Opinion Summarization

By:
Hyun Duk Kim
Kavita Ganesan
Parikshit Sondhi
ChengXiang Zhai
Opinion Mining - The Area of Study

Opinion Mining

- Sentiment Classification
- Comparative Mining
- Opinion Summarization
- Opinion Trustworthiness

Opinion Retrieval

Opinion Question Answering

Opinion Matching

Information Retrieval

Use one or combination
Opinion Mining - The Area of Study

Information Retrieval

Opinion Retrieval
Opinion Question Answering
Opinion Matching

Widely studied – since 2002

Opinion Mining

Sentiment Classification
Comparative Mining
Opinion Summarization
Opinion Trustworthiness

Sentence Level
Document Level
Feature Level

Use one or combination
Opinion Mining - The Area of Study

Opinion Mining

- Sentiment Classification
- Comparative Mining
- Opinion Summarization
- Opinion Trustworthiness

Use one or combination

Information Retrieval

Opinion Retrieval
Opinion Question Answering
Opinion Matching

NEW – Limited [Jindal & Liu 06]
Opinion Mining - The Area of Study

Opinion Mining

- Opinion Retrieval
- Opinion Question Answering
- Opinion Matching

Information Retrieval

Use one or combination

Opinion Mining

- Sentiment Classification
- Opinion Summarization
- Opinion Trustworthiness

NEW – Growing
[Lu & Zhai 2008, Lu et. al 2009, Mei et. al 2007.....]
Opinion Mining - The Area of Study

Information Retrieval

Opinion Retrieval
Opinion Question Answering
Opinion Matching

use one or combination

NEW – Growing
[Jindal & Liu 08, JinJing et. al 07, Kim et. al 06... ]

Opinion Mining

Opinion Classification
Opinion Summarization
Opinion Trustworthiness

Sentence Level
Document Level
Feature Level
Opinion Mining - The Area of Study

- Sentiment Classification
- Comparative Mining
- Opinion Summarization
- Opinion Trustworthiness

Information Retrieval

Opinion Retrieval
Opinion Question Answering
Opinion Matching

use one or combination

Sentence Level
Document Level
Feature Level
What is Opinion Summarization?

- **Aggregate and represent** sentiment information drawn from [Pang & Lee 08]:
  - individual document or
  - from a collection of documents

---

**Beyond All Expectations!**

40 of 46 people found the following review helpful:

🌟🌟🌟 Would be perfect, except...

102 of 104 people found the following review helpful:

🌟🌟🌟 iPod Touch replaced my PDA, February 8, 2008

By Richard Jacobson (New Berlin, WI USA) - [See all my reviews](#)

For those who may be interested in acquiring an iPod Touch as a replacement for a PDA, I have found it to be quite useful. The primary function used in regard is syncing with Microsoft Outlook, which works flawlessly. Now I can check appointments and find those phone numbers for early morning conference calls without waiting 10 minutes for my laptop to boot up.
Type 1: Overall score/star rating

From the Commercial World

- Amazon.com - manual

Apple iPod touch 8 GB (1st Generation) OLD MODEL

Price: $209.99

In Stock.
Ships from and sold by Electronica Direct. Gift-wrap available.

14 new from $179.99  44 used from $116.99  10 refurbished from $134.95

There is a newer model of this item:

Apple iPod touch 8 GB (3rd Generation) NEWEST MODEL  ★★★★★ (258 reviews)

$188.95
In Stock.

