

This tutorial has been given at AAAI-2011, EACL-2012, and Sentiment Analysis Symposium.

# Sentiment Analysis and Opinion Mining

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

## Introduction

- **Sentiment analysis** or **opinion mining**
  - Computational study of opinions, sentiments, evaluations, attitudes, appraisal, affects, views, emotions, subjectivity, etc., expressed in text.
    - Reviews, blogs, discussions, news, comments, feedback, or any other documents.
- **Terminology:**
  - **Sentiment analysis** is more widely used in industry.
  - Both are widely used in academia
- **But they can be used interchangeably.**

## Why are opinions important?

- **“Opinions” are key influencers of behaviors.**
- Our beliefs and perceptions of reality are largely conditioned on how others see the world.
- Whenever we need to **make a decision**, we often seek out the opinions of others. Past:
  - **Individuals**: ask opinions from friends and family
  - **Organizations**: use surveys, focus groups, opinion polls, consultants.

## Introduction – social media + beyond

- **Word-of-mouth on the Web**
  - Personal experiences and opinions about anything in reviews, forums, blogs, Twitter, micro-blogs, etc
  - Comments about articles, issues, topics, reviews, etc.
  - Postings at social networking sites, e.g., facebook.
- **Global scale**: No longer – one’s circle of friends
- **Organization internal data**
  - Customer feedback from emails, call centers, etc.
- **News and reports**
  - Opinions in news articles and commentaries

## Introduction – applications

### ■ Businesses and organizations

- Benchmark products and services; market intelligence.
  - Businesses spend a huge amount of money to find consumer opinions using consultants, surveys and focus groups, etc

### ■ Individuals

- Make decisions to buy products or to use services
- Find public opinions about political candidates and issues

### ■ Ads placements: Place ads in the social media content

- Place an ad if one praises a product.
- Place an ad from a competitor if one criticizes a product.

### ■ Opinion retrieval: provide general search for opinions.

## A fascinating problem!

### ■ Intellectually challenging & many applications.

- A popular research topic in NLP, text mining, and Web mining in recent years (Shanahan, Qu, and Wiebe, 2006 (edited book); Surveys - Pang and Lee 2008; Liu, 2010; 2011)
- It has spread from computer science to management science (Hu, Pavlou & Zhang, 2006; Archak, Ghose & Ipeirotis, 2007; Liu et al 2007; Park, Lee & Han, 2007; Dellarocas, Zhang & Awad, 2007; Chen & Xie 2007).
- 40-60 companies in USA alone
- It touches every aspect of NLP and yet is confined.
  - Little research in NLP/Linguistics in the past.
- Potentially a major technology from NLP.
  - But it is hard.

## A large problem space

- **Many names and tasks** with somewhat different objectives and models
  - Sentiment analysis
  - Opinion mining
  - Sentiment mining
  - Subjectivity analysis
  - Affect analysis
  - Emotion detection
  - Opinion spam detection
  - *Etc.*

Bing Liu, tutorial

7

## About this tutorial

- **Like a traditional tutorial**, I will introduce the research in the field.
  - Key topics, main ideas and approaches
  - Since there are a large number of papers, it is not possible to introduce them all, but a comprehensive reference list will be provided.
- **Unlike many traditional tutorials**, this tutorial is also based on my experience in working with clients in a startup, and in my consulting
  - **I focus more on practically important tasks** (IMHO)

Bing Liu, tutorial

8

## Roadmap

- ➔ ■ **Sentiment Analysis Problem**
  - Document sentiment classification
  - Sentence subjectivity & sentiment classification
  - Aspect-based sentiment analysis
  - Aspect-based opinion summarization
  - Sentiment lexicon generation
  - Mining comparative opinions
  - Some other problems
  - Opinion spam detection
  - Utility or helpfulness of reviews
  - Summary

Bing Liu, tutorial

9

## Structure the unstructured (Hu and Liu 2004)

- **Structure the unstructured**: Natural language text is often regarded as **unstructured data**.
- The problem definition should provide a structure to the unstructured problem.
  - **Key tasks**: Identify key tasks and their inter-relationships.
  - **Common framework**: Provide a common framework to unify different research directions.
  - **Understanding**: help us understand the problem better.

Bing Liu, tutorial

10

## Problem Statement: Abstraction

- It consists of two issues

- (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

- 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 opinion summarization.

## Abstraction (1): what is an opinion?

- **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 mother* was mad with me as I did not tell her before I bought the *phone*. She also thought the phone was too *expensive*, ...”

- One can look at this review/blog at the

- document level, i.e., is this review + or -?
- sentence level, i.e., is each sentence + or -?
- entity and feature/aspect level

## Entity and aspect/feature level

- **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 mother* was mad with me as I did not tell her before I bought the *phone*. She also thought the phone was too *expensive*, ...”
- **What do we see?**
  - **Opinion targets:** entities and their features/aspects
  - **Sentiments:** positive and negative
  - **Opinion holders:** persons who hold the opinions
  - **Time:** when opinions are expressed

## 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.”
- **We focus on regular opinions first, and just call them opinions.**

## A (regular) opinion

- *An opinion has the follow basic components*

$(g_j, so_{ijkl}, h_i, t_l)$ ,

where

- $g_j$  is a target
- $so_{ijl}$  is the sentiment value of the opinion from opinion holder  $h_i$  on target  $g_j$  at time  $t_l$ .  $so_{ijl}$  is positive, negative or neutral, or a rating score
- $h_i$  is an opinion holder.
- $t_l$  is the time when the opinion is expressed.

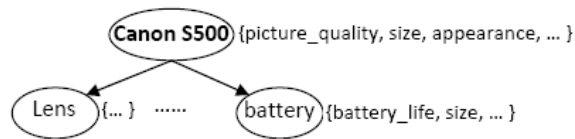
## Opinion target

- In some cases, **opinion target** is a single **entity** or **topic**.
  - “I love **iPhone**” and “I support **tax cut**.”
- But in many other cases, it is more complex.
  - “I bought an **iPhone** a few days ago. It is such a nice phone. **The touch screen is really cool.**”
    - Opinion target of the 3<sup>rd</sup> sentence is not just **touch screen**, but the “**touch screen of iPhone**”.
  - “I support **tax cut** for **the middle class**, but for **the rich**”
- **We decompose the opinion target**



## Entity and aspect (Hu and Liu, 2004; Liu, 2006)

- **Definition (entity):** An *entity*  $e$  is a product, person, event, organization, or topic.  $e$  is represented as
  - a hierarchy of *components*, *sub-components*, and so on.
  - Each node represents a component and is associated with a set of *attributes* of the component.



- An opinion can be expressed on any node or attribute of the node.
- For simplicity, we use the term *aspects (features)* to represent both components and attributes.

Bing Liu, tutorial

17

## Opinion definition (Liu, Ch. in NLP handbook, 2010)

- *An opinion is a quintuple*

$$(e_j, a_{jk}, so_{ijkl}, h_i, t_l),$$

where

- $e_j$  is a target entity.
- $a_{jk}$  is an aspect/feature of the entity  $e_j$ .
- $so_{ijkl}$  is the sentiment value of the opinion from the opinion holder  $h_i$  on aspect  $a_{jk}$  of entity  $e_j$  at time  $t_l$ .  $so_{ijkl}$  is +ve, -ve, or neu, or a more granular rating.
- $h_i$  is an opinion holder.
- $t_l$  is the time when the opinion is expressed.

Bing Liu, tutorial

18

## Some remarks about the definition

- Although introduced using a product review, the definition is generic
  - Applicable to other domains,
  - E.g., politics, social events, services, topics, etc.
- $(e_j, a_{jk})$  is also called the opinion target
  - Opinion without knowing the target is of limited use.
- The five components in  $(e_j, a_{jk}, so_{ijkl}, h_i, t_l)$  must correspond to one another. Very hard to achieve
- The five components are essential. Without any of them, it can be problematic in general.

## Some remarks (contd)

- Of course, one can add any number of other components to the tuple for more analysis. E.g.,
  - Gender, age, Web site, post-id, etc.
- The original definition of an entity is a hierarchy of parts, sub-parts, and so on.
  - The simplification can result in information loss.
    - E.g., “The ink of this printer is very expensive.”
    - “ink” is a part of the printer and “price” (expensive) is an aspect of “ink” (not the printer).
  - But it is usually sufficient for practical applications.
    - It is too hard without the simplification.

## “Confusing” terminologies

- **Entity** is also called **object**.
- **Aspect** is also called **feature**, **attribute**, **facet**, etc
- **Opinion holder** is also called **opinion source**
- Some researchers also use **topic** to mean **entity** and/or **aspect**.
  - Separating entity and aspect is preferable
- In specific applications, some specialized terms are also commonly used, e.g.,
  - Product features, political issues

## Reader's standing point

- See this sentence
  - “I am so happy that Google price shot up today.”
- Although the sentence gives an explicit sentiment, different readers may feel very differently.
  - If a reader sold his Google shares yesterday, he will not be that happy.
  - If a reader bought a lot of Google shares yesterday, he will be very happy.
- Current research either implicitly assumes a standing point, or ignores the issue.

## 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 mother* 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.

## Structure the unstructured

- **Goal:** Given an opinionated document,
  - Discover all quintuples  $(e_j, a_{jk}, so_{ijkl}, h_i, t_l)$ ,
  - Or, solve some simpler forms of the problem
    - E.g., sentiment classification at the document or sentence level.
- **With the quintuples,**
  - **Unstructured Text → Structured Data**
    - Traditional data and visualization tools can be used to slice, dice and visualize the results.
    - Enable qualitative and quantitative analysis.

## Two closely related concepts

- **Subjectivity** and **emotion**.
- **Sentence subjectivity**: An *objective sentence* presents some factual information, while a *subjective sentence* expresses some personal feelings, views, emotions, or beliefs.
- **Emotion**: Emotions are people's subjective feelings and thoughts.

## Subjectivity

- **Subjective expressions come in many forms, e.g., opinions, allegations, desires, beliefs, suspicions, and speculations** (Wiebe, 2000; Riloff et al 2005).
  - A subjective sentence may contain a positive or negative opinion
- **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 of people among researchers.
- Based on (Parrott, 2001), people have six main emotions,
  - love, joy, surprise, anger, sadness, and fear.
- Strengths of opinions/sentiments are related to certain emotions, e.g., joy, anger.
  - However, the concepts of emotions and opinions are not equivalent.

Bing Liu, tutorial

27

## Rational and emotional evaluations

- Rational evaluation: Many evaluation/opinion sentences express no emotion
  - e.g., “The voice of this phone is clear”
- Emotional evaluation
  - e.g., “I love this phone”
  - “The voice of this phone is crystal clear” (?)
- Some emotion sentences express no (positive or negative) opinion/sentiment
  - e.g., “I am so surprised to see you”.

Bing Liu, tutorial

28

## Sentiment, subjectivity, and emotion

- Although they are clearly related, these concepts are not the same
  - $\text{Sentiment} \neq \text{subjective} \neq \text{emotion}$
- Sentiment is not a subset of subjectivity (without implied sentiments by facts, it should be)
  - $\text{sentiment} \not\subset \text{subjectivity}$
- The following should hold
  - $\text{emotion} \subset \text{subjectivity}$
  - $\text{sentiment} \not\subset \text{emotion}, \dots$

## Abstraction (2): opinion summary

- **With a lot of opinions, a summary is necessary.**
  - A multi-document summarization task
- For factual texts, summarization is to select the most important facts and present them in a sensible order while avoiding repetition
  - $1 \text{ fact} = \text{any number of the same fact}$
- But for opinion documents, it is different because opinions have a quantitative side & have targets
  - $1 \text{ opinion} \neq \text{a number of opinions}$
  - **Aspect-based summary** is more suitable
    - Quintuples form the basis for opinion summarization

# 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

...

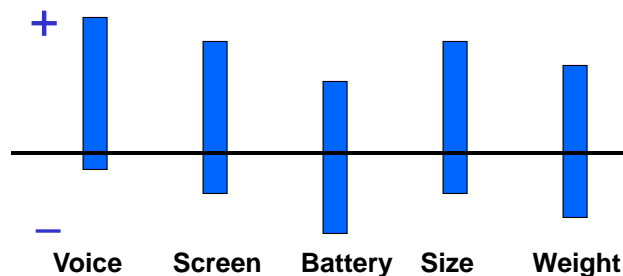
Note: We omit opinion holders

.... Bing Liu, tutorial

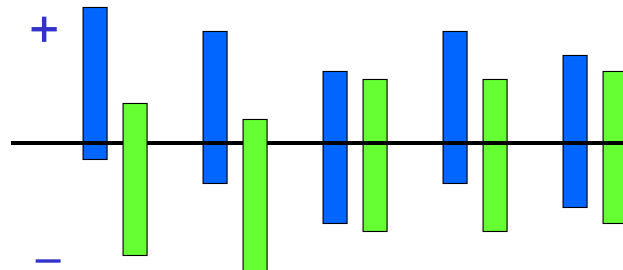
31

## Opinion Observer (Liu et al. 2005)

- Summary of reviews of
- Cell Phone 1



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



Bing Liu, tutorial

32



## Aspect-based opinion summary

The screenshot shows a Bing Shopping result for an HP LaserJet 1020 printer. On the left, a sidebar lists 'POPULAR FEATURES' with horizontal bar charts for Affordability, Speed, Print Quality, Reliability, Ease Of Use, Brand, Installation, Size, and Compatibility. The main content area shows the product name, a small image of the printer, and pricing information: 'from \$179 (2 stores)' and 'Bing cashback : 3%'. It also displays 'user reviews (177)' with a star rating. Below this, there are tabs for 'user reviews', 'product details', 'expert reviews', and 'compare prices'. The 'user reviews' tab is active, showing a 'view: positive comments (44)' and a 'speed' bar chart at 96%. Several user review snippets are visible, such as 'The quality is as good as any laserjet printer I've used and the speed is fast.' and 'Quick and fast transaction.'

