Web Mining
(網路探勘)

Opinion Mining and Sentiment Analysis
(意見探勘與情感分析)

1011WM11
TLMXM1A
Wed 8,9 (15:10-17:00) U705

Min-Yuh Day
戴敏育
Assistant Professor
專任助理教授

Dept. of Information Management, Tamkang University
淡江大學 資訊管理學系

http://mail.tku.edu.tw/myday/
2012-12-12
Outline

• Introduction to Opinion Mining and Sentiment Analysis
• Social Media Monitoring/Analysis
• Resources of Opinion Mining
<table>
<thead>
<tr>
<th>週次</th>
<th>日期</th>
<th>內容（Subject/Topics）</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>101/09/12</td>
<td>Introduction to Web Mining (網路探勘導論)</td>
</tr>
<tr>
<td>2</td>
<td>101/09/19</td>
<td>Association Rules and Sequential Patterns (關聯規則和序列模式)</td>
</tr>
<tr>
<td>3</td>
<td>101/09/26</td>
<td>Supervised Learning (監督式學習)</td>
</tr>
<tr>
<td>4</td>
<td>101/10/03</td>
<td>Unsupervised Learning (非監督式學習)</td>
</tr>
<tr>
<td>5</td>
<td>101/10/10</td>
<td>國慶紀念日 (放假一天)</td>
</tr>
<tr>
<td>6</td>
<td>101/10/17</td>
<td>Paper Reading and Discussion (論文研讀與討論)</td>
</tr>
<tr>
<td>7</td>
<td>101/10/24</td>
<td>Partially Supervised Learning (部分監督式學習)</td>
</tr>
<tr>
<td>8</td>
<td>101/10/31</td>
<td>Information Retrieval and Web Search (資訊檢索與網路搜尋)</td>
</tr>
<tr>
<td>9</td>
<td>101/11/07</td>
<td>Social Network Analysis (社會網路分析)</td>
</tr>
<tr>
<td>週次</td>
<td>日期</td>
<td>內容（Subject/Topics）</td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>----------------------------------------------------------</td>
</tr>
<tr>
<td>10</td>
<td>101/11/14</td>
<td>Midterm Presentation (期中報告)</td>
</tr>
<tr>
<td>11</td>
<td>101/11/21</td>
<td>Web Crawling (網路爬行)</td>
</tr>
<tr>
<td>12</td>
<td>101/11/28</td>
<td>Structured Data Extraction (結構化資料擷取)</td>
</tr>
<tr>
<td>13</td>
<td>101/12/05</td>
<td>Information Integration (資訊整合)</td>
</tr>
<tr>
<td>14</td>
<td>101/12/12</td>
<td>Opinion Mining and Sentiment Analysis (意見探勘與情感分析)</td>
</tr>
<tr>
<td>15</td>
<td>101/12/19</td>
<td>Paper Reading and Discussion (論文研讀與討論)</td>
</tr>
<tr>
<td>16</td>
<td>101/12/26</td>
<td>Web Usage Mining (網路使用挖掘)</td>
</tr>
<tr>
<td>17</td>
<td>102/01/02</td>
<td>Project Presentation 1 (期末報告1)</td>
</tr>
<tr>
<td>18</td>
<td>102/01/09</td>
<td>Project Presentation 2 (期末報告2)</td>
</tr>
</tbody>
</table>
Social Media and the Voice of the Customer

• Listen to the Voice of the Customer (VoC)
  – Social media can give companies a torrent of highly valuable customer feedback.
  – Such input is largely free
  – Customer feedback issued through social media is qualitative data, just like the data that market researchers derive from focus group and in-depth interviews
  – Such qualitative data is in digital form – in text or digital video on a web site.

Listen and Learn
Text Mining for VoC

• Categorization
  – Understanding what topics people are talking or writing about in the unstructured portion of their feedback.

• Sentiment Analysis
  – Determining whether people have positive, negative, or neutral views on those topics.

Opinion Mining and Sentiment Analysis

• Mining opinions which indicate positive or negative sentiments
• Analyzes people’s opinions, appraisals, attitudes, and emotions toward entities, individuals, issues, events, topics, and their attributes.

