Social Media and Sentiment Analysis
(社群媒體與情緒分析)

時間：2016/11/01 (二) (2:10-5:00pm)
地點：政治大學綜合院館270407，北棟407教室
主持人：陳恭主任

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/
2016-11-01
Sentiment Analysis on Social Media
(社群媒體情感分析)

Min-Yuh Day
助教
Dept. of Information Management, Tamkang University
http://mail.tku.edu.tw/myday/
2016-07
Outline

• Architectures of Sentiment Analytics on Social Media
• Social Media Monitoring/Analysis
• Sentiment Analytics on Social Media: Tools and Applications
Sentiment Analysis on Social Media
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 ____ phone.
(3) The ____ screen was really ____.
(4) The ____ quality was ____.
(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 ____ , and wanted me to return it to the shop. …”

Architectures of Sentiment Analytics
Bing Liu (2015),
Sentiment Analysis:
Mining Opinions, Sentiments, and Emotions,
Cambridge University Press

http://www.amazon.com/Sentiment-Analysis-Opinions-Sentiments-Emotions/dp/1107017890
Sentiment Analysis and Opinion Mining

- 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.

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

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.

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**”

# 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.

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.

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

Sentiment Analysis Architecture

Positive tweets  Negative tweets  Word features

Training set

Features extractor

Features extractor

Tweet

Positive

Negative

Sentiment Classification Based on Emoticons

Lexicon-Based Model

Preassembled Word Lists

Merged Lexicon

Tokenized Document Collection

Sentiment Scoring and Classification: Polarity

Sentiment Polarity

Generic Word Lists

Sentiment Analysis Tasks

Opinionated Document → Subjectivity Classification → Sentiment Classification

Subjectivity Classification:
- Opinion holder extraction
- Object/Feature extraction

## 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>
Levels of Sentiment Analysis

- Word level Sentiment Analysis
- Sentence level Sentiment Analysis
- Document level Sentiment Analysis
- Feature level Sentiment Analysis

Sentiment Analysis

Tasks

- Subjectivity Classification
- Polarity Determination
- Vagueness resolution in opinionated text
- Cross-domain SC
- Opinion Spam Detection
- Multi- & Cross-Lingual SC
- Review Usefulness Measurement
- Lexicon Creation
- Cross-domain SC
- Opinion Spam Detection
- Aspect Extraction
- Application

Approaches

- Machine Learning based
- Lexicon based
- Hybrid approaches
- Ontology based
- Non-Ontology based

Sentiment Classification Techniques

- Sentiment Analysis
  - Machine Learning Approach
    - Supervised Learning
      - Decision Tree Classifiers
      - Linear Classifiers
      - Rule-based Classifiers
      - Probabilistic Classifiers
    - Unsupervised Learning
      - Support Vector Machine (SVM)
      - Neural Network (NN)
      - Deep Learning (DL)
      - Naïve Bayes (NB)
      - Bayesian Network (BN)
      - Maximum Entropy (ME)
  - Lexicon-based Approach
    - Dictionary-based Approach
  - Corpus-based Approach
    - Statistical
    - Semantic

## 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 (SVM, 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 Huettner 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

<table>
<thead>
<tr>
<th>Word Vector (WV)</th>
<th>Polarity Score Vector (PSV)</th>
<th>Microstate Sequence (MS)</th>
<th>Probability Distribution (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>This</td>
<td>0 0 Neutral (0)</td>
<td>0</td>
<td>P(&quot;1&quot;)=3/17</td>
</tr>
<tr>
<td>book</td>
<td>0 0 Neutral (0)</td>
<td>0</td>
<td>P(&quot;-1&quot;)=3/17</td>
</tr>
<tr>
<td>is</td>
<td>0 0 Neutral (0)</td>
<td>0</td>
<td>P(&quot;0&quot;)=11/17</td>
</tr>
<tr>
<td>the</td>
<td>0 0 Neutral (0)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>best</td>
<td>0.75 0 Positive (0.75)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>written</td>
<td>0 0 Neutral (0)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>documentary</td>
<td>0.375 0 Positive (0.375)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>thus</td>
<td>0.375 0 Positive (0.375)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>far</td>
<td>0 0.125 Negative (0.125)</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>yet</td>
<td>0.25 0.5 Negative (0.25)</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>sadly</td>
<td>0 0 Neutral (0)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>there</td>
<td>0 0 Neutral (0)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>is</td>
<td>0 0 Neutral (0)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>no</td>
<td>0 0.75 Negative (0.75)</td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>soft</td>
<td>0 0 Neutral (0)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>cover</td>
<td>0 0 Neutral (0)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>edition</td>
<td>0 0 Neutral (0)</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

