



# SOCIAL MEDIA & NETWORK ANALYTICS

Sentiment  
Analysis



# Acknowledgments

- This slides based on
  - Tutorial by Bing Liu, AAAI 2011 Tutorial
  - Tutorial by Ronen Feldman

# Outline of Lecture

- Introduction to Sentiment Analysis
  - What is it?
  - Motivation
- Types of Sentiment Analysis
  - Sentiment & Subjectivity Classification
  - Aspect Extraction
  - Lexicon Generation

# Pixel 2 Reviews on Amazon

★★★★★ Glad I made the switch from Apple to Google!!

February 2, 2018

Color: N/A | Size: 64 GB | **Verified Purchase**

Great phone!! I finally decided to ditch Apple, and was torn between the Samsung Galaxy and the Pixel. After reading about how fragile the Samsung screens are, I decided on the Pixel. The phone is not as user friendly as the iPhone, but once you get it set up to your liking, it's amazing. The camera is terrific. And it charges extremely quickly. I've never had a phone that lasted all day (and evening!) on a single charge. I'll never go back to Apple!

[Comment](#) | 8 people found this helpful. Was this review helpful to you?   [Report abuse](#)



★☆☆☆☆ had to wait for the phone to cool down, it took hours

By [Amazon Customer](#) on February 19, 2018

Color: N/A | Size: 64 GB | **Verified Purchase**

this product has a heating problem. my battery over heated to many times, had to wait for the phone to cool down, it took hours. there explanation was to many apps or games being played. well i didn't i even deleted as many apps as possible. had a had time deleting. it didn't help also every time i would try to reach the phone app a new app would pop up trying to sell me on there products. half the time i couldn't answer the phone could not fine the app to answer the call. I am very disappointed in this product, i took the word of a good friend on how well this phone responseds. only thing that works good was the voice activator. i now have to fined out how to return this product and get my money back a phone that costs what it does should not have these problems. very in happy.

► [Comment](#) | 2 people found this helpful. Was this review helpful to you?   [Report abuse](#)

# What is Sentiment / opinion?

- What is sentiment?
  - A **view** or **opinion** or **subjective** statement **someone** holds for **something** at a given time (and location).
  - Sometimes sentiment called opinion (both terms are used interchangeably)

Entity (something)

Opinion holder (someone)

“Great phone! I finally decided to ditch Apple, and was torn between the Samsung Galaxy and the Pixel. After reading about how fragile the Samsung screens are, I decided on the Pixel. The phone is not as user friendly as the iPhone, but once you get it set up to your liking, it’s amazing. The camera is terrific. And it charges extremely quickly. I’ve never had a phone that lasted all day (and evening) on a single charge. I’ll neve go back to Apple!”

Sentiment value/orientation

# Sentiment Analysis

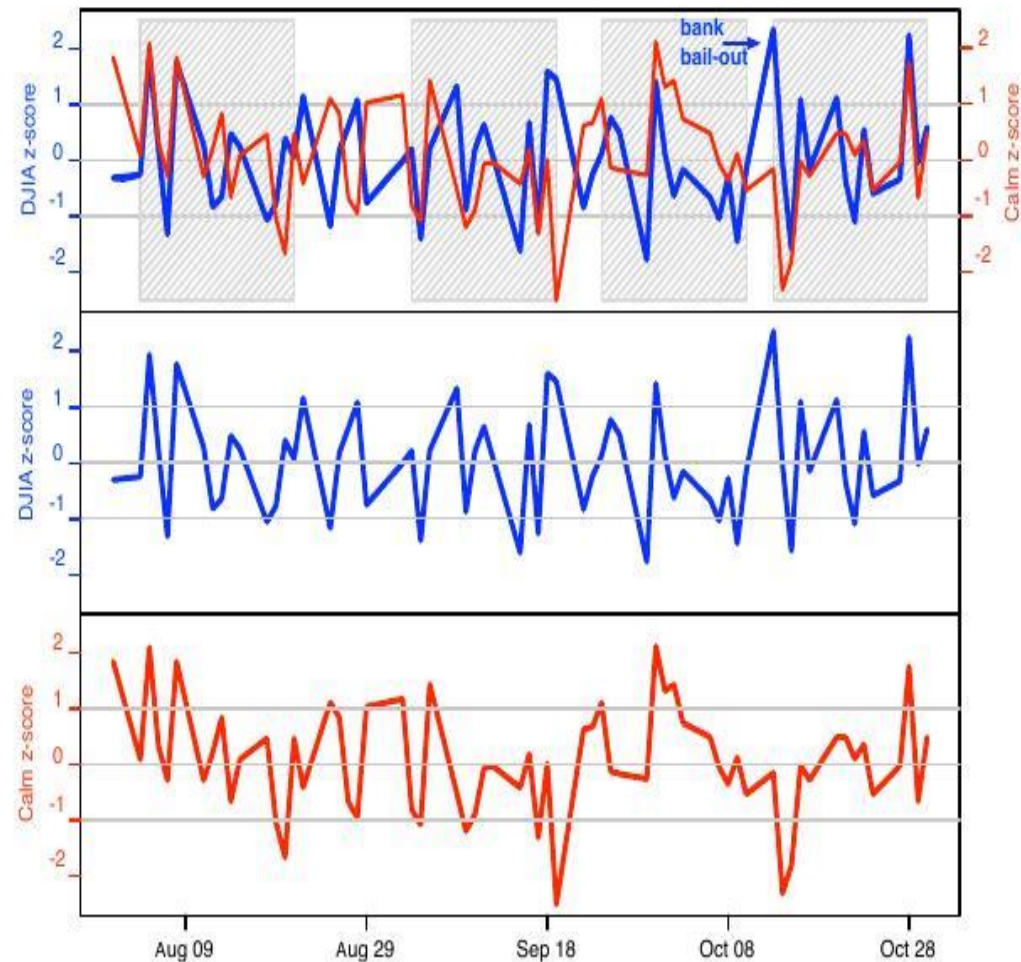
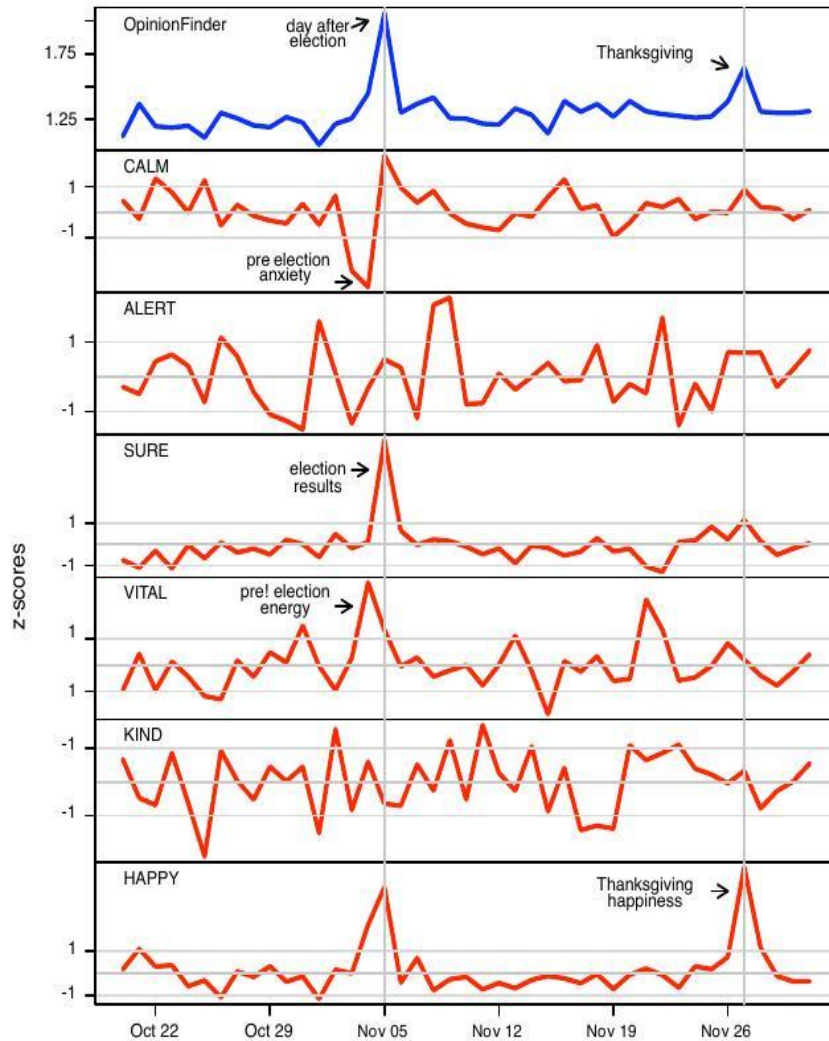
- What is Sentiment Analysis?
  - Use of computational approaches (NLP, text mining, machine learning) to identify, extract and study sentiment/opinions
- Identification of:
  - Who holds this opinion?
  - What is this sentiment/opinion towards?
  - What is the sentiment orientation?
- Other names known by:
  - Sentiment mining, opinion mining, opinion extraction, subjectivity analysis