Opinion Mining and 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
• Opinion mining / Sentiment Analysis are widely used in academia
• Opinion mining / Sentiment Analysis can be used interchangeably

Example of Opinion: review segment on iPhone

“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.
However, my mother was mad with me as I did not tell her before I bought it.
She also thought the phone was too expensive, and wanted me to return it to the shop. ... ”

Example of Opinion: review segment on iPhone

“(1) I bought an [ ] a few days ago.
(2) It was such a nice phone.
(3) The [ ] was really cool.
(4) The [ ] was clear too.
(5) However, my mother was mad with me as I did not tell her before I bought it.
(6) She also thought the phone was too expensive, and wanted me to return it to the shop. ...”

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 opinion of others.

In the past,

– Individuals
  • Seek opinions from friends and family

– Organizations
  • Use surveys, focus groups, opinion pools, consultants

Word-of-mouth on the Social media

• Personal experiences and opinions about anything in reviews, forums, blogs, micro-blog, Twitter.

• Posting at social networking sites, e.g., Facebook

• Comments about articles, issues, topics, reviews.

Social media + beyond

• **Global scale**
  – No longer – one’s circle of friends.

• **Organization internal data**
  – Customer feedback from emails, call center

• **News and reports**
  – Opinions in news articles and commentaries

Applications of Opinion Mining

• Businesses and organizations
  – Benchmark products and services
  – Market intelligence
    • Business spend a huge amount of money to find consumer opinions using consultants, surveys, and focus groups, etc.
• Individual
  – Make decision 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.

Research Area of Opinion Mining

• Many names and tasks with difference objective and models
  – Sentiment analysis
  – Opinion mining
  – Sentiment mining
  – Subjectivity analysis
  – Affect analysis
  – Emotion detection
  – Opinion spam detection

Existing Tools
(“Social Media Monitoring/Analysis")

• Radian 6
• Social Mention
• Overtone OpenMic
• Microsoft Dynamics Social Networking Accelerator
• SAS Social Media Analytics
• Lithium Social Media Monitoring
• RightNow Cloud Monitor

Source: Wiltrud Kessler (2012), Introduction to Sentiment Analysis
Existing Tools
(“Social Media Monitoring/Analysis"

• Radian 6
• Social Mention
• Overtone OpenMic
• Microsoft Dynamics Social Networking Accelerator
• SAS Social Media Analytics
• Lithium Social Media Monitoring
• RightNow Cloud Monitor

Source: Wiltrud Kessler (2012), Introduction to Sentiment Analysis
Word-of-mouth
Voice of the Customer

• 1. Attensity
  – Track social sentiment across brands and competitors

• 2. Clarabridge
  – Sentiment and Text Analytics Software
Attensity: Track social sentiment across brands and competitors

http://www.attensity.com/

http://www.youtube.com/watch?v=4goxmBEg2lw#!
http://www.radian6.com/

http://www.youtube.com/watch?feature=player_embedded&v=8i6Exg3Urg0
http://www.sas.com/software/customer-intelligence/social-media-analytics/
http://www.tweetfeel.com

Those are all the results available right now. Try again or try another term to see how people feel towards it.

Got questions? Read our FAQ.

RT @jiggljello: This 12 year old has an iPhone4s wtf

So my 9 year old little sister has a iPhone4s . Wtf bruh?!

This 12 year old has an iPhone4s wtf

So my sister has a android and i dont even have a phone and she gets a brand new iPhone4s -___- #Wtf

iPhone4s is funny ass a bitch

-Ohwell . a new iPhone4s won't hurt , aha
Niggie, if i dont jailbreak my Iphone4s, u mean i cant play Street fighter4? wtf @louistekneeq
1 minute ago by Flow__Show

リフォロー100％です！ #Arashi #韓fan #wstcg #桜野友美 #followme #アメブロ #kimiboku #ntb #autofollow #apple #mt2 #相互フォローの輪 #iphone4s #Arashi #twitter #前田敦子
1 minute ago by designer_sayaka

