion words.
  - A domain independent method.



### (3) Using supervised learning

- Using sequence labeling methods such as
  - **Hidden Markov Models** (HMM) (Jin and Ho, 2009)
  - **Conditional Random Fields** (Jakob and Gurevych, 2010).
  - Other supervised or semi-supervised learning.
- (Liu, Hu and Cheng 2005; Kobayashi et al., 2007; Li et al., 2010; Choi and Cardie, 2010; Yu et al., 2011; ...).

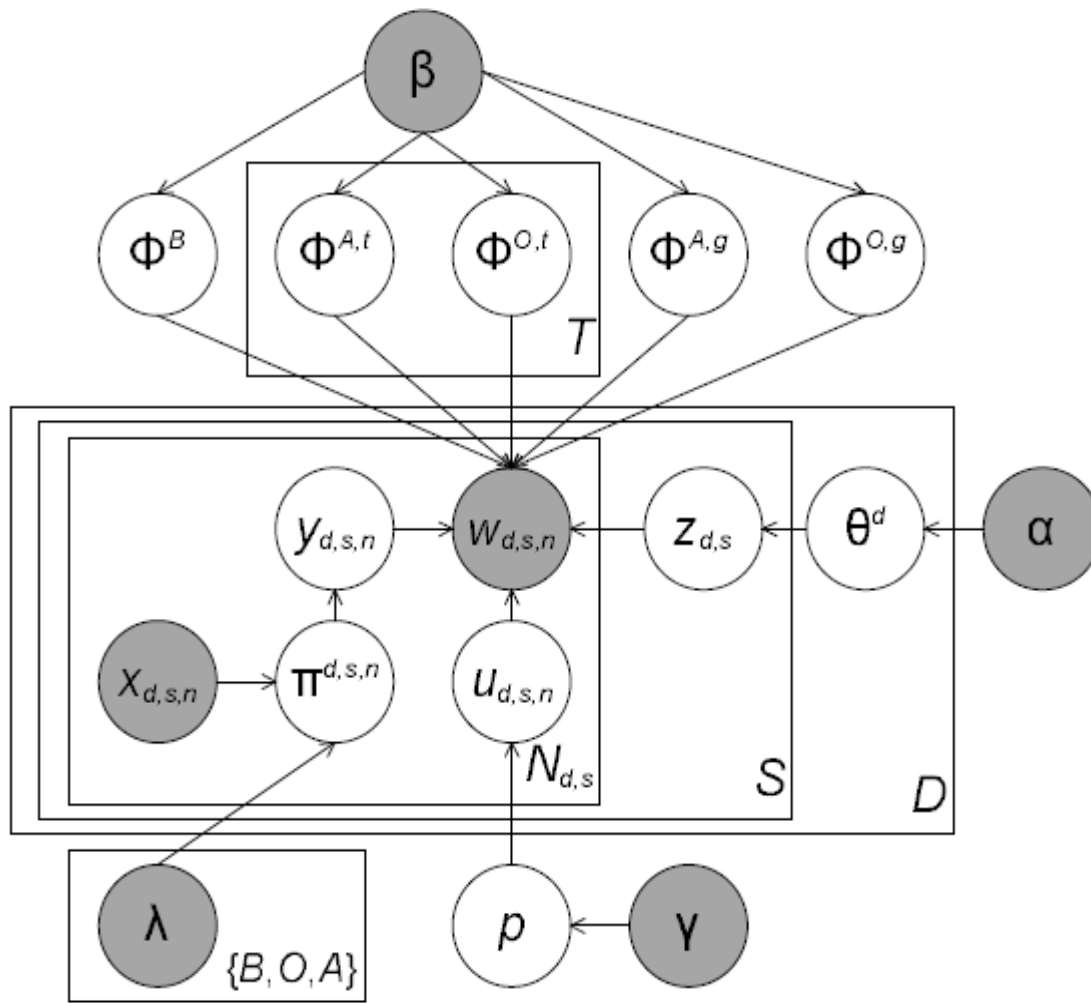
## (4) Topic modeling

- **Aspect extraction has two tasks:**
  - (1) extract aspect expressions
  - (2) cluster them (same: “picture,” “photo,” “image”)
- Top models such as **pLSA** (Hofmann 1999) and **LDA** (Blei et al 2003) perform both tasks at the same time. **A topic is basically an aspect.**
  - A document is a distribution over topics
  - **A topic is a distribution over terms/words, e.g.,**
    - *{price, cost, cheap, expensive, ...}*
      - Ranked based on probabilities (not shown).