## 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“</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“</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“</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;it was woefully inadequate“</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>
SenticNet

The car is very old but it is rather not expensive.

The car is very old but it is rather not expensive.

The car is very old but it is rather not expensive.

The car is very old but it is rather not expensive.

The car is **very** old but it is **rather** not expensive.

Polarity Detection with SenticNet

Polarity Detection with SenticNet

Polarity Detection with SenticNet

Polarity Detection with SenticNet

Evaluation of Text Mining and Sentiment Analysis

• Evaluation of Information Retrieval
• Evaluation of Classification Model (Prediction)
  – Accuracy
  – Precision
  – Recall
  – F-score
Deep Learning for Sentiment Analytics
Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank

Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng and Christopher Potts
Stanford University, Stanford, CA 94305, USA
richard@socher.org, {aperelyg, jcchuang, ang}@cs.stanford.edu
{jeaneis, manning, cgpotts}@stanford.edu

Abstract

Semantic word spaces have been very useful but cannot express the meaning of longer phrases in a principled way. Further progress towards understanding compositionality in tasks such as sentiment detection requires richer supervised training and evaluation resources and more powerful models of composition. To remedy this, we introduce a Sentiment Treebank. It includes fine grained sentiment labels for 215,154 phrases in the parse trees of 11,855 sentences and presents new challenges for sentiment compositionality. To address them, we introduce the Recursive Neural Tensor Network. When trained on the new treebank, this model out-

Figure 1: Example of the Recursive Neural Tensor Network accurately predicting 5 sentiment classes, very negative to very positive (−−, −, 0, +, ++), at every node of a parse tree and capturing the negation and its scope in this sentence.

Recursive Neural Tensor Network (RNTN)

Recursive Neural Network (RNN) models for sentiment

Recursive Neural Tensor Network (RNTN)

Roger Dodger is one of the **most** compelling variations on this theme.

Roger Dodger is one of the **least** compelling variations on this theme.

RNTN for Sentiment Analysis

Roger Dodger is one of the most compelling variations on this theme.

Roger Dodger is one of the least compelling variations on this theme.

### Accuracy for fine grained (5-class) and binary predictions at the sentence level (root) and for all nodes

<table>
<thead>
<tr>
<th>Model</th>
<th>Fine-grained</th>
<th></th>
<th>Positive/Negative</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Root</td>
<td>All</td>
<td>Root</td>
</tr>
<tr>
<td>NB</td>
<td>67.2</td>
<td>41.0</td>
<td>82.6</td>
<td>81.8</td>
</tr>
<tr>
<td>SVM</td>
<td>64.3</td>
<td>40.7</td>
<td>84.6</td>
<td>79.4</td>
</tr>
<tr>
<td>BiNB</td>
<td>71.0</td>
<td>41.9</td>
<td>82.7</td>
<td>83.1</td>
</tr>
<tr>
<td>VecAvg</td>
<td>73.3</td>
<td>32.7</td>
<td>85.1</td>
<td>80.1</td>
</tr>
<tr>
<td>RNN</td>
<td>79.0</td>
<td>43.2</td>
<td>86.1</td>
<td>82.4</td>
</tr>
<tr>
<td>MV-RNN</td>
<td>78.7</td>
<td>44.4</td>
<td>86.8</td>
<td>82.9</td>
</tr>
<tr>
<td>RNTN</td>
<td>80.7</td>
<td>45.7</td>
<td>87.6</td>
<td>85.4</td>
</tr>
</tbody>
</table>

