Social Word-of-Mouth and Web Content Analysis
(社群口碑與網路文本分析)

時間：2012/06/04(一)13:30-16:30
地點：財團法人商業發展研究院
台北市復興南路1段303號3樓(遠瑟大樓)301會議室

戴敏育
Min-Yuh Day
Assistant Professor
Dept. of Information Management, Tamkang University
http://mail.tku.edu.tw/myday/

2012-06-04

Outline

1. 社群媒體的特性
2. 社群媒體發展趨勢
3. 社群媒體商業應用
4. 社群口碑趨勢分析
5. 網路文本分析

#1 Activity on the Web?
Social Media

社群媒體的特性

Source: Social Media Business, http://www.youtube.com/watch?v=X9sTq3pzNQQ
Marketing 4P to 4C

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

Four Pillars of Social Media Strategy

- Communication
- Collaboration
- Education
- Entertainment

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

- Internal
- Innovation
- Trusted Network
- The World
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


Social Media Landscape

Social Media Services

Source: http://www.fredcavazza.net/2008/06/09/social-media-landscape/

Kaplan & Haenlein, 2010

Social media (social media)的定義

建立在Web 2.0概念與技術的基础上，
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
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.
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
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 33

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 34

EXHIBIT 7.4 The Top Nine Social Networking Sites

<table>
<thead>
<tr>
<th>Social Networking Site</th>
<th>Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>MySpace</td>
<td>252,145,404</td>
</tr>
<tr>
<td>Facebook</td>
<td>450,000,000</td>
</tr>
<tr>
<td>Windows Live Spaces</td>
<td>120,000,000</td>
</tr>
<tr>
<td>Friendster</td>
<td>90,000,000</td>
</tr>
<tr>
<td>hi5</td>
<td>80,000,000</td>
</tr>
<tr>
<td>Orkut</td>
<td>67,000,000</td>
</tr>
<tr>
<td>Classmates.com</td>
<td>50,000,000</td>
</tr>
<tr>
<td>Bebo</td>
<td>40,000,000</td>
</tr>
<tr>
<td>LinkedIn</td>
<td>70,000,000</td>
</tr>
</tbody>
</table>

Source: Turban et al. (2010), Introduction to Electronic Commerce 35

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 36
BUSINESS AND ENTERPRISE SOCIAL NETWORKS

- 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
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

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

SHOPPING IN SOCIAL NETWORKS

FEEDBACK FROM CUSTOMERS: CONVERSATIONAL MARKETING
  – Customer Feedback with Twitter

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
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

ENTERTAINMENT WEB 2.0 STYLE: FROM SOCIAL NETWORKS TO MARKETPLACES

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

社群媒體商業應用
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

Word-of-Mouth Marketing

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

Elements in the Communications Process

Source: Kotler and Keller (2008)
Field of Experience

Sender's field

Receiver's field

The Communications Process

Selective attention

Selective distortion

Selective retention

Social Media Marketing

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

Source: Kotler and Keller (2008)
Source: Kotler and Keller (2008)
Source: Kotler and Keller (2008)
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

Internal

Innovation

Trusted Network

The World

Examples of Social Media Selling Strategies in the Market Today

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

User Reviews

Social Media Wildfire

Creating Urgency/Spontaneous Selling

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

“Pass it along” promtions

Recruiting others/Group Sales

Customers as “Community Organizers”

Influencer-Led Development

Social Media Community

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

“Pro-sumers” collaboration

Engaging the Advocates

User Reviews

Word-of-Mouth
Social Media

Case Study: LenovoClub CareerLife 職場人生

http://www.google.com.tw/trends/?q=word+of+mouth,+social+media&ctab=0&geo=all&date=all&sort=0

http://www.lenovoclub.com.tw/careerlife/

http://www.youtube.com/watch?v=XRUv2hFiPig
ACM Categories and Subject Descriptors

- I.2.7 [Artificial Intelligence]
  - Natural Language Processing
    - Text analysis
- H.2.8 [Database Management]
  - Database Applications
    - Data mining
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

Web Mining and Social Networking

- Web Mining and Social Networking: Techniques and Applications

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

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

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

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. …”