# Many related Models and papers

- **Use topic models to model aspects** (Titov and McDonald, 2008; Chen et al., 2013; ...)
- **Jointly model both aspects and sentiments** (Mei et al 2007; Lu et al 2009; Wang et al. 2010; Brody and Elhadad. 2010; Jo and Oh 2011; Fang and Huang 2012)
- **Knowledge-based modeling**: Unsupervised models are often insufficient
  - ❑ Not producing coherent topics/aspects
  - ❑ To tackle the problem, *knowledge-based topic models* have been proposed

# MaxEnt-LDA hybrid (Zhao et al. 2010)



- $y_{d,s,n}$  indicates
  - Background word
  - Aspect word, or
  - Opinion word
- MaxEnt is used to train a model using training set
  - $\pi^{d,s,n}$
  - $x_{d,s,n}$  feature vector
- $u_{d,s,n}$  indicates
  - General or
  - Aspect-specific

# Knowledge-based topic models

(Mukherjee and Liu, 2012; Chen et al., 2013)

- Unsupervised models are often insufficient
  - because their objective functions may not correlate well with human judgments (Chang et al., 2009).
- To tackle the problem, *knowledge-based topic models* (KBTM) or semi-supervised topic model have been proposed
  - Guided by user-specified prior domain knowledge.
    - Seed terms (Mukherjee & Liu, 2012)
    - Constraints (Chen et al., 2013)

# Learn as humans do (Chen et al., 2014)

- **Knowledge-based modeling is good**, but
  - Knowledge provided by user and not comprehensive.
- **Lifelong learning**: retain knowledge (topics or aspects learned the past to help future learning).
- **Key observation**: Although every domain is different, there is a fair amount of aspect overlapping across domains
  - Every product review domain has the topic/aspect *price*,
  - Most electronic products share the topic/aspect *battery*
  - Some products share the topic/aspect *screen*.

# Topic & word sharing: an example

- We have reviews from 3 domains and each domain gives a topic related to *price*.
  - Domain 1: {*price*, *color*, *cost*, *life*}
  - Domain 2: {*cost*, *picture*, *price*, *expensive*}
  - Domain 3: {*price*, *money*, *customer*, *expensive*}
- If we require words appear in at least two domains, we get two sets (i.e., knowledge):
  - {*price*, *cost*} and {*price*, *expensive*}.
  - These can be used by a KBTM for better results.

# Aspect sentiment classification

- For each aspect, identify the sentiment about it
- Work based on sentences, but also consider,
  - A sentence can have multiple aspects with different opinions.
  - E.g., The **battery life** and **picture quality** are *great* (+), but the **viewfinder** is *small* (-).
- Almost all approaches make use of **opinion words and phrases**. But notice:
  - Some opinion words have **context independent orientations**, e.g., “good” and “bad” (almost)
  - Some other words have **context dependent orientations**, e.g., “long,” “quiet,” and “sucks” (+ve for vacuum cleaner)



# Aspect sentiment classification

*“Apple is doing very well in this poor economy”*

- **Lexicon-based approach:** Opinion words/phrases
  - **Parsing:** simple sentences, compound sentences, conditional sentences, questions, modality verb tenses, etc (Hu and Liu, 2004; Ding et al. 2008; Narayanan et al. 2009).
- **Supervised learning is tricky:**
  - **Feature weighting:** consider distance between word and target entity/aspect (e.g., Boiy and Moens, 2009)
  - **Use a parse tree** to generate a set of target dependent features (e.g., Jiang et al. 2011)

# A lexicon-based method (Ding et al. 2008)

- **Input:** A set of opinion words and phrases. A pair  $(a, s)$ , where  $a$  is an aspect and  $s$  is a sentence that contains  $a$ .
- **Output:** whether the opinion on  $a$  in  $s$  is +ve, -ve, or neutral.
- Two steps:
  - Step 1: split the sentence if needed based on BUT words (but, except that, etc).
  - Step 2: work on the segment  $s_f$  containing  $a$ . Let the set of opinion words in  $s_f$  be  $w_1, \dots, w_n$ . Sum up their orientations  $(1, -1, 0)$ , and assign the orientation to  $(a, s)$  based on:

$$\sum_{i=1}^n \frac{w_i.o}{d(w_i, a)}$$

where  $w_i.o$  is the opinion orientation of  $w_i$ .  $d(w_i, a)$  is the distance from  $a$  to  $w_i$ .