See a problem? Let us know
Type 2a: Aspect Based Star Rating

From the Commercial World

- Amazon.com - manual

Customer Reviews
Apple iPod touch 8 GB (1st Generation) OLD MODEL by Apple

Average Customer Rating

181 customer reviews

<table>
<thead>
<tr>
<th>Rating</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 star</td>
<td>109</td>
</tr>
<tr>
<td>4 star</td>
<td>41</td>
</tr>
<tr>
<td>3 star</td>
<td>10</td>
</tr>
<tr>
<td>2 star</td>
<td>4</td>
</tr>
<tr>
<td>1 star</td>
<td>17</td>
</tr>
</tbody>
</table>

- Appearance 5⭐️⭐️⭐️⭐️⭐️ (6)
- Battery life 5⭐️⭐️⭐️⭐️⭐️ (6)
- Features 5⭐️⭐️⭐️⭐️⭐️ (6)
- Portability 5⭐️⭐️⭐️⭐️⭐️ (6)

See and rate all 6 attributes.
**Type 2b: Aspect Based Scores**

From the Commercial World

- **Bing.com** – automatic

Percentage positive statements on a given aspect

- **Screen**
- **Speed**
- **Ease Of Use**
- **Affordability**
- **Appearance**
- **Sound Quality**
- **Battery Life**
- **Size**
- **Compatibility**

User reviews:

- **Speed**: 37% positive comments.

Positive comments:

- Both have loaded pages very quickly - much faster than my Moto Q!
- Safari is the best mobile webbrowser and works very well, despite the slow AT&T network speed.
- It also offers WI-FI which enables fast broadband speed for browsing and connectivity.

Negative comments:

- There are no delays between apps is very fast and there is no delay at all.
Type 3: Aspect Ratings + Representative Phrases
From the Commercial World

- alatest.com - automatic

Collect reviews from many sources on the web
Generate aspect based scores
Highlight pros and cons
Opinion Summaries already there in the Commercial World....

1. Overall ratings
2. Aspect based star rating/scores
3. Concise textual summary on good/bad
Other types of summaries in the Research World:

1. Contrastive Summary
2. Aspect ratings with representative phrases
3. Entity Based Summary
4. Graphical Visualization of Opinions
What other types of summaries would be useful to users?

1. Abstract type opinion summaries
   - Similar to traditional text summarization
   - Only summarize opinions
   - Ex: Instead of having 1000s of reviews – 1 representative review summary

Summary on Obama
(fictitious summary)

Obama understands the american people and he is fast with policy implementation. However, Obama believes in massive spending and proposes bad healthcare reform.

Hard Problem!!
What other types of summaries would be useful to users?

2. Confidence based opinion summaries
   - Before selecting an opinion for summary ➔ how confident are we about it?
   - Summarize the truth!

   *Obama understands the american people*

   Agree (30 sources) ➔ Disagree (40 sources)

   **Confidence: 40%**
Opinion Summarization Methods
Methods in Opinion Summarization

Feature/Entity Identification
Identify salient topics

Sentiment Prediction
Determine polarity of text containing topics

Summary Presentation
- Aggregate polarity ratings
- Present opinion summaries

Battery Life
Sound Quality
Barrack Obama

battery life is great ➔ +ve
long battery life ➔ +ve

Battery Life: ★★★★★
Sound Quality: ★★★★★
## Methods in Opinion Summarization

### Feature/Aspect/Entity Identification

<table>
<thead>
<tr>
<th></th>
<th>NLP techniques</th>
<th>Frequent itemset mining</th>
<th>Manual selection</th>
<th>Clustering</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ Archak et. al 07 ]</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>[ Lu et al 09 ]</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>[Stoyanov&amp; Cardie 06]</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[ Mei et al 07 ]</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>[ Popescu 05 ]</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[ Hu&amp;Liu 06 ]</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[ Hu&amp;Liu 04 ]</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Methods in Opinion Summarization

### Sentiment Prediction

<table>
<thead>
<tr>
<th>[Lu et al 09]</th>
<th>Lexicon/Rule Based</th>
<th>NLP techniques</th>
<th>Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| [Ku et al 06] | X                   |                |          |

| [Mei et al 07] |                    |                | X        |

| [Ku et al 06] |                    | X              |          |

| [Hu&Liu 04]  | X                   |                | X        |

| [Hu&Liu 06]  | X                   | X              |          |

| [Zhuang 06]  | X                   |                |          |

| [Popescu 05] | X                   | X              |          |
## Methods in Opinion Summarization