Rumor has it that #Apple #iPhone5 will have a 4" screen (compared 2 3.5" in #iPhone4S), giving direct competition 2... http://t.co/zslGPXbX
1 minute ago by abhay01007

RT @BallMe_Dollar: The iPhone4S look better then the iPhone5.
1 minute ago by SuckMy_TwitNuts

#iphoneography #iphonography #bahrain #iphone4s #photography #sunrise #scenery #beautiful #sun
http://t.co/Ee4VNnSL
2 minutes ago by thementaldawg
http://www.i-buzz.com.tw/
OpView Service

連上OpView，品牌形象，新聞觀點監測好輕鬆

什麼是OpView服務？
OpView是協助您蒐集、處理、分析網路資訊的雲端服務。

OpView服務，猶如提供您無盡網絡情報能量的發電廠。以雲端架構為基礎，OpView服務蒐集、處理、分析各類型網絡資訊與情緒，並以雲端服務平臺供客戶使用。OpView服務的資料涵蓋範圍包括台灣最具代表性的新聞網站、
Sentiment Analysis

- Sentiment
  - A thought, view, or attitude, especially one based mainly on emotion instead of reason

- Sentiment Analysis
  - opinion mining
  - use of natural language processing (NLP) and computational techniques to automate the extraction or classification of sentiment from typically unstructured text
Applications of Sentiment Analysis

• Consumer information
  – Product reviews

• Marketing
  – Consumer attitudes
  – Trends

• Politics
  – Politicians want to know voters’ views
  – Voters want to know politicians’ stances and who else supports them

• Social
  – Find like-minded individuals or communities
Sentiment detection

• How to interpret features for sentiment detection?
  – Bag of words (IR)
  – Annotated lexicons (WordNet, SentiWordNet)
  – Syntactic patterns

• Which features to use?
  – Words (unigrams)
  – Phrases/n-grams
  – Sentences
Problem statement of Opinion Mining

• Two aspects of abstraction
  – Opinion definition
    • What is an opinion?
    • What is the structured definition of opinion?
  – Opinion summarization
    • Opinion 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
    • Is this review + or -?
  – Sentence level
    • 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 opinion are expressed
Two main types of opinions

• **Regular opinions**: Sentiment/Opinion expressions on some target entities
  
  – **Direct opinions**: sentiment expressions on one object:
    
    • “The touch screen is really cool.”
    
    • “The picture quality of this camera is great”
  
  – **Indirect opinions**: comparisons, relations expressing similarities or differences (objective or subjective) of more than one object
    
    • “phone X is cheaper than phone Y.” (objective)
    
    • “phone X is better than phone Y.” (subjective)

• **Comparative opinions**: comparisons of more than one entity.
  
  – “iPhone is better than Blackberry.”

Subjective and Objective

• Objective
  – An objective sentence expresses some factual information about the world.
  – “I returned the phone yesterday.”
  – Objective sentences can implicitly indicate opinions
    • “The earphone broke in two days.”

• Subjective
  – A subjective sentence expresses some personal feelings or beliefs.
  – “The voice on my phone was not so clear”
  – Not every subjective sentence contains an opinion
    • “I wanted a phone with good voice quality”

• ➔ Subjective analysis

Sentiment Analysis vs. Subjectivity Analysis

<table>
<thead>
<tr>
<th>Sentiment Analysis</th>
<th>Subjectivity Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Subjective</td>
</tr>
<tr>
<td>Negative</td>
<td>Objective</td>
</tr>
<tr>
<td>Neutral</td>
<td></td>
</tr>
</tbody>
</table>
A (regular) opinion

• **Opinion** (a restricted definition)
  – An opinion (regular opinion) is simply a positive or negative sentiment, view, attitude, emotion, or appraisal about an entity or an aspect of the entity from an opinion holder.

• **Sentiment orientation of an opinion**
  – Positive, negative, or neutral (no opinion)
  – Also called:
    • Opinion orientation
    • Semantic orientation
    • Sentiment polarity

Entity and aspect

- **Definition of Entity:**
  - An *entity e* is a product, person, event, organization, or topic.
  - e is represented as
    - A hierarchy of components, sub-components.
    - 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.