# Why is Sentiment Analysis Important?

- Decision making
  - Our beliefs and perceptions of reality are conditioned on how others see the world, and we often seek opinions of others when we make a decision
  - E.g., buy products and services
- Market and Social Research (social sensing)
  - Business spend a huge amount of money to find consumer opinions, using consultants, surveys and focus groups, etc
  - Find public opinion about political candidates and issues
  - Predict elections, stock market etc.
- Ads placements:
  - Place an ad if a user praises product, place competitor ad if a user criticises product



# Using Twitter to Analyse Mood on Stock Market



J. Bollen, H. Mao, X. Zeng, "Twitter Mood predicts the stock market", arXiv:1010.3003v1 [cs.CE], 2010.



# Sources of Sentiment/Opinion in Social Media and Networks

- 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 on social networking sites, e.g, Facebook
- Organisation internal data
  - Customer feedback from emails, call centres, etc
- News and reports
  - Opinions in news articles and commentaries

# More Concise Sentiment/Opinion Definition

- Sentiment / Opinion
  - An opinion is a **positive or negative** sentiment, view, attitude, emotion or appraisal about **an entity** or **an aspect of an entity** from an **opinion holder**, held at a particular **time (and context)**
- Sentiment orientation of an opinion
  - Positive, negative, or neutral (no opinion)
  - Can be a numeric rating, e.g., -5 to 5
  - Also called opinion orientation, semantic orientation, sentiment polarity

# Definition of Opinion/Sentiment

- An opinion/sentiment is a 5 element tuple:

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

where

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

# Some remarks about the definition

- Although introduced using a product review, the definition is generic
- $(e_j, a_{jk})$  is also called the opinion target
  - Opinion without knowing the target **is of limited use**
- $t_l$  can be generalised to context
  - E.g., while on holiday in Cairns you like ice-cream, but perhaps not in winter in Greenland

# Example of the 5 element Tuple Identification

- ABC123 on 2/2/2018: "Great phone! I finally decided to ditch Apple, and was torn between the Samsung Galaxy and the Pixel. After reading about how fragile the Samsung screens are, I decided on the Pixel. The phone is not as user friendly as the iPhone, but once you get it set up to your liking, it's amazing. The camera is terrific. And it charges extremely quickly. I've never had a phone that lasted all day (and evening) on a single charge. I'll neve go back to Apple!"
- Tuples:
  - (Pixel, GENERAL, +ve, ABC123, 2/2/2018)
  - (iPhone, camera, +ve, ABC123, 2/2/2018)
  - ...

# Challenges of Sentiment Analysis

- People express sentiment and opinions in complex ways
- Lexical content alone can be misleading
  - E.g., "good" vs "not good"
- Direct vs indirect opinions
  - E.g., "The camera is terrific" vs "After taking the drug, my pain was gone"
- Regular vs comparative opinions
  - E.g., "The camera is terrific" vs "The phone is not as user friendly as the iPhone"
- Labels depend on humans who always have some inter-labeller disagreement (typically about 20%)

# Easier and harder problems

- Tweets from Twitter are the easiest
  - Short and thus usually straight to the point
  - Generally one sentiment for whole tweet/document (document level sentiment analysis)
- Reviews are next
  - Entities are given (almost) and there is little noise
  - Might have one sentiment per sentence (sentence level)
- Discussion, comments and blogs are hard
  - Multiple entities, comparisons, noisy, sarcasm, etc
- Determining sentiments is relatively easy
- Extracting entities and aspects is harder
- Combining both is even harder
  - Aspect level analysis



# Sentiment Analysis Resolution

- ABC123 on 2/2/2018: "Great phone! I finally decided to ditch Apple, and was torn between the Samsung Galaxy and the Pixel. After reading about how fragile the Samsung screens are, I decided on the Pixel. The phone is not as user friendly as the iPhone, but once you get it set up to your liking, it's amazing. The camera is terrific. And it charges extremely quickly. I've never had a phone that lasted all day (and evening) on a single charge. I'll neve go back to Apple!"
- 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**

# Outline of Lecture

- Introduction to Sentiment Analysis
  - What is it?
  - Motivation
- Types of Sentiment Analysis
  - Sentiment & Subjectivity Classification
  - Aspect Extraction
  - Lexicon Generation

# Document Sentiment Classification

- Classify a whole opinion document (e.g., a review) based on the overall sentiment of the opinion holder
  - Classes: +ve, -ve, (possibly neutral)
  - Generally neutral is hard, hence most current approaches 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 (task): +ve or -ve?
- Perhaps the most widely studied problem in sentiment analysis

# Assumption and Goal

- **Assumption:** the document is written by a single person and expresses opinion/sentiment on a single entity
  - E.g., Review data, author (reviewer), entity (product), only thing to predict is sentiment
- **Goal (most basic):** discover  $(-, -, SO, -, -)$   
Where e, a, h and t can be ignored
- Many forum posts and blogs do not satisfy this assumption
  - They can mention and compare multiple entities
  - Many posts express no sentiment

# Text classification task

- It is essentially a text classification problem
  - Features
  - Classifier (if data have labels)
  - Unsupervised (if no labels)
  - Semi-supervised (have partial labels)
- Features are very important for sentiment classification
  - E.g., individual words are a possible feature
  - But what works for topic-based text classification might not work here
  - But in sentiment classification, **opinion/sentiment words** are more important, e.g., great, excellent, horrible, bad, worst etc
  - Require more complex features

