

# Sentiment Analysis

**Dr. Korra Sathya Babu**

Assistant Professor  
Department of Computer Science & Engineering  
National Institute of Technology Rourkela



## Table of Contents

### 1 Introduction

- motivation
- Basics

### 2 Sentiment Analysis

- Sentimen Analysis Problem
- Classification
- Other topics

### 3 Selected References





# Motivation

## What other people think?

This has always been an important piece of information for most of us during the decision-making process. Long before awareness of the World Wide Web became widespread, many of us asked our friends to recommend an auto mechanic or to explain who they were planning to vote for in local elections, requested reference letters regarding job applicants from colleagues, or consulted Consumer Reports to decide what dishwasher to buy. But the Internet and the Web have now (among other things) made it possible to find out about the opinions and experiences of those in the vast pool of people that are neither our personal acquaintances nor well-known professional critics.



# Motivation

## Few Statistics

- 81% of Internet users have done online research on a product at least once
- Readers of online reviews of restaurants, hotels, and various services (e.g., travel agencies or doctors), between 73% and 87% report that reviews had a significant influence on their purchase
- consumers report being willing to pay from 20% to 99% more for a 5-star-rated item than a 4-star-rated item
- 32% have provided a rating on a product, service, or person via an online ratings system, and 30% (including 18% of online senior citizens) have posted an online comment or review regarding a product or service



# What is Sentiment ?

- Sentiment = Feelings
  - Attitudes
  - Opinions
  - Emotions



# What is Sentiment ?

- Sentiment = Feelings
  - Attitudes
  - Opinions
  - Emotions
- Subjective impressions, not facts



# What is Sentiment ?

- Sentiment = Feelings
  - Attitudes
  - Opinions
  - Emotions
- Subjective impressions, not facts
- In generally, a binary opposition in opinions is assumed
  - for/against
  - like/dislike
  - good/bad
  - yes/no



# What is Sentiment ?

- Sentiment = Feelings
  - Attitudes
  - Opinions
  - Emotions
- Subjective impressions, not facts
- In generally, a binary opposition in opinions is assumed
  - for/against
  - like/dislike
  - good/bad
  - yes/no
- Some sentiment analysis jargon:
  - Semantic orientation
  - Polarity



# Synonyms

They mean a judgment one holds as true

- **Opinion** implies a conclusion thought out yet open to dispute
- **View** suggests a subjective opinion
- **Belief** implies often deliberate acceptance and intellectual assent
- **Conviction** applies to a firmly and seriously held belief
- **Persuasion** suggests a belief grounded on assurance (as by evidence) of its truth
- **Sentiment** suggests a settled opinion reflective of one's feelings



# What is Sentiment Analysis

## Definition 1: Sentiment Analysis

- It is a part of natural language processing (NLP) that analyse ones' attitude and opinion in social media, including Facebook, Twitter, reviews, blogs etc.
- The main aim of sentiment analysis is to identify the overall polarity (positive, negative, or neutral) in a given piece of text.

## Definition 2: Sentiment Analysis

Using NLP, statistics, or machine learning methods to extract, identify, or otherwise characterize the sentiment content of a text unit



# Introduction

- **Opinion mining or sentiment analysis**
  - Computational study of opinions, sentiments, subjectivity, evaluations, attitudes, appraisal, affects, views, emotions, 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 our behaviors.
- Our beliefs and perceptions of reality are conditioned on how others see the world.
- Whenever we need to make a decision, we often seek out the opinions of others. In the past,
  - Individuals: seek opinions from friends and family
  - Organizations: use surveys, focus groups, opinion polls, consultants.



# Social Media and 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



# 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.



# Structure the Unstructured data

- **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.



# Problem Statement

■ It consists of two aspects of abstraction

(1) Opinion definition. What is an opinion?

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

(2) Opinion summarization. why?

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

- **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.



