Sentiment Analysis and Opinion Mining

Slides from Professor Bing Liu at University of Illinois, Chicago

Introduction – facts and opinions

- Two main types of textual information on the Web.
  - Facts and Opinions
- Current search engines search for facts (assume they are true)
  - Facts can be expressed with topic keywords.
- Search engines do not search for opinions
  - Opinions are hard to express with a few keywords
    - How do people think of Motorola Cell phones?
  - Current search ranking strategy is not appropriate for opinion retrieval/search.

Introduction – user generated content

- Word-of-mouth on the Web
  - One can express personal experiences and opinions on almost anything, at review sites, forums, discussion groups, blogs ... (called the user generated content.)
  - They contain valuable information
  - Web/global scale: No longer – one's circle of friends
- Our interest: to mine opinions (sentiments) expressed in the user-generated content
  - An intellectually very challenging problem.
  - Practically very useful.

Introduction – Applications

- Businesses and organizations: product and service benchmarking. Market intelligence.
  - Business spends a huge amount of money to find consumer sentiments and opinions.
    - Consultants, surveys and focus groups, etc
- Individuals: interested in other's opinions when
  - Purchasing a product or using a service,
  - Finding opinions on political topics,
- Ads placements: Placing ads in the user-generated content
  - Place an ad when one praises a product.
  - Place an ad from a competitor if one criticizes a product.
- Opinion retrieval/search: providing general search for opinions.

An Interesting Problem!

- Intellectually challenging & major applications.
  - A very popular research topic in recent years in NLP and Web data mining.
  - 20-60 companies in USA alone
- It touches everything aspect of NLP and yet is restricted and confined.
  - Little research in NLP/Linguistics in the past.
- Potentially a major technology from NLP.
  - But it is not easy!

Two types of evaluation

- Regular Opinions: sentiment expressions on some entities, e.g., products, events, topics, persons.
  - E.g., "the picture quality of this camera is great"
  - Subjective
- Comparisons: relations expressing similarities or differences of more than one entity. Usually expressing an ordering.
  - E.g., "car x is cheaper than car y."
  - Objective or subjective.

Opinion search (Liu, Web Data Mining book, 2007)

- Can you search for opinions as conveniently as general Web search?
- Whenever you need to make a decision, you may want some opinions from others,
  - Wouldn't it be nice? you can find them on a search system instantly, by issuing queries such as
    - Opinions: "Motorola cell phones"
    - Comparisons: "Motorola vs. Nokia"
- Cannot be done yet! (but could be soon …)

Typical opinion search queries

- Find the opinion of a person or organization (opinion holder) on a particular entity or an aspect of the entity.
  - E.g., what is Bill Clinton's opinion on abortion?
- Find positive and/or negative opinions on a particular entity (or some aspects of the entity), e.g.,
  - customer opinions on a digital camera.
  - public opinions on a political topic.
- Find how opinions on an entity change over time.
- How entity A compares with entity B?
  - Gmail vs. Hotmail

Find the opinion of a person on X

- In some cases, the general search engine can handle it, i.e., using suitable keywords.
  - Bill Clinton's opinion on abortion
- Reason:
  - One person or organization usually has only one opinion on a particular topic.
  - The opinion is likely contained in a single document.
  - Thus, a good keyword query may be sufficient.

Find opinions on an entity

We use product reviews as an example:

- Searching for opinions in product reviews is different from general Web search.
  - E.g., search for opinions on "Motorola RAZR V3"
- General Web search (for a fact): rank pages according to some authority and relevance scores.
  - The user views the first page (if the search is perfect).
  - **One fact = Multiple facts**
- Opinion search: rank is desirable, however
  - reading only the review ranked at the top is not appropriate because it is only the opinion of one person.
  - **One opinion ≠ Multiple opinions**

Search opinions (contd)

- Ranking:
  - produce two rankings
    - Positive opinions and negative opinions
    - Some kind of summary of both, e.g., # of each
  - Or, one ranking but
    - The top (say 30) reviews should reflect the natural distribution of all reviews (assume that there is no spam), i.e., with the right balance of positive and negative reviews.
- Questions:
  - Should the user reads all the top reviews? OR
  - Should the system prepare a summary of the reviews?

Reviews are similar to surveys

- Reviews can be regarded as traditional surveys.
  - In traditional survey, returned survey forms are treated as raw data.
  - Analysis is performed to summarize the survey results.
    - E.g., % against or for a particular issue, etc.
- In opinion search,
  - Can a summary be produced?
  - What should the summary be?

Roadmap

- **Opinion mining – problem definition**
- Document level sentiment classification
- Sentence level sentiment classification
- Opinion lexicon generation
- Aspect-based opinion mining
- Opinion mining of comparative sentences
- Opinion spam detection
- Summary

An Example Review

- *"I bought an iPhone a few days ago. It was such a nice phone. The touch screen was really cool. The voice quality was clear too. Although the battery life was not long, that is ok for me. 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, and wanted me to return it to the shop. …"*
- What do we see?
  - **Opinions, targets of opinions, and opinion holders**

