Social Word-of-Mouth and Web Mining

(社群口碑與網路探勘)

Time: 2012/10/31(Wed) 08:10-10:00
Place: 2F Rm.6203, 杏春樓
College of Oral Medicine, Taipei Medical University

Min-Yuh Day
Assistant Professor
Dept. of Information Management, Tamkang University

http://mail.tku.edu.tw/myday/
2012-10-31
Outline

1. Social Media (社群媒體)
2. Social Word-of-Mouth (社群口碑)
3. Web Mining (網路探勘)
Social Media
(社群媒體)
#1 Activity on the Web?

Social Media

Source: Social Media Business, [http://www.youtube.com/watch?v=X9sTq3pzNQQ](http://www.youtube.com/watch?v=X9sTq3pzNQQ)
THE SOCIAL MEDIA MANAGEMENT HANDBOOK
EVERYTHING YOU NEED TO KNOW TO GET SOCIAL MEDIA WORKING IN YOUR BUSINESS

NICK SMITH & ROBERT WOLLAN
WITH
CATHERINE ZHOU

Facebook Marketing An Hour A Day

INSIDE: Get your free ticket to an online Facebook Marketing Workshop with the authors!

Source: http://www.amazon.com/Facebook-Marketing-Hour-Chris-Treadaway/dp/0470569646
YouTube®
Online Video Marketing for Any Business
for Business
Second Edition

This latest edition is a must-read book for any business owner wanting to implement a successful inbound video marketing campaign.
—Rey Ybarra, Host/Producer of “The New Media Radio Hour”
www.newmediaradiohour.com

Michael Miller
BESTSELLER!

Social Media Management Pyramid

Source: http://www.infobarrel.com/Social_Media_Management:_Hiring_a_Social_Media_Manager
Marketing 4P to 4C

- **Product** → **Customer solution**
- **Price** → **Customer Cost**
- **Place** → **Convenience**
- **Promotion** → **Communication**

Source: Kotler and Keller (2008)
Four Pillars of Social Media Strategy

C²E²

Source: Safko and Brake (2009)
Social Media Can Help Orchestrate Three Spheres to Influence to Boost a Company’s Innovation Efforts

- Internal
- Trusted Network
- The World

Innovation

Examples of Social Media Selling Strategies in the Market Today

Strategy #1 – “Accessing social Consumers”: Use Social Media as a New Channel to Individuals

Strategy #2 – “Engaging the Hive”: Get Customers to Mobilize Their Personal Networks

Strategy #3 – “Appealing to Influencers”: Target Influencers Who Can Move the Masses

- “Pro-sumer” collaboration
- Influencer-Led Development
- Customers as “Community Organizers”
- Recruiting others/Group Sales
- “Pass it along” promtions

- Engaging the Advocates
- User Reviews
- Social Media Wildfire
- Creating Urgency/Spontaneous Selling

Social Media Landscape

http://www.fredcavazza.net/2008/06/09/social-media-landscape/
社會媒體 (social media) 的定義
(Kaplan & Haenlein, 2010)

建立在 Web 2.0 概念與技術的基礎上，
以網路為平台的應用系統
(Internet-based applications)，
讓網路使用者可以
方便產生與交流使用者建立的內容
(user generated content; UGC)。
社會媒體服務
(Social Media Services)

提供使用者在網絡環境中使用社會媒體應用系統的線上服務
(online services)

Google+, Youtube, Facebook, Plurk
THE WEB 2.0 REVOLUTION, SOCIAL MEDIA, AND INDUSTRY DISRUPTORS
Web 2.0

• The popular term for advanced Internet technology and applications, including blogs, wikis, RSS, and social bookmarking.

• One of the most significant differences between Web 2.0 and the traditional World Wide Web is greater collaboration among Internet users and other users, content providers, and enterprises.