Bing Liu, tutorial

33

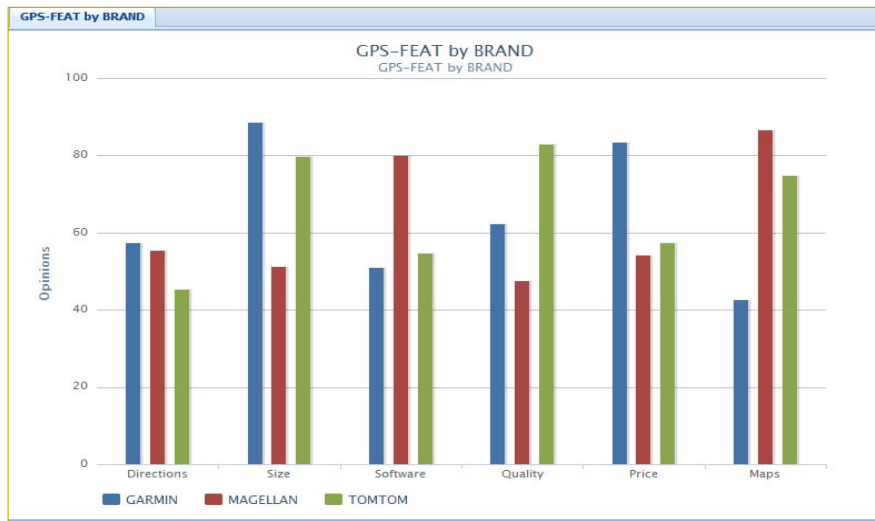
## Google Product Search (Blair-Goldensohn et al 2008 ?)

The screenshot shows Google Products search results for a Sony Cyber-shot DSC-W370 14.1 MP Digital Camera (Silver). The top section includes the Google Products logo, the search term 'sony camera', and a 'Search Products' button. Below this, the product name is displayed, followed by links for 'Overview', 'Online stores', 'Nearby stores', 'Reviews', 'Technical specifications', 'Similar items', and 'Accessories'. A product image is shown next to the price '\$140 online, \$170 nearby' and a star rating with '159 reviews'. A 'Reviews' section follows, showing a 'Summary - Based on 159 reviews' with a star distribution bar (1 star: 1, 2 stars: 2, 3 stars: 3, 4 stars: 4, 5 stars: 5). Below the summary, a section titled 'What people are saying' lists various aspects of the product with corresponding star ratings and user comments. The aspects listed are pictures, features, zoom/lens, design, video, battery life, and screen.

Bing Liu, tutorial

34

## Some examples from OpinionEQ

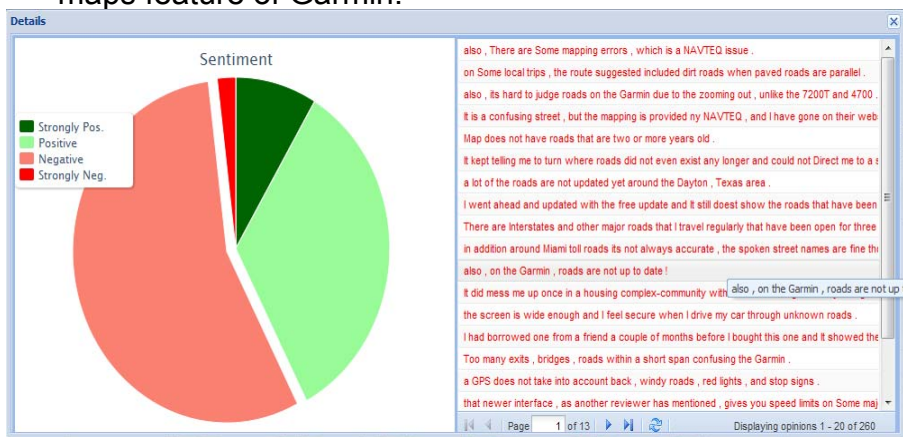


Bing Liu, tutorial

35

## Detail opinion sentences

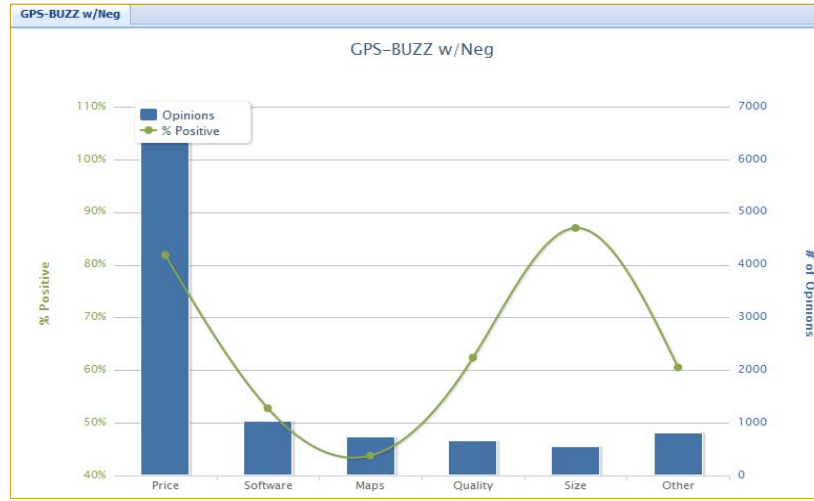
- Click on any bar (previous slide) to see the opinion sentences. Here are negative opinion sentences on the maps feature of Garmin.



Bing Liu, tutorial

36

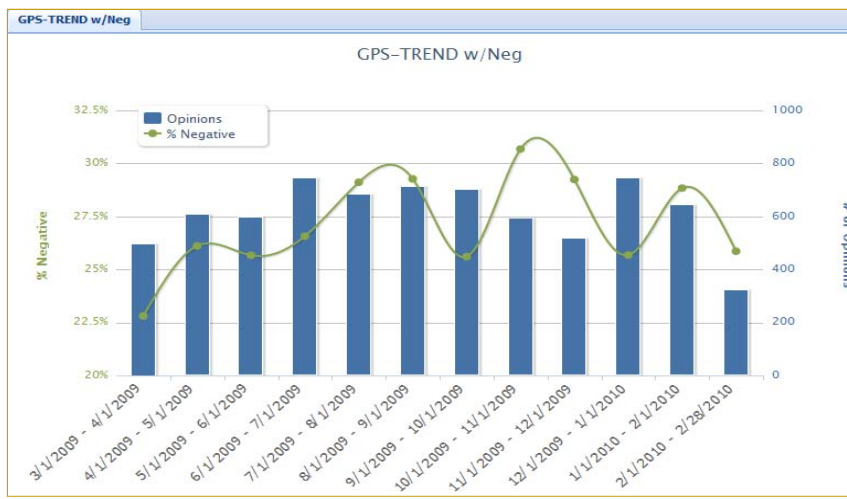
## % of +ve opinion and # of opinions



Bing Liu, tutorial

37

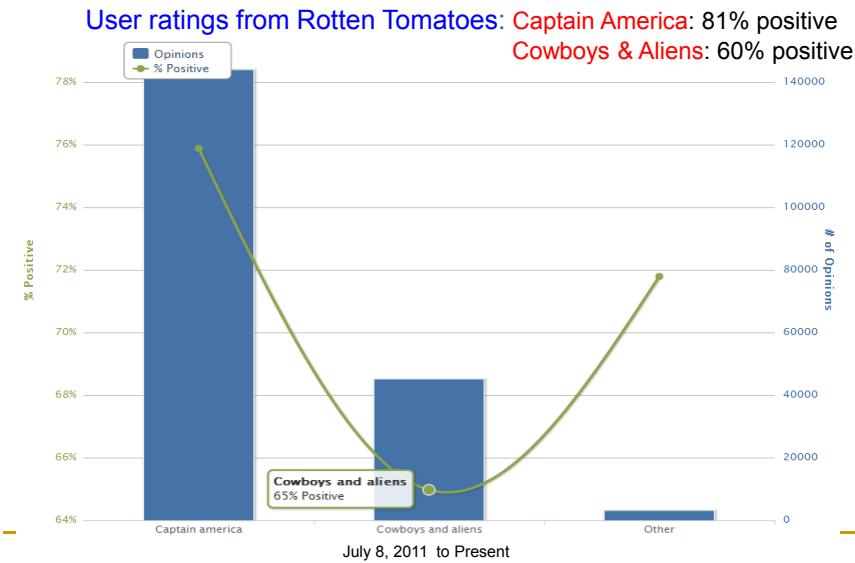
## Aggregate opinion trend



Bing Liu, tutorial

38

## Live tracking of two movies (Twitter)



Bing Liu, tutorial

39

## Not just ONE problem

- $(e_j, a_{jk}, so_{ijkl}, h_i, t_l)$ 
  - $e_j$  - a target entity: Named Entity Extraction (more)
  - $a_{jk}$  - an aspect of  $e_j$ : Information Extraction
  - $so_{ijkl}$  is sentiment: Sentiment Identification
  - $h_i$  is an opinion holder: Information/Data Extraction
  - $t_l$  is the time: Information/Data Extraction
  - 5 pieces of information must match
- Coreference resolution
- Synonym match (voice = sound quality)
- ...

Bing Liu, tutorial

40

## Opinion mining 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.”***

Bing Liu, tutorial

41

## Easier and harder problems

- Tweets from Twitter are probably the easiest
  - short and thus usually straight to the point
- Reviews are next
  - entities are given (almost) and there is little noise
- Discussions, comments, and blogs are hard.
  - Multiple entities, comparisons, noisy, sarcasm, etc
- **Determining sentiments seems to be easier.**
- **Extracting entities and aspects is harder.**
- **Combining them is even harder.**

Bing Liu, tutorial

42

## Opinion mining in the real world

- Source the data, e.g., reviews, blogs, etc
  - (1) Crawl all data, store and search them, or
  - (2) Crawl only the target data
- Extract the right entities & aspects
  - Group entity and aspect expressions,
    - Moto = Motorola, photo = picture, etc ...
- Aspect-based opinion mining (sentiment analysis)
  - Discover all quintuples  
(Store the quintuples in a database)
- Aspect based opinion summary

Bing Liu, tutorial

43

## Roadmap

- Sentiment Analysis Problem
- ➔ ■ **Document sentiment classification**
- Sentence subjectivity & sentiment classification
- Aspect-based sentiment analysis
- Aspect-based opinion summarization
- Sentiment lexicon generation
- Mining comparative opinions
- Some other problems
- Opinion spam detection
- Utility or helpfulness of reviews
- Summary

Bing Liu, tutorial

44

## Sentiment classification

- **Classify a whole opinion document** (e.g., a review) based on the overall sentiment of the opinion holder (Pang et al 2002; Turney 2002)
  - **Classes:** Positive, negative (possibly neutral)
  - Neutral or no opinion is hard. Most papers ignore it.
- **An example review:**
  - *"I bought an iPhone a few days ago. It is such a nice phone, although a little large. The touch screen is cool. The voice quality is clear too. I simply love it!"*
  - **Classification:** positive or negative?
- **Perhaps the most widely studied problem.**

Bing Liu, tutorial

45

## A text classification task

- **It is basically a text classification problem**
- **But different from topic-based text classification.**
  - In topic-based text classification (e.g., computer, sport, science), topic words are important.
  - But in sentiment classification, opinion/sentiment words are more important, e.g., great, excellent, horrible, bad, worst, etc.
- **Opinion/sentiment words**
  - Words and phrases that express desired or undesired states or qualities.

Bing Liu, tutorial

46

## Assumption and goal

- **Assumption:** The doc is written by a single person and express opinion/sentiment on a single entity.
- **Goal:** discover (\_\_, \_\_, so, \_\_, \_\_), where e, a, h, and t are ignored
- **Reviews usually satisfy the assumption.**
  - Almost all papers use reviews
  - Positive: 4 or 5 stars, negative: 1 or 2 stars
- **Many forum postings and blogs do not**
  - They can mention and compare multiple entities
  - Many such postings express no sentiments

## Some Amazon reviews

248 of 263 people found the following review helpful:

★★★★★ **This is one to get if you want 5MP**, April 14, 2004

By [Gadgester "No Time, No Money"](#) (Mother Earth) - [See all my reviews](#)

TOP 100 REVIEWER

Amazon Verified Purchase (What's this?)

This review is from: Canon PowerShot S500 5MP Digital Elph with 3x Optical Zoom (Electronics)

The new Canon PowerShot S500 is a 5MP upgrade to the immensely popular S400 model, which was a 4MP digital camera. The S500 produces excellent images, is easy to use, and is compact enough to carry in a pocket. 3X optical zoom is standard on these cameras. Besides shooting still photos, you can record low-res video clips as well as audio clips, but don't expect high quality on either.