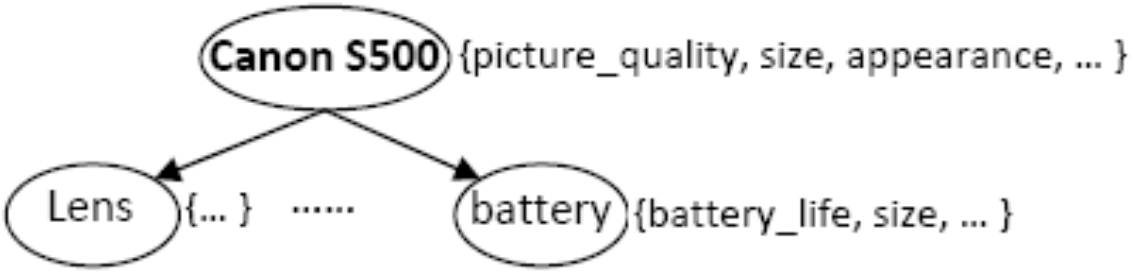
Opinion mining – the **abstraction**

(Hu and Liu, KDD-04; Liu, Web Data Mining book 2007)

- Basic components of an opinion
  - Opinion holder: The person or organization that holds a specific opinion on a particular entity.
  - Entity: on which an opinion is expressed
  - Opinion: a view, attitude, or appraisal on an entity from an opinion holder.
- Objectives of opinion mining: many ...
- **Let us abstract the problem**
  - put existing research into a common framework
- We use consumer reviews of products to develop the ideas. Other opinionated contexts are similar.

Target entity (Liu, Web Data Mining book, 2006)

- **Definition** (**entity**): An *entity e* is a product, person, event, organization, or topic. *o* 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.
- To simplify our discussion, we use the term ***aspect*** (*features)* to represent both components & attributes.

Model of an entity

- An entity e_i is represented with a finite set of aspects, $A = \{a_1, a_2, \ldots, a_n\}$.
  - The entity can be expressed with any one of a final set of *entity expressions* $EE_i = \{oe_{i1}, oe_{i2}, \ldots, oe_{ik}\}$.
  - Each aspect $a_{ij} \in A_i$ of the entity can be expressed by any one of a finite set of *aspect expressions* $AE_{ij} = \{ae_{ij1}, ae_{ij2}, \ldots, ae_{ijm}\}$.

Model of a review

- **Model of a review**: An opinion holder *j* comments on a subset of the aspects $S_j \subseteq A$ of entity *e*.
  - For each aspect $a_k \in S_j$ that *j* comments on, he/she
    - chooses a word or phrase from EE_i to describe the entity, and
    - chooses a word or phrase from AE_{ij} to describe the aspect, and
    - expresses a positive, negative or neutral opinion on a_k.

What is an Opinion? (Liu, Ch. in NLP handbook)

- An *opinion* is a quintuple

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

  where
  - e_j is a target entity.
  - a_k is a aspect of the entity e_j.
  - so_{ijkl} is the sentiment value of the opinion of 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.

Objective – structure the unstructured

- Objective: Given an opinionated document,
  - Discover all quintuples (e_j, a_k, so_{ijkl}, h_i, t_l),
    - i.e., mine the five corresponding pieces of information in each quintuple, and
  - Or, solve some simpler problems
- With the quintuples,
  - Unstructured Text → Structured Data
    - Traditional data and visualization tools can be used to slice, dice and visualize the results in all kinds of ways
    - Enable qualitative and quantitative analysis.

Aspect-Based Opinion Summary

(Hu & Liu, KDD-2004)

"I bought an iPhone a few days ago. It was such a nice phone. The touch screen was really cool. The voice quality was clear too. Although the battery life was not long, that is ok for me. 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, and wanted me to return it to the shop. …"

aspect Based Summary:

aspect1: **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.

…

aspect2: **battery life**

…

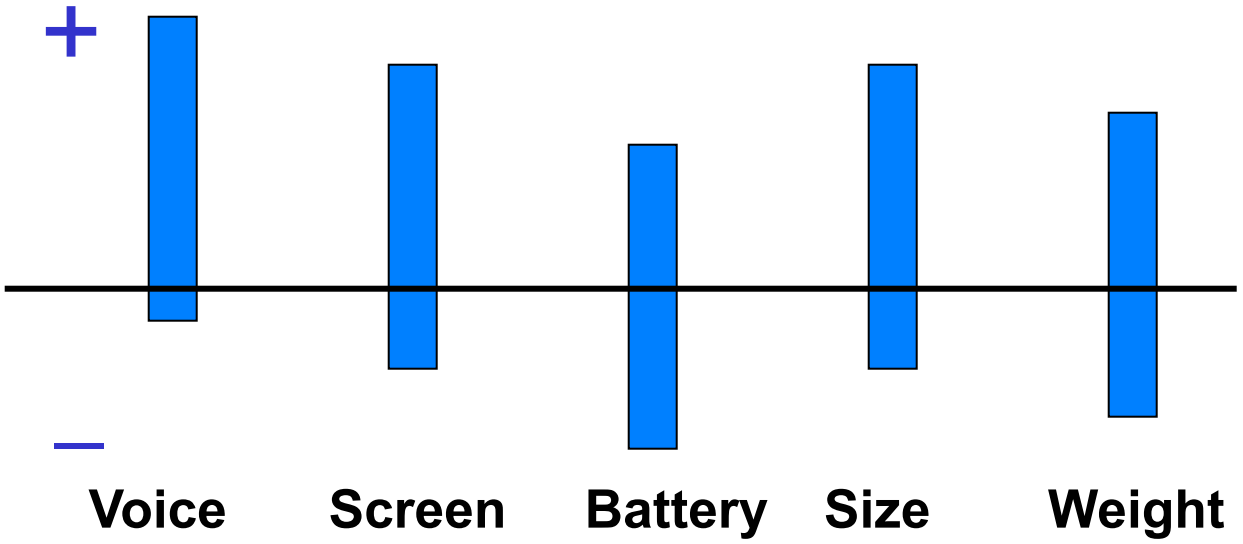
Note: We omit opinion holders

….