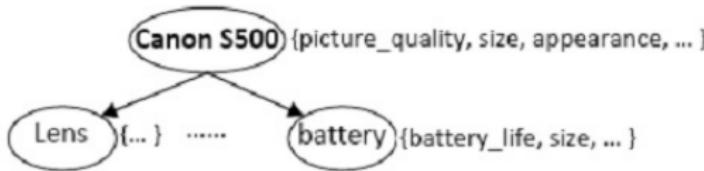
# A (regular) opinion

- Opinion (a restricted definition)
  - An opinion (or regular opinion) is simply a **positive or negative** sentiment, view, attitude, emotion, or appraisal about **an entity or an aspect of the entity** (Hu and Liu 2004; Liu 2006) from an **opinion holder** (Bethard et al 2004; Kim and Hovy 2004; Wiebe et al 2005).
- Sentiment orientation of an opinion
  - Positive, negative, or neutral (no opinion)
    - Also called *opinion orientation*, *semantic orientation*, *sentiment polarity*.



# Entity and Aspect

- **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.



# Opinion definition

- 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 feature  $a_{jk}$  of entity  $e_j$  at time  $t_l$ .  
 $so_{ijkl}$  is +ve, -ve, or neu, or more granular ratings.
- $h_i$  is an opinion holder.
- $t_l$  is the time when the opinion is expressed.



## Remarks on 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.



## Remarks continued.

- 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 **seat** of this car is really **ugly**.”
    - “**seat**” is a part of the car and “**appearance**” (implied by **ugly**) is an aspect of “**seat**” (not the car).
  - 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.



## 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  $(e_j, f_{jk}, so_{ijkl}, h_i, t_l)$ ,
  - 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, speculations (Wiebe 2000; Wiebe et al 2004; Riloff et al 2006).
  - 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.



# 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”.



# Sentiment, subjectivity, and emotion

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



## 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 fact = any number of the same fact
- But for opinion documents, it is different because opinions have a quantitative side & have targets
  - 1 opinion  $\neq$  a number of opinions
  - Aspect-based summary is more suitable
    - Quintuples form the basis for opinion summarization



# Aspect-based opinion summary

“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**, ...”

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

## Feature Based Summary of iPhone:

### Feature1: Touch screen

Positive: 212

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

...

Negative: 6

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

...

### Feature2: voice quality

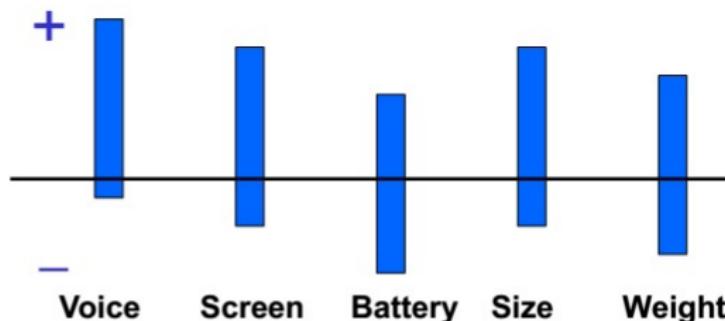
...