### Summarization

<table>
<thead>
<tr>
<th></th>
<th>Clustering</th>
<th>Sentence/Phrase Selection</th>
<th>Coreference Resolution</th>
<th>Visualization</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ Lu et al 09 ]</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[ Lu et al 08 ]</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[ Ku et al 06 ]</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>[ Stoyanov &amp; Cardie 06 ]</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>[ Mei et al 07 ]</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[ Kim &amp; Zhai 09 ]</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[ Ku et al 06 ]</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>[ Popescu 05 ]</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>[ Hu &amp; Liu 04 ]</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[ Hu &amp; Liu 06 ]</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. Aspect/Feature Identification Methods
Aspect/Feature Identification:

- **Method:** NLP techniques/heuristics [Archak et al 07], [Lu et al 09], [Stoyanov&Cardie 06], [Popescu 05], [Hu&Liu 06], [Hu&Liu 04]
  - Nouns and noun phrases describe features
  - Adjectives and adverbs express opinions

- **Pros:**
  - High accuracy

- **Cons:**
  - Low recall, e.g. “it charges fast”
  - Dependent on POS tagging accuracy
Aspect/Feature Identification:

- **Method:** Frequent itemset ([Archak et al 07], [Popescu 05], [Hu & Liu 06], [Hu & Liu 04])
  - Frequent subsets of words as candidate features
  - Pruning by compactness/redundancy/mutual information
  - Identifying infrequent features as words sharing opinions with frequent features

- **Pros:**
  - Simple heuristics work reasonably well for product reviews

- **Cons:**
  - May need different heuristics given a different domain
  - No normalization of features, e.g. shipping and delivery
  - Many heuristic parameters to tune, e.g. support
Aspect/Feature Identification:

- **Method: clustering** [Lu et al 09], [Mei et al 07]
  - Words co-occurring frequently tend to be describing the same feature

- **Pros:**
  - Natural normalization of features as clusters
  - Few parameter to tune
  - Easy to incorporate prior knowledge

- **Cons:**
  - Performance depends on the domain
2. Sentiment Prediction Methods
Sentiment Prediction

- **Method: Lexicon/Rule Based** [Ku et al 06], [Hu&Liu 2004], [Hu&Liu 2006], [Zhuang 06], [Popescu 05]
  - Based on sentiment word dictionary

- **Pros:**
  - Simple and fast dictionary matching algorithm
  - Performance is reasonably good

- **Cons:**
  - Performance depends on the quality of the dictionary
  - Performance depends on the application domain; context is ignored
Sentiment Prediction

- Linguistics heuristics:
  - Positive “and” Positive
  - Positive “but” Negative
  - “negation” Positive = Negative

- Pros:
  - Applicable to any English text; domain independent

- Cons:
  - Could have some exceptions
Sentiment Prediction

- **Method: Learning** [Lu et al 09] [Mei et al 07] [Chen et al 06]

- **Pros:**
  - Captures context to some extent
  - All kinds of information can potentially be cast as features, e.g. lexicons, heuristics

- **Cons:**
  - Dependent on the quality of labels
  - Hard to obtain sufficient labels for some domains
3. Summarization Methods
Summarization

- **Clustering** [Lu et al 08] [Mei et al 07] [Hu&Liu 04] [Hu&Liu 06] [Kim&Zhai 09]
  - Cluster similar topics together

- **Pros:**
  - Rather simple technique

- **Cons:**
  - Difficulties in deciding clustering level
    - How many clusters, what granularity
Summarization

- **Sentence Phrase Selection** [Lu et al 09] [Ku et al 06] [Popescu 05] [Kim&Zhai 09]
  - Select some part of original data and show

- **Pros:**
  - Smaller amount of information to show

- **Cons:**
  - Cannot show all details
  - Bad selection may lose information
Summarization

- Coreference Resolution [Stoyanov&Cardie 06]
  - Focus on ‘who’ says what about ‘whom’

- Pros:
  - Can recognize the opinion holder and target more clearly

- Cons:
  - Sometimes opinion holder is not important
    - Ex. Product review
  - Relatively complex computation
Summarization