# Sentiment Classification Features

Feature	Explanation and examples
N-Grams	<ul style="list-style-type: none"><li>• Individual words (unigrams)</li><li>• Multiple word patterns, n-grams</li><li>• E.g., “unpredictable steering” vs “unpredictable plot”</li></ul>
Frequency	<ul style="list-style-type: none"><li>• TF-IDF</li></ul>
Part of speech (POS) tags and patterns	<ul style="list-style-type: none"><li>• Noun, verb, adjective etc</li><li>• E.g. POS patterns, N followed by +ve Adj indicate positive sentiment</li></ul>
Modifiers	<ul style="list-style-type: none"><li>• Not, very, quite</li><li>• E.g., the movie is not interesting, this camera is very handy</li></ul>
All caps	<ul style="list-style-type: none"><li>• E.g., YES, COOL</li></ul>
Punctuation	<ul style="list-style-type: none"><li>• E.g, !, ?, !?</li></ul>
Emoticons	<ul style="list-style-type: none"><li>• E.g., 😊</li></ul>
Elongated words	<ul style="list-style-type: none"><li>• E.g., yayyyy</li></ul>

# Supervised Learning

- Directly apply supervised learning techniques to classify documents into +ve and -ve
- Example is Pang et al, 2002
- Features:
  - negation tag, unigram + bigram, POS tag, position, presence vs frequency of sentiment words (manually discovered)
- Several classification techniques were tried:
  - Naïve Bayes, SVM



# Supervised Learning

- Training and test data
  - Movie reviews with star ratings
    - 4-5 stars labelled as +ve
    - 1-2 stars labelled as -ve
  - Neutral is ignored
- SVM gave the best classification accuracy based on balanced training data
  - 83%

# Review Rating Prediction

- In addition to classification of +ve and –ve sentiment
  - Work has been done to predict the rating scores (e.g., 1-5 stars) of reviews (Pang and Lee, 2005)
  - Training and testing are reviews with star ratings
- Formulation: The problem is formulated as (ordinal) regression since the rating scores are ordinal
- Again, feature engineering and model building

# Break

- Trivia: Who penned the phrase "The pen is mightier than the sword"

# Unsupervised Sentiment Classification

- Determine average sentiment of sentiment phrases or words
  - If average is +ve, then document is +ve, otherwise negative
- Most basic, count number of positive words – number of negative words
  - “The show is **exciting** and **uplifting** but organisation is **poor**”
- More advanced: Turney, 2002
- **First:** Part-of-speech (PoS) tagging and patterns
  - Extract phrases containing adjectives & adverbs (good indicators of subjective sentences)
  - Single words don't provide context (e.g., unpredictable steering vs unpredictable plot)
  - Extract 2-word phrases if they confirm to certain patterns

# Unsupervised Sentiment Classification

First Word	Second Word	Third Word (not extracted)
JJ	NN or NNS	anything
RB, RBR, RBS	JJ	Not NN nor NNS
JJ	JJ	Not NN or NNS
NN or NNS	JJ	Nor NN nor NNS
RB, RBR, or RBS	VB, VBD, VBN, VBG	anything

Christopher Manning , Stanford NLP course

# Measuring polarity of phrase

- Positive phrases co-occur more often with "excellent"
- Negative phrases co-occur more often with "poor"

- Use Point wise mutual information

$$\text{PMI}(\text{word1}, \text{word2}) = \log_2 \frac{P(\text{word1}, \text{word2})}{P(\text{word1})P(\text{word2})}$$

- Measure how much more words *word1* and *word2* co-occur than if they were independent

# How to Estimate $P(\text{word})$ ?

- Use search engine (Altavista)
  - Why would we use this?
- $P(\text{word}) \sim \text{hits}(\text{word}) / N$
- $P(\text{word1}, \text{word2}) \sim \text{hits}(\text{word1 near word2}) / kN$ 
  - near operator returns search results where word1 is within k-words of word2



# Unsupervised Sentiment Classification

- **Second:** Determine semantic orientation (SO) of phrases using PMI:
  - Semantic orientation
$$\text{SO}(\text{phrase}) = \text{PMI}(\text{phrase}, \text{"excellent"}) - \text{PMI}(\text{phrase}, \text{"poor"})$$
- **Third:** Compute average SO across all phrases in document
  - Classify positive if average SO is +ve, otherwise classify as negative

# Example Sentiment Orientation

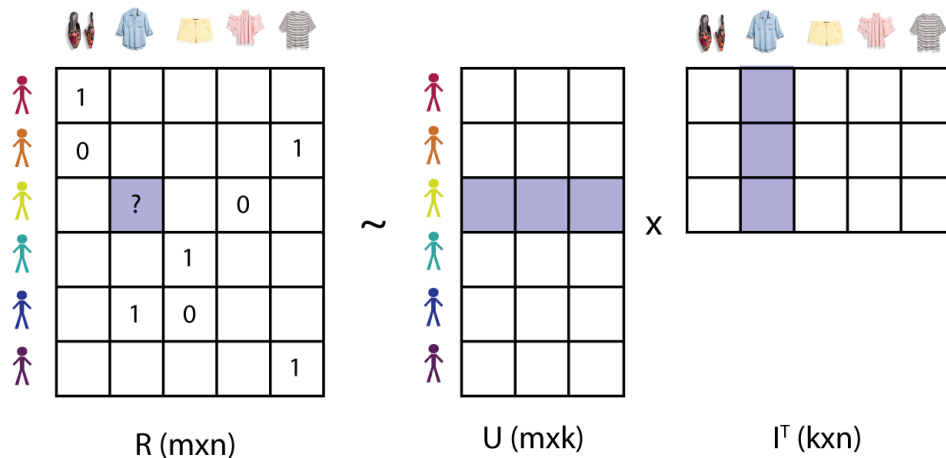
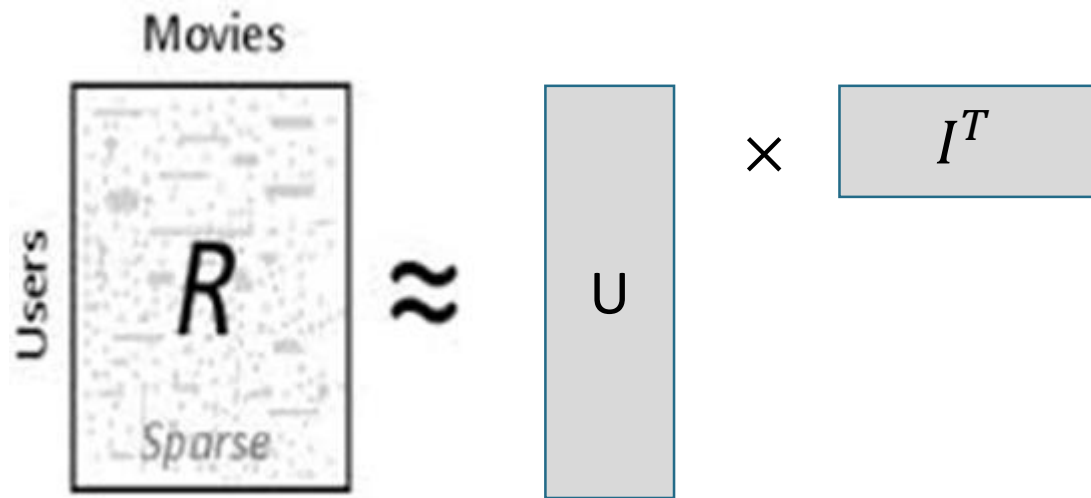
Phrase	POS tags	Polarity
direct deposits	JJ NNS	5 . 8
online web	JJ NN	1 . 9
very handy	RB JJ	1 . 4
...		
virtual monopoly	JJ NN	-2 . 0
lesser evil	RBR JJ	-2 . 3
other problems	JJ NNS	-2 . 8
low funds	JJ NNS	-6 . 8
unethical practices	JJ NNS	-8 . 5
<i>Average</i>		-1 . 2