# Sentiment shifters (e.g., Polanyi and Zaenen 2004)

- Sentiment/opinion shifters (also called **valence shifters** are words and phrases that can shift or change opinion orientations.
  - Negation words like *not*, *never*, *cannot*, etc., are the most common type.
  - Many other words and phrases can also alter opinion orientations. E.g., **modal auxiliary verbs** (e.g., *would*, *should*, *could*, etc)
    - “The brake could be improved.”
  - Many more ... (Liu 2012; 2015)

---

# Roadmap

- Sentiment analysis problem
- Document and sentence sentiment classification
- Aspect-based sentiment analysis
  - Aspect extraction
  - Aspect sentiment classification
- Summary

# Summary

- Defined the sentiment analysis (SA) problem
  - The definition provides a structure to the problem
    - Core: Sentiment and target
  - SA is a *restricted* semantic analysis problem
- Discussed some main research directions.
- It is a fascinating NLP or text mining problem.
  - None of the sub-problems is solved.
  - Many sub-problems hardly attempted.
- Despite challenges, applications are thriving!

# Advanced Topics

---

Some difficult problems

# Introduction

- Not discuss sophisticated algorithms
  - No algorithm is good enough yet.
- **Instead:** Focus on highlighting some important problems not studied much in the literature.
  - Researchers have worked in only limited domains
- One-size-fits-all is unlikely to work (Liu 2012, 2015)
  - Need to identify and deal with different types of sentences individually
- SA can contribute to general NLP significantly

# SA is a very rich problem

## ■ (*entity*, *aspect*, *sentiment*, *holder*, *time*)

- target *entity*: Entity extraction & resolution
- *aspect* of *entity*: Aspect extraction & resolution
- *sentiment*: Aspect sentiment classification
- opinion *holder*: Information/data extraction
- *time*: Information/data extraction

## ■ Other NLP problems

- Synonym grouping (voice = sound quality)
- Lexical semantics
- Coreference resolution
- .....



---

# Roadmap

- Entity set expansion
- Explicit and implicit aspects
- Resource usage aspect and sentiment
- Factual terms implying aspects and sentiment
- Some other difficult sentences
- Coreference resolution
- Comparative opinions
- Opinion spam detection
- Summary

# Find entities (entity set expansion)

- Although similar, it is somewhat different from the traditional named entity recognition (NER).
- E.g., one wants to study opinions about phones
  - given Motorola and Nokia, find all phone brands and models in a corpus, e.g., Samsung, Moto,
- **Formulation:** Given a set  $Q$  of seed entities of class  $C$ , and a set  $D$  of candidate entities, we wish to determine which of the entities in  $D$  belong to  $C$ .
  - A classification problem. It needs a binary decision for each entity in  $D$  (belonging to  $C$  or not)
  - But it's often solved as a ranking problem

# Some methods (Li et al., 2010, Zhang and Liu, 2011)

- **Distributional similarity**: A traditional method used in NLP. It compares the surrounding text of candidates using cosine or PMI.
  - It does not perform well.
- **PU learning**: Learning from positive and unlabeled examples.
  - S-EM algorithm (Liu et al. 2002)
- **Bayesian Sets**: We extended the method given in (Ghahramani and Heller, 2005).

# Roadmap

- Entity set expansion
- **Explicit and implicit aspects**
- Resource usage aspect and sentiment
- Factual terms implying aspects and sentiment
- Some other difficult sentences
- Coreference resolution
- Comparative opinions
- Opinion spam detection
- Summary

# Explicit and implicit aspects

(Hu and Liu, 2004)

- **Explicit aspects:** Aspects explicitly mentioned as nouns or noun phrases in a sentence
  - “The **picture quality** is of this phone is great.”
- **Implicit aspects:** Aspects not explicitly mentioned in a sentence but are implied
  - “This car is so **expensive**.”
  - “Included **16MB** is stingy.”
  - “This phone will not easily **fit in a pocket**.”
  - “The machine can **play DVD’s**, which is its best feature.”
- **Some work has been done** (Su et al. 2009; Hai et al 2011)

# Implicit aspect extraction

- Many types of implicit aspect expressions: Adjectives/adverbs are most common.
- **Lexical semantics problem** (Hartung and Frank, 2010)
  - Most adjectives modify specific attributes.
  - “expensive”  $\Rightarrow$  “price,” “beautiful”  $\Rightarrow$  “appearance”, “heavy”  $\Rightarrow$  “weight” (Fei et al., 2012)
- Different contexts, different meanings
  - “The traffic is very heavy.”
- **More complex implicit aspects not studied**

---

# Roadmap

- Entity set expansion
- Explicit and implicit aspects
- **Resource usage aspect and sentiment**
- Factual terms implying aspects and sentiment
- Some other difficult sentences
- Coreference resolution
- Comparative opinions
- Opinion spam detection
- Summary

# Aspect and sentiment: resource usage

(Zhang and Liu, 2011)

## ■ Resource usage

- E.g., “This washer uses a lot of water.”