- **Aspects (features):**
  - Represent both components and attribute.

Entity and aspect

Opinion definition

• An opinion is a quintuple

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

where

– $e_j$ is a target entity.
– $a_{jk}$ is an aspect/feature of the entity $e_j$.
– $s_{ijkl}$ is the sentiment value of the opinion from the opinion holder on feature of entity at time. $s_{ijkl}$ 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.

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 on feature of entity at time. \(so_{ijkl}\) 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.

• \((e_j, a_{jk})\) is also called opinion target

Terminologies

- **Entity**: object
- **Aspect**: feature, attribute, facet
- **Opinion holder**: opinion source
- **Topic**: entity, aspect
- **Product features, political issues**

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.

Emotion

• Six main emotions
  – Love
  – Joy
  – Surprise
  – Anger
  – Sadness
  – Fear

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 <> a number of opinions
  - Aspect-based summary is more suitable
  - Quintuples form the basis for opinion summarization

An aspect-based opinion summary

<table>
<thead>
<tr>
<th>Cellular phone 1:</th>
<th>Aspect: GENERAL</th>
<th>Positive: 125</th>
<th>&lt;individual review sentences&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Negative: 7</td>
<td></td>
<td>&lt;individual review sentences&gt;</td>
</tr>
<tr>
<td>Aspect: Voice quality</td>
<td>Positive: 120</td>
<td>&lt;individual review sentences&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Negative: 8</td>
<td>&lt;individual review sentences&gt;</td>
<td></td>
</tr>
<tr>
<td>Aspect: Battery</td>
<td>Positive: 80</td>
<td>&lt;individual review sentences&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Negative: 12</td>
<td>&lt;individual review sentences&gt;</td>
<td></td>
</tr>
</tbody>
</table>

Visualization of aspect-based summaries of opinions

Visualization of aspect-based summaries of opinions

Classification Based on Supervised Learning

• Sentiment classification
  – Supervised learning Problem
  – Three classes
    • Positive
    • Negative
    • Neutral

Opinion words in Sentiment classification

• topic-based classification
  – topic-related words are important
    • e.g., politics, sciences, sports

• Sentiment classification
  – topic-related words are unimportant
  – opinion words (also called sentiment words)
    • that indicate positive or negative opinions are important,
      e.g., great, excellent, amazing, horrible, bad, worst

Features in Opinion Mining

• Terms and their frequency
  – TF-IDF

• Part of speech (POS)
  – Adjectives

• Opinion words and phrases
  – beautiful, wonderful, good, and amazing are positive opinion words
  – bad, poor, and terrible are negative opinion words.
  – opinion phrases and idioms, e.g., cost someone an arm and a leg

• Rules of opinions

• Negations

• Syntactic dependency

Rules of opinions

Syntactic template

<subj> passive-verb
<subj> active-verb
active-verb <dobj>
noun aux <dobj>
passive-verb prep <np>

Example pattern

<subj> was satisfied
<subj> complained
endorsed <dobj>
fact is <dobj>
was worried about <np>