Note: We omit opinion holders

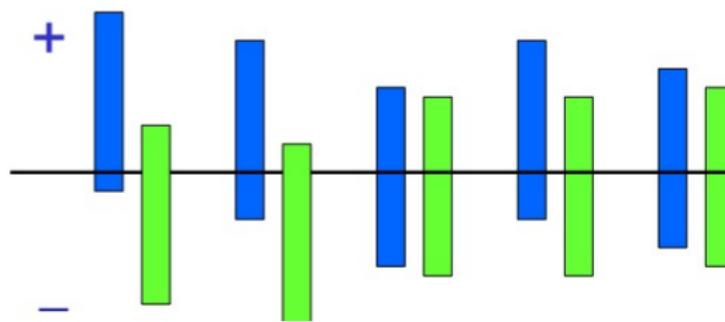


# Opinion Observer

- Summary of reviews of
- Cell Phone 1



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



# Aspect-based opinion summary

The screenshot shows a Bing search results page for the query "HP printer". The top navigation bar includes the Bing logo, a search bar with "HP printer", and a magnifying glass icon. Below the search bar, the word "SHOPPING" is displayed. The main result is for the "HP LaserJet 1020 - printer - B/W - laser, 15ppm, USB". It features a thumbnail image of the printer, a price of "from \$179 (2 stores)" with a "Bing cashback 3%", and a rating of "★★★★☆ user reviews (177)". A brief description follows: "The HP LaserJet 1020 Printer, an excellent laser printer for the cost-co...". To the left of the main result, there is a sidebar titled "ALL RESULTS" under "Shopping". It lists "POPULAR FEATURES" with horizontal bar charts: "Affordability" (green), "Speed" (blue), "Print Quality" (green), "Reliability" (green), "Ease Of Use" (green), "Brand" (green), "Installation" (green), "Size" (green), and "Compatibility" (green). At the bottom right of the page is the Indian Government logo.

ALL RESULTS

Shopping

POPULAR FEATURES

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

HP printer

SHOPPING

HP LaserJet 1020 - printer - B/W - laser, 15ppm, USB

from \$179 (2 stores) Bing cashback 3%

★★★★☆ user reviews (177)

The HP LaserJet 1020 Printer, an excellent laser printer for the cost-co...  
high-quality LaserJet printing in a compact size, and at a price you can...

user reviews | product details | expert reviews | compare prices |

user reviews view: positive comments (44)

speed 96%

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

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

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

# Google Product Search

Google products

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

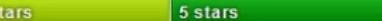
[Overview](#) - [Online stores](#) - [Nearby stores](#) - [Reviews](#) - [Technical specifications](#) - [Similar items](#) - [Accessories](#)

 \$140 [online](#), \$170 [nearby](#)

 159 reviews  0

### Reviews

Summary - Based on 159 reviews

#### What people are saying

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



# 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)
- ...



# 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



# Document Sentiment Classification



# 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.**



# 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.



# Assumption and goal

- **Assumption:** The doc is written by a single person and express opinion/sentiment on a single entity.
- **Goal:** discover  $(e, a, h, t)$ ,  
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 bucks 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

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

Was this review helpful to you?

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  
Johnathan Parker

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

- Data: reviews from opinions.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.



# Patterns of POS tags

First word	Second word	Third word <u>(Not Extracted)</u>
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



# Unsupervised classification

- 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.



# Unsupervised classification

- 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

- 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



## 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, Wei 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)



# Sentence subjectivity and sentiment classification



# 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).



# 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."



## 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 extend) 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.”



# Aspect-based sentiment analysis



# 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.



# 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.



# 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**.



# Explicit and Implicit aspects

- **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.



# 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” ⇒ aspect “price,” “beautiful” ⇒ aspect “appearance”, “heavy” ⇒ aspect “weight”
- Although manual mapping is possible, in different contexts, the meaning can be different.
  - E.g., “The computation is expensive”.



# Identify aspect synonyms

- 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.



# 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)



# A lexicon-based method

- **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  $w_i$ .



# Sentiment shifters

- 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

- 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

- 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



## 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



## 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

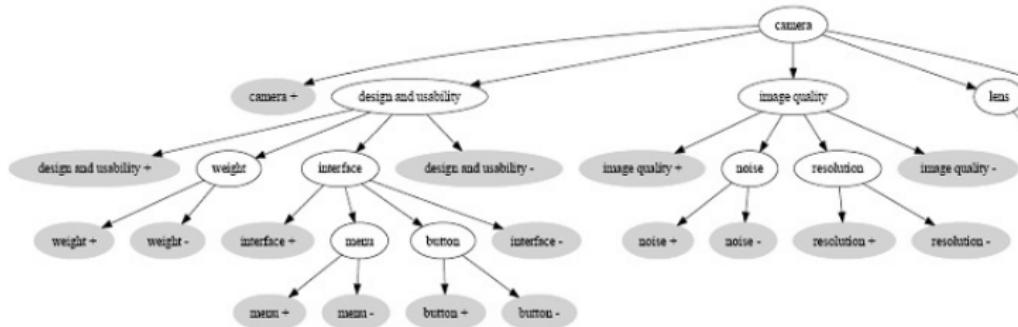
| consume a large quantity of or more resource

| consume no, little or less waste



# Sentiment ontology tree

- 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.