Visual Comparison (Liu et al. WWW-2005)

- Summary of reviews of

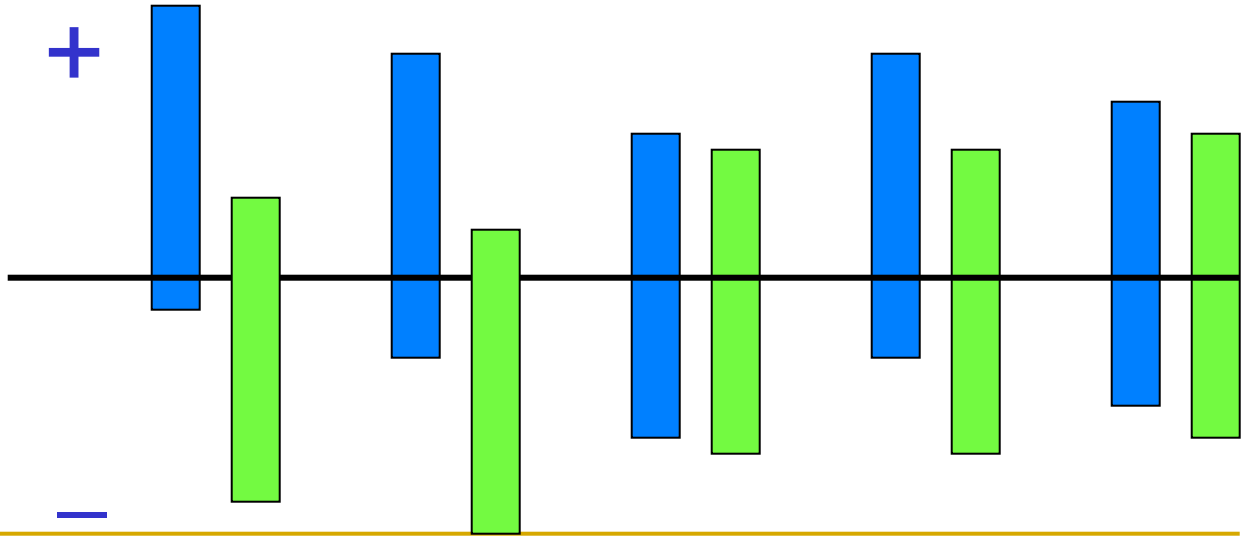
Cell Phone 1



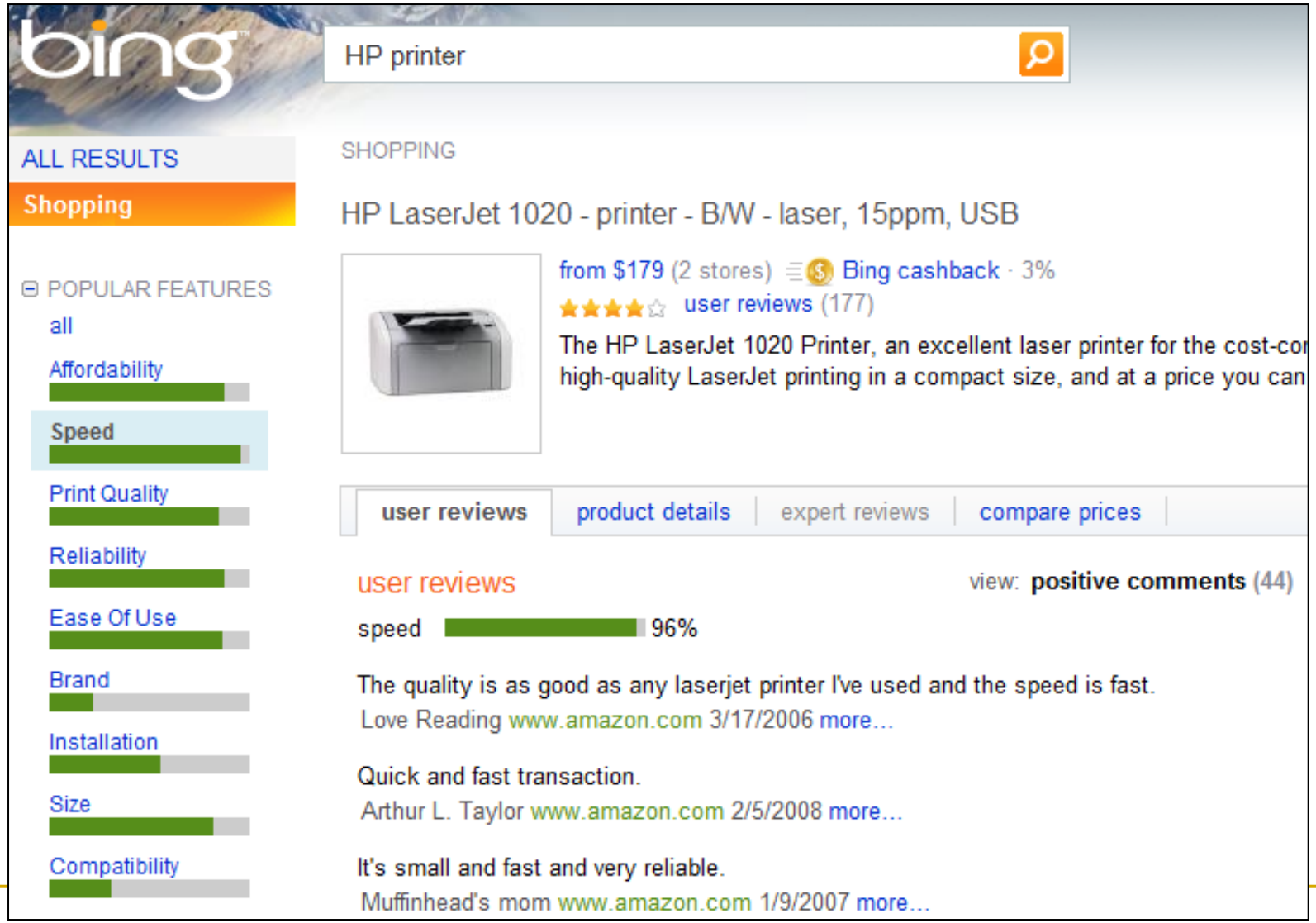
- Comparison of reviews of

Cell Phone 1

Cell Phone 2



Feat.-based opinion summary in Bing



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

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

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

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 phone. The battery life was long. 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.*"

It is not Just ONE Problem

- $(e_j, a_k, so_{ijkl}, h_i, t_l)$,
  - e_j - a target entity: Named Entity Extraction (more)
  - a_{jk} - a aspect of e_j: Information Extraction
  - so_{ijkl} is sentiment: Sentiment determination
  - h_i is an opinion holder: Information/Data Extraction
  - t_l is the time: Data Extraction
- Co-reference resolution
- Synonym match (voice = sound quality) …
- None of them is a solved problem!

Opinion mining tasks

- At the document (or review) level:

  Task: sentiment classification of reviews
  - Classes: positive, negative, and neutral
  - Assumption: each document (or review) focuses on a single entity (not true in many discussion posts) and contains opinion from a single opinion holder.
- At the sentence level:

  Task 1: identifying subjective/opinionated sentences
  - Classes: objective and subjective (opinionated)

  Task 2: sentiment classification of sentences
  - Classes: positive, negative and neutral.
  - Assumption: a sentence contains only one opinion
    - not true in many cases.
  - Then we can also consider clauses or phrases.

Opinion mining tasks (contd)

- At the aspect level:
  - **Task 1** (entity extraction and grouping): Extract all entity expressions, and group synonymous entity expressions into entity clusters. Each cluster indicates a unique entity e_i.
  - **Task 2** (aspect extraction and grouping): Extract all aspect expressions of the entities, and group synonymous aspect expressions into clusters. Each aspect expression cluster of entity e_i indicates a unique aspect a_{ij}.

Opinion mining tasks (contd)

- **Task 3** (opinion holder and time extraction): Extract these pieces of information from the text or structured data.
- **Task 4** (aspect sentiment classification): Determine whether each opinion on an aspect is positive, negative or neutral.