A Brief Summary of **Sentiment Analysis** Methods

<table>
<thead>
<tr>
<th>Study</th>
<th>Analysis Task</th>
<th>Sentiment Identification Method</th>
<th>Level</th>
<th>Sentiment Aggregation Method</th>
<th>Level</th>
<th>Nature of Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hu and Li, 2011</td>
<td>Polarity</td>
<td>ML (Probabilistic model)</td>
<td>Snippet</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Li and Wu, 2010</td>
<td>Polarity</td>
<td>Lexicon/Rule</td>
<td>Phrase</td>
<td>Sum</td>
<td>Snippet</td>
<td>Valence</td>
</tr>
<tr>
<td>Thelwall et al., 2010</td>
<td>Polarity</td>
<td>Lexicon/Rule</td>
<td>Sentence</td>
<td>Max &amp; Min</td>
<td>Snippet</td>
<td>Range</td>
</tr>
<tr>
<td>Boiy and Moens, 2009</td>
<td>Both</td>
<td>ML (Cascade ensemble)</td>
<td>Sentence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chung 2009</td>
<td>Polarity</td>
<td>Lexicon</td>
<td>Phrase</td>
<td>Average</td>
<td>Sentence</td>
<td>Valence</td>
</tr>
<tr>
<td>Wilson, Wiebe, and Hoffmann, 2009</td>
<td>Both</td>
<td>ML (SVM, AdaBoost, Rule, etc.)</td>
<td>Phrase</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Zhang et al., 2009</td>
<td>Polarity</td>
<td>Lexicon/Rule</td>
<td>Sentence</td>
<td>Weighted average</td>
<td>Snippet</td>
<td>Valence</td>
</tr>
<tr>
<td>Abbasi, Chen, and Salem, 2008</td>
<td>Polarity</td>
<td>ML (GA + feature selection)</td>
<td>Snippet</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Subrahmanian and Reforgiato, 2008</td>
<td>Polarity</td>
<td>Lexicon/Rule</td>
<td>Phrase</td>
<td>Rule</td>
<td>Snippet</td>
<td>Valence</td>
</tr>
<tr>
<td>Tan and Zhang 2008</td>
<td>Polarity</td>
<td>ML (SVM, Winnow, NB, etc.)</td>
<td>Snippet</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Airoldi, Bai, and Padman, 2007</td>
<td>Polarity</td>
<td>ML (Markov Blanket)</td>
<td>Snippet</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Das and Chen, 2007</td>
<td>Polarity</td>
<td>ML (Bayesian, Discriminate, etc.)</td>
<td>Snippet</td>
<td>Average</td>
<td>Daily</td>
<td>Valence</td>
</tr>
<tr>
<td>Liu et al., 2007</td>
<td>Polarity</td>
<td>ML (PLSA)</td>
<td>Snippet</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Kennedy and Inkpen, 2006</td>
<td>Polarity</td>
<td>Lexicon/Rule, ML (SVM)</td>
<td>Phrase</td>
<td>Count</td>
<td>Snippet</td>
<td>Valence</td>
</tr>
<tr>
<td>Mishne 2006</td>
<td>Polarity</td>
<td>Lexicon</td>
<td>Phrase</td>
<td>Average</td>
<td>Snippet</td>
<td>Valence</td>
</tr>
<tr>
<td>Mishne 2005</td>
<td>Polarity</td>
<td>ML (SVM)</td>
<td>Snippet</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Popescu and Etzioni 2005</td>
<td>Polarity</td>
<td>Lexicon/Rule</td>
<td>Phrase</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Efron 2004</td>
<td>Polarity</td>
<td>ML (SVN, NB)</td>
<td>Snippet</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Wilson, Wiebe, and Hwa, 2004</td>
<td>Both</td>
<td>ML (SVM, AdaBoost, Rule, etc.)</td>
<td>Sentence</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Nigam and Hurst 2004</td>
<td>Polarity</td>
<td>Lexicon/Rule</td>
<td>Chunk</td>
<td>Rule</td>
<td>Sentence</td>
<td>Valence</td>
</tr>
<tr>
<td>Dave, Lawrence, and Pennock, 2003</td>
<td>Polarity</td>
<td>ML (SVM, Rainbow, etc.)</td>
<td>Snippet</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Nasukawa and Yi 2003</td>
<td>Polarity</td>
<td>Lexicon/Rule</td>
<td>Phrase</td>
<td>Rule</td>
<td>Sentence</td>
<td>Valence</td>
</tr>
<tr>
<td>Yi et al., 2003</td>
<td>Polarity</td>
<td>Lexicon/Rule</td>
<td>Phrase</td>
<td>Rule</td>
<td>Sentence</td>
<td>Valence</td>
</tr>
<tr>
<td>Yu and Hatzivassiloglou 2003</td>
<td>Both</td>
<td>ML (NB) + Lexicon/Rule</td>
<td>Phrase</td>
<td>Average</td>
<td>Sentence</td>
<td>Valence</td>
</tr>
<tr>
<td>Pang, Lee, and Vaithyanathan 2002</td>
<td>Polarity</td>
<td>ML (SVM, MaxEnt, NB)</td>
<td>Snippet</td>
<td></td>
<td></td>
<td>Valence</td>
</tr>
<tr>
<td>Subasic and Huefner 2001</td>
<td>Polarity</td>
<td>Lexicon/Fuzzy logic</td>
<td>Phrase</td>
<td>Average</td>
<td>Snippet</td>
<td>Valence</td>
</tr>
<tr>
<td>Turney 2001</td>
<td>Polarity</td>
<td>ML (ML, Rule)</td>
<td>Phrase</td>
<td>Average</td>
<td>Snippet</td>
<td>Valence</td>
</tr>
</tbody>
</table>