## Sentiment ontology tree continued..

- The leaves of the tree are positive or negative sentiments.
- It then uses a hierarchical classification model to learn to assign a 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.



# 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.



## Text summary of opinions

- One can also generate a summary in the **traditional 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.



## Select and order sentences

- 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 mention fewest aspects (the sentence is focused).
  - Order the sentences: It uses an ontology to map sentences to the ontology nodes (domain concepts).



# Opinion lexicon

- Opinion 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



# Corpus-based approaches

- Rely on syntactic patterns in large corpora.
  - Can find domain dependent orientations (positive, negative, or neutral).
- 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



# 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.



# Opinions implied by resource usage

- 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



# Dictionary-based methods

- 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.



## 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.



# Comparative Opinions



# Comparative Opinions

## ■ *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*”



# 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.



# 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 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

- 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?”



# Some interesting sentences

- 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 interesting sentences

- 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

- 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?



# Opinion Spam

- 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)



## Utility or quality of reviews

- **Goal:** Determining the usefulness, helpfulness, or utility of each review.
  - 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?*”



# Available tools for text processing

Name of the tool	Purpose
TweetMotif	Tokenization of tweets
POS tagger	Twitter POS tagger
TweetNLP <sup>a</sup>	Twitter natural language processing
Lancaster stemming algorithm	Stemmer
GNU Aspell	Spell checker
Snowball	English stemmer
Stanford Log-linear Part-Of-Speech Tagger	POS tagger
Tweeboparser	Tweet dependency parser