# Accuracy of negation detection

<table>
<thead>
<tr>
<th>Model</th>
<th>Negated Positive</th>
<th>Negated Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>biNB</td>
<td>19.0</td>
<td>27.3</td>
</tr>
<tr>
<td>RNN</td>
<td>33.3</td>
<td>45.5</td>
</tr>
<tr>
<td>MV-RNN</td>
<td>52.4</td>
<td>54.6</td>
</tr>
<tr>
<td>RNTN</td>
<td>71.4</td>
<td>81.8</td>
</tr>
</tbody>
</table>

# Deep Learning for Sentiment Analysis

**CNN** **RNTN** **LSTM**

<table>
<thead>
<tr>
<th>Model</th>
<th>Fine (5-class)</th>
<th>Binary</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCNN (Blunsom, et al. 2014)</td>
<td>0.485</td>
<td>0.868</td>
</tr>
<tr>
<td>RNTN (Socher, et al. 2013)</td>
<td>0.457</td>
<td>0.854</td>
</tr>
<tr>
<td>CNN-non-static (Kim, 2014)</td>
<td>0.480</td>
<td>0.872</td>
</tr>
<tr>
<td>CNN-multi-channel (Kim, 2014)</td>
<td>0.474</td>
<td>0.881</td>
</tr>
<tr>
<td>DRNN w. pretrained word-embeddings (Irsoy and Cardie, 2014)</td>
<td>0.498</td>
<td>0.866</td>
</tr>
<tr>
<td>Paragraph Vector (Le and Mikolov. 2014)</td>
<td>0.487</td>
<td>0.878</td>
</tr>
<tr>
<td>Dependency Tree-LSTM (Tai, et al, 2015)</td>
<td>0.484</td>
<td>0.857</td>
</tr>
<tr>
<td>Constituency Tree-LSTM (Tai, et al, 2015)</td>
<td>0.439</td>
<td>0.820</td>
</tr>
<tr>
<td>Constituency Tree-LSTM (Glove vectors) (Tai, et al, 2015)</td>
<td>0.510</td>
<td>0.880</td>
</tr>
<tr>
<td>Paragraph Vector</td>
<td>0.391</td>
<td>0.798</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.456</td>
<td>0.843</td>
</tr>
<tr>
<td>Deep Recursive-NN</td>
<td>0.469</td>
<td>0.847</td>
</tr>
</tbody>
</table>

Source: https://cs224d.stanford.edu/reports/HongJames.pdf
# Performance Comparison of Sentiment Analysis Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Data Set</th>
<th>Acc.</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine Learning</td>
<td>SVM</td>
<td>86.40%</td>
<td>Pang, Lee[23]</td>
</tr>
<tr>
<td></td>
<td>CoTraining SVM</td>
<td>82.52%</td>
<td>Liu[14]</td>
</tr>
<tr>
<td></td>
<td>Deep learning</td>
<td>80.70%</td>
<td>Richard[18]</td>
</tr>
<tr>
<td>Lexical based</td>
<td>Corpus</td>
<td>74.00%</td>
<td>Turkey</td>
</tr>
<tr>
<td></td>
<td>Dictionary</td>
<td>---</td>
<td>Taboada[20]</td>
</tr>
<tr>
<td>Cross-lingual</td>
<td>Ensemble</td>
<td>81.00%</td>
<td>Wan,X[16]</td>
</tr>
<tr>
<td></td>
<td>Co-Train</td>
<td>81.30%</td>
<td>Wan,X[16]</td>
</tr>
<tr>
<td></td>
<td>EWGA</td>
<td>&gt;90%</td>
<td>Abbasi,A.</td>
</tr>
<tr>
<td></td>
<td>CLMM</td>
<td>83.02%</td>
<td>Mengi</td>
</tr>
<tr>
<td>Cross-domain</td>
<td>Active Learning</td>
<td>80% (avg)</td>
<td>Li, S</td>
</tr>
<tr>
<td></td>
<td>Thesaurus</td>
<td></td>
<td>Bollegala[22]</td>
</tr>
<tr>
<td></td>
<td>SFA</td>
<td></td>
<td>Pan S J[15]</td>
</tr>
</tbody>
</table>