For a hundred bux less, you can get the 4MP S410 model which is otherwise identical to the S500. Should you go for this or the S410? I think for most consumers 4MP is plenty enough, with room for cropping and enlargements. 5MP is only necessary if you really crop a lot \*and\* plan to blow up the cropped images. The S410 strikes a great balance between pixel count and price -- it's a better value.

Help other customers find the most helpful reviews

Was this review helpful to you? ☐ Yes ☐ No

[Report abuse](#) | [Permalink](#)

[Comment](#)

41 of 41 people found the following review helpful:

★★★★★ **E18 Error / problem with the lens**, September 29, 2004

By [Johnathan Parker](#) (Springdale, AR USA) - [See all my reviews](#)

REAL NAME

This review is from: Canon PowerShot S500 5MP Digital Elph with 3x Optical Zoom (Electronics)

This is my second Canon digital elph camera. Both were great cameras. Recently upgraded to the S500. About 6 months later I get the dreaded E18 error. I searched the Internet and found numerous people having problems. When I determined the problem to be the lens not fully extending I decided to give it a tug. It clicked and the camera came on,



## Unsupervised classification

(Turney, 2002)

- Data: reviews from epinions.com on automobiles, banks, movies, and travel destinations.
- The approach: Three steps
- Step 1:
  - Part-of-speech (POS) tagging
  - Extracting two consecutive words (**two-word phrases**) from reviews if their tags conform to some given patterns, e.g., (1) JJ, (2) NN.

Bing Liu, tutorial

49

## Patterns of POS tags

	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

Bing Liu, tutorial

50

- Step 2: Estimate the sentiment orientation (SO) of the extracted phrases

- Use Pointwise mutual information

$$PMI(word_1, word_2) = \log_2 \left( \frac{P(word_1 \wedge word_2)}{P(word_1)P(word_2)} \right)$$

- Semantic orientation (SO):

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

- Using AltaVista near operator to do search to find the number of hits to compute PMI and SO.

- Step 3: Compute the average SO of all phrases

- classify the review as **positive** if average SO is positive, **negative** otherwise.

- Final classification accuracy:

- automobiles - 84%
- banks - 80%
- movies - 65.83
- travel destinations - 70.53%

## Supervised learning (Pang et al, 2002)

- **Directly apply supervised learning techniques** to classify reviews into positive and negative.
  - Like a text classification problem
- **Three classification techniques** were tried:
  - Naïve Bayes
  - Maximum entropy
  - Support vector machines
- **Pre-processing:**
  - **Features:** negation tag, unigram (single words), bigram, POS tag, position.

## Supervised learning

- **Training and test data**
  - Movie reviews with star ratings
    - 4-5 stars as **positive**
    - 1-2 stars as **negative**
- **Neutral is ignored.**
- **SVM** gives the best classification accuracy based on balance training data
  - 83%
  - **Features:** unigrams (bag of individual words)

## Features for supervised learning

- The problem has been studied by numerous researchers subsequently
  - Probably the most extensive studied problem
    - Including domain adaption and cross-lingual, etc.
- **Key:** feature engineering. A large set of features have been tried by researchers. E.g.,
  - Terms frequency and different IR weighting schemes
  - Part of speech (POS) tags
  - Opinion words and phrases
  - Negations
  - Syntactic dependency

Bing Liu, tutorial

55

## A large number of related papers

- Bickerstaffe and Zukerman (2010) used a hierarchical multi-classifier considering inter-class similarity
- Burfoot, Bird and Baldwin (2011) sentiment-classified congressional floor debates
- Cui et al. (2006) evaluated some sentiment classification algorithms
- Das and Chen (2001) extracted market sentiment from stock message boards
- Dasgupta and Ng (2009) used semi-supervised learning
- Dave, Lawrence & Pennock (2003) designed a custom function for classification
- Gamon (2004) classified customer feedback data

Bing Liu, tutorial

56

## A large number of related papers

- Goldberg and Zhu (2006) used semi-supervised learning.
- Kim, Li and Lee (2009) and Paltoglou and Thelwall (2010) studied different IR term weighting schemes
- Li et al (2010) made use of different polarity shifting.
- Li, Huang, Zhou and Lee (2010) used personal (I, we) and impersonal (they, it, this product) sentences to help
- Maas et al (2011) used word vectors which are latent aspects of the words.
- Mullen and Collier (2004) used PMI, syntactic relations and other attributes with SVM.
- Nakagawa, Inui and Kurohashi (2010) used dependency relations and CRF.

Bing Liu, tutorial

57

## A large number of related papers

- Ng, Dasgupta and Arifin (2006) identified reviews and classified sentiments of reviews
- Pang and Lee (2004) used minimum cuts
- Qiu, Zhang, Hu and Zhao (2009) proposed a lexicon-based and self-supervision approach
- Tong (2001) used a set of domain specific phrases
- Yessenalina, Choi and Cardie (2010) automatically generated annotator rationales to help classification
- Yessenalina, Yue and Cardie (2010) found subjective sentences and then used them for model building
- Zhou, Chen and Wang (2010) used semi-supervised and active learning

Bing Liu, tutorial

58

## Review rating prediction

- Apart from classification of positive or negative sentiments,
  - research has also been done to **predict the rating scores** (e.g., 1–5 stars) of reviews (Pang and Lee, 2005; Liu and Seneff 2009; Qu, Ifrim and Weikum 2010; Long, Zhang and Zhu, 2010).
  - Training and testing are reviews with star ratings.
- **Formulation**: The problem is formulated as regression since the rating scores are ordinal.
- Again, feature engineering and model building.

## Domain adaptation (transfer learning)

- **Sentiment classification is sensitive to the domain of the training data.**
  - A classifier trained using reviews from one domain often performs poorly in another domain.
    - words and even language constructs used in different domains for expressing opinions can be quite different.
    - same word in one domain may mean positive but negative in another, e.g., “*this vacuum cleaner really **sucks**.*”
- Existing research has used labeled data from one domain and unlabeled data from the target domain and general opinion words for learning (Aue and Gamon 2005; Blitzer et al 2007; Yang et al 2006; Pan et al 2010; Wu, Tan and Cheng 2009; Bollegala, Weir and Carroll 2011; He, Lin and Alani 2011).

## Cross-lingual sentiment classification

- Useful in the following scenarios:
  - E.g., there are many English sentiment corpora, but for other languages (e.g. Chinese), the annotated sentiment corpora may be limited.
  - Utilizing English corpora for Chinese sentiment classification can relieve the labeling burden.
- Main approach: use available language corpora to train sentiment classifiers for the target language data. Machine translation is typically employed
  - (Banea et al 2008; Wan 2009; Wei and Pal 2010; Kim et al. 2010; Guo et al 2010; Mihalcea & Wiebe 2010; Boyd-Graber and Resnik 2010; Banea et al 2010; Duh, Fujino & Nagata 2011; Lu et al 2011)

Bing Liu, tutorial

61

## Roadmap

- Sentiment Analysis Problem
- Document sentiment classification
- ➔ ■ **Sentence subjectivity & sentiment classification**
- Aspect-based sentiment analysis
- Aspect-based opinion summarization
- Sentiment lexicon generation
- Mining comparative opinions
- Some other problems
- Opinion spam detection
- Utility or helpfulness of reviews
- Summary

Bing Liu, tutorial

62

## Subjectivity classification

- Document-level sentiment classification is too coarse for most applications.
- We now move to the sentence level.
- Much of the early work on sentence level analysis focuses on identifying **subjective sentences**.
- **Subjectivity classification**: classify a sentence into one of the **two classes** (Wiebe et al 1999)
  - Objective and subjective.
- Most techniques use supervised learning.
  - E.g., a naïve Bayesian classifier (Wiebe et al. 1999).

Bing Liu, tutorial

63

## Sentence sentiment analysis

- **Usually consist of two steps**
  - **Subjectivity classification**
    - To identify subjective sentences
  - **Sentiment classification** of subjective sentences
    - Into two classes, positive and negative
- **But bear in mind**
  - Many objective sentences can imply sentiments
  - Many subjective sentences do not express positive or negative sentiments/opinions
    - E.g., "I believe he went home yesterday."

Bing Liu, tutorial

64



## As an intermediate step

- We do not use the quintuple  $(e, a, so, h, t)$  to define the problem here because
  - sentence classification is an intermediate step.
- Knowing that some sentences have positive or negative opinions are not sufficient.
- However, it helps
  - filter out sentences with no opinions (mostly)
  - determine (to some extent) if sentiments about entities and their aspects are positive or negative.
    - But not enough

## Assumption

- **Assumption:** Each sentence is written by a single person and expresses a single positive or negative opinion/sentiment.
- True for simple sentences, e.g.,
  - “I like this car”
- But not true for compound and “complex” sentences, e.g.,
  - “I like the picture quality but battery life sucks.”
  - “Apple is doing very well in this lousy economy.”

## Subjectivity classification using patterns

(Riloff and Wiebe, 2003)

- **A bootstrapping approach.**
  - A high precision classifier is first used to automatically identify some subjective and objective sentences.
    - Two high precision (but low recall) classifiers are used,
      - a high precision subjective classifier
      - A high precision objective classifier
      - Based on manually collected lexical items, single words and n-grams, which are good subjective clues.
  - A set of patterns are then learned from these identified subjective and objective sentences.
    - Syntactic templates are provided to restrict the kinds of patterns to be discovered, e.g., <subj> passive-verb.
  - The learned patterns are then used to extract more subject and objective sentences (the process can be repeated).

Bing Liu, tutorial

67

## Subjectivity and sentiment classification

(Yu and Hazivassiloglou, 2003)

- **Subjective sentence identification:** a few methods were tried, e.g.,
  - Sentence similarity.
  - Naïve Bayesian classification.
- **Sentiment classification (positive, negative or neutral)** (also called **polarity**): it uses a similar method to (Turney, 2002), but
  - with more seed words (rather than two) and based on log-likelihood ratio (LLR).
  - For classification of each word, it takes the average of LLR scores of words in the sentence and use cutoffs to decide positive, negative or neutral.

Bing Liu, tutorial

68

## Segmentation and classification

- Since a single sentence may contain multiple opinions and subjective and factual clauses
- A study of automatic clause sentiment classification was presented in (Wilson et al 2004)
  - to classify clauses of every sentence by the *strength* of opinions being expressed in individual clauses, down to four levels
    - *neutral, low, medium, and high*
- Clause-level may not be sufficient
  - “Apple is doing very well in this lousy economy.”

Bing Liu, tutorial

69

## Some other related work

- Abdul-Mageed, Diab and Korayem (2011) carried out subjectivity and sentiment analysis of Arabic sentences
- Alm (2011) analyzed subjectivity research motivations, applications, characteristics, etc
- Barbosa and Feng (2010) and Davidov, Tsur and Rappoport (2010) performed Twitter subjectivity and sentiment classification using many features, hashtags, and smileys
- Eguchi and Lavrendo (2006) studied sentiment sentence retrieval
- Gamon et al. (2005) used semi-supervised learning
- Hassan, Qazvinian, Radev (2010) found attitude sentences
- Kim and Hovy (2004) summed up orientations of opinion words in a sentence (or within some word window).
- Hatzivassiloglou & Wiebe (2000) considered gradable adjectives

Bing Liu, tutorial

70

## Some other related work

- Johansson and Moschitti (2011) extracted opinion expressions and sentiments
- Joshi and Penstein-Rose (2009) used dependency triples with “back-off” using POS tags rather than words
- Kim and Hovy (2006a) automatically identified pro and con reasons
- Kim and Hovy (2006b) Identified judgment opinions
- Kim and Hovy (2007) mined predictive opinions in election postings
- Kim, Li and Lee (2010) compared subjectivity analysis tools
- McDonald et al (2007) performed sentence to document sentiment classification
- Mukund and Srihari (2010) performed subjectivity classification with co-training

Bing Liu, tutorial

71

## Some other related work

- Nasukawa and Yi (2003) captured favorability
- Nigam and Hurst (2005) classified subjective and topic sentences
- Tackstrom & McDonald (2011) performed sentence sentiment classification
- Wiebe et al (2004) learned subjective language
- Wiebe and Riloff (2005) used semi-supervised learning with a initial training set identified by some strong patterns
- Wiebe and Mihalcea (2006) studied word sense and subjectivity
- Wilson, Wiebe and Hwa (2006) recognized strong and weak opinion clauses
- Wilson et al. (2004, 2005) found strength of sentiments/opinions in clauses

Bing Liu, tutorial

72

## Roadmap

- Sentiment Analysis Problem
- Document sentiment classification
- Sentence subjectivity & sentiment classification
- ➔ ■ **Aspect-based sentiment analysis**
- Aspect-based opinion summarization
- Sentiment lexicon generation
- Mining comparative opinions
- Some other problems
- Opinion spam detection
- Utility or helpfulness of reviews
- Summary

Bing Liu, tutorial

73

## We need to go further

- Sentiment classification at both the document and sentence (or clause) levels are useful, **but**
  - They do not find what people liked and disliked.
- **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.

Bing Liu, tutorial

74

## Recall an opinion is a quintuple

- An *opinion* is a quintuple

$$(e_j, a_{jk}, so_{ijkl}, h_i, t_l),$$

where

- $e_j$  is a target entity.
- $a_{jk}$  is an aspect/feature of the entity  $e_j$ .
- $so_{ijkl}$  is the sentiment value of the opinion of the opinion holder  $h_i$  on feature  $a_{jk}$  of entity  $e_j$  at time  $t_l$ .  $so_{ijkl}$  is +ve, -ve, or neu, or a more granular rating.
- $h_i$  is an opinion holder.
- $t_l$  is the time when the opinion is expressed.