Source: Turban et al. (2010), Introduction to Electronic Commerce
THE WEB 2.0 REVOLUTION, SOCIAL MEDIA, AND INDUSTRY DISRUPTORS

• REPRESENTATIVE CHARACTERISTICS OF WEB 2.0

  – The ability to tap into the collective intelligence of users
  – Data is made available in new or never-intended ways
  – Web 2.0 relies on user-generated and user-controlled content and data
  – The virtual elimination of software-upgrade cycles makes everything a work in progress and allows rapid prototyping

Source: Turban et al. (2010), Introduction to Electronic Commerce
THE WEB 2.0 REVOLUTION, SOCIAL MEDIA, AND INDUSTRY DISRUPTORS

– Users can access applications entirely through a browser
– An architecture of participation encourages users to add value to the application
– A major emphasis on social networks and computing
– Strong support of information sharing and collaboration
– Rapid and continuous creation of new business models

Source: Turban et al. (2010), Introduction to Electronic Commerce
THE WEB 2.0 REVOLUTION, SOCIAL MEDIA, AND INDUSTRY DISRUPTORS

• WEB 2.0 COMPANIES AND NEW BUSINESS MODELS

• social media

The online platforms and tools that people use to share opinions, experiences, insights, perceptions, and various media, including photos, videos, and music, with each other.

Source: Turban et al. (2010), Introduction to Electronic Commerce
EXHIBIT 7.1 The Emergence and Rise of Mass Social Media

Source: Turban et al. (2010), Introduction to Electronic Commerce
THE WEB 2.0 REVOLUTION, SOCIAL MEDIA, AND INDUSTRY DISRUPTORS

• INDUSTRY AND MARKET DISRUPTORS
  – disruptors
    Companies that introduce a significant change in their industries, thus causing a disruption in normal business operations.

Source: Turban et al. (2010), Introduction to Electronic Commerce
ONLINE SOCIAL NETWORKING: BASICS AND EXAMPLES

• **social networking**

  Social networks and activities conducted in social networks. It also includes activities conducted using Web 2.0 (e.g., wikis, microblogs) not within social networks.

  – The Size of Social Network Sites
  – New Business Models

Source: Turban et al. (2010), Introduction to Electronic Commerce
EXHIBIT 7.4  The Top Nine Social Networking Sites

- MySpace: 253,145,404
- Facebook: 450,000,000
- Windows Live Spaces: 120,000,000
- Friendster: 90,000,000
- hi5: 80,000,000
- Orkut: 67,000,000
- Classmates.com: 50,000,000
- Bebo: 40,000,000
- LinkedIn: 70,000,000

Source: Turban et al. (2010), Introduction to Electronic Commerce
ONLINE SOCIAL NETWORKING: BASICS AND EXAMPLES

– social network analysis (SNA)

The mapping and measuring of relationships and information flows among people, groups, organizations, computers, and other information- or knowledge-processing entities. The nodes in the network are the people and groups, whereas the links show relationships or flows between the nodes. SNAs provide both visual and a quantitative analysis of relationships.

Source: Turban et al. (2010), Introduction to Electronic Commerce
The major reasons to use or deploy a business social network are to:

- Build better customer relationships
- Improve knowledge management
- Facilitate recruiting and retention
- Increase business opportunities
- Build a community
- Gain expert advice
- Improve trade show experiences
- Improve communication and collaboration

Source: Turban et al. (2010), Introduction to Electronic Commerce
THE FUTURE: WEB 3.0 AND WEB 4.0

• Web 3.0

A term used to describe the future of the World Wide Web. It consists of the creation of high-quality content and services produced by gifted individuals using Web 2.0 technology as an enabling platform.

Source: Turban et al. (2010), Introduction to Electronic Commerce
THE FUTURE: WEB 3.0 AND WEB 4.0

– Semantic Web

An evolving extension of the Web in which Web content can be expressed not only in natural language, but also in a form that can be understood, interpreted, and used by intelligent computer software agents, permitting them to find, share, and integrate information more easily.

Source: Turban et al. (2010), Introduction to Electronic Commerce
THE FUTURE: WEB 3.0 AND WEB 4.0

– Web 4.0
The Web generation after Web 3.0. It is still mostly an unknown entity. However, it is envisioned as being based on islands of intelligence and as being ubiquitous.

– Future Threats
  • Security concerns
  • Lack of Net neutrality
  • Copyright complaints
  • Choppy connectivity

Source: Turban et al. (2010), Introduction to Electronic Commerce
COMMERCIAL ASPECTS OF WEB 2.0 AND SOCIAL NETWORKING APPLICATIONS

• WHY IS THERE AN INTEREST?