- **Visualization [Ku et al. 06]**
  - Graph presentation

- **Pros:**
  - More intuitive representation

- **Cons:**
  - Cannot read detail opinions
Summarization Presentation

- Text summary
  - Feature based text summary
    - General format
    - Opinion integration
    - Rated aspect summary
  - Contrastive summary

- Visual summary
  - Feature based graph summary
  - Opinion Summary with time line
  - Entity based summary
  - Other visualizations
Feature based text summary
[Hu&Liu 04, Hu&Liu 06, Zhuang 06, Mei et al 07]

- **Product: Camera**
  - **Feature:** picture
    - **Positive:** 12
      - Overall this is a good camera with a really good picture clarity.
      - The pictures are absolutely amazing - the camera captures the minutest of details.
      - After nearly 800 pictures I have found that this camera takes incredible pictures.
      ...
    - **Negative:** 2
      - The pictures come out hazy if your hands shake even for a moment during the entire process of taking a picture.
      - Focusing on a display rack about 20 feet away in a brightly lit room during day time, pictures produced by this camera were blurry and in a shade of orange.
      ...

---

Support

Raw review Sentences

Sentiment

Feature
Opinion Integration [Lu&Zhai 08]

- Based on the expert review
- Similar opinions/ Supplementary opinion from non-expert reviews

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Review</th>
<th>Similar Opinions</th>
<th>Supplementary Opinions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Background</td>
<td>Even with the new $399 price for the 8GB model (down from an original price of $599), it’s still a lot to ask for a phone that lacks so many features and locks you into an iPhone-specific two-year contract with AT&amp;T.</td>
<td>[support=13] The iPhone will come in two versions, a 4GB 499 model, and an 8GB 599 model with a two year contract.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[support=16] The Price: 499 (4GB) or 599(8GB) with a two year contract, by the time the contract is over your iPhone will probably be scratched all over like the Nano or be made obsolete by better phone on the market.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[support=12] Recently, Apple decided to cut down price of iPhone from 399 to 200, giving rise to much rage from consumers bought the phone before.</td>
<td></td>
</tr>
<tr>
<td>Activation</td>
<td>You can make emergency calls, but you can’t use any other functions, including the iPod music player.</td>
<td>[support=10] Several other methods for unlocking the iPhone have emerged on the Internet in the past few weeks, although they involve tinkering with the iPhone hardware or more complicated ways of bypassing the protections for AT T’s exclusivity.</td>
<td></td>
</tr>
<tr>
<td>Battery</td>
<td>Battery life The Apple iPhone has a rated battery life of 8 hours talk time, 24 hours of music playback, 7 hours of video playback, and 6 hours on Internet use.</td>
<td>[support=19] iPhone will Feature Up to 8 Hours of Talk Time, 6 Hours of Internet Use, 7 Hours of Video Playback or 24 Hours of Audio Playback</td>
<td>[support=7] Playing relatively high bitrate VGA H.264 videos, our iPhone lasted almost exactly 9 freaking hours of continuous playback with cell and WiFi on (but Bluetooth off).</td>
</tr>
</tbody>
</table>
Rated Aspect Summary [Lu&Zhai 09]

- Aspect ratings with representative text
  - Users want to know why an aspect got a particular rating
  - Short comments: easy to generate phrases
  - Verbose text: much harder to generate concise summaries
## Rated Aspect Summary [Lu&Zhai 09]

