Outline

- Sentiment Analysis
  - Motivation
  - Task
  - Machine Learning Approach
  - Rule-based Approach
Outline

• Sentiment Analysis
  - Motivation
  - Task
  - Machine Learning Approach
  - Rule-based Approach
Product reviews

Customer Reviews
Speech and Language Processing, 2nd Edition

15 Reviews

<table>
<thead>
<tr>
<th>Rating</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 star</td>
<td>8</td>
</tr>
<tr>
<td>4 star</td>
<td>3</td>
</tr>
<tr>
<td>3 star</td>
<td>3</td>
</tr>
<tr>
<td>2 star</td>
<td>0</td>
</tr>
<tr>
<td>1 star</td>
<td>1</td>
</tr>
</tbody>
</table>

Average Customer Review ★★★★★ (15 customer reviews)

Share your thoughts with other customers

Create your own review

The most helpful favorable review

4 of 4 people found the following review helpful

★★★★★★ Great introductions and reference book
I read the first edition of that book and it is terrific. The second edition is much more adapted to current research. Statistical methods in NLP are more detailed and some syntax-based approaches are presented. My specific interest is in machine translation and dialogue systems. Both chapters are extensively rewritten and much more elaborated. I believe this book is...

Read the full review>
Published on August 9, 2008 by carheg

See more 5 star, 4 star reviews

The most helpful critical review

37 of 37 people found the following review helpful

★★★★☆ Good description of the problems in the field, but look elsewhere for practical solutions
The authors have the challenge of covering a vast area, and they do a good job of highlighting the hard problems within individual sub-fields, such as machine translation. The availability of an accompanying Web site is a strong plus, as is the extensive bibliography, which also includes links to freely available software and resources.

Now for the...

Read the full review>
Published on April 2, 2009 by P. Nadkarni

See more 3 star, 2 star, 1 star reviews
Social Media

Streamcrab
Realtime, Twitter sentiment analysis engine

Contact: Email  Github  LinkedIn  Xing

Tweet count in %
Negative (%)
Positive (%)

Polarity sums
Negative
Positive

(http://www.streamcrab.com/)
## Event Analysis and Prediction

### Impact Feed

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Sentiment</th>
<th>Stock Name</th>
<th>Summary</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>+0.94</td>
<td>Positive</td>
<td>Prologis Inc</td>
<td>Prologis Raised to &quot;Buy&quot; at Vetr Inc.</td>
<td>mideasttime.com</td>
</tr>
<tr>
<td>+0.90</td>
<td>Positive</td>
<td>Cabot Oil &amp; Gas Corporation</td>
<td>Cabot Oil &amp; Gas Co. Upgraded to &quot;Overweight&quot; at Simmons</td>
<td>lulegacy.com</td>
</tr>
<tr>
<td>+0.90</td>
<td>Positive</td>
<td>Actavis Inc</td>
<td>Barclays Raises Actavis plc Price Target to $300.00</td>
<td>wkrb13.com</td>
</tr>
<tr>
<td>+0.90</td>
<td>Positive</td>
<td>Agilent Technologies Inc</td>
<td>Agilent Technologies Raised to Buy at Vetr Inc.</td>
<td>lulegacy.com</td>
</tr>
<tr>
<td>+0.90</td>
<td>Positive</td>
<td>Ameriprise Financial Inc</td>
<td>Traders Sell Shares of Ameriprise Financial on Strength Following Insider ...</td>
<td>lulegacy.com</td>
</tr>
<tr>
<td>+0.87</td>
<td>Positive</td>
<td>Texas Instruments Incorporated</td>
<td>Texas Instruments Incorporated Declares $0.34 Quarterly Dividend</td>
<td>wkrb13.com</td>
</tr>
</tbody>
</table>

(http://www.thestocksonar.com/Sentiment-Analysis)
Event Analysis and Prediction

Negative 2011-2013: 64% out of 9644 docs

Food term: chicken RELATED depression

Get Sentiment Statistics
Outline

- Sentiment Analysis
  - Motivation
  - Task
  - Machine Learning Approach
  - Rule-based Approach
Other names

- Opinion mining
- Opinion extraction
- Sentiment mining
- Subjectivity detection
- Subjectivity analysis
Sentiment Analysis Levels

Text

Opinion

Fact

+ 

- 

angry, afraid, ...