  – Web 2.0 applications are spreading rapidly, and many of them cater to a specific segment of the population (e.g., music lovers, travelers, game lovers, and car fans), enabling segmented advertising

  – Many users of Web 2.0 tools are young, and they will grow older and have more money to spend

Source: Turban et al. (2010), Introduction to Electronic Commerce
COMMERCIAL ASPECTS OF WEB 2.0 AND SOCIAL NETWORKING APPLICATIONS

• ADVERTISING USING SOCIAL NETWORKS, BLOGS, AND WIKIS
  – Viral (Word-of-Mouth) Marketing
    • viral blogging
      Viral (word-of-mouth) marketing done by bloggers.
    – Classified Ads, Job Listings, and Recruitment
    – Special Advertising Campaigns
    – Mobile Advertising

Source: Turban et al. (2010), Introduction to Electronic Commerce
COMMERCIAL ASPECTS OF WEB 2.0 AND SOCIAL NETWORKING APPLICATIONS

• SHOPPING IN SOCIAL NETWORKS
• FEEDBACK FROM CUSTOMERS: CONVERSATIONAL MARKETING
  – Customer Feedback with Twitter

Source: Turban et al. (2010), Introduction to Electronic Commerce
COMMERCIAL ASPECTS OF WEB 2.0 AND SOCIAL NETWORKING APPLICATIONS

• COMMERCIAL ACTIVITIES IN BUSINESS AND ENTERPRISE SOCIAL NETWORKS
  – Finding and Recruiting Workers
  – Management Activities and Support
  – Training
  – Knowledge Management and Expert Location
  – Enhancing Collaboration
  – Using Blogs and Wikis Inside the Enterprise

Source: Turban et al. (2010), Introduction to Electronic Commerce
EXHIBIT 7.5  Generating Revenue from Web 2.0 Applications

Tenets
- Web as platform
- Beyond single device
- Data as the next “Intel Inside”
- Lightweight models
- Rich user experiences
- Harnessing collective intelligence
- Leverage the long tail

○ = Monetization methods

Source: Turban et al. (2010), Introduction to Electronic Commerce
COMMERCIAL ASPECTS OF WEB 2.0 AND SOCIAL NETWORKING APPLICATIONS

• REVENUE-GENERATION STRATEGIES IN SOCIAL NETWORKS
  – Increased Revenue and Its Benefit

• RISKS AND LIMITATIONS WHEN INTERFACING WITH SOCIAL NETWORKS

• JUSTIFYING SOCIAL MEDIA AND NETWORKING

Source: Turban et al. (2010), Introduction to Electronic Commerce
ENTERTAINMENT WEB 2.0 STYLE: FROM SOCIAL NETWORKS TO MARKETPLACES

• MOBILE WEB 2.0 DEVICES FOR ENTERTAINMENT AND WORK
  – iPhone and Its Clones

Source: Turban et al. (2010), Introduction to Electronic Commerce
Social Word-of-Mouth
(社群口碑)
Social Media
Word-of-Mouth Marketing
How to Start Buzz

• Identify influential individuals and companies and devote extra effort to them
• Supply key people with product samples
• Work through community influentials
• Develop word-of-mouth referral channels to build business
• Provide compelling information that customers want to pass along

Source: Kotler and Keller (2008)
Word-of-Mouth Marketing

• Person-to-person
• Chat rooms
• Blogs
• Twitter, Plurk
• Facebook
• Youtube

Source: Kotler and Keller (2008)
Elements in the Communications Process

Source: Kotler and Keller (2008)
Field of Experience

Sender’s field

Receiver’s field

Source: Kotler and Keller (2008)
The Communications Process

Selective attention
Selective distortion
Selective retention

Source: Kotler and Keller (2008)
Social Media Marketing

• Scorecard for Social Media
  – 4 - Extremely Valuable
  – 3 - Very Valuable
  – 2 - Somewhat Valuable
  – 1 - Not Very Valuable
  – 0 - No Value