### Example output

<table>
<thead>
<tr>
<th>No.</th>
<th>Aspects</th>
<th>Ratings</th>
<th>Positive Phrase</th>
<th>Negative Phrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>described, promised</td>
<td>4.8457</td>
<td>as described (3993)</td>
<td>than expected (68)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>as promised (323)</td>
<td>than described (43)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>as advertised (149)</td>
<td>i ordered (10)</td>
</tr>
<tr>
<td>2</td>
<td>shipped, arrived</td>
<td>4.3301</td>
<td>quickly shipped (162)</td>
<td>open box (39)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>great thanks (149)</td>
<td>wrong sent (29)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>quickly arrived (138)</td>
<td>back sent (15)</td>
</tr>
<tr>
<td>3</td>
<td>recommended, was</td>
<td>3.9322</td>
<td>highly recommended (236)</td>
<td>back be (42)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>highly recommend (115)</td>
<td>defective was (40)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>exactly was (84)</td>
<td>not have (37)</td>
</tr>
<tr>
<td>4</td>
<td>shipping, delivery</td>
<td>4.7875</td>
<td>fast shipping (5354)</td>
<td>good shipping (170)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>quick shipping (879)</td>
<td>slow shipping (81)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>fast delivery (647)</td>
<td>reasonable shipping (32)</td>
</tr>
</tbody>
</table>
Contrastive Summary [Kim&Zhai 09]

\[ X = \{x_1, \ldots, x_n\} \]

Output:

\[ U = \{u_1\}_{i=1}^{k} \subset X, \quad V = \{v_1\}_{i=1}^{k} \subset Y \]

\[ Y = \{y_1, \ldots, y_m\} \]

Contrastive Sentence Pairs, which is subset of original sentence sets
## Contrastive Summary [Kim&Zhai 09]

### Example output

<table>
<thead>
<tr>
<th>No</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>oh ... and <strong>file transfers</strong> are fast &amp; easy.</td>
<td>you need the <strong>software</strong> to actually <strong>transfer files</strong></td>
</tr>
<tr>
<td>2</td>
<td>i noticed that the micro adjustment knob and collet are well made and work well too.</td>
<td>the adjustment knob seemed ok, but when lowering the router, i have to practically pull it down while turning the knob.</td>
</tr>
</tbody>
</table>
Feature based Graph Summary

[Hu & Liu 06, Hu et al 05]
Summary with Timeline

[Ku et al 06, Mei et al 07]
Entity based summary
[Stoyanov & Cardie 06a, 06b]

- Focused on ‘who’ talk what to ‘whom’
  cf) Product review analysis not care the teller
Entity based summary
[Stoyanov & Cardie 06a, 06b]

- Used coreference resolution
- Example output
Other Visualizations [Chen et al 06]

- Clustering, time analysis, terms extracted, coordinated view, distributed of selected terms, decision tree, etc.
Other Visualizations

[Chen et al 06, Mishne et. al 07]
Other Visualizations [Chen et al 06]

Terms extracted monthly
Other Visualizations [Chen et al 06]
Other Visualizations [Chen et al 06]

Coordinated Term View
Evaluation Methods
Data set

- **Corpus**
  - TREC Blog, NTCIR [Ku et al 2006],...
  - MPQA [Stoyanov & Cardie 06a, 06b, 08]

- **Task specific data**
  - Crawled, manually annotated
  - Product reviews, movie reviews
  - [Hu & Liu 04, Zhuang 06, Archak et al. 07, Lu & Zhai 08, Lu & Zhai 09, Kim & Zhai 09]
Measures

- Precision & Recall (aspect coverage)
  - Sometimes with F1
  - [Hu & Liu 04, Zhuang 06, Kim & Zhai 09]

- Classification
  - Prediction precision, accuracy [Archak et al. 07]

- Others
  - Rank loss (error rate) [Lu & Zhai 09]
  - Co-reference resolution metrics [Stoyanov & Cardie 08]
Future Directions
Future Directions

Many challenges in sentiment analysis

- Context sensitive sentiment analysis
  - “large screen” vs “large ipod”
Future Directions

Textual summaries - limited (but necessary)
- Most work focus on providing an overall rating on aspect/feature
- What is the biggest complaint on the iPod ‘screen’

Validity/Confidence of generated summaries
- Use multiple sources
- Extend concepts from contrastive opinion summarization [Kim & Zhai 2009]
Future Directions

Which summary format most effective?
- Provide different interfaces
- User survey?

Scalability vs. effectiveness of different methods
- Which method can scale to large amounts of reviews?
Thank you