Social Media Monitoring/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
  – 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#!
SAS® Social Media Analytics
Integrate, archive, analyze and act on online conversations

Overview  Benefits  Features  Demos & Screenshots  System Requirements

SAS Social Media Analytics is an enterprise-hosted, on-demand solution that integrates, archives, analyzes and enables organizations to act on intelligence gleaned from online conversations on professional and consumer-generated media sites. It enables you to attribute online conversations to specific parts of your business, allowing accelerated responses to marketplace shifts.

Based on your unique business challenges and enterprise goals, SAS can provide a tailored implementation that’s hosted and managed by SAS Solutions OnDemand.

Benefits
- Analyze conversation data.
- Identify advocates of and threats to, corporate reputation and brand.
- Quantify interaction among traditional media/campaigns and social media activity.
- Establish a platform for social CRM strategy.

“

The great thing about SAS is that it’s so powerful and has such a broad offering.

—Jonathan Pranther
Manager of Statistics

Organic

Read full story

Product Demo

Questions?
- Phone
- Contact Form

White Paper
Text Analytics for Social Media: Evolving Tools for an Evolving Environment!
Download Now

SAS® Social Media Analytics
» Overview

RESOURCES
- Fact Sheet (PDF)
- Solution Brief (PDF)
- White Papers
- Webinars

http://www.tweetfeel.com

iPhone 4s

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

40 41 = 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 @jigglinjello: 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
http://www.eland.com.tw/
OpView

http://www.opview.com.tw/
母親節好禮大比拼 聽聽網友怎麼說

這個周末就是母親節了，大家有想好要如何慶祝了嗎？吃大餐、送好禮已成了節慶的基本盤，再加上百貨針對母親節紛紛推出特賣優惠，不僅讓孝子孝女省下荷包，也讓平常有在觀望當期活動的網友們紅了眼，更增添了其口碑豐富性...
Resources of Opinion Mining
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“</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“</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“</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“</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><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，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

Opinion Spam Detection
Opinion Spam Detection

- Opinion Spam Detection: Detecting Fake Reviews and Reviewers
  - Spam Review
  - Fake Review
  - Bogus Review
  - Deceptive review
  - Opinion Spammer
  - Review Spammer
  - Fake Reviewer
  - Shill (Stooge or Plant)

Source: http://www.cs.uic.edu/~liub/FBS/fake-reviews.html
Opinion Spamming

- Opinion Spamming
  - "illegal" activities
    - e.g., writing fake reviews, also called shilling
  - try to mislead readers or automated opinion mining and sentiment analysis systems by giving undeserving positive opinions to some target entities in order to promote the entities and/or by giving false negative opinions to some other entities in order to damage their reputations.

Source: [http://www.cs.uic.edu/~liub/FBS/fake-reviews.html](http://www.cs.uic.edu/~liub/FBS/fake-reviews.html)
Forms of Opinion spam

- fake reviews (also called bogus reviews)
- fake comments
- fake blogs
- fake social network postings
- deceptions
- deceptive messages

Source: [http://www.cs.uic.edu/~liub/FBS/fake-reviews.html](http://www.cs.uic.edu/~liub/FBS/fake-reviews.html)
Fake Review Detection

• Methods
  – supervised learning
  – pattern discovery
  – graph-based methods
  – relational modeling

• Signals
  – Review content
  – Reviewer abnormal behaviors
  – Product related features
  – Relationships

