Opinion Integration Through Semi-supervised Topic Modeling

Yue Lu and Chengxiang Zhai
University of Illinois at Urbana-Champaign
Why Opinion Integration?

• Web 2.0 ➔ huge amount of opinions
• What have been said about Hillary Clinton?

How to digest all?

190,451 posts

4,773,658 results
Two Kinds of Opinions

Expert opinions
- CNET editor’s review
- Wikipedia article
- Well-structured
- Easy to access
- Maybe biased
- Outdated soon

Ordinary opinions
- Forum discussions
- Blog articles
- Fragmental
- Hard to access
- Represent the majority
- Up to date

How to benefit from both?
• How do we formalize the problem of opinion integration?
• How do we solve the problem in a general way?
• How do we evaluate it?
Problem Definition

Input

Topic: iPod
Expert review with aspects
Weblogs

Output

Integrated Summary

Design Battery
Price

Similar opinions
- cute... tiny...
- last many hrs
- die out soon
- could afford it

Supplementary opinions
- thicker...
- still expensive

iTunes
- ... easy to use...

warranty
- ... better to extend...

Review Aspects
Extra Aspects
Challenges

1. How to align opinions to expert aspects or extra aspects?

Integrated Summary

Review Aspects

- Design
- Battery
- Price

Extra Aspects

- iTunes
- Warranty

Similar opinions:
- cute... tiny...
- last many hrs
- could afford it

Supplementary opinions:
- thicker...
- die out soon
- still expensive

3. How to extract representative opinions with support?
Two Major Steps

• Step 1: opinion sentences retrieval

General Weblogs

Weblogs on iPod

Query = “iPod”

• Step 2: opinion integration using probabilistic topic models (3 subtasks)
Subtask 1: Aspect Alignment

Align opinion sentences to aspects

- **Design**
  - cute... tiny...
  - thicker...

- **Battery**
  - last many hrs
  - die out soon

- **Price**
  - could afford it
  - still expensive

- **iTunes**
  - ... easy to use...

- **warranty**
  - ...better to extend..
Separate sim opinions from supp ones

<table>
<thead>
<tr>
<th>Review Aspects</th>
<th>Extra Aspects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Design</strong></td>
<td><strong>iTunes</strong></td>
</tr>
<tr>
<td><strong>Battery</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Price</strong></td>
<td><strong>warranty</strong></td>
</tr>
<tr>
<td>cute... tiny...</td>
<td>easy to use...</td>
</tr>
<tr>
<td>..thicker...</td>
<td></td>
</tr>
<tr>
<td>last minute...</td>
<td></td>
</tr>
<tr>
<td>it's out soon</td>
<td></td>
</tr>
<tr>
<td>could afford it</td>
<td></td>
</tr>
<tr>
<td>still expensive</td>
<td></td>
</tr>
<tr>
<td>...better to extend..</td>
<td></td>
</tr>
</tbody>
</table>
Summarize each block with representative sentences and support

<table>
<thead>
<tr>
<th>Review Aspects</th>
<th>Design</th>
<th>Battery</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>cute... tiny...</td>
<td>..thicker..</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>last</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>many hrs</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>could afford it</td>
<td>still expensive</td>
</tr>
</tbody>
</table>

**Representative Opinion (RO)**

- Representative sentence: cute... tiny...
- Support = 3

**Extra Aspects**

<table>
<thead>
<tr>
<th>Extra Aspects</th>
<th>iTunes</th>
<th>warranty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>... easy...</td>
<td>...better to extend..</td>
</tr>
</tbody>
</table>

iTunes: ... easy...
Warranty: ...better to extend..
Basic PLSA: Generation Process

[Hofmann 99], [Zhai et al. 04]

Topic model
= unigram language model
= multinomial distribution

\[ p_d(w) = \lambda_B p(w|\theta_B) + (1 - \lambda_B) \sum_{j=1}^{k} [\pi_{d,j} p(w|\theta_j)] \]

Generate a word in a document

Battery 0.3
Life 0.2
Design 0.1
Screen 0.05
Price 0.2
Purchase 0.15

Is 0.05
The 0.04
A 0.03
Basic PLSA: Estimation

[Hofmann 99], [Zhai et al. 04]

\[ p_d(w) = \lambda_B p(w|\theta_B) + (1 - \lambda_B) \sum_{j=1}^{k} [\pi_{d,j} p(w|\theta_j)] \]

Log-likelihood of the collection

\[ \log p(C_O|\Lambda) = \sum_{d\in C_O} \sum_{w\in V} \cdot \{c(w,d) \times \log p_d(w)\} \]

\[ \hat{\Lambda} = \arg \max_{\Lambda} \log p(C_O|\Lambda) \]

Estimated with Maximum Likelihood Estimator (MLE) through an EM algorithm
Basic PLSA: Problem?