Christopher Manning , Stanford NLP course

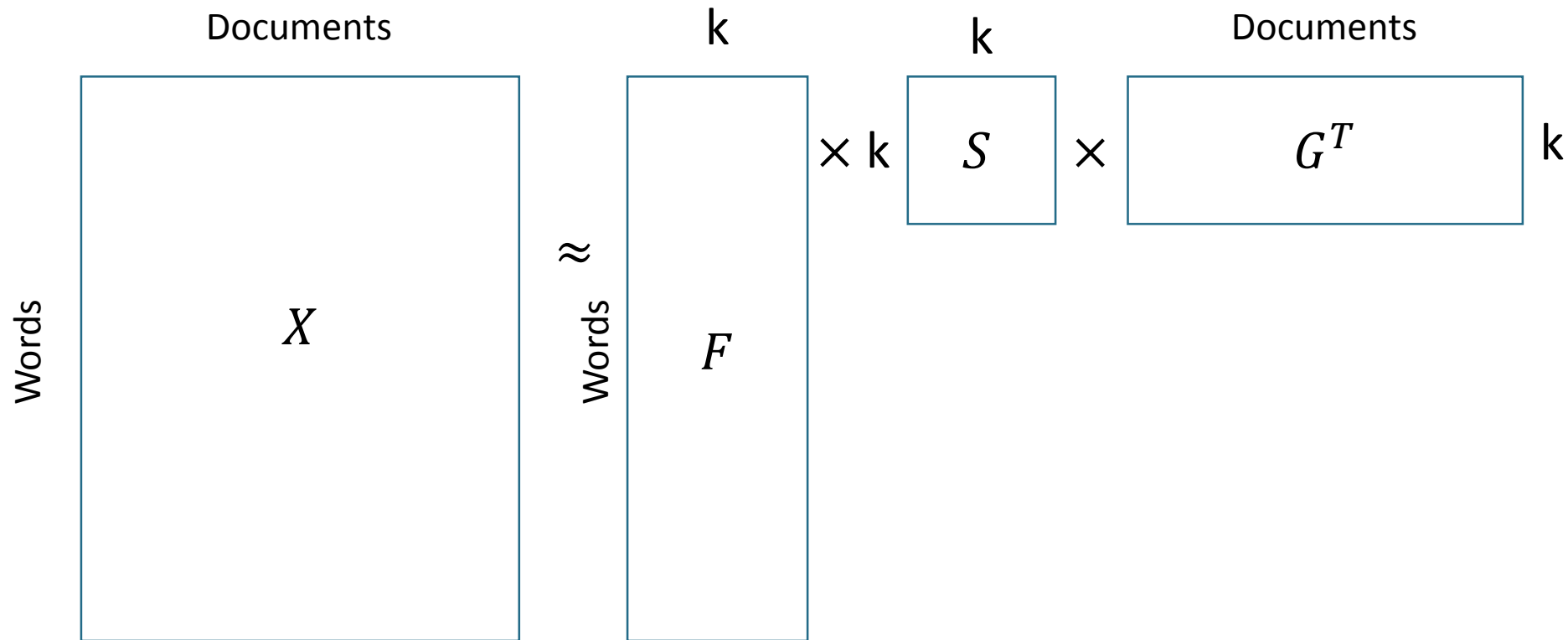
# Semi-supervised Sentiment Classification

- Labels are important, but expensive to obtain (typically human labelled).
- Best of both worlds: Semi-supervised
  - Assume some instances have labels given
  - Rest do not
- One example, Li et al [2009]
  - Latent space model (matrix factorisation)
  - $X \sim FS G^T$