Source: http://www.cs.uic.edu/~liub/FBS/fake-reviews.html
Professional Fake Review Writing Services
(some Reputation Management companies)

• Post positive reviews
• Sponsored reviews
• Pay per post
• Need someone to write positive reviews about our company (budget: $250-$750 USD)
• Fake review writer
• Product review writer for hire
• Hire a content writer
• Fake Amazon book reviews (hiring book reviewers)
• People are just having fun (not serious)

Source: http://www.cs.uic.edu/~liub/FBS/fake-reviews.html
SponsoredReviews connects bloggers with SEO's, Marketers, and Advertisers looking to build Links, Traffic and Buzz.

**Direct Traffic.**
Millions of people read blogs every day. Paying for posts puts the spotlight on your company and will generate tons of targeted traffic.

**Buzz & Branding.**
The more bloggers talk about your site the better. Many blogs syndicate stories they see on other sites. A couple well timed sponsored posts has the potential to generate a flurry of other post being written.

**Search Engine Rankings.**
Every post has links back to your site. Getting links from quality blogs will increase your link popularity and will help your site rank better in the search engines.

**Valuable Feedback.**
Getting Reviewed by bloggers will provide you with valuable feedback that you can use to better understand your audience and customers.

---

### Advertisers
Start Here.

- Announce your products, services, websites, and ideas to the world!
- Tap into the power of the blogosphere to build traffic, links and valuable feedback.

[Free Sign Up] [Read More]

### Bloggers
Earn Cash.

- Earn cash by writing honest posts about our advertiser’s products and services.
- Write posts in your own tone and style, and gear them to your audience's interest.

[Free Sign Up] [Read More]

advertisers

Hire bloggers to blog about your company, service or website. PayPerPost gives you access to a diverse pool of bloggers from all over the world. Make offers, negotiate deals and approve posts.

signup now

bloggers

Make money blogging! PayPerPost lets you pick your advertisers, name your own price and negotiate your own deals. You can get paid to blog on virtually any subject. Sign up below!

signup now

customer testimonial

"PayPerPost has been instrumental in helping our company streamline our various product awareness campaigns."

-C. Litchfield
Need someone to write and post positive reviews

Project Description:
We need someone to write and post positive reviews about our company on websites. Please send an example of a review you would post for any company. We can also send examples of comments our customers have sent us to use and refer too as well.

This is a long term project, so if it works out there will be a healthy amount of work. Please reply back with all your experience and how much you would charge per post.

thank you.

Skills required:
Publicación en foros, Opiniones

Project posted by:
dvel
5.0 (1 Review)
Follow

Your ad could
From $100/week
Papers on Opinion Spam Detection


2. Guan Wang, Sihong Xie, Bing Liu, Philip S. Yu. Identify Online Store Review Spammers via Social Review Graph. ACM Transactions on Intelligent Systems and Technology, accepted for publication, 2011.


Source: [http://www.cs.uic.edu/~liub/FBS/fake-reviews.html](http://www.cs.uic.edu/~liub/FBS/fake-reviews.html)
Summary

- Architectures of Sentiment Analytics on Social Media
- Social Media Monitoring/Analysis
- Sentiment Analytics on Social Media: Tools and Applications
References

  [http://www.cs.uic.edu/~liub/WebMiningBook.html](http://www.cs.uic.edu/~liub/WebMiningBook.html)
- Bing Liu (2013), Opinion Spam Detection: Detecting Fake Reviews and Reviewers, 
  [http://www.cs.uic.edu/~liub/FBS/fake-reviews.html](http://www.cs.uic.edu/~liub/FBS/fake-reviews.html)
- 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)
References

• Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In the 26th International Conference on Computational Linguistics (COLING), Osaka. 2016.


• Steven Struhl (2015), Practical Text Analytics: Interpreting Text and Unstructured Data for Business Intelligence (Marketing Science), Kogan Page

• Bing Liu (2015), Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Cambridge University Press