## Aspect-based sentiment analysis

- Much of the research is based on online reviews
- For reviews, aspect-based sentiment analysis is easier because the entity (i.e., product name) is usually known
  - Reviewers simply express positive and negative opinions on different aspects of the entity.
- For blogs, forum discussions, etc., it is harder:
  - both entity and aspects of entity are unknown,
  - there may also be many comparisons, and
  - there is also a lot of irrelevant information.

## 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 on 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

Bing Liu, tutorial

77

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

- **Distributional similarity:** This is the traditional method used in NLP. It compares the surrounding text of candidates using cosine or PMI.
  - It performs poorly.
- **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).

Bing Liu, tutorial

78

## Aspect extraction

- **Goal:** Given an opinion corpus, extract all aspects
- **A frequency-based approach** (Hu and Liu, 2004): nouns (NN) that are frequently talked about are likely to be true **aspects** (called frequent aspects) .
- **Why the frequency based approach?**
  - Different reviewers tell different stories (irrelevant)
  - When product aspects/features are discussed, the words they use converge.
  - They are the main aspects.
- Sequential/association pattern mining finds **frequent nouns and noun phrases**.

Bing Liu, tutorial

79

## An example review

**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. ...

....

Bing Liu, tutorial

80



## Infrequent aspect extraction

- To improve recall due to loss of infrequent aspects. It uses opinion words to extract them
- **Key idea:** **opinions have targets**, i.e., opinion words are used to modify aspects and entities.
  - “The pictures are absolutely **amazing**.”
  - “This is an **amazing** software.”
- The modifying relation was approximated with the nearest noun to the opinion word.
- The idea was generalized to dependency in (Zhuang et al 2006) and double propagation in (Qiu et al 2009).
  - It has been used in many papers and practical systems

Bing Liu, tutorial

81

## Using part-of relationship and the Web (Popescu and Etzioni, 2005)

- Improved (Hu and Liu, 2004) by removing those frequent noun phrases that may not be aspects: better precision (a small drop in recall).
- 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:

$$PMI(a, d) = \frac{hits(a \wedge d)}{hits(a)hits(d)},$$

Bing Liu, tutorial

82

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

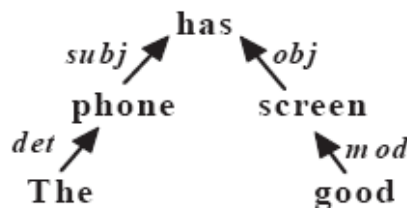
- A *double propagation (DP)* approach proposed
- Based on the definition earlier, **an opinion should have a target**, entity or aspect.
- Use dependency of opinions & aspects to extract both aspects & opinion words.
  - *Knowing one helps find the other.*
  - E.g., “The **rooms** are *spacious*”
-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

Bing Liu, tutorial

83

## The DP method

- **DP is a bootstrapping method**
  - **Input**: a set of seed opinion words,
  - no aspect seeds needed
- **Based on dependency grammar** (Tesniere 1959).
  - “This phone has good screen”



Bing Liu, tutorial

84

## Rules from dependency grammar

	Relations and Constraints	Output	Examples
R1 <sub>1</sub>	$O \rightarrow O-Dep \rightarrow F$ s.t. $O \in \{O\}$ , $O-Dep \in \{MR\}$ , $POS(F) \in \{NN\}$	$f = F$	The phone has a <u>good</u> "screen". $good \rightarrow mod \rightarrow screen$
R1 <sub>2</sub>	$O \rightarrow O-Dep \rightarrow H \leftarrow F-Dep \leftarrow F$ s.t. $O \in \{O\}$ , $O/F-Dep \in \{MR\}$ , $POS(F) \in \{NN\}$	$f = F$	"iPod" is the <u>best</u> mp3 player. $best \rightarrow mod \rightarrow player \leftarrow subj \leftarrow iPod$
R2 <sub>1</sub>	$O \rightarrow O-Dep \rightarrow F$ s.t. $F \in \{F\}$ , $O-Dep \in \{MR\}$ , $POS(O) \in \{JJ\}$	$o = O$	same as R1 <sub>1</sub> with <i>screen</i> as the known word and <i>good</i> as the extracted word
R2 <sub>2</sub>	$O \rightarrow O-Dep \rightarrow H \leftarrow F-Dep \leftarrow F$ s.t. $F \in \{F\}$ , $O/F-Dep \in \{MR\}$ , $POS(O) \in \{JJ\}$	$o = O$	same as R1 <sub>2</sub> with <i>iPod</i> is the known word and <i>best</i> as the extract word.
R3 <sub>1</sub>	$F_{i(j)} \rightarrow F_{i(j)}-Dep \rightarrow F_{j(i)}$ s.t. $F_{j(i)} \in \{F\}$ , $F_{i(j)}-Dep \in \{CONJ\}$ , $POS(F_{i(j)}) \in \{NN\}$	$f = F_{i(j)}$	Does the player play dvd with <u>audio</u> and "video"? $video \rightarrow conj \rightarrow audio$
R3 <sub>2</sub>	$F_i \rightarrow F_i-Dep \rightarrow H \leftarrow F_j-Dep \leftarrow F_j$ s.t. $F_i \in \{F\}$ , $F_i-Dep = F_j-Dep$ , $POS(F_j) \in \{NN\}$	$f = F_j$	Canon "G3" has a great <u>len</u> . $len \rightarrow obj \rightarrow has \leftarrow subj \leftarrow G3$
R4 <sub>1</sub>	$O_{i(j)} \rightarrow O_{i(j)}-Dep \rightarrow O_{j(i)}$ s.t. $O_{j(i)} \in \{O\}$ , $O_{i(j)}-Dep \in \{CONJ\}$ , $POS(O_{i(j)}) \in \{JJ\}$	$o = O_{i(j)}$	The camera is <u>amazing</u> and "easy" to use. $easy \rightarrow conj \rightarrow amazing$
R4 <sub>2</sub>	$O_i \rightarrow O_i-Dep \rightarrow H \leftarrow O_j-Dep \leftarrow O_j$ s.t. $O_i \in \{O\}$ , $O_i-Dep = O_j-Dep$ , $POS(O_j) \in \{JJ\}$	$o = O_j$	If you want to buy a <u>sexy</u> , "cool", accessory- available mp3 player, you can choose iPod. $sexy \rightarrow mod \rightarrow player \leftarrow mod \leftarrow cool$

Bing Liu, tutorial

85

## 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."
  - "This phone will not easily fit in a pocket."
  - "Included 16MB is stingy"
- Not much work has been done on mining or mapping implicit aspects.

Bing Liu, tutorial

86

## Implicit aspect mapping

- There are many types of implicit aspect expressions. Adjectives and adverbs are perhaps the most common type.
  - Most adjectives modify or describe some specific attributes of entities.
  - “expensive”  $\Rightarrow$  aspect “price,” “beautiful”  $\Rightarrow$  aspect “appearance”, “heavy”  $\Rightarrow$  aspect “weight”
- Although manual mapping is possible, in different contexts, the meaning can be different.
  - E.g., “The computation is expensive”.

Bing Liu, tutorial

87

## A mutual reinforcement method

(Su et al. 2009)

- It proposed an unsupervised approach which exploits the mutual reinforcement relationship between aspects and opinion words.
  - Specifically, it uses the co-occurrence of aspect and opinion word pair in a sentence.
- The algorithm iteratively clusters the set of aspects and the set of opinion words separately,
  - but before clustering each set, clustering results of the other set is used to update the pairwise weight of the set.
  - The model is based on a bipartite graph.

Bing Liu, tutorial

88

## Other papers on aspect extraction

We will discuss topic modeling based methods later.

- Carvalho et al (2011) annotated political debates with aspects and others.
- Choi and Cardie (2010) used a CRF based approach.
- Jin and Ho (2009) proposed a HMM-based method
- Jakob and Gurevych (2010) used anaphora (or coreference) resolution to help find aspects that are mentioned in previous sentences but are referred to as pronouns in the next sentences.
  - E.g., "I took a few pictures yesterday. They look great."
  - There is almost no improvement with anaphora resolution, higher recall but lower precision.

## Other papers on aspect extraction

- Jakob and Gurevych (2010) used CRF to train on review sentences from different domains for a more domain independent extraction. A set of domain independent features were used, e.g. tokens, POS tags, dependency, word distance, and opinion sentences.
- Kobayashi et al (2006) extracted subject-attribute-value
- Kobayashi et al (2007) extracted aspect-evaluation and aspect-of relations using a tree-based classifier with different features.
- Ku et al. (2006a, 2006b) performed the extraction from Chinese reviews and news.

## Other papers on aspect extraction

- Li et al (coling-2010) integrated Skip-CRF and Tree-CRF to extract aspects and opinions. It was able to exploit structure features
- Long, Zhang and Zhu (2010) extracted aspects (nouns) based on frequency and the information distance, and dependent words (adjectives). These words are then used to select reviews which discuss an aspect most.
- Ma and Wan (2010) used centering theory for extraction in news comments. It also exploited aspects in the news title and contents.
- Meng and Wang (2009) extracted aspects from product specifications, which are usually structured data.

Bing Liu, tutorial

91

## Other papers on aspect extraction

- Scaffidi et al (2007) extracted frequent nouns and noun phrases but compare their frequency in a review corpus with their occurrence rates in generic English to identify true aspects
- Somasundaran and Wiebe (2009) also used syntactic dependency for aspect and opinion extraction.
- Toprak, Jakob and Gurevych (2010) designed a comprehensive annotation scheme for aspect-based opinion annotation. Earlier annotations are partial and mainly for individual papers.
- Yi et al (2003) used language models to extract product features.

Bing Liu, tutorial

92

## Other papers on aspect extraction

- Yu et al (2011) ranked aspects by considering their frequency and contribution to the overall review rating
- Zhu et al (CIKM-2009) used a method for finding multi-word terms, called cvalue, to find aspects.
  - The method also segments a sentence with multiple aspects.

## Identify aspect synonyms (Carenini et al 2005)

- Once aspect expressions are discovered, group them into aspect categories.
  - E.g., power usage and battery life are the same.
- It proposed a method based on some similarity metrics, but it needs a taxonomy of aspects.
  - The system merges each discovered aspect to a aspect node in the taxonomy.
  - Similarity metrics: string similarity, synonyms and other distances measured using WordNet.
- Many ideas in Web information integration are applicable.

## Multilevel latent categorization

(Guo et al 2009)

- This method performs multilevel latent semantic analysis to group aspects expressions.
  - At the first level, all the words in aspect expressions are grouped into a set of concepts using LDA. The results are used to build *latent topic structures* for aspect expressions, e.g.,
    - *touch screen*: topic-1, topic-2
  - At the second level, aspect expressions are grouped by LDA again according to
    - their latent topic structures produced from level 1 and
    - context snippets in reviews.

Bing Liu, tutorial

95

## Group aspect synonyms (Zhai et al. 2011a, b)

- A variety of information/similarities are used to cluster aspect expressions into aspect categories.
  - Lexical similarity based on WordNet
  - Distributional information (surrounding words context)
  - Syntactical constraints (sharing words, in the same sentence)
- Two unsupervised learning methods were used:
  - **Clustering**: EM-based.
  - **Constrained topic modeling**: Constrained-LDA
    - By intervening Gibbs sampling.

Bing Liu, tutorial

96



## The EM method

- WordNet similarity

$$Jcn(w_1, w_2) = \frac{1}{IC(w_1) + IC(w_2) - 2 \times Res(w_1, w_2)}$$

- EM-based probabilistic clustering

$$P(w_t | c_j) = \frac{1 + \sum_{i=1}^{|D|} N_{ti} P(c_j | d_i)}{|V| + \sum_{m=1}^{|V|} \sum_{i=1}^{|D|} N_{mi} P(c_j | d_i)}$$

$$P(c_j) = \frac{1 + \sum_{i=1}^{|D|} P(c_j | d_i)}{|C| + |D|}$$

$$P(c_j | d_i) = \frac{P(c_j) \prod_{k=1}^{|d_i|} P(w_{d_i,k} | c_j)}{\sum_{r=1}^{|C|} P(c_r) \prod_{k=1}^{|d_i|} P(w_{d_i,k} | c_r)}$$

Bing Liu, tutorial

97

## Aspect sentiment classification

- For each aspect, identify the sentiment or opinion expressed on 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 **view founder** 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., “small” and “sucks” (+ve for vacuum cleaner)

Bing Liu, tutorial

98

## Some approaches

### ■ Supervised learning

- Sentence level classification can be used, but ...
- Need to consider target and thus to segment a sentence (e.g., Jiang et al. 2011)

### ■ Lexicon-based approach (Ding, Liu and Yu, 2008)

- **Need parsing to deal with:** Simple sentences, compound sentences, comparative sentences, conditional sentences, questions; different verb tenses, etc.
- Negation (not), contrary (but), comparisons, etc.
- A large opinion lexicon, context dependency, etc.
- **Easy:** “*Apple is doing well in this bad economy.*”

Bing Liu, tutorial

99

## A lexicon-based method (Ding, Liu and Yu 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  $(a, s)$  accordingly.