(1) I bought an iPhone 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
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
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

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 components

• 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
  $(e_j, a_{ijk}, s_{ijkl}, h_i, t_i)$
  where
  – $e_j$ is a target entity.
  – $a_{ijk}$ 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.
  – $t_i$ is the time when the opinion is expressed.
Opinion definition

• An opinion is a quintuple
  \((e_j, a_{jk}, s_{ijkl}, h_i, t_i)\)
  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.
  – \(t_i\) 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

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

<table>
<thead>
<tr>
<th>Syntactic template</th>
<th>Example pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;subj&gt; passive-verb</td>
<td>&lt;subj&gt; was satisfied</td>
</tr>
<tr>
<td>&lt;subj&gt; active-verb</td>
<td>&lt;subj&gt; complained</td>
</tr>
<tr>
<td>active-verb &lt;dobj&gt;</td>
<td>endorsed &lt;dobj&gt;</td>
</tr>
<tr>
<td>noun aux &lt;dobj&gt;</td>
<td>fact is &lt;dobj&gt;</td>
</tr>
<tr>
<td>passive-verb prep &lt;np&gt;</td>
<td>was worried about &lt;np&gt;</td>
</tr>
</tbody>
</table>
A Brief Summary of Sentiment Analysis Methods

<table>
<thead>
<tr>
<th>Study</th>
<th>Analytic Task</th>
<th>Sentiment Identification</th>
<th>Sentiment Aggregation</th>
<th>Nature of Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li and Li, 2011</td>
<td>Tactic</td>
<td>ML (Probabilistic model)</td>
<td>Support</td>
<td>Value</td>
</tr>
<tr>
<td>Li and Wu, 2010</td>
<td>Tactic</td>
<td>Levenshtein</td>
<td>Phrase</td>
<td>Support</td>
</tr>
<tr>
<td>Theodorescu et al., 2010</td>
<td>Tactic</td>
<td>Levenshtein</td>
<td>Phrase</td>
<td>Max &amp; Min</td>
</tr>
<tr>
<td>Dong and Miao, 2009</td>
<td>Tactic</td>
<td>Both</td>
<td>Levenshtein</td>
<td>Average</td>
</tr>
<tr>
<td>Zhang, 2009</td>
<td>Tactic</td>
<td>Both</td>
<td>ML (W), AdaBoost, Rule, etc.</td>
<td>Sentence</td>
</tr>
<tr>
<td>Zhang et al., 2009</td>
<td>Tactic</td>
<td>Levenshtein</td>
<td>Phrase</td>
<td>Support</td>
</tr>
<tr>
<td>Alhusen, Chen, and Salemi, 2003</td>
<td>Tactic</td>
<td>ML (GA + feature selection)</td>
<td>Support</td>
<td>Sentence</td>
</tr>
<tr>
<td>Tahmooressi and Zhang, 2008</td>
<td>Tactic</td>
<td>ML (HMM, Winner, NB, etc.)</td>
<td>Support</td>
<td>Sentence</td>
</tr>
<tr>
<td>Aylward, D. and Padamsee, 2007</td>
<td>Tactic</td>
<td>ML (Markov, Blakel)</td>
<td>Support</td>
<td>Average</td>
</tr>
<tr>
<td>Das and Das, 2007</td>
<td>Tactic</td>
<td>ML (Bayesian, Discriminate, etc.)</td>
<td>Support</td>
<td>Average</td>
</tr>
<tr>
<td>Lin et al., 2007</td>
<td>Tactic</td>
<td>ML (PLSA)</td>
<td>Support</td>
<td>Sentence</td>
</tr>
<tr>
<td>Kennedy and Iakten, 2006</td>
<td>Tactic</td>
<td>Levenshtein, ML (WVM)</td>
<td>Support</td>
<td>Support</td>
</tr>
<tr>
<td>Milana, 2006</td>
<td>Tactic</td>
<td>Levenshtein</td>
<td>Phrase</td>
<td>Support</td>
</tr>
<tr>
<td>Lin et al., 2005</td>
<td>Tactic</td>
<td>Levenshtein</td>
<td>Phrase</td>
<td>Support</td>
</tr>
<tr>
<td>Miczke, 2003</td>
<td>Tactic</td>
<td>ML (WVM)</td>
<td>Support</td>
<td>Sentence</td>
</tr>
<tr>
<td>Oussama and Enes, 1999</td>
<td>Tactic</td>
<td>Levenshtein, Rule</td>
<td>Support</td>
<td>Sentence</td>
</tr>
<tr>
<td>Richmond, 2004</td>
<td>Tactic</td>
<td>PL (VNN, ND)</td>
<td>Support</td>
<td>Sentence</td>
</tr>
<tr>
<td>Wilson, Worbe, and Pazsa, 2004</td>
<td>Tactic</td>
<td>ML (WVM, AdaBoost, Rule, etc.)</td>
<td>Support</td>
<td>Sentence</td>
</tr>
<tr>
<td>Nagam and Hunt, 2004</td>
<td>Tactic</td>
<td>Levenshtein, Rule</td>
<td>Support</td>
<td>Sentence</td>
</tr>
<tr>
<td>Dav, Lawrence, and Penning, 2003</td>
<td>Tactic</td>
<td>ML (WVM, Random, etc.)</td>
<td>Support</td>
<td>Sentence</td>
</tr>
<tr>
<td>Nastakhit and V, 2003</td>
<td>Tactic</td>
<td>Levenshtein</td>
<td>Phrase</td>
<td>Support</td>
</tr>
<tr>
<td>Yu and Hattori, 2002</td>
<td>Tactic</td>
<td>ML (C200) + Levenshtein</td>
<td>Support</td>
<td>Average</td>
</tr>
<tr>
<td>Peng, Lee, and Vickers, 2002</td>
<td>Tactic</td>
<td>ML (WVM, MacToR, NB)</td>
<td>Support</td>
<td>Sentence</td>
</tr>
<tr>
<td>Sydow and Horster, 2001</td>
<td>Tactic</td>
<td>Levenshtein, Random, Log</td>
<td>Support</td>
<td>Average</td>
</tr>
<tr>
<td>Herter, 2001</td>
<td>Tactic</td>
<td>Both</td>
<td>Levenshtein</td>
<td>Sentence</td>
</tr>
</tbody>
</table>