Source: Safko and Brake (2009)
### Scorecard for Social Media

<table>
<thead>
<tr>
<th>Social Media Tool</th>
<th>Internal Value</th>
<th>External Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>4 3 2 1 0</td>
<td>4 3 2 1 0</td>
</tr>
<tr>
<td>LinkedIn</td>
<td>4 3 2 1 0</td>
<td>4 3 2 1 0</td>
</tr>
<tr>
<td>Blogger</td>
<td>4 3 2 1 0</td>
<td>4 3 2 1 0</td>
</tr>
<tr>
<td>SlideShare</td>
<td>4 3 2 1 0</td>
<td>4 3 2 1 0</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>4 3 2 1 0</td>
<td>4 3 2 1 0</td>
</tr>
<tr>
<td>Flickr</td>
<td>4 3 2 1 0</td>
<td>4 3 2 1 0</td>
</tr>
<tr>
<td>Picasa</td>
<td>4 3 2 1 0</td>
<td>4 3 2 1 0</td>
</tr>
<tr>
<td>iTunes</td>
<td>4 3 2 1 0</td>
<td>4 3 2 1 0</td>
</tr>
<tr>
<td>Podcast</td>
<td>4 3 2 1 0</td>
<td>4 3 2 1 0</td>
</tr>
<tr>
<td>Youtube</td>
<td>4 3 2 1 0</td>
<td>4 3 2 1 0</td>
</tr>
<tr>
<td>Twitter</td>
<td>4 3 2 1 0</td>
<td>4 3 2 1 0</td>
</tr>
<tr>
<td>Plurk</td>
<td>4 3 2 1 0</td>
<td>4 3 2 1 0</td>
</tr>
</tbody>
</table>

*Scorecard for Social Media*

4 - Extremely Valuable, 3 - Very Valuable, 2 – Somewhat Valuable, 1 - Not Very Valuable, 0 - No Value

Source: Safko and Brake (2009)
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.

Accenture’s SLOPE Model for Listening to the Social Voice of the Customer

Social Voice of the Customer

Synchronize → Listen & Learn → Optimize & Operationalize → Personalize & Propagate → Execution & Expectations
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.

Customers’ Opinions About Operational versus Customer Experience Issues

Reactive, Reputation Management

Operational Issue

Customer Experience

Multiple Customers

Individual Customer

Urgency

Social Media Can Help Orchestrate Three Spheres to Influence to Boost a Company’s Innovation Efforts

Examples of Social Media Selling Strategies in the Market Today

Strategy #1 – “Accessing social Consumers”: Use Social Media as a New Channel to Individuals

Strategy #2 – “Engaging the Hive”: Get Customers to Mobilize Their Personal Networks

Strategy #3 – “Appealing to Influencers”: Target Influencers Who Can Move the Masses

- “Pro-sum”er collaboration
- Influencer-Led Development
- Customers as “Community Organizers”
- Recruiting others/Group Sales
- “Pass it along” promotions
- User Reviews
- Social Media Wildfire
- Creating Urgency/Spontaneous Selling

Word-of-Mouth

Social Media

word of mouth 1.00 social media 7.40

http://www.google.com.tw/trends/?q=word+of+mouth,+social+media&ctab=0&geo=all&date=all&sort=0
Case Study: LenovoClub CareerLife 職場人生