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

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

Bing Liu, tutorial

100

## 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.”

## Sentiment shifters (contd)

- Some **presuppositional** items also can change opinions, e.g., *barely* and *hardly*
  - “It hardly works.” (comparing to “it works”)
  - It presupposes that better was expected.
- Words like *fail*, *omit*, *neglect* behave similarly,
  - “This camera fails to impress me.”
- Sarcasm changes orientation too
  - “What a great car, it did not start the first day.”
- Jia, Yu and Meng (2009) designed some rules based on parsing to find the scope of negation.

## Basic rules of opinions (Liu, 2010)

- Opinions/sentiments are governed by many rules, e.g.,
  - *Opinion word or phrase, ex: "I love this car"*
    - P ::= a positive opinion word or phrase
    - N ::= an negative opinion word or phrase
  - *Desirable or undesirable facts, ex: "After my wife and I slept on it for two weeks, I noticed a mountain in the middle of the mattress"*
    - P ::= desirable fact
    - N ::= undesirable fact

Bing Liu, tutorial

103

## Basic rules of opinions

- *High, low, increased and decreased quantity of a positive or negative potential item, ex: "The battery life is long."*
  - PO ::= no, low, less or decreased quantity of NPI  
| large, larger, or increased quantity of PPI
  - NE ::= no, low, less, or decreased quantity of PPI  
| large, larger, or increased quantity of NPI
  - NPI ::= a negative potential item
  - PPI ::= a positive potential item

Bing Liu, tutorial

104

## Basic rules of opinions

- *Decreased and increased quantity of an opinionated item, ex: “This drug reduced my pain significantly.”*

PO ::= less or decreased N

| more or increased P

NE ::= less or decreased P

| more or increased N

- *Deviation from the desired value range: “This drug increased my blood pressure to 200.”*

PO ::= within the desired value range

NE ::= above or below the desired value range

## Basic rules of opinions

- *Producing and consuming resources and wastes, ex: “This washer uses a lot of water”*

PO ::= produce a large quantity of or more resource

| produce no, little or less waste

| consume no, little or less resource

| consume a large quantity of or more waste

NE ::= produce no, little or less resource

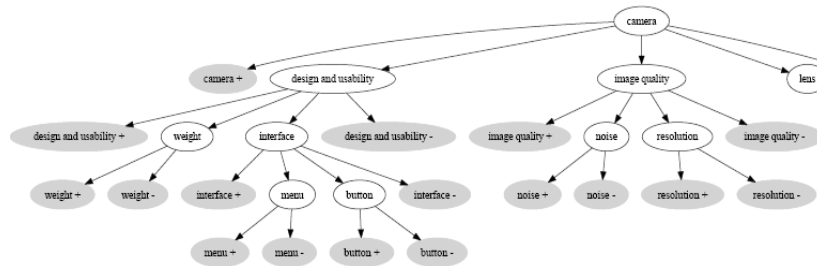
| produce some or more waste

| consume a large quantity of or more resource

| consume no, little or less waste

## Sentiment ontology tree (Wei and Gulla, 2010)

- Recall in the definition of opinions, we simplified the tree structure to two levels (entity & aspects).
- This paper uses a full tree ontology to denote the relationships of aspects of a product.



Bing Liu, tutorial

107

## Sentiment ontology tree (contd)

- The leaves of the tree are positive or negative sentiments.
- It then uses a hierarchical classification model to learn to assign an sentiment to each node, which is reflected as a child leaf node.
  - Hierarchical classifier is useful here because it considers parents when classifying children.
- However, the ontology for each product has to be built manually.

Bing Liu, tutorial

108

## Aspect-sentiment statistical models

- This direction of research is mainly based on **topic models**:
  - **pLSA**: Probabilistic Latent Semantic Analysis (Hofmann 1999)
  - **LDA**: Latent Dirichlet allocation (Blei, Ng & Jordan, 2003; Griffiths & Steyvers, 2003; 2004)
- Topic models:
  - documents are mixtures of topics
  - a topic is a probability distribution over words.
- A topic model is a document **generative model**
  - it specifies a simple probabilistic procedure by which documents can be generated.

Bing Liu, tutorial

109

## Aspect-sentiment model (Mei et al 2007)

- This model is based on pLSA (Hofmann, 1999).
- It builds a topic (aspect) model, a positive sentiment model, and a negative sentiment model.
- A training data is used to build the initial models.
  - Training data: topic queries and associated positive and negative sentences about the topics.
- The learned models are then used as priors to build the final models on the target data.
- Solution: log likelihood and EM algorithm

Bing Liu, tutorial

110

## Multi-Grain LDA to extract aspects

(Titov and McDonald, 2008a, 2008b)

- Unlike a diverse document set used for traditional topic modeling. All reviews for a product talk about the same topics/aspects. It makes applying PLSA or LDA in the traditional way problematic.
- Multi-Grain LDA (MG-LDA) models global topics and local topics (Titov and McDonald, 2008a).
  - Global topics are entities (based on reviews)
  - Local topics are aspects (based on local context, sliding windows of review sentences)
- MG-LDA was extended to MAS model to give aspect rating (Titov and McDonald, 2008b).

Bing Liu, tutorial

111

## Aspect-rating of short text (Lu et al 2009)

- This work makes use of short phrases, head terms ( $w_h$ ) and their modifiers ( $w_m$ ), i.e.
  - ( $w_m, w_h$ )
  - E.g., great shipping, excellent seller
- Objective: (1) extract aspects and (2) compute their ratings in each short comment.
- It uses pLSA to extract and group aspects
- It uses existing rating for the full post to help determine aspect ratings.

Bing Liu, tutorial

112



## Aspect-rating regression

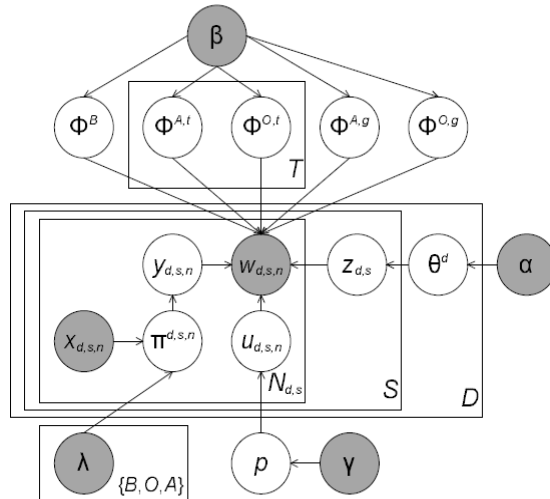
(Wang, Lu, and Zhai, 2010)

- In this work, some seed aspects are given. Its first step finds more aspect words using a heuristic bootstrapping method.
- Its regression model makes use of the review rating and assumes the overall review rating is a linear combination of its aspect ratings.
- The problem is model as a Bayesian regression problem.
  - It is solved using log-likelihood and EM.

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

We now describe the generative process of the model. First, we draw several multinomial word distributions from a symmetric Dirichlet prior with parameter  $\beta$ : a background model  $\phi^B$ , a general aspect model  $\phi^{A,g}$ , a general opinion model  $\phi^{O,g}$ ,  $T$  aspect models  $\{\phi^{A,t}\}_{t=1}^T$  and  $T$  aspect-specific opinion models  $\{\phi^{O,t}\}_{t=1}^T$ . All these are multinomial distributions over the vocabulary, which we assume has  $V$  words. Then for each review document  $d$ , we draw a topic distribution  $\theta^d \sim \text{Dir}(\alpha)$  as in standard LDA. For each sentence  $s$  in document  $d$ , we draw an aspect assignment  $z_{d,s} \sim \text{Multi}(\theta^d)$ .

## Graphical model (plate)



- $y_{d,s,n}$  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

Bing Liu, tutorial

115

# Topic model of snippets

(Sauper, Haghighi and Barzilay, 2011)

- This method works on short snippets already extracted from reviews.
  - “battery life is the best I’ve found”
- The model is a variation of LDA but with seeds for sentiment words as priors,
  - but it also has HMM for modeling the sequence of words with types (aspect word, sentiment word, or background word).
- Inference: variational technique

Bing Liu, tutorial

116

## Considering both syntax and semantics

(Lakkaraju et al. 2011)

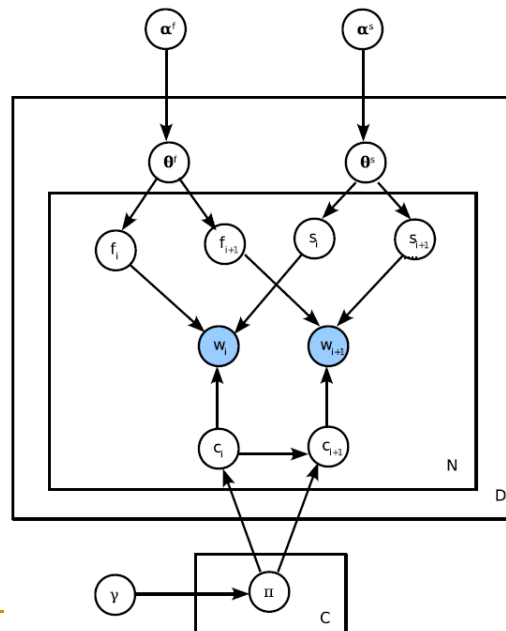
- This work is based the composite model of HMM-LDA of Griffiths et al. (2005), which consider both word sequence and word-bag
- It captures both syntactic structure and semantic dependencies (similar to the previous paper)
- A class label is used for each word to represent the syntactic category of the word, whether it is
  - an aspect word,
  - a sentiment word, or
  - some other category.

Bing Liu, tutorial

117

## FACTS model

- **Words:**  $w_{d,1}, w_{d,2} \dots w_{d,N}$
- **Hidden variables**
  - **Class:**  $c_{d,i}$ 
    - 1: aspect word
    - 2: sentiment word
    - others
  - **Aspect cat.:**  $f_{d,i}$
  - **Sentiment cat.:**  $s_{d,i}$
- **It also has more sophisticated models**
  - **CFACTS:** consider neighboring windows
  - **CFACTS-R:** consider ratings



Bing Liu, tutorial

118

## About topic model based methods

- There several other similar topic model based methods (e.g., Brody and Elhadad, 2010; Lu et al. 2011; Jo and Oh, 2011; Lin and He 2009; Liu et al, 2007; Moghaddam et al. 2011; Mukherjee and Liu, 2012; ).
- These methods tend to need a large number reviews (10000 and more) to make it statistically stable. They are hard to use for most specific products, which often have <100 reviews.
- They also need a lot of parameter tuning.
- The results usually are quite coarse, not precise enough for practical needs.

## Roadmap

- Sentiment Analysis Problem
- Document sentiment classification
- Sentence subjectivity & sentiment classification
- Aspect-based sentiment analysis
- ➔ ■ **Aspect-based opinion summarization**
- Sentiment lexicon generation
- Mining comparative opinions
- Some other problems
- Opinion spam detection
- Utility or helpfulness of reviews
- Summary

## Aspect-based opinion summarization

- A multi-document summarization problem.
  - An opinion from a single person is usually not sufficient for action unless from a VIP (e.g., President)
- Key Idea: Use aspects as basis for a summary
  - Not done in traditional multi-document summarization.
- We have discussed the aspect-based summary using quintuples earlier (Hu and Liu 2004; Liu, 2010).
  - Also called: *Structured Summary*
- Similar approaches are also taken in
  - (e.g., Ku et al 2006; Carenini, Ng and Paul 2006) and
  - By most topic modeling based methods

Bing Liu, tutorial

121

## Text summary of opinions

- One can also generate a summary in the tradition fashion, e.g., producing a short text summary (Lerman et al 2009), by extracting some important sentences, etc.
  - Weakness: It is only qualitative but not quantitative.
- One can generate sentences based on aspects and opinions using some templates.
  - E.g., 60% of the people like the picture quality.

Bing Liu, tutorial

122

## Select and order sentences

(Tata and Di Eugenio, 2010)

- If we produce summary as a list of sentences for each aspect and each sentiment (+ or –), it is useful to
  - Select a representative sentence for each group: it selects a sentence that mentions fewest aspects (the sentence is focused).
  - Order the sentences: It uses an ontology to map sentences to the ontology nodes (domain concepts).

Bing Liu, tutorial

123

## Informativeness and Readability

(Nishikawa et al. 2010)

- It summarizes by considering both informativeness and readability.

