Expert-Guided Contrastive Opinion Summarization

Kenney Guo, Yujie Lu, Tatsunori Mori, Catherine Blake

5/8/2015
Motivation

Gay Marriage

Gun Control

Euthanasia

Obamacare

Abortion

Gay Marriage

Animal Testing
Motivation

Gay marriage can bring financial gain to governments and boost the economy.

People should not have their tax dollars used to support something they believe is wrong.

Marriage is a privilege, not a right.

Children need both a mother and a father.

Gay marriage undermines the institutions that exist.

Gay couples make good parents.

Marriage is civil right.

Gay marriage is not redefining marriage, considering the history.

Gay marriage can bring financial gain to governments and boost the economy.
### Goal: Summarization for Controversial Topic

**Topic:** "gay marriage"

### Contrastive expert arguments

<table>
<thead>
<tr>
<th>Arg 1</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Arg 2</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Ordinary opinions (Twitter)

### Expert Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Pros/Cons</th>
<th>Similar Opinions from Twitter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tradition</td>
<td>traditional marriage definition is inaccurate</td>
<td>gay marriage is not redefining it, considering the history.</td>
</tr>
<tr>
<td></td>
<td>marriage is traditionally defined between a man and a woman</td>
<td>gay marriage really just to undermine the institutions that exist</td>
</tr>
<tr>
<td>Children</td>
<td>Gay couples make good parents</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>Children need both a mother and a father</td>
<td>opposition is more about harm to children in gay marriage, not sole purpose of breeding.</td>
</tr>
</tbody>
</table>

### Extra Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liberalism</td>
<td>None</td>
<td>extreme conservatives love to tout out the slippery slope that gay marriage would bring were it put in place. the one we're on is far scarier</td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>damn slippery slopes, just like the gays, first you legalize gay marriage, then we all start practicing bestiality</td>
</tr>
</tbody>
</table>
Outline

• Current Approaches on Opinion Summarization
• Expert-Guided Contrastive Opinion Summarization (ECOS)
• Use ECOS on Contrastive Opinion Summarization (COS) task
  • Semi-supervised PLSA model for argument clustering
  • Contrastive Similarity
  • Experiment
• Conclusion
Outline

• **Current Approaches on Opinion Summarization**
• Expert-Guided Contrastive Opinion Summarization (ECOS)
• Use ECOS on Contrastive Opinion Summarization (COS) task
  • Semi-supervised PLSA model for argument clustering
  • Contrastive Similarity
  • Experiment
• Conclusion
Current Approaches on Opinion Summarization

(Lu & Zhai, 2008)

(Paul et al., 2010)

(Kim & Zhai, 2009)

(Zheng & Fang, 2010)
Our Approach on Opinion Summarization
Outline

• Current Approaches on Opinion Summarization

• **Expert-Guided Contrastive Opinion Summarization (ECOS)**

• Use ECOS on Contrastive Opinion Summarization (COS) task
  - Semi-supervised PLSA model for argument clustering
  - Contrastive Similarity
  - Experiment

• Conclusion
Expert-Guided Contrastive Opinion Summarization (ECOS)

Design your output

Exp pro 1  Exp con 1
Exp pro 2  Exp con 2
Exp pro 3  Exp con 3
NA        Exp con 4
Exp pro 5  NA

Exp pro 1
Exp con 1

Exp pro 2
Exp con 2

Exp pro 3
Exp con 3

NA
Exp con 4

Exp pro 5
NA

Exp pro 1
Exp con 1

Exp pro 2
Exp con 2

Exp pro 3
Exp con 3

NA
Exp con 4

Exp pro 5
NA

Exp pro 1
Exp con 1

Exp pro 2
Exp con 2

Exp pro 3
Exp con 3

NA
Exp con 4

Exp pro 5
NA

OR

...
Outline

• Current Approaches on Opinion Summarization
• Expert-Guided Contrastive Opinion Summarization (ECOS)
• Use ECOS on Contrastive Opinion Summarization (COS) task
  • Semi-supervised PLSA model for argument clustering
  • Contrastive Similarity
  • Experiment
• Conclusion
Problem Definition for COS

• Contrastive Opinion Summarization (COS) (Kim & Zhai, 2009)