http://www.lenovoclub.com.tw/careerlife/
Case Study: LenovoClub CareerLife 職場人生

第一集 按怎擲骰免驚 熱烈上映中

[職場人生] 第一集 不管怎擲骰免驚

拉NO佛國際集團空出副總職缺，
平日是死對頭的行銷部經理貳可陳，和業務部經理尖妮ㄛㄛ／
為求升官互相惡鬥，
壕聲過程中有人順勢把手中一杯水潑出……

http://www.lenovoclub.com.tw/careerlife/
Case Study: LenovoClub CareerLife 職場人生

http://www.youtube.com/watch?v=XRUvbfEnPig

重要探案活動

<table>
<thead>
<tr>
<th>日期</th>
<th>活動</th>
<th>觀看次數</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>精選影片初次觀看次數</td>
<td>7,250</td>
</tr>
<tr>
<td>B</td>
<td>初次嵌入於 static.ak.facebook.com</td>
<td>6,380</td>
</tr>
<tr>
<td>C</td>
<td>初次嵌入於 share.youthwant.com.tw</td>
<td>2,946</td>
</tr>
<tr>
<td>D</td>
<td>首次透過廣告播放</td>
<td>7,727</td>
</tr>
<tr>
<td>E</td>
<td>初次嵌入於 <a href="http://www.mobile01.com">www.mobile01.com</a></td>
<td>2,820</td>
</tr>
<tr>
<td>F</td>
<td>初次嵌入於 <a href="http://www.lenovoclub.com.tw">www.lenovoclub.com.tw</a></td>
<td>14,749</td>
</tr>
<tr>
<td>G</td>
<td>來自 YouTube搜尋的首次推薦 – 職場人生</td>
<td>9,522</td>
</tr>
<tr>
<td>H</td>
<td>首次透過行動裝置播放</td>
<td>4,700</td>
</tr>
<tr>
<td>I</td>
<td>首次推薦媒介 – <a href="http://www.facebook.com">www.facebook.com</a></td>
<td>4,583</td>
</tr>
<tr>
<td>J</td>
<td>來自相關影片的首次推薦 – 職場人生</td>
<td>8,749</td>
</tr>
</tbody>
</table>

觀眾

此影片最受下列人士歡迎：

<table>
<thead>
<tr>
<th>性別</th>
<th>年齡</th>
</tr>
</thead>
</table>

此影片在以下地點最受歡迎：
觀眾
此影片最受下列人士歡迎：

<table>
<thead>
<tr>
<th>性別</th>
<th>年齡</th>
</tr>
</thead>
<tbody>
<tr>
<td>男性</td>
<td>35-44</td>
</tr>
<tr>
<td>男性</td>
<td>25-34</td>
</tr>
<tr>
<td>男性</td>
<td>18-24</td>
</tr>
</tbody>
</table>

職場人生

http://www.youtube.com/watch?v=XRUVbFEnPig

個讀者澄清說，平日是對不切實際的行銷部經理麥可陳和業務部經理尖妮丫互聯網...
http://www.lenovoclub.com.tw/careerlife/?utm_source=Youtube&utm_medium=vi...

顯示更多

最佳評論

不簡單的推薦片段!! 裡面完全沒畫半點廣告氣息，亦完全沒有感到是在播廣告!!
而且他完全不會想在電台求救!! 因為在電台大家就會以為是在看廣告!! 但在電台
世界大家就會以為是產品的體現!! 這個產品，不會在廣告時說那是
什麼產品!! 兩個字已經完全地帶出廣告意思!!! 不簡單的廣告!! 不簡單的姓沛和
拉NO佛
nomuchmore 7個月以前 14

“我就跟你姓朱” 笑死XDDD
p90349 6個月以前 10

http://www.youtube.com/watch?v=XRUVbFEnPig
Web Mining
(網路探勘)
ACM Categories and Subject Descriptors

• I.2.7 [Artificial Intelligence]
  – Natural Language Processing
    • Text analysis

• H.2.8 [Database Management]
  – Database Applications
    • Data mining
Web Mining

Web mining (or Web data mining) is the process of discovering intrinsic relationships from Web data (textual, linkage, or usage)

- **Web Content Mining**
  Source: unstructured textual content of the Web pages (usually in HTML format)

- **Web Structure Mining**
  Source: the unified resource locator (URL) links contained in the Web pages

- **Web Usage Mining**
  Source: the detailed description of a Web site's visits (sequence of clicks by sessions)

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems
Web Content/Structure Mining

• Mining of the textual content on the Web
• Data collection via Web crawlers

• Web pages include hyperlinks
  – Authoritative pages
  – Hubs
  – hyperlink-induced topic search (HITS) alg

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems
Web Usage Mining

• Extraction of information from data generated through Web page visits and transactions...
  – data stored in server access logs, referrer logs, agent logs, and client-side cookies
  – user characteristics and usage profiles
  – metadata, such as page attributes, content attributes, and usage data

• Clickstream data

• Clickstream analysis

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems
Web Usage Mining

• Web usage mining applications
  – Determine the lifetime value of clients
  – Design cross-marketing strategies across products.
  – Evaluate promotional campaigns
  – Target electronic ads and coupons at user groups based on user access patterns
  – Predict user behavior based on previously learned rules and users' profiles
  – Present dynamic information to users based on their interests and profiles...