# Publicly available datasets

S#	Data set	Type	Lang.	Web resource	Details
1	Stanford large movie data set	Movie Reviews	English	<a href="http://ai.stanford.edu/~amaas/data/sentiment/">http://ai.stanford.edu/~amaas/data/sentiment/</a>	Movie Reviews
2	COAE2008	Product Reviews	Chinese	<a href="http://ir-china.org.cn/coae2008.html">http://ir-china.org.cn/coae2008.html</a>	
3	Boacar	Car Reviews	Chinese	<a href="http://www.riche.com.cn/boacar/">http://www.riche.com.cn/boacar/</a>	
4	[187]	Reviews, forums	English	<a href="http://ufaka.cs.uiuc.edu/~wang296/Data/">http://ufaka.cs.uiuc.edu/~wang296/Data/</a>	
5	[188]	Reviews	English	<a href="http://julab.kaist.ac.kr/resources/WSDM11">http://julab.kaist.ac.kr/resources/WSDM11</a>	
6	Movie-v2.0	Movie Reviews	English	<a href="http://www.cs.cornell.edu/people/pabo/movie-review-data/">http://www.cs.cornell.edu/people/pabo/movie-review-data/</a>	
7	Multi-domain	Multi-domain	English	<a href="http://www.cs.jhu.edu/~ndreze/datasets/sentiment">http://www.cs.jhu.edu/~ndreze/datasets/sentiment</a>	
8	SkyDrive de Hermit Dave	Spanish Word Lists	Spanish	<a href="https://skydrive.live.com/?cid=3732e80b128d016fb&amp;id=3732e80b128d016fb&amp;t=213584">https://skydrive.live.com/?cid=3732e80b128d016fb&amp;id=3732e80b128d016fb&amp;t=213584</a>	
9	TripAdvisor	Reviews	Spanish	<a href="http://cltc.ub.edu/courses/node/106">http://cltc.ub.edu/courses/node/106</a>	18,000 customer reviews on hotels and restaurants from Hopinion
10	[38]	Multi-Domain	English	<a href="http://www2.cs.cmu.edu/~lrb/FBS/sentiment-analysis.html">http://www2.cs.cmu.edu/~lrb/FBS/sentiment-analysis.html</a>	6800 opinion words on 10 different products
11	TBOD [144]	Reviews	English		Product Review on Cars, Headphones, Hotels
12	[68]	Product Reviews	English	<a href="http://www.lsi.us.es/_fermin/index.php/Datasets">http://www.lsi.us.es/_fermin/index.php/Datasets</a>	Product Reviews from <a href="#">Opinion.com</a> on headphones 587 reviews, hotels 988 reviews and cars 972 reviews
13	[148]	Movie Reviews	Turkish	<a href="http://www.win.tue.nl/~mpechen/projects/smm/#Datasets">http://www.win.tue.nl/~mpechen/projects/smm/#Datasets</a>	5331 positive and 5331 negative reviews on movie
14	[148]	Product Reviews	Turkish	<a href="http://www.win.tue.nl/~mpechen/projects/smm/#Datasets">http://www.win.tue.nl/~mpechen/projects/smm/#Datasets</a>	700 +ve & 700 -ve reviews on books, DVD, electronics, kitchen appliances
15	ISEAR	English sentences	English	<a href="http://www.affective-sciences.org/system/files/page/2636/ISEAR.zip">www.affective-sciences.org/system/files/page/2636/ISEAR.zip</a>	The dataset contains 7666 such statements, which include 18,146 sentences, 449,060 running words.
16	[149]	Product Reviews	English	<a href="http://www.cs.jhu.edu/~mdredze/datasets/sentiment/">http://www.cs.jhu.edu/~mdredze/datasets/sentiment/</a>	Amazon reviews on 4 domain (books, DVDs, electronics, kitchen appliances)
17	DUC data, NIST	Texts	English	<a href="http://www.nlpri.nist.gov/projects/duc/data.html, http://www.nist.gov/tac/data/index.html">http://www.nlpri.nist.gov/projects/duc/data.html, http://www.nist.gov/tac/data/index.html</a>	Text summarization data
18	[70]	Restaurant and Hotel Reviews	English	<a href="http://julab.kaist.ac.kr/research/WSDM11">http://julab.kaist.ac.kr/research/WSDM11</a>	Restaurant and Hotel Reviews from Amazon and Yelp
19	[114]	Restaurant Reviews	Cantonese	<a href="http://www.opentrace.com">http://www.opentrace.com</a>	Reviews on restaurant
20	[125]	Biographical Articles	Dutch	<a href="http://www.iisg.nl/bwsa">http://www.iisg.nl/bwsa</a>	574 Biographical articles
21	Spin3r dataset	Multi-Domain	English	<a href="http://www.icwsm.org/2011/data.php">http://www.icwsm.org/2011/data.php</a>	
22	[86]	Ironic Dataset	English	<a href="http://users.dsic.upv.es/grupos/nle/">http://users.dsic.upv.es/grupos/nle/</a>	3163 ironic reviews on five products
23	HASH [179]	Tweets	English	<a href="http://demeter.inf.ed.ac.uk">http://demeter.inf.ed.ac.uk</a>	31,861 Pos tweets, 64,850 Neg tweets, 125,859 Neu tweets
24	EMOT [179]	Tweets and Emotions	English	<a href="http://twittersentiment.appspot.com">http://twittersentiment.appspot.com</a>	230,811 Pos & 150,570 Neg tweets
25	ISIEVE [179]	Tweets	English	<a href="http://www.i-sieve.com">http://www.i-sieve.com</a>	1520 Pos tweets, 200 Neg tweets, 2295 Neu tweets
26	[177]	Tweets	English	<a href="mailto:apoorn@cs.columbia.edu">e-mail: apoorn@cs.columbia.edu</a>	11,875 tweets
27	[52]	Opinions	English	<a href="http://patientopinion.org.uk">http://patientopinion.org.uk</a>	2000 patient opinions
28	[96]	Tweets	English	<a href="http://goo.gl/lQvdv">http://goo.gl/lQvdv</a>	667 tweets
29	[39]	Movie Reviews	English	<a href="http://ai.stanford.edu/~amaas/data/sentiment/">http://ai.stanford.edu/~amaas/data/sentiment/</a>	50,000 movie reviews
30	[164]	Tweets	English	<a href="http://cscs.stanford.edu/people/alecmgo/trainingandtestdata.zip">http://cscs.stanford.edu/people/alecmgo/trainingandtestdata.zip</a>	400 deceptive and 400 truthful reviews in positive and negative category. Last
31	[210]	Spam Reviews	English	<a href="http://mylerott.com/op_spam">http://mylerott.com/op_spam</a>	Accessed by: 12 April, 2015
32	[230]	Sarcasm and nasty reviews	English	<a href="https://nlds.soe.ucsc.edu/jac">https://nlds.soe.ucsc.edu/jac</a>	1000 discussions, ~390,000 posts, and some ~73,000,000 words