Expert review with aspects

Weblogs on iPod

Basic PLSA

Extracted topics may not align with expert review aspects

Solution: conjugate priors
Semi-supervised PLSA
Semi-supervised PLSA

Add Dirichlet priors

Maximum Likelihood Estimation (MLE)
\[ \hat{\Lambda} = \arg\max_{\Lambda} \log p(C_O|\Lambda) \]

Maximum A Posterior (MAP) Estimation
\[ \hat{\Lambda} = \arg\max_{\Lambda} \log p(C_O|\Lambda) p(\Lambda) \]

• Confidence in priors
• Regularization
### Subtask 1: Aspect Alignment

#### Design
- cute... tiny...
- last many hrs
- could afford it

#### Battery
- ..thicker..
- die out soon
- still expensive

#### Price
- can
- could afford it

#### Weblogs
- 

#### Extra Aspects
- 

#### iTunes
- ... easy to use...

#### Warranty
- ...better to extend..
Subtask 1: Aspect Alignment

“aspect words” (nouns)

Topics

Align to review aspects

Documents

Opinion sentence aligned to the most relevant aspect

Discover extra aspects

Battery life

design screen

\[ p(\Lambda) \]

\[ \theta_1 \]

\[ \theta_2 \]

\[ \theta_k \]

\[ \pi_{d2} \]

\[ \pi_{dk} \]

\[ \lambda_B \]

Collection background

Is 0.05 the 0.04 a 0.03 ..
## Subtask 2: Opinions Separation

<table>
<thead>
<tr>
<th>Review Aspects</th>
<th>Extra Aspects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design</td>
<td>cute... tiny... ..thicker..</td>
</tr>
<tr>
<td>Battery</td>
<td>last many... die out soon</td>
</tr>
<tr>
<td>Price</td>
<td>could afford it... still expensive</td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>iTunes</td>
<td>...easy to use...</td>
<td></td>
</tr>
<tr>
<td>warranty</td>
<td>...better to extend..</td>
<td></td>
</tr>
</tbody>
</table>
Subtask 2: Opinions Separation

Sub-Collection: Battery

“opinion words” about Battery

long
many
### Representative Opinion (RO)

- **Design**
  - cute... tiny...
- **Battery**
  - ..thicker..
- **Price**
  - could afford it
  - still expensive

### Extra Aspects

<table>
<thead>
<tr>
<th>Aspects</th>
<th>Opinion</th>
</tr>
</thead>
<tbody>
<tr>
<td>iTunes</td>
<td>... easy...</td>
</tr>
<tr>
<td>warranty</td>
<td>...better to extend...</td>
</tr>
</tbody>
</table>

- **Representative sentence:**
  - Support = 3
Subtask 3: Opinion Summary

Centroid Sentence 1

Centroid Sentence 2

\[ \pi_{d1} \]

\[ \pi_{d2} \]

\[ \pi_{d_k} \]

Support = cluster size

\[ \text{Is } 0.05 \]

\[ \text{the } 0.04 \]

\[ \text{a } 0.03 .. \]
Experiment Setup

• Expert review data:

<table>
<thead>
<tr>
<th>Topic</th>
<th>Source</th>
<th># words</th>
<th># aspects</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPhone</td>
<td>CNET</td>
<td>4434</td>
<td>19</td>
</tr>
<tr>
<td>Barack Obama</td>
<td>Wikipedia</td>
<td>312</td>
<td>14</td>
</tr>
</tbody>
</table>

• Ordinary opinion data:

<table>
<thead>
<tr>
<th>Topic</th>
<th>Query Terms</th>
<th># articles</th>
<th># sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPhone</td>
<td>iPhone</td>
<td>552</td>
<td>3000</td>
</tr>
<tr>
<td>Barack Obama</td>
<td>Barack+Obama</td>
<td>639</td>
<td>1000</td>
</tr>
</tbody>
</table>
Results: Product (iPhone)

- Opinion Integration with review aspects

<table>
<thead>
<tr>
<th>Review article</th>
<th>Similar opinions</th>
<th>Supplementary opinions</th>
</tr>
</thead>
<tbody>
<tr>
<td>You can make emergency calls, but you can't use any other functions...</td>
<td>N/A</td>
<td>Unlock/hack iPhone... methods for unlocking the iPhone have emerged on the Internet in the past few weeks, although they involve tinkering with the iPhone hardware...</td>
</tr>
<tr>
<td>rated battery life, 8 hours talk time, 24 hours of music playback, 7 hours of video playback, and 6 hours on Internet use.</td>
<td>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>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)...</td>
</tr>
</tbody>
</table>

Additional info under real usage
### Opinions on extra aspects

<table>
<thead>
<tr>
<th>Support</th>
<th>Supplementary opinions on extra aspects</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>You may have heard of <strong>iASign</strong>, an iPhone Dev. Wiki tool that allows you to <strong>activate</strong> your phone without going through the iTunes rigamarole.</td>
</tr>
<tr>
<td>13</td>
<td><strong>Cisco</strong> has owned the <strong>trademark</strong> on the name &quot;<strong>iPhone</strong>&quot; since 2000, when it acquired InfoGear Technology Corp., which originally registered the name discarding it.</td>
</tr>
<tr>
<td>13</td>
<td>With the imminent availability of the <strong>iPhone</strong>, a look at 10 things current smartphones like the <strong>Nokia N95</strong> have been able to do for a while that the <strong>iPhone</strong> can't currently <strong>match</strong>...</td>
</tr>
</tbody>
</table>

Another way to activate iPhone

iPhone trademark originally owned by Cisco

A better choice for smart phones?
Results: Product (iPhone)

- Support statistics for review aspects

- People care about price
- Controversy: activation requires contract with AT&T
- People comment a lot about the unique wi-fi feature
Quantitative Evaluation

• Goal
  – Evaluate human agreement (how hard is opinion integration?)
  – Evaluate how our approach could reproduce human choice (how well is our method doing?)

• Method
  – Ask 3 users to perform 3 tasks
  – Tasks designed from the Obama example
**Task 1: Distinguish Extra Aspects**

- **Result**
  - Low human agreement (1/7)
  - Our method recovers **52.4%** of user choices on avg

<table>
<thead>
<tr>
<th>Design</th>
<th>Battery</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>cute... tiny...</td>
<td>..thicker..</td>
<td></td>
</tr>
<tr>
<td>hour</td>
<td>die soon</td>
<td></td>
</tr>
<tr>
<td>could afford it</td>
<td>still expensive</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>iTunes</th>
<th>warranty</th>
</tr>
</thead>
<tbody>
<tr>
<td>... easy...</td>
<td>better to extend..</td>
</tr>
</tbody>
</table>

7 extra aspects

34 opinions
Task 2: Aspect Alignment

- Mix 27 opinions
- Label each with one of 14 aspects

Results:
- Users agree on 13/27 = 48% sentences
- Our method recovers 10.67/27 = 40% sentences on avg.
### Quantitative Evaluation: Task 3

#### Review Aspects

<table>
<thead>
<tr>
<th>Design</th>
<th>Battery</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>cute... tiny...</td>
<td>..thicker..</td>
<td></td>
</tr>
<tr>
<td>only hrs</td>
<td>die soon</td>
<td></td>
</tr>
<tr>
<td>could afford it</td>
<td>still expensive</td>
<td></td>
</tr>
</tbody>
</table>

- Mix one sim opinion with many supp opinions
- Select one opinion most similar to the review opinion
- Result: recovers $60\%$ of human choice
Summary

- **Novel problem**: opinion integration
- **Unified approach**: semi-supervised probabilistic topic modeling
- Many potential *interesting applications*
- **Future Work**
  - More rigorous evaluation
  - More general setup: many expert reviews instead of one
Thank you!