# Non-negative Matrix Factorisation

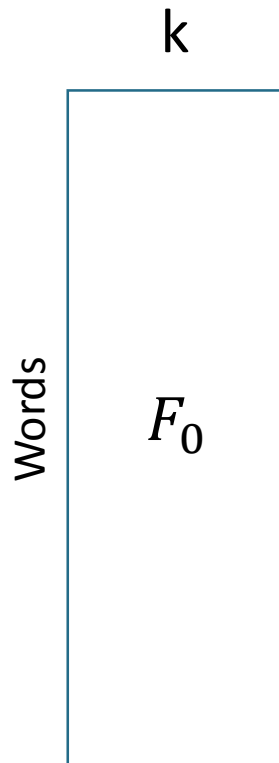


# Non-negative Matrix Tri-Factorisation

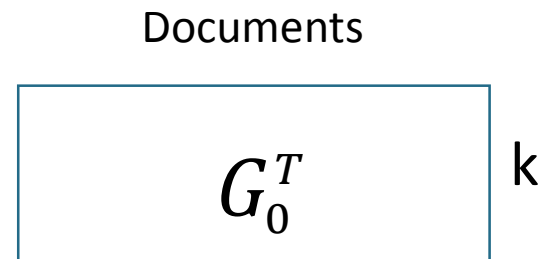


# Semi-supervision of Words and Documents

Semi-supervision of words



Semi-supervision of documents



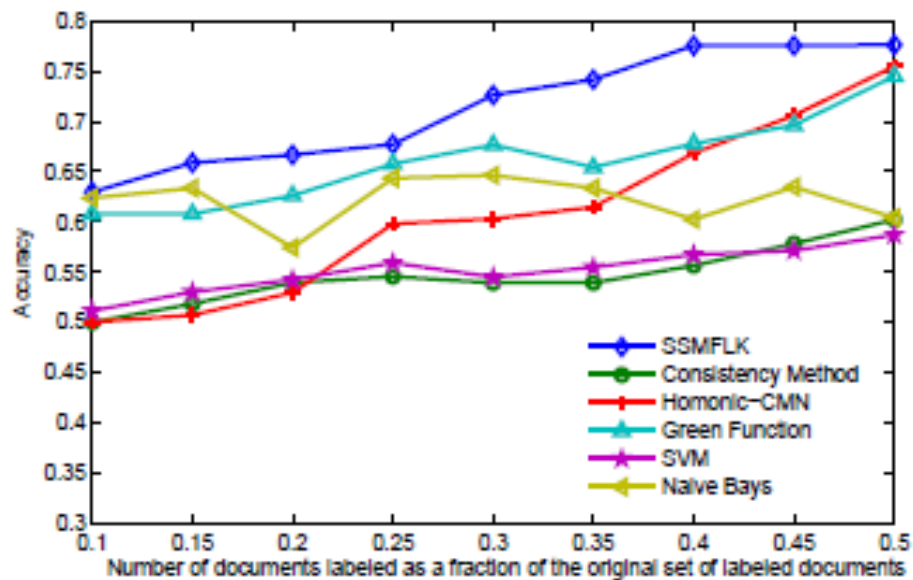
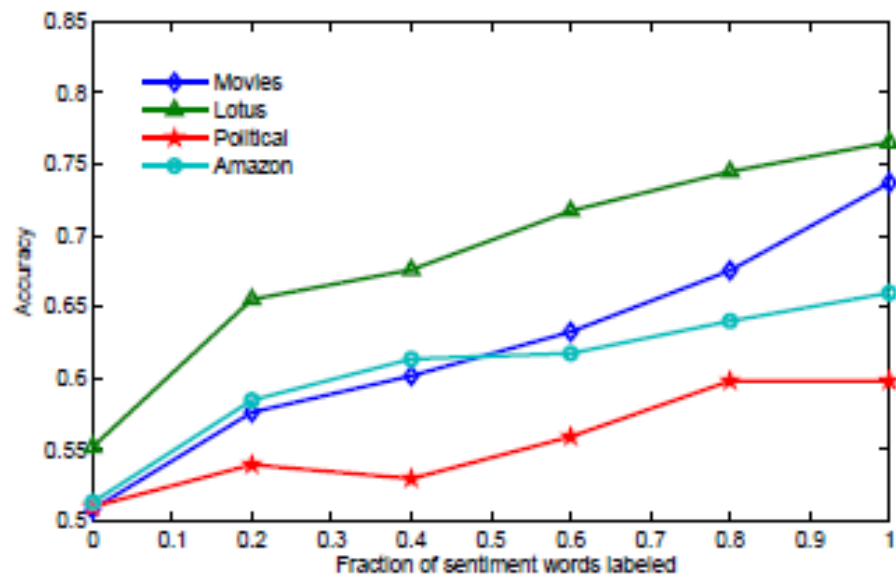
# Semi-supervised Non-negative Tri-Factorisation

$$X \approx F \times S \times G^T$$

$$+ \text{funcWords}(F_0, F) + \text{funcDocs}(G^T, G_0^T)$$

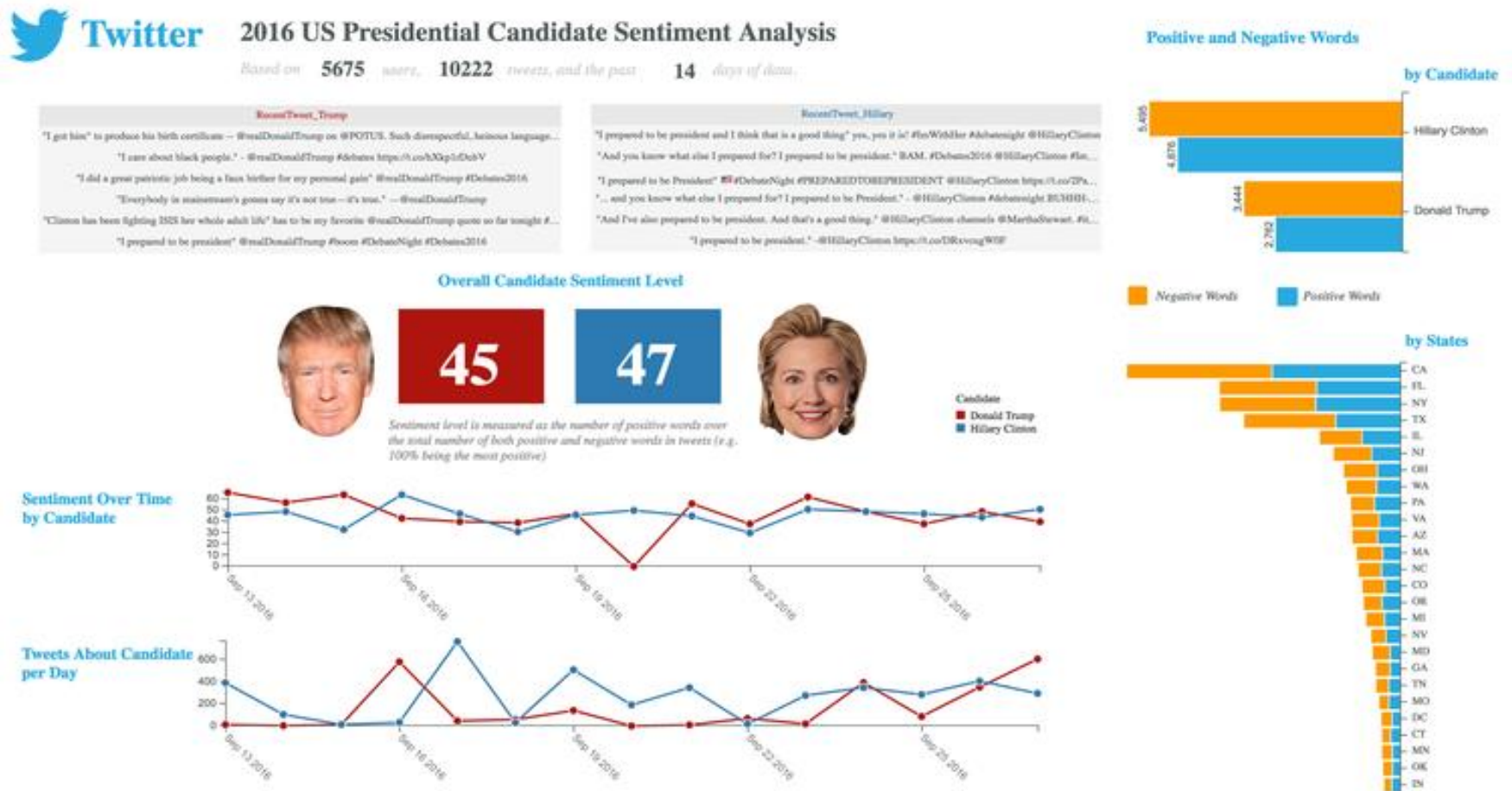


# Semi-supervised Sentiment Classification



# Example of Sentiment Classification

- From TechRepublic article about Tweeter sentiment classification before first presidential debate (<https://www.techrepublic.com/article/what-twitter-sentiment-analysis-is-saying-about-the-first-presidential-debate/>)



# Outline of Lecture

- Introduction to Sentiment Analysis
  - What is it?
  - Motivation
- Types of Sentiment Analysis
  - Sentiment & Subjectivity Classification
  - Aspect Extraction
  - Lexicon Generation

# Subjectivity classification (Sentence level)

- Document level sentiment classification is too coarse for most applications
- So do sentence level analysis
  - Assumes a single sentiment per sentence
  - Not always true, so one can classify clauses instead
- Usually consists of two steps
  - Subjectivity classification
    - To identify subjective sentences
    - Either use classification based approaches, or seed with a set of subjective and objective key words (unsupervised)
  - Sentiment classification of subjective sentences
    - As positive or negative

# 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 or disliked
- They do not identify the target of opinion, i.e.,
  - Entities and their aspects
  - Not all aspects have same sentiment orientation
- We need to go to the aspect level

# Recall an opinion is a tuple

- An opinion is a 5 element tuple:

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

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

# Aspect Extraction

- **Goal:** Given an opinion corpus, extract all aspects
- Many different approaches
- Frequent nouns and noun phrase
  - Nouns that are frequently talked about across reviews about the same product or product type are likely to be true aspects (frequent aspects)
  - E.g., think of what aspects are frequently discussed for phones – UI, battery, screen, camera etc
  - First use POS tagging to identify nouns
  - Then use sequential/association pattern mining to find these frequent nouns/phrases

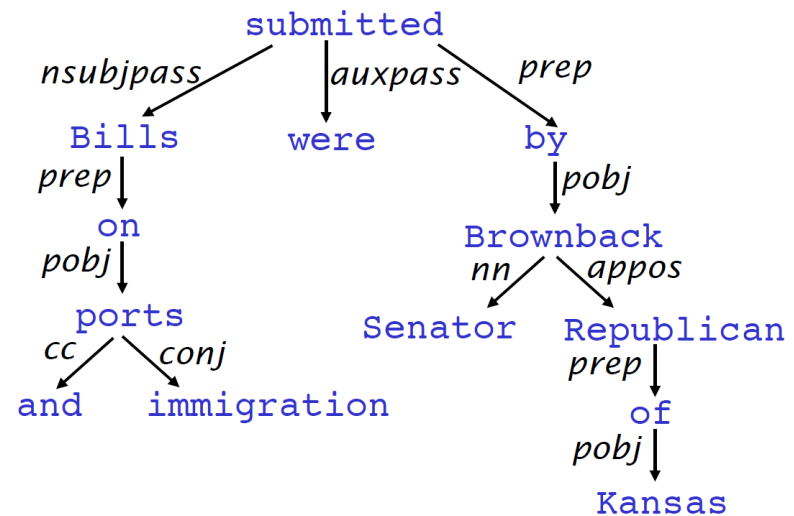
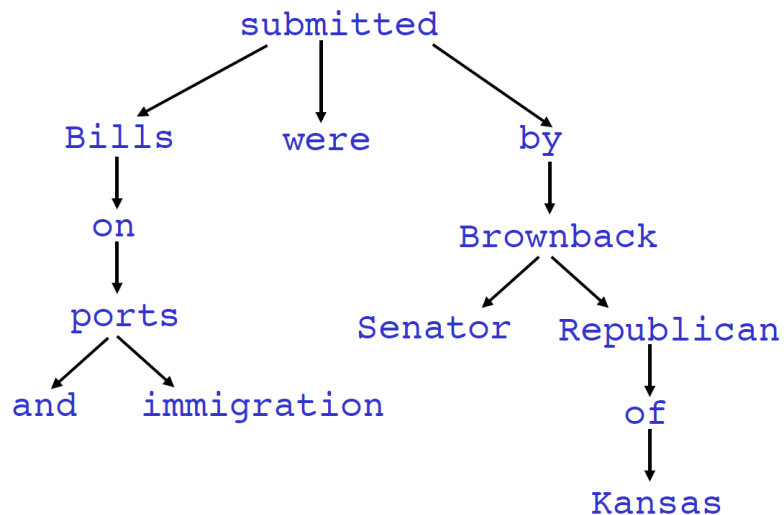
# Frequent Pattern Mining

- Seek patterns of text that are frequent across a number of instances/documents
- E.g., mobile phones
  - "The **battery** is long-lasting"
  - "I can do so much as the **battery** goes on and on"
  - "The **battery** on these phones sucks"
  - "...has diamond casing"



# Dependency Grammar

- Recall we briefly discussed syntactic analysis of text



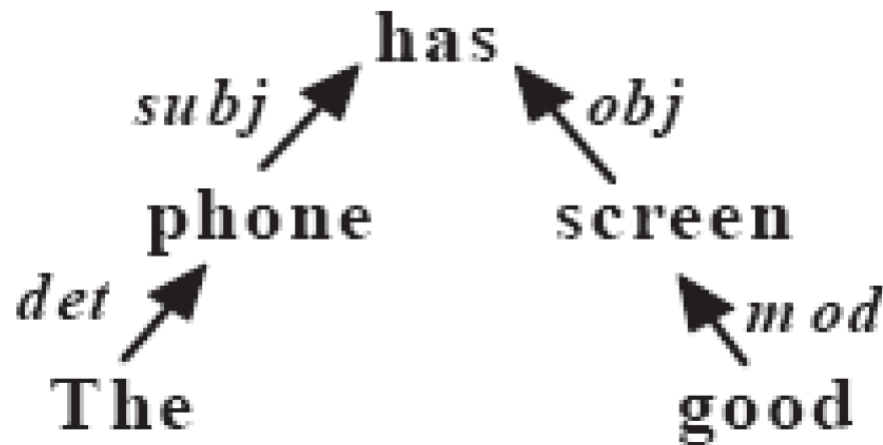
Christopher Manning , Stanford NLP course

# Infrequent Aspect Extraction

- What if the aspects are not frequently discussed across opinion corpus?
- **Key idea:** opinion have targets, i.e, opinion words are used to modify aspects and entities:
  - "The pictures are absolutely amazing"
  - "This is an amazing piece of software"
- One solution is the double propagation (DP) approach
- Use dependency of opinions & aspects to extract both
  - Knowing one helps find the other
  - E.g., "The **rooms** are spacious"
- DP extracts both aspects and opinion words
  - A domain independent method

# 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"



# Rules from Dependency Grammar

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

# Aspect Sentiment Classification

- For each aspect, identify the sentiment or opinion expressed on it
- Work based on identifying segments/phrases and opinion words
  - E.g., "The **battery life** and **picture quality** are **great** (+ve), but the **view** founder is **small** (-ve)"
- Challenge: Finding the necessary segments
  - use "but", "except that" etc to segment sentences
  - **Supervised:** Apply sentence level sentiment analysis
  - **Unsupervised:** Use DP approach

# Example of Aspect Extraction and Sentiment Analysis

Google products

sony camera

Search Products

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

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



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

★★★★☆ 159 reviews +1 0

#### Reviews

Summary - Based on 159 reviews

1

2

3 stars

4 stars

5 stars

What people are saying

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

# Outline of Lecture

- Introduction to Sentiment Analysis
  - What is it?
  - Motivation
- Types of Sentiment Analysis
  - Sentiment & Subjectivity Classification
  - Aspect Extraction
  - Lexicon Generation

# Sentiment/Opinion Lexicon Generation

- **Opinion lexicon:** lists of **words** and **expressions** used to express people's subjective feelings and sentiment/opinion
  - Not only individual words, but also phrases and idioms, e.g., "cost an arm and a leg"
- Fundamental for sentiment analysis
- Many sentiment lexica can be found on the web
  - They often have thousands of terms and expression are quite useful
- Many of opinion words are context dependent, not just application domain dependent
  - E.g., "The ventilation system is very cool" (Summer in Cairns vs Winter in Greenland)
- Three main ways to compile such lists:
  - Manual approach
  - Corpus-based approach
  - Dictionary based approach