- **Task 5** (opinion quintuple generation): Produce all opinion quintuples (e_i, a_{ij}, oo_{ijkl}, h_k, t_l) expressed in D.

Roadmap

- Opinion mining – problem definition
- **Document level sentiment classification**
- Sentence level sentiment classification
- Opinion lexicon generation
- Aspect-based opinion mining
- Opinion mining of comparative sentences
- Opinion spam detection
- Summary

Sentiment classification

- Classify documents (e.g., reviews) based on the overall sentiments expressed by opinion holders (authors),
  - Positive, negative, and (possibly) neutral
  - Since in our model an entity *e* itself is also a aspect, then sentiment classification essentially determines the opinion expressed on *e* in each document (e.g., review).
- Similar but different from topic-based text classification.
  - In topic-based text classification, topic words are important.
  - In sentiment classification, sentiment words are more important, e.g., great, excellent, horrible, bad, worst, etc.

Unsupervised review classification
(Turney, ACL-02)

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

- Step 2: Estimate the semantic 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 recommended if average SO is positive, not recommended otherwise.

- Final classification accuracy:
  - automobiles - 84%
  - banks - 80%
  - movies - 65.83
  - travel destinations - 70.53%

Sentiment classification using machine learning methods (Pang et al, EMNLP-02)

- This paper directly applied several machine learning techniques to classify movie reviews into positive and negative.
- Three classification techniques were tried:
  - Naïve Bayes
  - Maximum entropy
  - Support vector machine
- Pre-processing settings: negation tag, unigram (single words), bigram, POS tag, position.
- SVM: the best accuracy 83% (unigram)

Review classification by scoring features

(Dave, Lawrence and Pennock, WWW-03)

- It first selects a set of features $F = f_1, f_2, \ldots\ldots$
  - Note: machine learning features, not product features.
- Score the features
  - C and C' are classes

$$score(f_i) = \frac{P(f_i \mid C) - P(f_i \mid C')}{P(f_i \mid C) + P(f_i \mid C')}$$

- Classification of a review d_j (using sign):

$$class(d_j) = \begin{cases} C & eval(d_j) > 0 \\ C' & eval(d_j) < 0 \end{cases}$$

$$eval(d_j) = \sum_i score(f_i)$$

- Accuracy of 84-88%.