(Both = Subjectivity and Polarity; ML = Machine Learning; Lexicon/Rule = Lexicon enhanced by linguistic rules)

Word-of-Mouth (WOM)

• “This book is the best written documentary thus far, yet sadly, there is no soft cover edition.”

• “This book is the best written documentary thus far, yet sadly, there is no soft cover edition.”

This book is the best written documentary thus far, yet there is no soft cover edition.

<table>
<thead>
<tr>
<th>Word</th>
<th>POS</th>
</tr>
</thead>
<tbody>
<tr>
<td>This</td>
<td>DT</td>
</tr>
<tr>
<td>book</td>
<td>NN</td>
</tr>
<tr>
<td>is</td>
<td>VBZ</td>
</tr>
<tr>
<td>the</td>
<td>DT</td>
</tr>
<tr>
<td>best</td>
<td>JJS</td>
</tr>
<tr>
<td>written</td>
<td>VBN</td>
</tr>
<tr>
<td>documentary</td>
<td>NN</td>
</tr>
<tr>
<td>thus</td>
<td>RB</td>
</tr>
<tr>
<td>far</td>
<td>RB</td>
</tr>
<tr>
<td>,</td>
<td></td>
</tr>
<tr>
<td>yet</td>
<td>RB</td>
</tr>
<tr>
<td>sadly</td>
<td>RB</td>
</tr>
<tr>
<td>,</td>
<td></td>
</tr>
<tr>
<td>there</td>
<td>EX</td>
</tr>
<tr>
<td>is</td>
<td>VBZ</td>
</tr>
<tr>
<td>no</td>
<td>DT</td>
</tr>
<tr>
<td>soft</td>
<td>JJ</td>
</tr>
<tr>
<td>cover</td>
<td>NN</td>
</tr>
<tr>
<td>edition</td>
<td>NN</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>
Conversion of text representation

Datasets of Opinion Mining

- Blog06
  - 25GB TREC test collection
  - [http://ir.dcs.gla.ac.uk/test collections/access to data.html](http://ir.dcs.gla.ac.uk/test collections/access to data.html)
- Cornell movie-review datasets
- Customer review datasets
- Multiple-aspect restaurant reviews
  - [http://people.csail.mit.edu/bsnyder/naacl07](http://people.csail.mit.edu/bsnyder/naacl07)
- NTCIR multilingual corpus
  - NTCIR Multilingual Opinion-Analysis Task (MOAT)

Lexical Resources of Opinion Mining

- SentiWordnet
  - [http://sentiwordnet.isti.cnr.it/](http://sentiwordnet.isti.cnr.it/)
- General Inquirer
  - [http://www.wjh.harvard.edu/~inquirer/](http://www.wjh.harvard.edu/~inquirer/)
- OpinionFinder’s Subjectivity Lexicon
  - [http://www.cs.pitt.edu/mpqa/](http://www.cs.pitt.edu/mpqa/)
- NTU Sentiment Dictionary (NTUSD)
- Hownet Sentiment
## Example of SentiWordNet