happy, surprised, ...
Advanced Sentiment Analysis

• Opinion holder and Opinion target/aspect
  - Students [OP HOLDER] like Wikipedia [TARGET] because it is easy to use and it sounds authoritative.
  - I had a nice stay in this hotel and the rooms [ASPECT] were very clean.
Advanced Sentiment Analysis

- Mixed opinions
  - The restaurant has an amazing view but it is very dirty.
Sentiment Analysis Approaches

- Machine learning methods
  - classification

- Rule-based methods
  - dictionary-oriented
Outline

• Sentiment Analysis
  - Motivation
  - Task
  - Machine Learning Approach
  - Rule-based Approach
Machine Learning Approach

Training

\[ T_1 \rightarrow C_1 \]
\[ T_2 \rightarrow C_2 \]
\[ \ldots \]
\[ T_n \rightarrow C_n \]

Testing

\[ T_{n+1} \rightarrow ? \]

\[ F_1 \]
\[ F_2 \]
\[ \ldots \]
\[ F_n \]

Model(F,C)

\[ F_{n+1} \]

\[ C_{n+1} \]
Sentiment Classification

• Using any kinds of supervised classifiers
  - K Nearest Neighbor
  - Support Vector Machines
  - Naïve Bayes
  - Maximum Entropy
  - Logistic Regression
  - ...

Features

• All words or just adjectives?
  - All words works better than adjectives only
Features

- Word occurrence or frequency?
  - Word occurrence is more useful than frequency
    - Using binary value for words
    - Replace all word counts higher than 0 in each text by 1
Features

- Negation
  - Negation words change the text polarity
    - Adding prefix NOT- to every word between negation and next punctuation

- „I did not like the restaurant location, but the food ...“

- I did not NOT-like NOT-the NOT-restaurant NOT-location, but the food ...
Features

- Other emotions
  - Considering emoticons as additional features
    - :)  
    - :(  
  - As well as smilies
    - 😊  
    - 😝  
    - 😞
Fine-grained analysis

- Dealing with finer classes of sentiment
  -3,-2,-1,+1,+2,+3

(from SAP HANA database)
Fine-grained Analysis

- Approaches
  - Using multiclass classifier (6 classes in this case)
  - Using two level classifier
    - First level: polarity classifier (positive or negative)
    - Second level: strength classifier (1 or 2 or 3)
Outline

- Sentiment Analysis
  - Motivation
  - Task
  - Machine Learning Approach
  - Rule-based Approach
Rule-based Approach

Training

\[ T_1 \rightarrow C_1 \]

\[ T_2 \rightarrow C_2 \]

\[ \cdots \]

\[ T_n \rightarrow C_n \]

Testing

\[ T_{n+1} \rightarrow ? \]

\[ C_{n+1} \]

bad

hate

lie

ugly

poor

\[ \cdots \]

good

love

brave

intelligent

nice

\[ \cdots \]
Rule-based Approach

• Looking for opinionated words in each text

• Classifying the text based on the number of positive and negative words
Rule-based Approach

- Considering different rules for classification
  - Fine-grained dictionary
  - Negation words
  - Booster words
  - Idioms
  - Emoticons
  - Mixed opinions
  - Linguistic features of the language
Rule-based Approach

- Fine-grained Dictionary
  - “It was a good song.”
  - “The song was excellent.”
Rule-based Approach

• Negation Words

  - „It was a good song.“ ✅

  - „The song was not good.“ 🚫
Rule-based Approach

- Booster Words

  - „The song was interesting.“ ✅✅

  - „The song was very interesting.“ ✅✅✅

  - „The song was somewhat interesting.“ ✅
Rule-based Approach

- Idioms
  - „shock horror“ 😞😞😞
Rule-based Approach

- Mixed Opinions

„The song was good, but I think its title was strange.“ 😊👎👎
Opinion Dictionary

- English
  - Subjectivity Clues (2005)
  - SentiSpin (2005)
  - SentiWordNet (2006)
  - Polarity Enhancement (2009)
  - SentiStrength (2010)
Opinion Dictionary

- German
  - GermanPolarityClues (2010)
  - SentiWortSchatz (2010)
  - GermanSentiStrength (2012)
Machine Learning with Opinion Dictionary

- Using opinion words as a feature in the algorithms
- Ignoring other words in the text