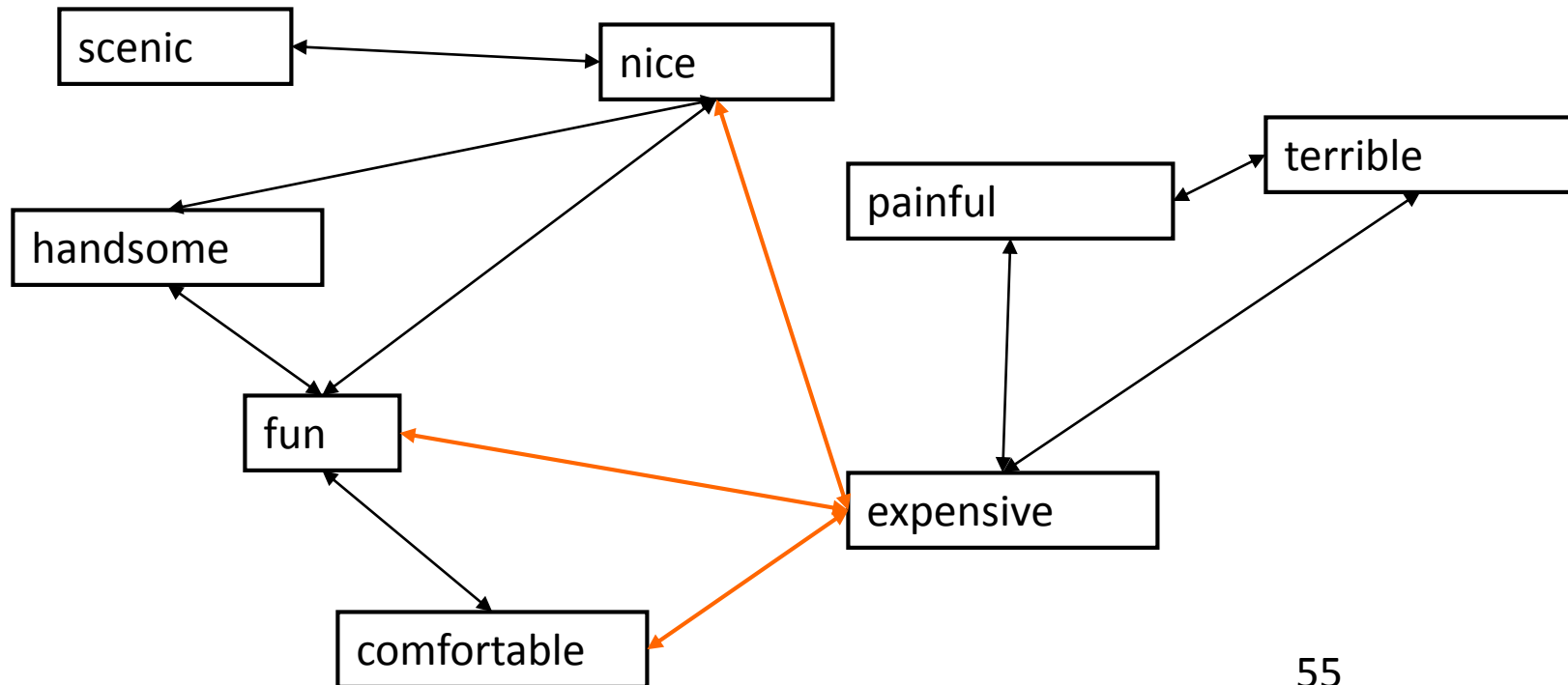
# Lexicon Construction Approaches

- Manual approach
  - Useful for one off
- Corpus based approaches
  - Use a (large) corpus to extract opinion words
  - Often use a double propagation between opinion words and the items they modify
  - Can be domain and context specific
- Dictionary based approaches
  - Typically use WordNet's synsets and hierarchies to acquire opinion words
  - Usually do not give domain or context dependent meanings

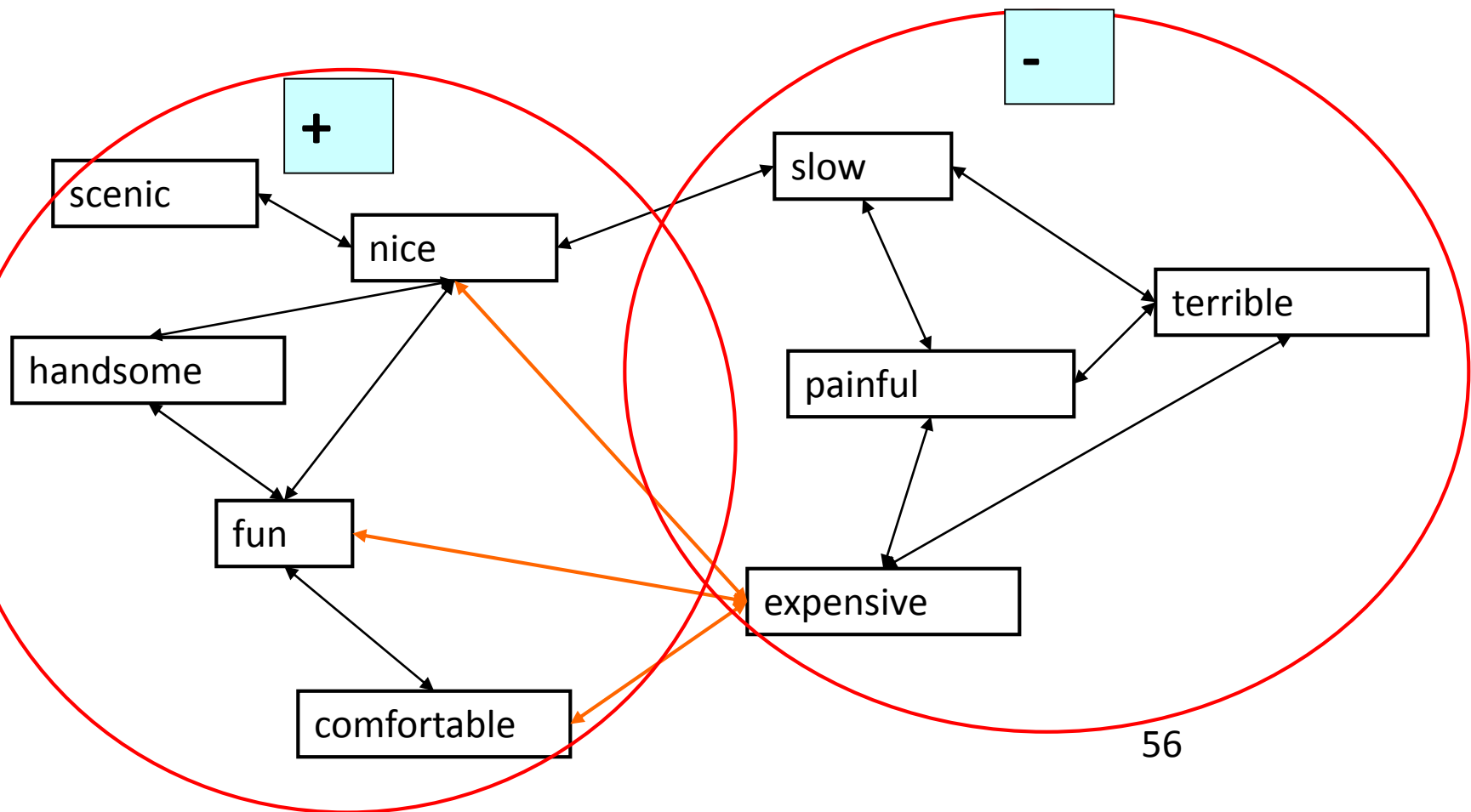
# Corpus based Approaches: Sentiment Consistency