## ■ Two key roles played by resource usage:

- An important **aspect** of an entity, e.g., water usage.
- Imply a positive or negative **opinion**

## ■ Resource usages that imply opinions can often be described by a *context triple*.

(verb, quantifier, noun\_term),

- **Verb**: uses, **quantifier**: “a lot of “, **noun\_term**: water



# One initial technique

- The method is graph-based.
  - Stage 1: Identifying Some Global Resource Verbs
    - Identify and score common resource usage verbs used in almost any domain, e.g., “use” and “consume”
  - Stage 2: Discovering Resource Terms in each Domain Corpus
    - Use a graph-based method considering occurrence probabilities.
    - With resource verbs identified from stage 1 as the seeds.
    - Score domain resource usage verbs and resource terms.
- Accuracy not good and doesn't do classification

# Roadmap

- Entity set expansion
- Explicit and implicit aspects
- Resource usage aspect and sentiment
- **Factual terms implying aspects and sentiment**
- Some other difficult sentences
- Coreference resolution
- Comparative opinions
- Opinion spam detection
- Summary

# Opinions implied by factual terms

(Zhang and Liu, 2011)

- Subjective statements **often** express/imply sentiment using words: good, bad, hate, love, etc
- **Objective sentences imply aspect and sentiment**
  - Desirable and undesirable facts
  - Highly domain dependent
- E.g.,
  - “After sleeping on the mattress for one month, a **valley/body impression** has formed in the middle.”
  - “This player **skips frames**.”
  - “The mouse **doubleclicks** on a single click.”

# One initial method for noun phrases

- Sentiment analysis to determine whether the context is +ve or –ve.
  - E.g., “I saw a **valley** in two days, which is terrible.”
  - This is a negative context.
- Statistical test to find +ve and –ve candidates.

$$Z = \frac{P - P_0}{\sqrt{\frac{P_0(1 - P_0)}{n}}}$$

- Pruning to move those unlikely ones though *sentiment homogeneity*.

# Pruning

- For an aspect with an implied opinion, it has a fixed opinion, either +ve or -ve, but not both.
- Find two direct modification relations using a dependency parser.
  - Type 1:  $O \rightarrow O\text{-}Dep \rightarrow A$ 
    - e.g. “ *This TV has **good** **picture** quality.*”
  - Type 2:  $O \rightarrow O\text{-}Dep \rightarrow H \leftarrow A\text{-}Dep \leftarrow A$ 
    - e.g. “ *The **springs** of the mattress **are bad**.*”
- If an aspect has mixed opinions based on the two dependency relations, prune it.

# One initial technique for verb phrases

(Li et al. 2015)

- “The mouse doubleclicks on a single click.”
- An initial technique
  - Step 1: extract verb phrases using a chunker.
  - Step 2: use Markov Networks to infer the sentiment orientation or polarity class of each candidate verb phrase.
    - The network represents grammatical dependency between words in a phrase.
    - Initial data with labels are extracted from titles of negative (1 star) and positive (5 star) reviews.

# Roadmap

- Entity set expansion
- Explicit and implicit aspects
- Resource usage aspect and sentiment
- Factual terms implying aspects and sentiment
- **Some other difficult sentences**
- Coreference resolution
- Comparative opinions
- Opinion spam detection
- Summary

# Sentiment based on clauses?

- “Trying out Chrome because Firefox keeps crashing.”
  - Firefox - negative; no opinion about chrome.
  - We need to segment the sentence into clauses to decide that “crashing” only applies to Firefox(?).
- But how about these
  - “I changed to Audi because BMW is so expensive.”
  - “I did not buy BWM because of the high price.”
  - “I am so happy that my iPhone is nothing like my old ugly Droid.”