(Both = Subjectivity and Polarity, ML = Machine Learning, Levenshtein = Levenshtein enhanced by linguistic rules)

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>
<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>JJ</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>yet</td>
<td>RB</td>
</tr>
<tr>
<td>sadly</td>
<td>RB</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>
</tbody>
</table>

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.”

Datasets of Opinion Mining

- Blog06
  - 25GB TREC test collection
    - http://ir.dcs.gla.ac.uk/test collections/access to data.html
- Cornell movie-review datasets
  - http://www.cs.cornell.edu/people/pabo/movie-review-data/
- Customer review datasets
- Multiple-aspect restaurant reviews
  - 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/
- General Inquirer
  - http://www.wjh.harvard.edu/~inquirer/
- OpinionFinder’s Subjectivity Lexicon
  - http://www.cs.pitt.edu/mpqa/
- NTU Sentiment Dictionary (NTUSD)
  - http://nlg18.csie.ntu.edu.tw:8080/opinion/
- 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 <code>good</code>) 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>0.625</td>
<td>0</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>0.875</td>
<td>0</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>0.25</td>
<td>0</td>
<td>sadly#2</td>
<td>with sadness; in a sad manner; &quot;She died last night,&quot; he said sadly</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

| 中文正面情感詞語 | 836 |
| 中文負面情感詞語 | 1254 |
| 中文正面評價詞語 | 3730 |
| 中文負面評價詞語 | 3116 |
| 中文程度級別詞語 | 219 |
| 中文主張詞語 | 38 |
| Total | 9193 |


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. “极其 | 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

1. 社群媒體的特性
2. 社群媒體發展趨勢
3. 社群媒體商業應用
4. 社群口碑趨勢分析
5. 網路文本分析

References


Social Word-of-Mouth and Web Content Analysis
(社群口碑與網路文本分析)

Q & A

戴敏育
Min-Yuh Day
Assistant Professor
Dept. of Information Management, Tamkang University
http://mail.tku.edu.tw/myday/