Roadmap

- Opinion mining – problem definition
- Document level sentiment classification
- **Sentence level sentiment classification**
- Opinion lexicon generation
- Aspect-based opinion mining
- Opinion mining of comparative sentences
- Opinion spam detection
- Summary

Sentence-level sentiment analysis

- Document-level sentiment classification is too coarse for most applications.
- Let us move to the sentence level.
- Much of the work on sentence level sentiment analysis focuses on identifying subjective sentences in news articles.
  - Classification: objective and subjective.
  - All techniques use some forms of machine learning.
  - E.g., using a naïve Bayesian classifier with a set of data features/attributes extracted from training sentences (Wiebe et al. ACL-99).

Using learnt patterns (Rilloff and Wiebe, EMNLP-03)

- 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).

Subjectivity and polarity (orientation)
(Yu and Hazivassiloglou, EMNLP-03)

- For subjective or opinion sentence identification, three methods are tried:
  - Sentence similarity.
  - Naïve Bayesian classification.
  - Multiple naïve Bayesian (NB) classifiers.
- For opinion orientation (positive, negative or neutral) (also called polarity) classification, it uses a similar method to (Turney, ACL-02), 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.

Let us go further?

- Sentiment classification at both document and sentence (or clause) levels are useful, but
  - They do not find what the opinion holder liked and disliked.
- An negative sentiment on an entity
  - does not mean that the opinion holder dislikes everything about the entity.
- A positive sentiment on an entity
  - does not mean that the opinion holder likes everything about the entity.
- **We need to go to the entity level.**

Roadmap

- Opinion mining – problem definition
- Document level sentiment classification
- Sentence level sentiment classification
- **Opinion lexicon generation**
- Aspect-based opinion mining
- Opinion mining of comparative sentences
- Opinion spam detection
- Summary

But before we go further

- Let us discuss **Opinion Words or Phrases** (also called polar words, opinion bearing words, etc). E.g.,
  - Positive: beautiful, wonderful, good, amazing,
  - Negative: bad, poor, terrible, cost someone an arm and a leg (idiom).
- They are instrumental for opinion mining (obviously)
- Three main ways to compile such a list:
  - Manual approach: not a bad idea, only a one-time effort
  - Corpus-based approaches
  - Dictionary-based approaches
- **Important to note:**
  - **Some opinion words are context independent (e.g., good).**
  - **Some are context dependent (e.g., long).**

Corpus-based approaches

- Rely on syntactic or co-occurrence patterns in large corpora. (Hazivassiloglou and McKeown, ACL-97; Turney, ACL-02; Yu and Hazivassiloglou, EMNLP-03; Kanayama and Nasukawa, EMNLP-06; Ding and Liu SIGIR-07)
  - Can find domain (not context!) dependent orientations (positive, negative, or neutral).
- (Turney, ACL-02) and (Yu and Hazivassiloglou, EMNLP-03) are similar.
  - Assign opinion orientations (polarities) to words/phrases.
  - (Yu and Hazivassiloglou, EMNLP-03) is different from (Turney, ACL-02)
    - use more seed words (rather than two) and use log-likelihood ratio (rather than PMI).

Corpus-based approaches (contd)

- Use constraints (or conventions) on connectives to identify opinion words (Hazivassiloglou and McKeown, ACL-97; Kanayama and Nasukawa, EMNLP-06; Ding and Liu, 2007). E.g.,
- Conjunction: conjoined adjectives usually have the same orientation (Hazivassiloglou and McKeown, ACL-97).
  - 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
  - Corpus: 21 million word 1987 Wall Street Journal corpus.

Corpus-based approaches (contd)

- (Kanayama and Nasukawa, EMNLP-06) takes a similar approach to (Hazivassiloglou and McKeown, ACL-97) but for Japanese words:
  - Instead of using learning, it uses two criteria to determine whether to add a word to positive or negative lexicon.
  - Have an initial seed lexicon of positive and negative words.
- (Ding and Liu, 2007) also exploits constraints on connectives, but with two differences
  - It uses them to assign opinion orientations to product aspects (more on this later).
    - One word may indicate different opinions in the same domain.
      - "The battery life is *long*" (+) and "It takes a *long* time to focus" (-).
    - **Find domain opinion words is insufficient.**
  - It can be used without a large corpus.

Corpus-based approaches (contd)

- A double propagation method is proposed in [Qiu et al. IJCAI-2009]
- It exploits dependency relations of opinions and aspects to extract opinion words.
  - Opinion words modify entity aspects/features, e.g.,
  - "This camera has *long battery life*"
- The algorithm essentially bootstraps using a set of seed opinion words
  - With the help of some dependency relations.