- Sentiment consistency: use conventions on connectives to identify opinion words (Hazivassiloglou and McKeown, 1997)
  - Conjunction: conjoined adjectives usually have the same orientation
    - E.g., this car is beautiful and spacious
    - The other car is fast but expensive
  - Learning:
    - Find all conjoined adjectives
    - Determine if two conjoined adjectives are of the same or different orientations (using a ML classifier) and build a graph
    - Cluster the graph: produce two sets of words
    - Cluster of words with higher average frequency is labelled +ve
    - Tested on 1987 Wall Street Journal Corpus

# Sentiment consistency



# Sentiment consistency

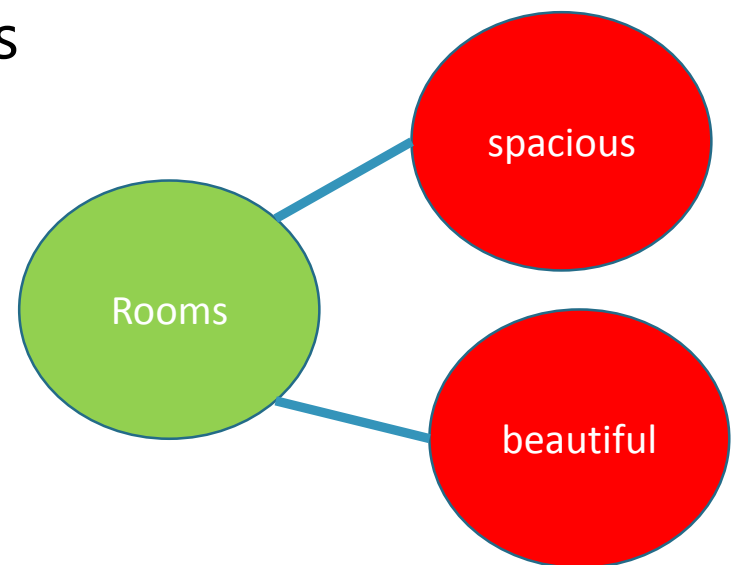


# Corpus based Approaches

- Turney 2002 and Yu and Hazivassiloglou, 2003
  - Similar, but instead of binary decision, assign opinion orientations (polarities) to words/phrases
  - Both use seed words, and a measure of similarity between the seed words and new words/phrases, e.g., log-likelihood ratio and PMI

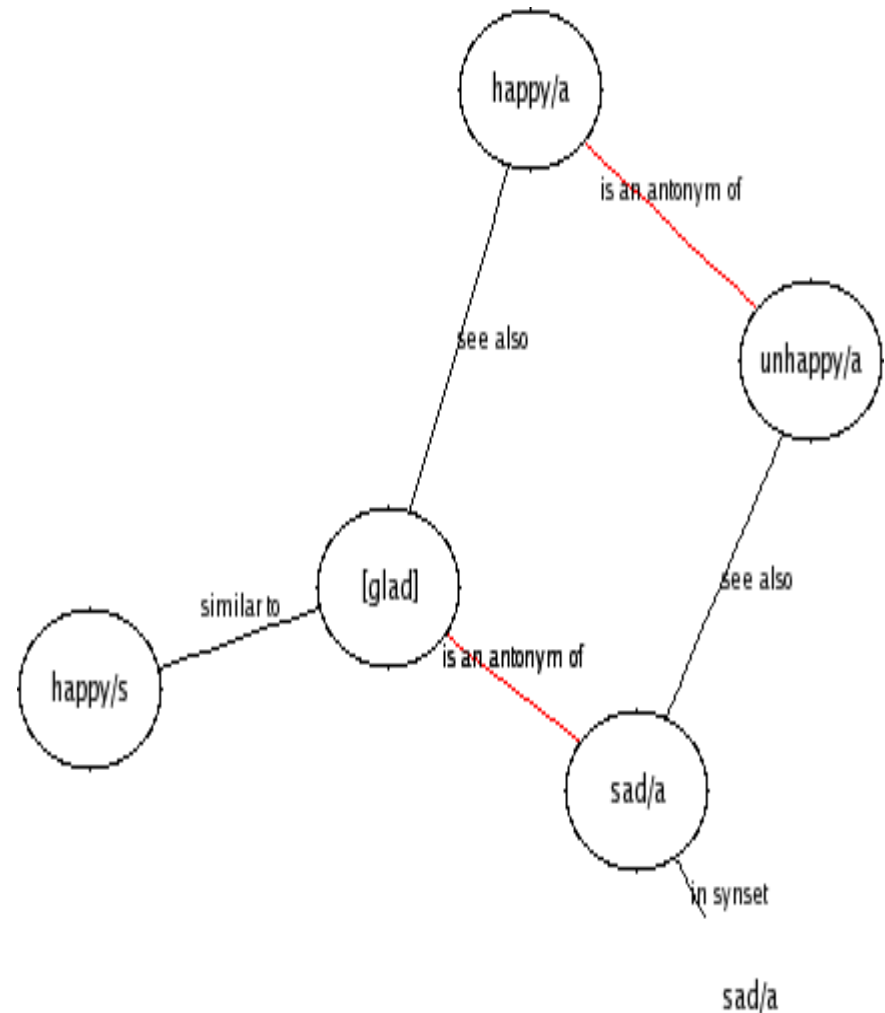
# Double Propagation method

- The DP method can also use dependency of opinions and 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**"



# Dictionary based Approaches: WordNet

- Typically use **WordNet's** synsets and hierarchies to acquire opinion words
- WordNet is a lexical database of nouns, verbs, adjectives and adverbs
  - Synonyms are linked, forming synsets (groups of synonyms expressing a concept)
  - Synsets are interlinked if similar in semantics or lexical relations
  - Similar to a thesaurus, but distinguish senses also



# Dictionary based Approaches

- Typically use WordNet's synsets and hierarchies to acquire opinion words
  - Start with a small seed of opinion words
  - Bootstrap the set to search for synonyms and antonyms in WordNet iteratively (Hu and Liu, 2004; Kim and Hovy, 2004)
  - E.g., if sad is a -ve seed opinion word, then we infer that its antonym is a +ve opinion word
  - Recursively propagate



# Which Lexicon Generation 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

# Open Challenges

- Sarcasm, irony
  - E.g., "Alcohol free. Gluten free. Dairy free. Predominantly meat free. Loving life!"
- Building an accurate domain specific system is possible
  - But accurate generic system is still open question
- Intentions & Understanding
  - "I come to see my doctor to get drugs because of severe pain in my stomach"
  - "After taking the drug, I got severe pain in my stomach"
  - "For paint X, one coat can cover the wood colour."
  - "For paint Y, we need three coats to cover the wood colour."

# Summary

- Sentiment Analysis
  - Introduction
  - Applications
- Model of sentiment analysis
- Sentiment analysis (documents)
- Sensitivity analysis (sentence)
- Aspect extraction and classification (aspects)
- Lexicon generation