# Indirect opinion sentences

- See these two sentences in a medical domain:
  - “I come to see my doctor because of severe stomach pain”
  - “After taking the drug, I got severe stomach pain”
- If we want opinions about a drug or doctor, the first sentence has no opinion, but the second implies an negative opinion about the drug.
  - Some understanding is needed

# Conditions and questions

- Conditional sentences are hard to deal with (Narayanan et al. 2009)
  - “If I can find a good Sony camera, I will buy it.”
  - But conditional sentences can have opinions
    - “If you are looking for a good phone, buy Nokia”
- Questions are also hard to handle
  - “Are there any great perks for employees?”
  - “Any idea how to fix this lousy Sony camera?”

# Sarcasm (Tsur, et al. 2010; Riloff *et al.* 2013)

- Sarcastic sentences, etc
  - “*What a great car, it stopped working the 2nd day*”
  - “*Can he read?*” (?)
  - “*[I] Love The Cover*” (book)
  - “*Where am I?*” (GPS device)
  - “*Be sure to save your purchase receipt*” (smart phone)
  - “*Are these iPods designed to die after two years?*” (music player)
  - “*Great for insomniacs*” (book)
- The meaning can be highly dependent on context.

# Discourse analysis

- Intra-sentence and inter-sentence
  - “When I first got the airbed a couple of weeks ago it was wonderful as all new things are, however as the weeks progressed I liked it less and less.”
  - “I'm not tryna be funny, but I'm scared for this country. Romney is winning.”

# Intention sentences (Chen and Liu 2013)

- Many intention sentences contain sentiment words and may or may not express sentiment
  - “My goal is to get a TV with good picture quality”
  - “I am looking a cheap health insurance for my family.”
  - “I am dying to see *Life of Pi*”
  - “I want to throw this phone out of the window.”

# Roadmap

- Entity set expansion
- Explicit and implicit aspects
- Resource usage aspect and sentiment
- Factual terms implying aspects and sentiment
- Some other difficult sentences
- **Coreference resolution**
- Comparative opinions
- Opinion spam detection
- Summary

# Coreference resolution: semantic level?

- Coreference resolution (Ding and Liu, 2010)
  - “This Sharp tv’s picture quality is so bad. Our old Sony tv is much better. *It is also very expensive.*”
    - “it” means “Sharp”
  - “This Sharp tv’s picture quality is so bad. Our old Sony tv is much better. *It is also very reliable.*”
    - “it” means “Sony”
- Sentiment consistency.

# Coreference resolution (contd)

- “This **phone**’s **sound** is great. **It** is cheap too.”
  - “It” means “This phone”
- “This **phone**’s **sound** is great. **It** is crystal clear.”
  - “it” means “sound”
- For coreference resolution, we need to
  - do sentiment analysis first, and
  - mine adjective-noun associations



# This is hard!

- *“This past Saturday, I bought a **Nokia** phone and my girlfriend bought a **Motorola** phone with **Bluetooth**. We called each other when we got home. **The voice on my phone was not so clear, worse than my previous Samsung phone.** **The battery life was short too.** **My girlfriend was quite happy with her phone.** **I wanted a phone with good sound quality.** **So my purchase was a real disappointment.** **I returned the phone yesterday.”***

# Roadmap

- Entity set expansion
- Explicit and implicit aspects
- Resource usage aspect and sentiment
- Factual terms implying aspects and sentiment
- Some other difficult sentences
- Coreference resolution
- **Comparative opinions**
- Opinion spam detection
- Summary

# Comparative opinions

(Jindal and Liu, 2006)

## ■ *Gradable*

- *Non-Equal Gradable*: Relations of the type *greater or less than*
  - *“The sound of phone A is better than that of phone B”*
- *Equative*: Relations of the type *equal to*
  - *“Camera A and camera B both come in 7MP”*
- *Superlative*: Relations of the type *greater or less than all others*
  - *“Camera A is the cheapest in market”*

# Analyzing comparative opinions

- **Objective:** Given an opinionated document  $d$ ,  
**Extract comparative opinions:**  
 $(E_1, E_2, A, po, h, t)$ ,  
 $E_1$  and  $E_2$ : entity sets being compared  
 $A$ : their shared aspects - the comparison is based on  
 $po$ : preferred entity set  
 $h$ : opinion holder  
 $t$ : time when the comparative opinion is posted.
- *“Canon’s optics is better than those of Sony & Nikon.”*
  - $(\{\text{Canon}\}, \{\text{Sony, Nikon}\}, \{\text{optics}\}, \text{preferred:}\{\text{Canon}\}, ?, ?)$