Rules from dependency grammar

	Relations and Constraints	Output	Examples
$R1_1$	$O \rightarrow O\text{-}Dep \rightarrow F$ *s.t.* $O \in \{O\}$, $O\text{-}Dep \in \{MR\}$, $POS(F) \in \{NN\}$	$f = F$	*The phone has a good "screen".* *good* →**mod**→ *screen*
$R1_2$	$O \rightarrow O\text{-}Dep \rightarrow H \leftarrow F\text{-}Dep \leftarrow F$ *s.t.* $O \in \{O\}$, $O/F\text{-}Dep \in \{MR\}$, $POS(F) \in \{NN\}$	$f = F$	*"iPod" is the best mp3 player.* *best* →**mod**→ *player* ←**subj**← *iPod*
$R2_1$	$O \rightarrow O\text{-}Dep \rightarrow F$ *s.t.* $F \in \{F\}$, $O\text{-}Dep \in \{MR\}$, $POS(O) \in \{JJ\}$	$o = O$	same as $R1_1$ with *screen* as the known word and *good* as the extracted word
$R2_2$	$O \rightarrow O\text{-}Dep \rightarrow H \leftarrow F\text{-}Dep \leftarrow F$ *s.t.* $F \in \{F\}$, $O/F\text{-}Dep \in \{MR\}$, $POS(O) \in \{JJ\}$	$o = O$	same as $R1_2$ with *iPod* is the known word and *best* as the extract word.
$R3_1$	$F_{i(j)} \rightarrow F_{i(j)}\text{-}Dep \rightarrow F_{j(i)}$ *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 audio and "video"?* *video* →**conj**→ *audio*
$R3_2$	$F_i \rightarrow F_i\text{-}Dep \rightarrow H \leftarrow F_j\text{-}Dep \leftarrow F_j$ *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 len.* *len* →**obj**→ *has* ←**subj**← *G3*
$R4_1$	$O_{i(j)} \rightarrow O_{i(j)}\text{-}Dep \rightarrow O_{j(i)}$ *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 amazing and "easy" to use.* *easy* →**conj**→ *amazing*
$R4_2$	$O_i \rightarrow O_i\text{-}Dep \rightarrow H \leftarrow O_j\text{-}Dep \leftarrow O_j$ *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 sexy, "cool", accessory-available mp3 player, you can choose iPod.* *sexy* →**mod**→ *player* ←**mod**← *cool*

Dictionary-based approaches

- Typically use WordNet's synsets and hierarchies to acquire opinion words
  - Start with a small seed set of opinion words.
  - Use the set to search for synonyms and antonyms in WordNet (Hu and Liu, KDD-04; Kim and Hovy, COLING-04).
  - Manual inspection may be used afterward.
- Use additional information (e.g., glosses) from WordNet (Andreevskaia and Bergler, EACL-06) and learning (Esuti and Sebastiani, CIKM-05).
- Weakness of the approach: Do not find context dependent opinion words, e.g., small, long, fast.

Roadmap

- Opinion mining – problem definition
- Document level sentiment classification
- Sentence level sentiment classification
- Opinion lexicon generation
- **Aspect-based opinion mining**
- Opinion mining of comparative sentences
- Opinion spam detection
- Summary

Aspect-based opinion mining and summarization (Hu and Liu, KDD-04)

- Again focus on reviews (easier to work in a concrete domain!)
- Objective: find what reviewers (opinion holders) liked and disliked
  - Product aspects and opinions on the aspects
- Since the number of reviews on an entity can be large, an opinion summary should be produced.
  - Desirable to be a structured summary.
  - Easy to visualize and to compare.
  - Analogous to but different from multi-document summarization.

The tasks

- We have 5 tasks, but only focus on two.
  - **Task 2** (aspect extraction and grouping): Extract all aspect expressions of the entities, and group synonymous aspect expressions into clusters. Each aspect expression cluster of entity e_i indicates a unique aspect a_{ij}.
  - **Task 4** (aspect sentiment classification): Determine whether each opinion on an aspect is positive, negative or neutral.

Aspect extraction (Hu and Liu, KDD-04; Liu, Web Data Mining book 2007)

- Frequent aspects (called features before): those aspects that have been talked about by many reviewers.
- Use sequential pattern mining
- Why the frequency based approach?
  - Different reviewers tell different stories (irrelevant)
  - When product aspects are discussed, the words that they use converge.
  - They are main aspects.
- Sequential pattern mining finds frequent phrases.
- Many companies implemented the approach (no POS restriction).

Using part-of relationship and the Web (Popescu and Etzioni, EMNLP-05)

- Improved (Hu and Liu, KDD-04) 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.
  - The part discriminators for the scanner class are, “of scanner”, “scanner has”, “scanner comes with”, etc, which are used to find components or parts of scanners by searching on the Web: the KnowItAll approach, (Etzioni et al, WWW-04).

Infrequent aspects extraction

- How to find the infrequent aspects?
- Observation: the same opinion word can be used to describe different aspects and entities.
  - "The pictures are absolutely amazing."
  - "The software that comes with it is amazing."

Using dependency relations

- A same *double propagation* approach in (Qiu et al. IJCAI-2009) is applicable here.
- It exploits the dependency relations of opinions and aspects to extract aspects.
  - Opinions words modify entity/aspect, e.g.,
  - "This camera has *long battery life*"
- The algorithm bootstraps using a set of seed opinion words (no aspect input).
  - To extract aspects (and also opinion words)