<table>
<thead>
<tr>
<th>POS</th>
<th>ID</th>
<th>PosScore</th>
<th>NegScore</th>
<th>SynsetTerms</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>00217728</td>
<td>0.75</td>
<td>0</td>
<td>beautiful#1</td>
<td>delighting the senses or exciting intellectual or emotional admiration; &quot;a beautiful child&quot;; &quot;beautiful country&quot;; &quot;a beautiful painting&quot;; &quot;a beautiful theory&quot;; &quot;a beautiful party&quot;</td>
</tr>
<tr>
<td>a</td>
<td>00227507</td>
<td>0.75</td>
<td>0</td>
<td>best#1</td>
<td>(superlative of `good') having the most positive qualities; &quot;the best film of the year&quot;; &quot;the best solution&quot;; &quot;the best time for planting&quot;; &quot;wore his best suit&quot;</td>
</tr>
<tr>
<td>r</td>
<td>00042614</td>
<td>0</td>
<td>0.625</td>
<td>unhappily#2 sadly#1</td>
<td>in an unfortunate way; &quot;sadly he died before he could see his grandchild&quot;</td>
</tr>
<tr>
<td>r</td>
<td>00093270</td>
<td>0</td>
<td>0.875</td>
<td>woefully#1 sadly#3 lamentably#1 deplorably#1</td>
<td>in an unfortunate or deplorable manner; &quot;he was sadly neglected&quot;; &quot;it was woefully inadequate&quot;</td>
</tr>
<tr>
<td>r</td>
<td>00404501</td>
<td>0</td>
<td>0.25</td>
<td>sadly#2</td>
<td>with sadness; in a sad manner; &quot;`She died last night,' he said sadly&quot;</td>
</tr>
</tbody>
</table>
《知網》情感分析用詞語集（beta版）

• “中英文情感分析用詞語集”
  - 包含詞語約 17887

• “中文情感分析用詞語集”
  - 包含詞語約 9193

• “英文情感分析用詞語集”
  - 包含詞語 8945

<table>
<thead>
<tr>
<th>中文情感分析用詞語集</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>中文正面情感詞語</td>
<td>836</td>
</tr>
<tr>
<td>中文正面評價詞語</td>
<td>1254</td>
</tr>
<tr>
<td>中文正面評價詞語</td>
<td>3730</td>
</tr>
<tr>
<td>中文負面評價詞語</td>
<td>3116</td>
</tr>
<tr>
<td>中文程度級別詞語</td>
<td>219</td>
</tr>
<tr>
<td>中文主張詞語</td>
<td>38</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>9193</strong></td>
</tr>
</tbody>
</table>
中文情感分析用詞語集

• “正面情感”詞語
  - 如：
    愛，讚賞，快樂，感同身受，好奇，喝彩，魂牽夢縈，嘉許 ...

• “負面情感”詞語
  - 如：
    哀傷，半信半疑，鄙視，不滿意，不是滋味兒，後悔，大失所望 ...

中文情感分析用詞語集

• “正面評價” 詞語
  - 如：
    不可或缺，部優，才高八斗，沉魚落雁，
    催人奮進，動聽，對勁兒 ...

• “負面評價” 詞語
  - 如：
    醜，苦，超標，華而不實，荒涼，混濁，
    畸輕畸重，價高，空洞無物 ...

中文情感分析用詞語集

• “程度級別”詞語
  – 1. “極其 | extreme / 最 | most”
    • 非常，極，極度，無以倫比，為
  – 2. “很 | very”
    • 多麼，分外，格外，著實
  – ...

• “主張”詞語
  – 1. {perception | 感知}
    • 感覺，覺得，預感
  – 2. {regard | 認為}
    • 認為，以為，主張

Summary

• Introduction to Opinion Mining and Sentiment Analysis
• Social Media Monitoring/Analysis
• Resources of Opinion Mining
References


• Wiltrud Kessler (2012), Introduction to Sentiment Analysis, [http://www.ims.uni-stuttgart.de/~kesslewd/lehre/sentimentanalysis12s/introduction_sentimentanalysis.pdf](http://www.ims.uni-stuttgart.de/~kesslewd/lehre/sentimentanalysis12s/introduction_sentimentanalysis.pdf)