- $S^*$  is the summary

$$S^* = \operatorname{argmax}_{S \in T} [\operatorname{Info}(S) + \lambda \operatorname{Read}(S)]$$
$$\text{s.t. } \operatorname{length}(S) \leq K$$

- It uses frequency  $f(\cdot)$  of (aspect, opinion), but it is more like a traditional summary.

$$\operatorname{Info}(S) = \sum_{e \in E(S)} f(e)$$

- It is not quantitative. Note: Lerman et al (2009) used +ve/-ve proportions.

$$\operatorname{Read}(S) = \sum_{i=0}^n \mathbf{w}^T \phi(s_i, s_{i+1})$$

Bing Liu, tutorial

124

## Summarization using an ontology

(Lu et al. Coling-2010)

- This work uses existing online ontologies of entities and aspects to organize opinions
  - Given an entity and an online ontology of the entity
  - **Goal**: Generate a structured summary of opinions.
- It performs
  - Aspect selection to capture major opinions
  - Aspect ordering that is natural for human viewing
  - Suggest new aspects to add to ontology

## Summarization using an ontology (contd)

- Aspect selection
  - E.g., by frequency, by opinion coverage (no redundancy), or by conditional entropy
- Ordering aspects and their corresponding sentences based on their appearance in their original posts, called **coherence**

$$Co(A_i, A_j) = \frac{\sum_{S_{i,k} \in S_i, S_{j,l} \in S_j} Co(S_{i,k}, S_{j,l})}{|S_i||S_j|}$$

$$\hat{\pi}(A') = \arg \max_{\pi(A')} \sum_{A_i, A_j \in A', A_i \prec A_j} Co(A_i, A_j)$$

## Some other summarization papers

- Carenini, Ng and Pauls (2006) evaluated different summarization methods using human judges.
- Huang, Wan and Xiao (2011) generated contrast summaries of news.
- Kim and Zhai (2009) generated contrast opinion sentence pairs.
- Lerman and McDonald (2009) generated summaries to contrast opinions about two different products.
- Lerman, Blair-Goldensohn and McDonald (2009) designed three summarizers and evaluated them with human raters.
- Paul, Zhai and Girju (2010) found opposing views.
- Park, Lee and Song (2011) also found opposing views
- Wang and Liu (2011) generated opinion summary for conversations.

Bing Liu, tutorial

127

## Roadmap

- Opinion Mining Problem
- Document sentiment classification
- Sentence subjectivity & sentiment classification
- Aspect-based sentiment analysis
- Aspect-based opinion summarization
- ➔ ■ **Sentiment lexicon generation**
- Mining comparative opinions
- Some other problems
- Opinion spam detection
- Utility or helpfulness of reviews
- Summary

Bing Liu, tutorial

128



## Sentiment (or opinion) lexicon

- **Sentiment lexicon**: lists of words and expressions used to express people's subjective feelings and sentiments/opinions.
  - Not just individual words, but also phrases and idioms, e.g., "cost an arm and a leg"
- They are instrumental for sentiment analysis.
- There seems to be endless variety of sentiment bearing expressions.
  - We have compiled more than 6,700 individual words.
  - There are also a large number of phrases.

Bing Liu, tutorial

129

## Sentiment lexicon

- **Sentiment words or phrases** (also called polar words, opinion bearing words, etc). E.g.,
  - **Positive**: beautiful, wonderful, good, amazing,
  - **Negative**: bad, poor, terrible, cost an arm and a leg.
- Many of them are context dependent, not just application domain dependent.
- Three main ways to compile such lists:
  - **Manual approach**: not a bad idea, only an one-time effort
  - **Corpus-based approach**
  - **Dictionary-based approach**

Bing Liu, tutorial

130

## Corpus-based approaches

- **Rely on syntactic patterns in large corpora.**  
(Hazivassiloglou and McKeown, 1997; Turney, 2002; Yu and Hazivassiloglou, 2003; Kanayama and Nasukawa, 2006; Ding, Liu and Yu, 2008)
  - **Can find domain dependent orientations** (positive, negative, or neutral).
- (Turney, 2002) and (Yu and Hazivassiloglou, 2003) are similar.
  - Assign opinion orientations (polarities) to words/phrases.
  - (Yu and Hazivassiloglou, 2003) is slightly different from (Turney, 2002)
    - use more seed words (rather than two) and use log-likelihood ratio (rather than PMI).

Bing Liu, tutorial

131

## Corpus-based approaches (contd)

- **Sentiment consistency: Use conventions on connectives** to identify opinion words (Hazivassiloglou and McKeown, 1997). E.g.,
  - **Conjunction**: conjoined adjectives usually have the same orientation.
    - E.g., "This car is *beautiful* **and** *spacious*." (conjunction)
  - AND, OR, BUT, EITHER-OR, and NEITHER-NOR have similar constraints.
  - **Learning using**
    - **log-linear model**: determine if two conjoined adjectives are of the same or different orientations.
    - **Clustering**: produce two sets of words: positive and negative

Bing Liu, tutorial

132

## Find domain opinion words

- A similar approach was also taken in (Kanayama and Nasukawa, 2006) but for Japanese words:
  - Instead of only based on **intra-sentence sentiment consistency**, the new method also looks at the previous and next sentence, i.e., **inter-sentence sentiment consistency**.
  - Have an initial seed lexicon of positive and negative words.

## Context dependent opinion

- **Find domain opinion words is insufficient.** A word may indicate different opinions in same domain.
  - “The battery life is **long**” (+) and “It takes a **long** time to focus” (-).
- Ding, Liu and Yu (2008) and Ganapathibhotla and Liu (2008) exploited sentiment consistency (both inter and intra sentence) based on contexts
  - **It finds context dependent opinions.**
  - **Context:** (adjective, aspect), e.g., (long, battery\_life)
  - It assigns an opinion orientation to the pair.

## The Double Propagation method

(Qiu et al 2009, 2011)

- The same DP method can also use dependency of opinions & aspects to extract new opinion words.
- **Based on dependency relations**
  - Knowing an aspect can find the opinion word that modifies it
    - E.g., “The **rooms** are **spacious**”
  - Knowing some opinion words can find more opinion words
    - E.g., “The **rooms** are **spacious** and **beautiful**”

Bing Liu, tutorial

135

## Opinions implied by objective terms

(Zhang and Liu, 2011)

- Most opinion words are adjectives and adverbs, e.g., good, bad, etc
  - There are also many subjective and opinion verbs and nouns, e.g., hate (VB), love (VB), crap (NN).
- **But objective nouns can imply opinions too.**
  - E.g., “After sleeping on the mattress for one month, a **valley/body impression** has formed in the middle.”
- How to discover such nouns in a domain or context?

Bing Liu, tutorial

136

## The technique

- 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*.

Bing Liu, tutorial

137

## Pruning

- For an aspect with an implied opinion, it has a fixed opinion, either +ve or -ve, but not both.
- We 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.

Bing Liu, tutorial

138

## Opinions implied by resource usage

(Zhang and Liu, 2011)

- Resource usage descriptions may also imply opinions (as mentioned in rules of opinions)
  - E.g., “This washer uses a lot of water.”
- Two key roles played by resources 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 triple.  
(verb, quantifier, noun\_term),
  - Verb: uses, quantifier: “a lot of”, noun\_term: water

Bing Liu, tutorial

139

## The proposed technique

- The proposed 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 specific resource usage verbs and resource terms.

Bing Liu, tutorial

140

## Dictionary-based methods

- Typically use WordNet's synsets and hierarchies to acquire opinion words
  - Start with a small seed set of opinion words.
  - Bootstrap the set to search for synonyms and antonyms in WordNet iteratively (Hu and Liu, 2004; Kim and Hovy, 2004; Kamps et al 2004).
- Use additional information (e.g., glosses) from WordNet (Andreevskaia and Bergler, 2006) and learning (Esuti and Sebastiani, 2005). (Dragut et al 2010) uses a set of rules to infer orientations.

## Semi-supervised learning

(Esuti and Sebastiani, 2005)

- Use supervised learning
  - Given two seed sets: positive set P, negative set N
  - The two seed sets are then expanded using synonym and antonymy relations in an online dictionary to generate the expanded sets P' and N'.
- P' and N' form the training sets.
- Using all the glosses in a dictionary for each term in  $P' \cup N'$  and converting them to a vector
- Build a binary classifier
  - Tried various learners.

## Multiple runs of bootstrapping

(Andreevskaia and Bergler, 2006)

- Basic bootstrapping with given seeds sets (adjectives)
  - First pass: seed sets are expanded using synonym, antonymy, and hyponyms relations in WordNet.
  - Second pass: it goes through all WordNet glosses and identifies the entries that contain in their definitions the sentiment-bearing words from the extended seed set and adds these head words to the corresponding category (+ve, -ve, neutral)
  - Third pass: clean up using a POS tagger to make sure the words are adjectives and remove contradictions.

Bing Liu, tutorial

143

## Multiple runs of bootstrapping (contd)

- Each word is then assigned a fuzzy score reflecting the degree of certainty that the word is opinionated (+ve/-ve).
- The method performs multiple runs of bootstrapping using non-overlapping seed sets.
  - A net overlapping score for each word is computed based on how many times the word is discovered in the runs as +ve (or -ve)
  - The score is normalized based on the fuzzy membership.

Bing Liu, tutorial

144



## Which approach to use?

- Both corpus and dictionary based approaches are needed.
- Dictionary usually does not give domain or context dependent meaning
  - Corpus is needed for that
- Corpus-based approach is hard to find a very large set of opinion words
  - Dictionary is good for that
- In practice, corpus, dictionary and manual approaches are all needed.

Bing Liu, tutorial

145

## Some other related papers

- Choi and Cardie (2009) adapting a lexicon to domain specific need using integer linear programming
- Du and Tan (2009) and Du, Tan, Cheng and Yun (2010) clustered sentiment words
- Hassan and Radev (2010) built a word graph based on synonyms and then used a number of random walks to hit known seed words
- Hassan et al. (2011) found sentiment orientations of foreign words. It first created a multilingual word network and then did random walk similar to the above paper.
- Jijkoun, Rijke and Weerkamp (2010) used target and sentiment word relationship. Similar to that in (Qiu et al 2009).

Bing Liu, tutorial

146

## Some other related papers

- Kaji and Kitsuregawa (2006, 2007) and Velikovich et al (2010) used text on the web to generate lexicons.
- Lu et al (2011) dealt with the same problem as (Ding et al 2008) but used various constraints in optimization.
- Mohammad, Dunne, and Dorr, (2009) used seeds and thesaurus.
- Rao and Ravichandran (2009) used WordNet and OpenOffice thesaurus and semi-supervised learning
- Wu and Wen (2010) found context adjectives like *large* and *small* by mining the web using lexico-syntactic patterns. They solved the same problem as (Ding et al 2008)

## Roadmap

- Sentiment Analysis Problem
- Document sentiment classification
- Sentence subjectivity & sentiment classification
- Aspect-based sentiment analysis
- Aspect-based opinion summarization
- Sentiment lexicon generation
- ➔ ■ **Mining comparative opinions**
- Some other problems
- Opinion spam detection
- Utility or helpfulness of reviews
- Summary

## Comparative Opinions

(Jindal and Liu, 2006)

### ■ Gradable

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

Bing Liu, tutorial

149

## Analyzing Comparative Opinions

- **Objective**: Given an opinionated document  $d$ ,  
Extract comparative opinions:

$(E_1, E_2, A, po, h, t)$ ,

where  $E_1$  and  $E_2$  are the entity sets being compared based on their shared aspects  $A$ ,  $po$  is the preferred entity set of the opinion holder  $h$ , and  $t$  is the time when the comparative opinion is expressed.

- **Note**: not positive or negative opinions.

Bing Liu, tutorial

150

## An example

- Consider the comparative sentence
  - “*Canon’s optics is better than those of Sony and Nikon.*”
  - Written by John in 2010.
- The extracted comparative opinion/relation:
  - ({Canon}, {Sony, Nikon}, {optics},  
*preferred*:{Canon}, John, 2010)

## Common comparatives

- In English, comparatives are usually formed by adding *-er* and superlatives are formed by adding *-est* to their **base adjectives** and **adverbs**
- Adjectives and adverbs with two syllables or more and not ending in *y* do not form comparatives or superlatives by adding *-er* or *-est*.
  - Instead, *more*, *most*, *less*, and *least* are used before such words, e.g., *more beautiful*.
- Irregular comparatives and superlatives, i.e., *more*, *most*, *less*, *least*, *better*, *best*, *worse*, *worst*, etc

## Some techniques (Jindal and Liu, 2006, Ding et al, 2009)

- Identify comparative sentences
  - Using class sequential rules as attributes in the data, and then perform
  - Supervised learning
- Extraction of different items
  - Label sequential rules
  - conditional random fields
- Determine preferred entities (opinions)
  - Parsing and opinion lexicon
- (Yang and Ko, 2011) is similar to (Jindal and Liu 2006)

Bing Liu, tutorial

153

## Analysis of comparative opinions

- 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:** “*optics of camera A*”
  - **Negative:** “*optics of camera B*”
- **Difficulty:** recognize non-standard comparatives
  - E.g., “I am so happy because my new iPhone is nothing like my old slow ugly Droid.”
  - ?

Bing Liu, tutorial

154

## Identifying preferred entities

(Ganapathibhotla and Liu, 2008)

- The following rules can be applied

Comparative Negative ::= increasing comparative N  
| decreasing comparative P

Comparative Positive ::= increasing comparative P  
| decreasing comparative N

- E.g., “Coke tastes better than Pepsi”
- “Nokia phone’s battery life is longer than Moto phone”

- Context-dependent comparative opinion words

- Using context pair: (aspect, JJ/JJR)
- Deciding the polarity of (battery\_life, longer) in a corpus

Bing Liu, tutorial

155

## Some other work

- Bos and Nissim (2006) proposed a method to extract items from superlative sentences, but does not study sentiments.
- Fiszman et al (2007) tried to identify which entity has more of a certain property in comparisons.
- Li et al (2010) finds comparative questions and compared entities using sequence patterns.
- Yang and Ko (2009, 2011) worked on Korean comparative sentences.
- Zhang, Narayanan and Choudhary (2010) found comparative sentences based on a set of rules, and the sentences must also mention at least two product names explicitly or implicitly (comparing with the product being reviewed).