Rules from dependency grammar

	Relations and Constraints	Output	Examples
$R1_1$	$O \rightarrow O\text{-}Dep \rightarrow F$ *s.t.* $O \in \{O\}$, $O\text{-}Dep \in \{MR\}$, $POS(F) \in \{NN\}$	$f = F$	*The phone has a good "screen".* *good* →**mod**→ *screen*
$R1_2$	$O \rightarrow O\text{-}Dep \rightarrow H \leftarrow F\text{-}Dep \leftarrow F$ *s.t.* $O \in \{O\}$, $O/F\text{-}Dep \in \{MR\}$, $POS(F) \in \{NN\}$	$f = F$	*"iPod" is the best mp3 player.* *best* →**mod**→ *player* ←**subj**← *iPod*
$R2_1$	$O \rightarrow O\text{-}Dep \rightarrow F$ *s.t.* $F \in \{F\}$, $O\text{-}Dep \in \{MR\}$, $POS(O) \in \{JJ\}$	$o = O$	same as $R1_1$ with *screen* as the known word and *good* as the extracted word
$R2_2$	$O \rightarrow O\text{-}Dep \rightarrow H \leftarrow F\text{-}Dep \leftarrow F$ *s.t.* $F \in \{F\}$, $O/F\text{-}Dep \in \{MR\}$, $POS(O) \in \{JJ\}$	$o = O$	same as $R1_2$ with *iPod* is the known word and *best* as the extract word.
$R3_1$	$F_{i(j)} \rightarrow F_{i(j)}\text{-}Dep \rightarrow F_{j(i)}$ *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 audio and "video"?* *video* →**conj**→ *audio*
$R3_2$	$F_i \rightarrow F_i\text{-}Dep \rightarrow H \leftarrow F_j\text{-}Dep \leftarrow F_j$ *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 len.* *len* →**obj**→ *has* ←**subj**← *G3*
$R4_1$	$O_{i(j)} \rightarrow O_{i(j)}\text{-}Dep \rightarrow O_{j(i)}$ *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 amazing and "easy" to use.* *easy* →**conj**→ *amazing*
$R4_2$	$O_i \rightarrow O_i\text{-}Dep \rightarrow H \leftarrow O_j\text{-}Dep \leftarrow O_j$ *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 sexy, "cool", accessory-available mp3 player, you can choose iPod.* *sexy* →**mod**→ *player* ←**mod**← *cool*

Identify aspect synonyms (grouping)

- Liu et al (WWW-05) made an attempt using only WordNet.
- Carenini et al (K-CAP-05) proposed a more sophisticated method based on similarity metrics, but it requires a taxonomy of aspects to be given.
  - The system merges each discovered aspect to a aspect node in the taxonomy.
  - The similarity metrics are defined based on string similarity, synonyms and other distances measured using WordNet.
- (Zhai et al Coling-2010; Zhai et al WSDM-2011) proposed a semi-supervised learning method and a unsupervised learning method together with linguistic constraints.

Aspect sentiment classification

- For each aspect, we identify the sentiment or opinion orientation expressed by a reviewer.
- We work based on sentences, but also consider,
  - A sentence can contain multiple aspects.
  - Different aspects may have 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 again:
  - Some opinion words have context independent orientations, e.g., "great".
  - Some other opinion words have context dependent orientations, e.g., "small"
- Many ways to use them.

Aggregation of opinion words

(Hu and Liu, KDD-04; Ding and Liu, 2008)

- Input: a pair (*f*, *s*), where *f* is a product feature (aspect) and *s* is a sentence that contains *f*.
- Output: whether the opinion on *f* in *s* is positive, negative, 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 *f*. Let the set of opinion words in s_f be w_1, .., w_n. Sum up their orientations (1, -1, 0), and assign the orientation to (*f*, *s*) accordingly.
- In (Ding and Liu, SIGIR-07), step 2 is changed to $\sum_{i=1}^{n} \frac{w_i.o}{d(w_i, f)}$

  with better results. $w_i.o$ is the opinion orientation of w_i. $d(w_i, f)$ is the distance from *f* to w_i.

Context dependent opinions

- Popescu and Etzioni (EMNLP-05) used
  - constraints of connectives in (Hazivassiloglou and McKeown, ACL-97), and some additional constraints, e.g., morphological relationships, synonymy and antonymy, and
  - relaxation labeling to propagate opinion orientations to words and features.
- Ding and Liu (2008) used
  - constraints of connectives both at intra-sentence and inter-sentence levels, and
  - additional constraints of, e.g., TOO, BUT, NEGATION, ….

  to directly assign opinions to (*f*, *s*) with good results (> 0.85 of F-score).

Basic Opinion Rules (Liu, Ch. in NLP handbook)

Opinions are governed by some rules, e.g.,

1. Neg → Negative
2. Pos → Positive
3. Negation Neg → Positive
4. Negation Pos → Negative
5. Desired value range → Positive
6. Below or above the desired value range → Negative

Basic Opinion Rules (Liu, Ch. in NLP handbook)

7. Decreased Neg → Positive
8. Decreased Pos → Negative
9. Increased Neg → Negative
10. Increased Pos → Positive
11. Consume resource → Negative
12. Produce resource → Positive
13. Consume waste → Positive
14. Produce waste → Negative

Divide and Conquer

- Most current techniques seem to assume one-technique-fit-all solution. Unlikely??
  - "The picture quality of this camera is great."
  - "Sony cameras take better pictures than Nikon".
  - "If you are looking for a camera with great picture quality, buy Sony."
  - "If Sony makes good cameras, I will buy one."
- Narayanan, et al (2009) took a divide and conquer approach to study conditional sentences

Roadmap

- Opinion mining – problem definition
- Document level sentiment classification
- Sentence level sentiment classification
- Opinion lexicon generation
- Aspect-based opinion mining
- **Opinion mining of comparative sentences**
- Opinion spam detection
- Summary

Extraction of Comparatives

(Jinal and Liu, SIGIR-06, AAAI-06; Liu's Web Data Mining book)

- Recall: Two types of evaluation
  - regular opinions: "This car is bad"
  - Comparisons: "Car X is not as good as car Y"
- They use different language constructs.
- Direct expression of sentiments are good. Comparison may be better.
  - Good or bad, compared to what?
- Comparative Sentence Mining
  - Identify comparative sentences, and
  - extract comparative relations from them.

Two Main Types of Opinions

- Regular Opinions: direct sentiment expressions on some target entities, e.g., products, events, topics, persons.
  - E.g., "the picture quality of this camera is great."
- Comparative Opinions: Comparisons expressing similarities or differences of more than one entity. Usually stating an ordering or preference.
  - E.g., "car x is cheaper than car y."