# What is Sarcasm

## Sentiment Analysis

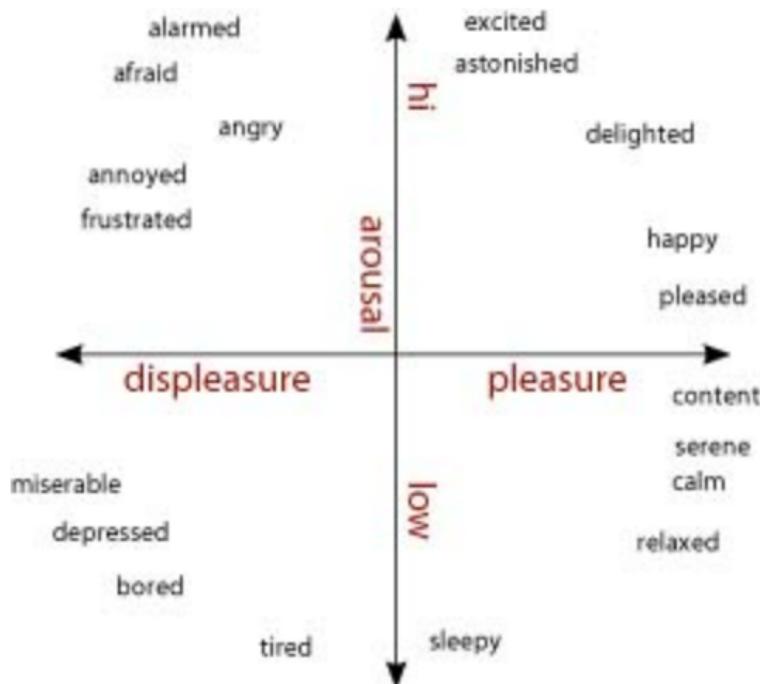
- It is a part of natural language processing (NLP) that analyse ones' attitude and opinion in social media, including Facebook, Twitter, reviews, blogs etc.
- The main aim of sentiment analysis is to identify the overall polarity (positive, negative, or neutral) in a given piece of text.

## Sarcasm

Sarcasm is a special type of sentiment that comprise of words which mean the opposite of what you really want to say (especially in order to insult or wit someone, to show irritation or to be funny). Ex: Nothing I love more than a crowded library with no seats :-) #sarcasm.



# Dimensions of Emotions



## The final remarks

At the end of the day, sentiment analysis techniques are going to be measured not by the traditional metrics that scientists use to measure them, accuracy, speed, scalability, etc., but by their ability to deliver value, and to enable people to make the decisions that help their company or their organization do better. – VS Subrahmanian (Keynote at Sentiment Analysis Symposium 2013)



# Selected References

① Kumar Ravi and Vadlamani Ravi

A survey on opinion mining and sentiment analysis: Tasks, approaches and applications.

Knowledge-Based Systems 89 (2015) 14 - 46.

② Bo Pang and Lillian Lee

Opinion mining and sentiment analysis.

Foundations and Trends in Information Retrieval.

③ Bing Liu,

Sentiment Analysis and Opinion Mining.

Morgan & Claypool Publishers.

④ Steven Bird, Ewan Klein, and Edward Loper

Natural Language Processing with Python

O'Reilly Media, Inc.



# Thank You !