Bing Liu, tutorial

156

## Roadmap

- Sentiment Analysis Problem
- Document sentiment classification
- Sentence subjectivity & sentiment classification
- Aspect-based sentiment analysis
- Aspect-based opinion summarization
- Sentiment lexicon generation
- Mining comparative opinions
- ➔ ■ **Some other problems**
- Opinion spam detection
- Utility or helpfulness of reviews
- Summary

## Coreference resolution: semantic level?

- **Coreference resolution** (Ding and Liu, 2010)
  - “I bought the Sharp tv a month ago. The picture quality is so bad. Our other Sony tv is much better than this Sharp. **It is also so expensive**”.
  - “it” means “Sharp”
  - “I bought the Sharp tv a month ago. The picture quality is so bad. Our other Sony tv is much better than this Sharp. **It is also very reliable**.”
  - “it” means “Sony”
- Sentiment consistency.

## Coreference resolution (contd)

- “The picture quality of this Canon camera is very good. *It* is not expensive either.”
  - Does “it” mean “Canon camera” or “Picture Quality”?
    - Clearly it is Canon camera because picture quality cannot be expensive.
    - Commonsense knowledge, but can be discovered.
- For coreference resolution, we actually need to
  - do sentiment analysis first, and
  - mine adjective-noun associations using dependency
- Finally, use supervised learning

## Some interesting sentences

- “Trying out Google chrome because Firefox keeps crashing.”
  - The opinion about Firefox is clearly negative, but for Google chrome, there is no opinion.
  - We need to segment the sentence into clauses to decide that “crashing” only applies to Firefox.
  - “Trying out” also indicates no opinion.
- How about this
  - “I changed to Audi because BMW is so expensive.”



## Some interesting sentences (contd)

- Conditional sentences are hard to deal with (Narayanan et al. 2009)
  - “If I can find a good camera, I will buy it.”
  - But conditional sentences can have opinions
    - “If you are looking for a good phone, buy Nokia”
- Questions may or may not have opinions
  - No sentiment
    - “Are there any great perks for employees?”
  - With sentiment
    - “Any idea how to repair this lousy Sony camera?”

Bing Liu, tutorial

161

## Some interesting sentences (contd)

- Sarcastic sentences
  - “What a great car, it stopped working in the second day.”
- Sarcastic sentences are very common in political blogs, comments and discussions.
  - They make political blogs difficult to handle
  - Many political aspects can also be quite complex and hard to extract because they cannot be described using one or two words.
- Some initial work by (Tsur, Davidov, Rappoport 2010)

Bing Liu, tutorial

162

## Some interesting sentences (contd)

- See these two sentences in a medical domain:
  - “I come to see my doctor because of severe pain in my stomach”
  - “After **taking the drug**, I got severe pain in my stomach”
- If we are interested in opinions on a drug, the first sentence has no opinion, but the second implies negative opinion on the drug.
  - Some understanding seems to be needed?

## Some interesting sentences (contd)

- The following two sentences are from reviews in the paint domain.
  - “For paint\_X, one coat can cover the wood color.”
  - “For paint\_Y, we need three coats to cover the wood color.”
- We know that paint\_X is good and Paint\_Y is not, but how by a system.
  - Do we need commonsense knowledge and understanding of the text?

## Some more interesting/hard sentences

- “My goal is to have a high quality tv with decent sound”
- “The top of the picture was much brighter than the bottom.”
- “Google steals ideas from Bing, Bing steals market shares from Google.”
- “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.”

Bing Liu, tutorial

165

## Roadmap

- Sentiment Analysis Problem
- Document sentiment classification
- Sentence subjectivity & sentiment classification
- Aspect-based sentiment analysis
- Aspect-based opinion summarization
- Sentiment lexicon generation
- Mining comparative opinions
- Some other problems
- ➔ ■ **Opinion spam detection**
- Utility or helpfulness of reviews
- Summary

Bing Liu, tutorial

166

## Incentives for Opinion Spamming

- **Opinions from social media are increasingly used by individuals and organizations for**
  - making purchase decisions
  - marketing and product design
  - making choices at elections
- **Positive opinions often mean profits and fames for businesses and individuals,**
  - Unfortunately, this gives strong incentives for people to game the system by posting **fake opinions and reviews**.

Bing Liu, tutorial

167

## Opinion spam detection

(Jindal and Liu, 2007, 2008)

- **Opinion spamming refers to people giving fake or untruthful 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.
- **Opinion spamming has become a business in recent years.**
- **Increasing number of customers are wary of fake reviews (biased reviews, paid reviews)**

Bing Liu, tutorial

168

## Problem is wide-spread

### Professional Fake Review Writing Services

- [Post positive reviews](#)
- [Fake review writer](#)
- [Product review writer for hire](#)
- [Hire a content writer](#)

### Manipulating Social Media (sock puppets - fake identities - fake personas)

- [Revealed: US spy operation that manipulates social media](#), Guardian.co.uk, Thursday 17 March 2011.
- [America's absurd stab at systematising sock puppetry](#), Guardian.co.uk, Thursday 17 March 2011.

### China's Internet "Water Army" (Shuijun) - Opinion Spammers

- You can hire people to write and post fake reviews or comments, and even bribe staff at review, forum
- ["Water Army" Whistleblower Threatened](#), January 7, 2011, People's Daily.
- [The Chinese Online "Water Army"](#), June 25, 2010, Wired.com.
- If you read Chinese, see [this description](#) from Baidu Baike at baidu.com.

Bing Liu, tutorial

169

## An example practice of review spam

### Belkin International, Inc

- Top networking and peripherals manufacturer | Sales ~ \$500 million in 2008
- **Posted an ad for writing fake reviews on amazon.com (65 cents per review)**

Timer: 00:00:00 of 60 minutes

Want to work on this HIT? [Accept HIT](#) Want to see other HITs? [Skip HIT](#)

Write Product Reviews 25-50 Words  
Requester: Mike Bayard  
Qualifications Required: HIT approval rate (%) is not less than 95

**Jan 2009**

**Write a Positive 5/5 Review for Product on Website**

Positive review writing.

- Use your best possible grammar and write in US English only
- Always give a 100% rating (as high as possible)
- Keep your entry between 25 and 50 words
- Write as if you own the product and are using it
- Tell a story of why you bought it and how you are using it
- Thank the website for making you such a great deal
- Mark any other negative reviews as "not helpful" once you post yours

Instructions:

The link below leads to a product on a website. Read-through the product's features and write a positive review for it using the guidelines above to the best of your ability. I have also provided the part number for this product and you can click on the links below to see it on several alternative websites. In order to post some reviews you will need to create an account on the site. You can use your own email address or open a new free webmail account (gmail, yahoo...) and use it to post with.

Bing Liu, tutorial

170

## 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!

Bing Liu, tutorial

171

## 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.

Bing Liu, tutorial

172

## One more?

Cameraworld is on my list of top photography/video equipment e-tailers. Their reps answer phones from early in the morning through late at night. The service is also first rate and the staff there is knowledgeable on the products they sell. Prices are competitive, although not always the best, but they do price match should you find it cheaper.

I have noticed that some of the products they carry, only a select few that are rare, are not listed on the website even though Cameraworld either stocks or is willing to get for you. This is only a minor inconvenience, and isn't really a bother to me as I normally have other questions that I can get answered when calling.

They also have a "Bonus Bucks" program in which online purchases receive a percentage credit towards a future purchase. I have yet to make a purchase online (always phoned in orders), so no experience with the program.

Bing Liu, tutorial

173

## Detecting fake review is hard

- Different from Web spam and email
  - Web spam: link spam and content spam
  - Email spam: mostly commercial ads
- For such spam, when you see it, you know it.
  - Easy to find training data for model building
  - Easy to evaluate the resulting models
- Fake reviews (opinion spam in general)
  - No link or content spam
  - Almost no commercial ads

Bing Liu, tutorial

174

## Detecting fake review is hard (contd)

- Fake reviews
  - When you see it, you do not know it.
  - Can only be reliably identified by their authors!
- If one writes carefully, there is almost no way to identify them by their content.
- Logically impossible!
  - I write a truthful 5-star review for a good hotel.
  - But I post the review to another hotel that I want to promote.

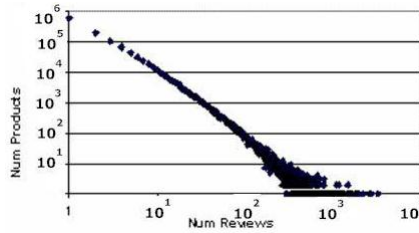
## A Study of Amazon Reviews

- June 2006
  - 5.8mil reviews, 1.2mil products and 2.1mil reviewers.
- A review has 8 parts
  - *<Product ID>*
  - *<Reviewer ID>*
  - *<Rating>*
  - *<Date>*
  - *<Review Title> <Review Body>*
  - *<Number of Helpful feedbacks> <Number of Feedbacks>*

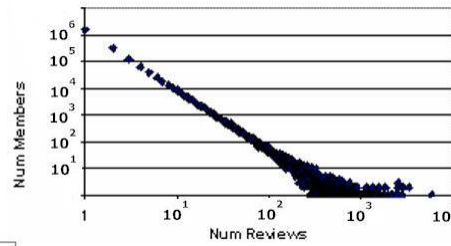


## Log-log plot

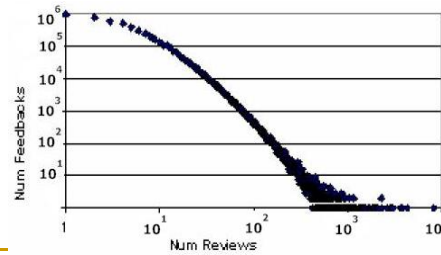
(Jindal and Liu, 2008)



■ Fig. 2 reviews and products



■ Fig. 1 reviews and reviewers



■ Fig. 3 reviews and feedbacks

Bing Liu, tutorial

177

## Star Ratings vs. Percent of Reviews

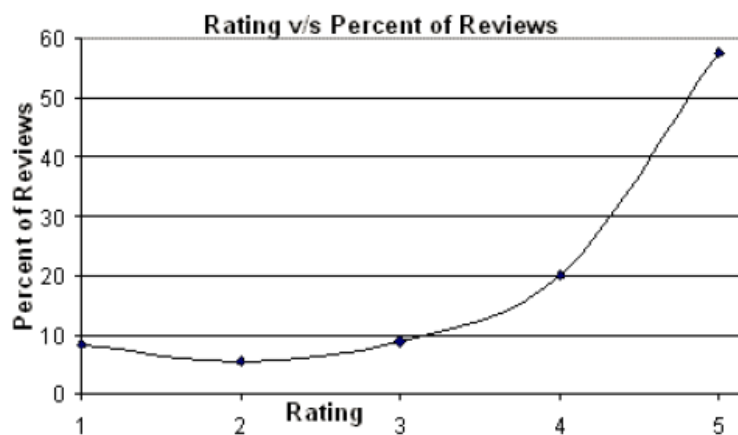


Figure 4. Rating vs. percent of reviews

Bing Liu, tutorial

178

## Categorization of opinion spam

(Jindal and Liu 2008)

- Type 1 (fake reviews)

Ex:

- Type 2 (Reviews on Brands Only)

Ex: "I don't trust HP and never bought anything from them"

- Type 3 (Non-reviews)

- Advertisements

Ex: "Detailed product specs: 802.11g, IMR compliant, ..."

"...buy this product at: compuplus.com"

- Other non-reviews

Ex: "What port is it for"

"The other review is too funny"

"Go Eagles go"

## Fake reviews vs. product quality

Table 4. Spam reviews vs. product quality

	Positive spam review	Negative spam review
Good quality product	1	2
Bad quality product	3	4
Average quality product	5	6

Harmful Regions

## Type of spammers

- Individual spammers:
  - The spammer does not work with anyone. He/she just writes fake reviews him/herself using a single user-id, e.g., the author of a book.
- Group spammers
  - A group of spammers (persons) works in collusion
  - A single person registers multiple user-ids (called *sock puppeting*)

## Type of data and clues

- Review content:
  - The actual text content of each review, linguistic features and style features
- Meta-data about each reviewer:
  - star rating, user-id,
  - time when a review was posted, and time taken to write/post the review,
  - host IP address and MAC address
  - geo-location of the reviewer
  - sequence of clicks at the review site

## Type of data and clues (contd)

- Product information:

- Information about the entity being reviewed, e.g.,
  - the product description,
  - sales volume
  - sales rank.

- Public data vs. site private (internal) data

- Site private data, very useful
- But hard to get by outsiders

## Spam detection (Jindal and Liu 2008)

- Type 2 and Type 3 spam reviews are relatively easy to detect

- Supervised learning, e.g., logistic regression
- It performs quite well, and not discuss it further.

- Type 1 spam (fake) reviews

- Manual labeling is extremely hard
- Propose to use duplicate and near-duplicate reviews as positive training data

## Four types of duplicates

1. Same userid, same product
  2. Different userid, same product
  3. Same userid, different products
  4. Different userid, different products
- The last three types are very likely to be fake!