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 camera available in market*"

Types of comparatives: non-gradable

- Non-Gradable: Sentences that compare aspects of two or more entities, but do not grade them. Sentences which imply:
  - Entity A is similar to or different from entity B with regard to some aspects.
  - Entity A has aspect F_1, entity B has aspect F_2 (F_1 and F_2 are usually substitutable).
  - Entity A has aspect F, but entity B does not have.

Mining 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.

Roadmap

- Opinion mining – problem definition
- Document level sentiment classification
- Sentence level sentiment classification
- Opinion lexicon generation
- Aspect-based opinion mining
- Opinion mining of comparative sentences
- **Opinion spam detection**
- Summary

Opinion Spam Detection (Jindal and Liu, 2007)

- Fake/untruthful reviews:
  - Write undeserving positive reviews for some target entities in order to promote them.
  - Write unfair or malicious negative reviews for some target entities to damage their reputations.
- Increasing number of customers wary of fake reviews (biased reviews, paid reviews)

An Example of 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)

Is this review fake or not?

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

What about this?

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

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.

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

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.

Experiments with 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> <Number of Helpful Feedbacks>*
- Industry manufactured products "*mProducts*"

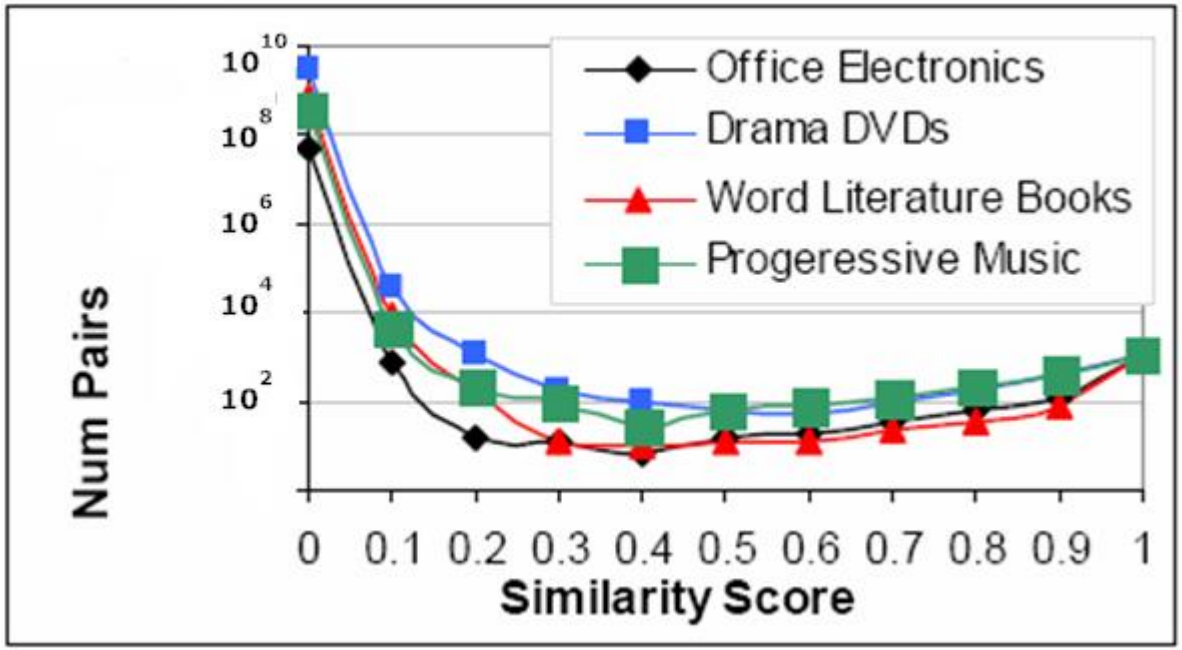
  e.g. electronics, computers, accessories, etc
  - 228K reviews, 36K products and 165K reviewers.

Deal with fake/untruthful reviews

- We have a problem: because
  - It is extremely hard to recognize or label fake/ untruthful reviews manually.
  - Without training data, we cannot do supervised learning.
- Possible solution:
  - Can we make use certain duplicate reviews as fake reviews (which are almost certainly untruthful)?

Duplicate Reviews

Two reviews which have similar contents are called duplicates



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 data attributes
  - Review centric features (content)
    - Features about reviews
  - Reviewer centric features
    - Features about the reviewers
  - Product centric features
    - Features about products reviewed.

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:

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

The unexpectedness is computed as follows (information gain):

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

Confidence unexpectedness

Rule: reviewer-1, brand-1 → 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.

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

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

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.

Definitions and equations

- Trustiness of a reviewer *r*

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

- Honesty of a review *v*

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

- Reliability of store *s*

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

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 (frustrtaing). 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

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)

Roadmap

- Opinion mining – problem definition
- Document level sentiment classification
- Sentence level sentiment classification
- Opinion lexicon generation
- Aspect-based opinion mining
- Opinion mining of comparative sentences
- Opinion spam detection
- **Summary**

Summary

- We briefly defined and introduced
  - Regular opinions: document, sentence and aspect level
  - Comparative opinions: different types of comparisons
  - Opinion spam detection: fake reviews.
- There are already many applications.
- Technical challenges are still huge.
  - Accuracy of all tasks is still a major issue
- But I am optimistic. Accurate solutions will be out in the next few years. Maybe it's already there.
  - A lot of unknown methods from industry.