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems
Web Usage Mining
(clickstream analysis)

- Pre-Process Data
  - Collecting
  - Merging
  - Cleaning
  - Structuring
    - Identify users
    - Identify sessions
    - Identify page views
    - Identify visits

- Extract Knowledge
  - Usage patterns
  - User profiles
  - Page profiles
  - Visit profiles
  - Customer value

How to better the data
How to improve the Web site
How to increase the customer value

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems
Web Mining Success Stories

- Amazon.com, Ask.com, Scholastic.com, ...
- Website Optimization Ecosystem

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems
Text and Web Mining

• Text Mining: Applications and Theory
• Web Mining and Social Networking
• Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites
• Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data
• Search Engines – Information Retrieval in Practice
Text Mining

Web Mining and Social Networking

Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites

http://www.amazon.com/Mining-Social-Web-Analyzing-Facebook/dp/1449388345
Web Data Mining
Exploring Hyperlinks, Contents, and Usage Data

1. Introduction
2. Association Rules and Sequential Patterns
3. Supervised Learning
4. Unsupervised Learning
5. Partially Supervised Learning
6. Information Retrieval and Web Search
7. Social Network Analysis
8. Web Crawling
9. Structured Data Extraction: Wrapper Generation
10. Information Integration
11. Opinion Mining and Sentiment Analysis
12. Web Usage Mining

Source: http://www.cs.uic.edu/~liub/WebMiningBook.html
Text Mining

• Text mining (text data mining)
  – the process of deriving high-quality information from text
• Typical text mining tasks
  – text categorization
  – text clustering
  – concept/entity extraction
  – production of granular taxonomies
  – sentiment analysis
  – document summarization
  – entity relation modeling
  • i.e., learning relations between named entities.

http://en.wikipedia.org/wiki/Text_mining
Web Mining

• Web mining
  – discover useful information or knowledge from the Web hyperlink structure, page content, and usage data.

• Three types of web mining tasks
  – Web structure mining
  – Web content mining
  – Web usage mining

Natural Language Processing (NLP)

• Structuring a collection of text
  – **Old approach**: bag-of-words
  – **New approach**: natural language processing

• NLP is ...
  – a very important concept in text mining
  – a subfield of artificial intelligence and computational linguistics
  – the studies of "understanding" the natural human language

• **Syntax versus semantics** based text mining

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems
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, ets., 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 touch screen was really cool.
(4) The voice quality 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
Word-of-mouth
Voice of the Customer

• 1. Attensity
  – Track social sentiment across brands and competitors
  – http://www.attensity.com/home/

• 2. Clarabridge
  – Sentiment and Text Analytics Software
  – http://www.clarabridge.com/
Attensity: Track social sentiment across brands and competitors

http://www.attensity.com/

http://www.youtube.com/watch?v=4goxmBEg2lw#!
Clarabridge: Sentiment and Text Analytics Software

http://www.clarabridge.com/

http://www.youtube.com/watch?v=IDHudt8M9P0
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

FAQ | Contact Us

(tweetfeel)

iPhone 4s

Search

Try some Twitter trends: Tomorrow is June H&M Defense of Marriage Act Diddy’s Bloomberg UCLA ESPN

mış 40

Mağ 41

 equals 51%

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 @jigginjello: This 12 year old has an iPhone 4s WTF

So my 9 year old little sister has a iPhone 4s. WTF bruh?!

This 12 year old has an iPhone 4s WTF

So my sister has a android and i dont even have a phone and she gets a brand new iPhone 4s ----- #WTF

iPhone 4s is funny ass a bitch

-Ohwell . a new iPhone 4s won't hurt . aha
Niggee, if i dont jailbreak my Iphone4s, u mean i cant play Street fighter4? wtf @louistekneeq
1 minute ago by Flow__Show

リフフォロー100％です！ #Arashi #韓fan #wstcg #板野友美 #followme #アメブロ #kimiboku #ntb #autofollow #apple #mt2 #相互フォローレの輪 #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/zs1GPXbX
1 minute ago by abhay01007

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

#iphoneography #iphoneography #bahrain #iphone4s #photography #sunset #scenery #beautiful #sun
http://t.co/Ee4V NimSL
2 minutes ago by thementaldawg

http://tweetsentiments.com/
OpView Service

連上OpView，品牌形象，輿論觀點監測好輕鬆

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報，並以雲端服務平台供客戶使用，OpView服務的資料涵蓋範圍包括台灣最具代表性的新聞網站。

http://www.eland.com.tw/solutions

http://opview-eland.blogspot.tw/2012/05/blog-post.html
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 components and is associated with a set of attributes of the components

• An opinion can be expressed on any node or attribute of the node

• Aspects(features)
  – represent both components and attribute

Entity and aspect

Canon S500
(picture_quality, size, appearance, ...)