## Supervised model building

- Logistic regression
  - Training: duplicates as spam reviews (positive) and the rest as non-spam reviews (negative)
- Use the follow features (clues)
  - Review centric features (content)
    - About reviews (contents (n-grams), ratings, etc)
  - Reviewer centric features
    - About reviewers (different unusual behaviors, etc)
  - Product centric features
    - Features about products reviewed (sale rank, etc)

## Predictive power of duplicates

- Representative of all kinds of spam
- Only 3% duplicates accidental
- Duplicates as positive examples, rest of the reviews as negative examples

Table 5. AUC values on duplicate spam reviews.

Features used	AUC
All features	78%
Only review features	75%
Only reviewer features	72.5%
Without feedback features	77%
Only text features	63%

- reasonable predictive power
- Maybe we can use duplicates as type 1 spam reviews(?)

## Tentative classification results

- Negative outlier reviews tend to be heavily spammed
- Those reviews that are the only reviews of products are likely to be spammed
- Top-ranked reviewers are more likely to be spammers
- Spam reviews can get good helpful feedbacks and non-spam reviews can get bad feedbacks
- ...

## Other Supervised Methods

- Li et al. (2011) built a model similar to that in (Jindal and Liu 2008), but
  - Also use sentiment and some other features
  - Manually labeled data
- Ott et al (2011) also used supervised learning.
  - Use Mechanical Turk to write fake reviews
  - Use n-grams as features
- Yoo and Gretzel (2009) also studied deceptive reviews.

## Finding unexpected reviewer behavior

- Move “behind the scenes”
  - to uncover the “secrets” of reviewers by profiling them based on their posted reviews and behaviors
- Lim et al (2010) and Nitin et al (2010) analyze the behavior of reviewers
  - identifying *unusual review patterns* which may indicate suspicious behaviors of reviewers.
- The problem is formulated as finding **unexpected rules and rule groups.**

## Spam behavior models (Lim et al 2010)

- Several unusual reviewer behavior models were identified.
  - Targeting products
  - Targeting groups
  - General rating deviation
  - Early rating deviation
- Their scores for each reviewer are then combined to produce the final spam score.
- Ranking and user evaluation

## Finding unexpected rules (Jindal, Liu, Lim 2010)

- For example, if a reviewer wrote all positive reviews on products of a brand but all negative reviews on a competing brand ...
  - Finding unexpected rules,
    - Data: *reviewer-id*, *brand-id*, *product-id*, and a *class*.
    - Mining: class association rule mining
    - Finding unexpected rules and rule groups, i.e., showing atypical behaviors of reviewers.
- Rule1: Reviewer-1, brand-1 -> positive (confid=100%)
- Rule2: Reviewer-1, brand-2 -> negative (confid=100%)



## The example (cont.)

**Expectation:** Let the subset of data with  $A_j = v_{jk}$  be  $D^{jk}$ . We have

$$E(v_{jk}, A_g \rightarrow C) = \text{entropy}(D^{jk}) \quad (24)$$

**Attribute unexpectedness:** To compute attribute unexpectedness, we first compute the entropy after adding the  $A_g$  attribute:

$$\text{entropy}_{A_g}(D^{jk}) = - \sum_{h=1}^{|A_g|} \frac{|D^{jk}_h|}{|D^v|} \text{entropy}(D^{jk}_h) \quad (25)$$

The unexpectedness is computed as follows (information gain):

$$Au(v_{jk}, A_g \rightarrow C) = \text{entropy}(D^{jk}) - \text{entropy}_{A_g}(D^{jk}) \quad (26)$$

## Confidence unexpectedness

**Rule:** reviewer-1, brand-1  $\rightarrow$  positive [sup = 0.1, conf = 1]

- If we find that on average reviewers give brand-1 only 20% positive reviews (expectation), then reviewer-1 is quite unexpected.

$$Cu(v_{jk} \rightarrow c_i) = \frac{\Pr(c_i | v_{jk}) - E(\Pr(c_i | v_{jk}))}{E(\Pr(c_i | v_{jk}))}$$

$$E(\Pr(c_i | v_{jk}, v_{gh})) = \frac{\Pr(c_i | v_{jk}) \Pr(c_i | v_{gh})}{\Pr(c_i) \sum_{r=1}^m \frac{\Pr(c_r | v_{jk}) \Pr(c_r | v_{gh})}{\Pr(c_r)}}$$

## Support unexpectedness

Rule: reviewer-1, product-1 -> positive [sup = 5]

- Each reviewer should write only one review on a product and give it a positive or negative rating (expectation).
- This unexpectedness can detect those reviewers who review the same product multiple times, which is unexpected.
  - These reviewers are likely to be spammers.
- Can be defined probabilistically as well.

## Detection using review graph

(Wang et al., 2011)

- This study was based on a snapshot of all reviews from resellerratings.com, which were crawled on Oct. 6th, 2010.
  - 343603 reviewers, 408470 reviews, 14561 store
- Form a heterogeneous review graph with three types of nodes,
  - reviewers, reviews and stores,
  - The graph captures their relationships and was used model spamming clues.

## The Relationships

- Three concepts were defined and computed,
  - *trustiness* of reviewers,
  - *honesty* of reviews, and
  - *reliability* of stores.
- A reviewer is more trustworthy if he/she has written more honesty reviews
- A store is more reliable if it has more positive reviews from trustworthy reviewers
- A review is more honest if it is supported by many other honest reviews.

Bing Liu, tutorial

197

## Definitions and equations

- Trustiness of a reviewer  $r$

$$T(r) = \frac{2}{1 + e^{-H_r}} - 1$$

- Honesty of a review  $v$

$$H(v) = |R(\Gamma_v)| A_n(v, \Delta t)$$

- Reliability of store  $s$

$$R(s) = \frac{2}{1 + e^{-\theta}} - 1$$

Bing Liu, tutorial

198

## Detecting group spam (Mukherjee et al 2011, 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

## Big John's Profile

1 of 1 people found the following review helpful:  
★★★★★ Practically FREE music, December 4, 2004  
This review is from: [Audio Xtract \(CD-ROM\)](#)  
I can't believe for \$10 (after rebate) I got a program that gets me free unlimited music. I was hoping it did half what was ....

3 of 8 people found the following review helpful:  
★★★★★ Yes – it really works, December 4, 2004  
This review is from: [Audio Xtract Pro \(CD-ROM\)](#)  
See my review for Audio Xtract - this PRO is even better. This is the solution I've been looking for. After buying iTunes, ....

5 of 5 people found the following review helpful:  
★★★★★ My kids love it, December 4, 2004  
This review is from: [Pond Aquarium 3D Deluxe Edition](#)  
This was a bargain at \$20 - better than the other ones that have no above water scenes. My kids get a kick out of the ....

## Cletus' Profile

2 of 2 people found the following review helpful:

★★★★★ Like a tape recorder..., December 8, 2004

This review is from: [Audio Xtract \(CD-ROM\)](#)

This software really rocks. I can set the program to record music all day long and just let it go. I come home and my ....

3 of 10 people found the following review helpful:

★★★★★ This is even better than..., December 8, 2004

This review is from: [Audio Xtract Pro \(CD-ROM\)](#)

Let me tell you, this has to be one of the coolest products ever on the market. Record 8 internet radio stations at once, ....

5 of 5 people found the following review helpful:

★★★★★ For the price you..., December 8, 2004

This review is from: [Pond Aquarium 3D Deluxe Edition](#)

This is one of the coolest screensavers I have ever seen, the fish move realistically, the environments look real, and the ....

## Jake's Profile

★★★★★ Wow, internet music! ..., December 4, 2004

This review is from: [Audio Xtract \(CD-ROM\)](#)

I looked forever for a way to record internet music. My way took a long time and many steps (frustrating). Then I found Audio Xtract. With more than 3,000 songs downloaded in ...

2 of 9 people found the following review helpful:

★★★★★ Best music just got ..., December 4, 2004

This review is from: [Audio Xtract Pro \(CD-ROM\)](#)

The other day I upgraded to this TOP NOTCH product. Everyone who loves music needs to get it from Internet ....

3 of 3 people found the following review helpful:

★★★★★ Cool, looks great..., December 4, 2004

This review is from: [Pond Aquarium 3D Deluxe Edition](#)

We have this set up on the PC at home and it looks GREAT. The fish and the scenes are really neat. Friends and family ....

## Finding candidate groups

- Frequent itemset mining
  - Items → Reviewer Ids (rids).
  - Transaction → set of rids for a product
- Frequent itemsets give us
  - “reviewer groups” that reviewed multiple products together
- Using reviews of manufactured products,
  - Found 7052 candidate groups
  - Minimum support count = 3

Bing Liu, tutorial

203

## A set of clues (or features)

- Group Time Window (GTW)
- Group Deviation (GD)
- Group Content Similarity (GCS)
- Group Member Content similarity (GMCS)
- Group Early Time Frame (GETF)
- Group Size Ratio (GSR)
- Group Size (GS)
- Group Support Count (GSUP)

Bing Liu, tutorial

204

## A relational model and algorithm

### Algorithm: GSRank

Input: Weight matrices  $W_{PG}$ ,  $W_{MP}$ , and  $W_{GM}$

Output: Ranked list of candidate spam groups

1. Initialize  $V_G^0 \leftarrow [0.5]_{|G|}$ ;  $t \leftarrow 1$ ;
2. Iterate:
  - i.  $V_P \leftarrow W_{PG} V_G^{(t-1)}$ ;  $V_M \leftarrow W_{MP} V_P$ ;
  - ii.  $V_G \leftarrow W_{GM} V_M$ ;  $V_M \leftarrow W_{GM}^T V_G$ ;
  - iii.  $V_P \leftarrow W_{MP}^T V_M$ ;  $V_G^{(t)} \leftarrow W_{PG}^T V_P$ ;
  - iv.  $V_G^{(t)} \leftarrow V_G^{(t)} / \|V_G^{(t)}\|_1$ ;until  $\|V_G^{(t)} - V_G^{(t-1)}\|_\infty < \delta$
- 3 Output the ranked list of groups,  $V_G^*$

## Roadmap

- Sentiment Analysis Problem
- Document sentiment classification
- Sentence subjectivity & sentiment classification
- Aspect-based sentiment analysis
- Aspect-based opinion summarization
- Sentiment lexicon generation
- Mining comparative opinions
- Some other problems
- Opinion spam detection
- ➡ ■ **Utility or helpfulness of reviews**
- Summary

## Utility or quality of reviews

- **Goal:** Determining the helpfulness, or utility of each review (not necessarily fake)
  - It is desirable to rank reviews based on utilities or qualities when showing them to users, with the highest quality review first.
- Many review aggregation sites have been practicing this, e.g., amazon.com.
  - “*x of y people found the following review helpful.*”
  - Voted by user - “*Was the review helpful to you?*”

Bing Liu, tutorial

207

## Application motivations

- Although review sites use helpfulness feedback to rank their reviews,
  - A review takes a long time to gather enough feedback.
    - New reviews will not be read.
  - Some sites do not provide feedback information.
- It is thus beneficial to score each review once it is submitted to a site.

Bing Liu, tutorial

208



## Regression formulation

(Zhang and Varadarajan, 2006; Kim et al. 2006)

- **Formulation:** Determining the utility of reviews is usually treated as a **regression** problem.
  - A set of features is engineered for model building
  - The learned model assigns an utility score to each review, which can be used in review ranking.
- Unlike fake reviews, the ground truth data used for both training and testing are available
  - Usually the user-helpfulness feedback given to each review.

## Features for regression learning

- Example features include
  - review length, review rating, counts of some POS tags, opinion words, tf-idf scores, wh-words, product aspect mentions, comparison with product specifications, timeliness, etc (Zhang and Varadarajan, 2006; Kim et al. 2006; Ghose and Ipeirotis 2007; Liu et al 2007)
- Subjectivity classification was applied in (Ghose and Ipeirotis 2007).
- Social context was used in (O'Mahony and Smyth 2009; Lu et al. 2010).

## Classification formulation

- **Binary classification:** Instead of using the original helpfulness feedback as the target or dependent variable,
  - Liu et al (2007) performed manual annotation of two classes based on whether the review evaluates many product aspects or not.
- Binary class classification is also used in (O'Mahony and Smyth 2009)
  - Classes: Helpful and not helpful
  - Features: helpfulness, content, social, and opinion

Bing Liu, tutorial

211

## Roadmap

- Sentiment Analysis Problem
- Document sentiment classification
- Sentence subjectivity & sentiment classification
- Aspect-based sentiment analysis
- Aspect-based opinion summarization
- Sentiment lexicon generation
- Mining comparative opinions
- Some other problems
- Opinion spam detection
- Utility or helpfulness of reviews

➔ ■ **Summary**

Bing Liu, tutorial

212

## Summary

- **This tutorial presented**
  - The problem of sentiment analysis and opinion mining
    - It provides a structure to the unstructured text.
    - It shows that summarization is crucial.
  - Main research directions and their representative techniques.
  - By no means exhaustive, a large body of work.
- Still many problems not attempted or studied.
- **None of the problem is solved.**

## Summary (contd)

- **A fascinating NLP or text mining problem.**
  - Every sub-problem is highly challenging.
  - But it is also highly restricted (semantically).
- **Despite the challenges, applications are flourishing!**
  - It is useful to every organization and individual.
- The general NLP is probably too hard, but **can we solve this highly restricted problem?**
  - **I am optimistic**

## References

All references are in the

### New Book

- Bing Liu. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers. May 2012.
- *Book Draft*:  
<http://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>