# Analysis of comparative opinions

(Jindal and Liu, 2006, Ding et al, 2009)

- Gradable comparative sentences can be dealt with *almost* as normal opinion sentences.
  - E.g., “*optics of camera A is better than that of camera B*”
  - **Positive:** (camera A, *optics*)
  - **Negative:** (*camera B, optics*)
- **Difficulty:** recognize non-standard comparatives
  - E.g., “*I am so happy because my new iPhone is *nothing like* my old slow ugly Droid.*”

# Roadmap

- Entity set expansion
- Explicit and implicit aspects
- Resource usage aspect and sentiment
- Factual terms implying aspects and sentiment
- Some other difficult sentences
- Coreference resolution
- Comparative opinions
- **Opinion spam detection**
- Summary

# Opinion spamming

(Jindal and Liu, 2007, 2008)

- **Opinion spamming** refers to people giving **fake or deceptive** reviews/opinions, e.g.,
  - Write undeserving positive reviews for some target entities in order to promote them.
  - Write unfair or malicious negative reviews for some target entities in order to damage their reputations.
- **Motivation:** positive opinions mean profits and fame for businesses and individuals
- Writing fake reviews has become a business
  - e.g., reputation management firms

---

# Is this review fake or not?

I want to make this review in order to comment on the excellent service that my mother and I received on the Serenade of the Seas, a cruise line for Royal Caribbean. There was a lot of things to do in the morning and afternoon portion for the 7 days that we were on the ship. We went to 6 different islands and saw some amazing sites! It was definitely worth the effort of planning beforehand. The dinner service was 5 star for sure. One of our main waiters, Muhammad was one of the nicest people I have ever met. However, I am not one for clubbing, drinking, or gambling, so the nights were pretty slow for me because there was not much else to do. Either than that, I recommend the Serenade to anyone who is looking for excellent service, excellent food, and a week full of amazing day-activities!



---

# What about this?

The restaurant is located inside of a hotel, but do not let that keep you from going! The main chef, Chef Chad, is absolutely amazing! The other waiters and waitresses are very nice and treat their guests very respectfully with their service (i.e. napkins to match the clothing colors you are wearing). We went to Aria twice in one weekend because the food was so fantastic. There are so many wonderful Asian flavors. From the plating of the food, to the unique food options, to the fresh and amazing nan bread and the tandoori oven that you can watch as the food is being cooked, all is spectacular. The atmosphere and the space are great as well. I just wished we lived closer and could dine there more frequently because it is quite expensive.

---

# State-of-the-art techniques

- Supervised learning (Jindal & Liu, 2008; Li et al., 2011; Ott et al., 2011; Feng et al 2012; Mukherjee & Liu 2013).
- Abnormal patterns (Jindal, Liu & Lim 2010)
- Graph-based methods (Wang et al, 2013; Akoglu et al 2013; Li et al. 2014)
- Using review bursts for detection (Xie et al 2013; Fei et al 2013)
- Identifying reviewers with multiple userids (Qian and Liu 2013)
- Finding spammer groups (Mukherjee et al. 2012)

# Detecting group spam (Mukherjee et al WWW-2012)

- A group of people (could be a single person with multiple ids) work together to promote a product or to demote a product.
- Such spam can be very damaging as
  - they can take total control of sentiment on a product
- The algorithm has three steps
  - Frequent pattern mining: find groups of people who reviewed a number of products together.
  - A set of feature indicators are identified
  - Ranking is performed using a relational model

# Roadmap

- Entity set expansion
- Explicit and implicit aspects
- Resource usage aspect and sentiment
- Factual terms implying aspects and sentiment
- Some other difficult sentences
- Coreference resolution
- Comparative opinions
- Opinion spam detection
- **Summary**

# Summary

- Sentiment analysis is very challenging.
  - Many inter-related sub-problems, many of which have hardly been attempted by researchers.
  - To see the complexity of the problem, we need to work in a large number of domains
- Breakthrough techniques are needed to go to the next level.
- Solving the general NLP is hopeless, but can we solve this highly focused problem of SA?
  - I am optimistic

# Books

- ❑ *(New)* B. Liu. *Sentiment Analysis: Mining Opinions, Sentiments and Emotions*. Cambridge University Press, March 2015.
- ❑ B. Liu. *Sentiment Analysis and Opinion Mining*. Morgan and Claypool publishers. May, 2012.