Lens
(...)

battery
(battery_life, size, ...)

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

\[(e_j, a_{jk}, s_{ijkl}, h_i, t_l)\]

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.

• \((e_j, a_{jk})\) 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

**Cellular phone 1:**

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GENERAL</strong></td>
<td>125</td>
<td>7</td>
</tr>
<tr>
<td><strong>Voice quality</strong></td>
<td>120</td>
<td>8</td>
</tr>
<tr>
<td><strong>Battery</strong></td>
<td>80</td>
<td>12</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>Valence</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>Liu et al., 2005</td>
<td>Polarity</td>
<td>Lexicon/Rule</td>
<td>Phrase</td>
<td>Distribution</td>
<td>Object</td>
<td>Range</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>Yu 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 Huettnner 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>Lexicon/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 sadly, there is no soft cover edition.
Conversion of text representation

Word Vector (WV)

<table>
<thead>
<tr>
<th>Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>This</td>
</tr>
<tr>
<td>book</td>
</tr>
<tr>
<td>is</td>
</tr>
<tr>
<td>the</td>
</tr>
<tr>
<td>best</td>
</tr>
<tr>
<td>written</td>
</tr>
<tr>
<td>documentary</td>
</tr>
<tr>
<td>thus</td>
</tr>
<tr>
<td>far</td>
</tr>
<tr>
<td>,</td>
</tr>
<tr>
<td>yet</td>
</tr>
<tr>
<td>sadly</td>
</tr>
<tr>
<td>,</td>
</tr>
<tr>
<td>there</td>
</tr>
<tr>
<td>is</td>
</tr>
<tr>
<td>no</td>
</tr>
<tr>
<td>soft</td>
</tr>
<tr>
<td>cover</td>
</tr>
<tr>
<td>edition</td>
</tr>
</tbody>
</table>

Polarity Score Vector (PSV)

<table>
<thead>
<tr>
<th>pscore</th>
<th>nscore</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0.75</td>
<td>0</td>
<td>Positive (0.75)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0.375</td>
<td>0</td>
<td>Positive (0.375)</td>
</tr>
<tr>
<td>0.375</td>
<td>0</td>
<td>Positive (0.375)</td>
</tr>
<tr>
<td>0</td>
<td>0.125</td>
<td>Negative (0.125)</td>
</tr>
<tr>
<td>0.25</td>
<td>0.5</td>
<td>Negative (0.25)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0</td>
<td>0.75</td>
<td>Negative (0.75)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>Neutral (0)</td>
</tr>
</tbody>
</table>

Microstate Sequence (MS)

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Probability Distribution (P)

P(“1”) = 3/17
P(“-1”) = 3/17
P(“0”) = 11/17

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

• OpinionFinder’s Subjectivity Lexicon

• 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;;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&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</td>
<td>sadly#1 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</td>
<td>sadly#3 lamentably#1 deplorably#1 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

中文情感分析用詞語集

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>中文正面情感詞語</td>
<td>836</td>
</tr>
<tr>
<td>中文負面情感詞語</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>Total</td>
<td>9193</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語集

• “程度級別”詞語
  – 1. “極其|extreme / 最|most”
    • 非常，極，極度，無以倫比，最為
  – 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

1. Social Media (社群媒體)
2. Social Word-of-Mouth (社群口碑)
3. Web Mining (網路探勘)
References


Social Word-of-Mouth and Web Mining
(社群口碑與網路探勘)

Q & A

Min-Yuh Day
Assistant Professor
Dept. of Information Management, Tamkang University

http://mail.tku.edu.tw/myday/
2012-10-31