<table>
<thead>
<tr>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_1$</td>
<td>$v_1$</td>
</tr>
<tr>
<td>$u_2$</td>
<td>$v_2$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$u_x$</td>
<td>$v_y$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contradictory Aspect</th>
<th>Positive (Pros)</th>
<th>Negative (Cons)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contradictory 1</td>
<td>$u_1$</td>
<td>$v_1$</td>
</tr>
<tr>
<td>Contradictory 2</td>
<td>$u_2$</td>
<td>$v_2$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Contradictory $k$</td>
<td>$u_k$</td>
<td>$v_k$</td>
</tr>
</tbody>
</table>
Optimization Framework

- **Content Similarity**
  - \( \phi(s_1, s_2) \in [0,1] \)

- **Contrastive Similarity**
  - \( \varphi(s_1, s_2) \in [0,1] \)

- **Representativeness**
  - \( r(S) = \frac{1}{|X|} \sum_{x \in X} \max_{i \in [1,k]} \phi(x, u_{\downarrow i}) + \frac{1}{|Y|} \sum_{y \in Y} \max_{i \in [1,k]} \phi(y, v_{\downarrow i}) \)

- **Contrastiveness**
  - \( c(S) = \frac{1}{k} \sum_{i=1}^{k} \max_{i \in [1,k]} \varphi(u_{\downarrow i}, v_{\downarrow i}) \)

- **Optimization function**
  - \( S^{*} = \arg \max_{S} (\mu r(S) + (1-\mu)c(S)) \)
  - \( = \arg \max_{S} (\mu/|X|) \sum_{x \in X} \max_{i \in [1,k]} \phi(x, u_{\downarrow i}) + \mu/|Y| \sum_{y \in Y} \max_{i \in [1,k]} \phi(y, v_{\downarrow i}) + 1-\mu/k \sum_{i=1}^{k} \max_{i \in [1,k]} \varphi(u_{\downarrow i}, v_{\downarrow i}) \)
Approximation Algorithms (1)

- Representativeness-First Approximation

Choose \( \{u_i, v_i\} \) from Clusters to Optimize Objective Function \( S^* \)

<table>
<thead>
<tr>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>( u_1 )</td>
<td>( v_1 )</td>
</tr>
<tr>
<td>( u_2 )</td>
<td>( v_2 )</td>
</tr>
<tr>
<td>( u_3 )</td>
<td>( v_3 )</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>( u_x )</td>
<td>( v_y )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>Cluster 1</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>Cluster 2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Cluster k</td>
<td>Cluster k</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contradictory Aspect</th>
<th>Positive (Pros)</th>
<th>Negative (Cons)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contradictory 1</td>
<td>( u_1 )</td>
<td>( v_1 )</td>
</tr>
<tr>
<td>Contradictory 2</td>
<td>( u_2 )</td>
<td>( v_2 )</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Contradictory k</td>
<td>( u_k )</td>
<td>( v_k )</td>
</tr>
</tbody>
</table>
Approximation Algorithms (2)

• Contrastiveness-First Approximation

Add \{u_i, v_i\} iteratively to Maximize the Increase of Objective Function $S^*$
Intuition

• The Optimization Framework is “optimal” not “best”
• Intuitively clustering is useful to pick up representativeness
• Representativeness-first strategy not as well as Contrastiveness-first strategy (Kim & Zhai, 2009)
• What if we guide the clustering process with themes/aspects?
Use ECOS to Solve the COS Problem

- Exp pro 1
- Exp con 1
- Exp pro 2
- Exp con 2
- Exp pro 3
- Exp con 3
- Exp pro 4
- Exp con 4
- Exp pro 5
- NA

Contrastive Similarity

Aligned Pairs

Unaligned Pool

Traditional COS Method
Outline

• Current Approaches on Opinion Summarization
• Expert-Guided Contrastive Opinion Summarization (ECOS)
• Use ECOS on Contrastive Opinion Summarization (COS) task
  • Semi-supervised PLSA model for argument clustering
  • Contrastive Similarity
  • Experiment
• Conclusion
Semi-supervised PLSA for Clustering

(1) Basic PLSA (Probabilistic Latent Semantic Analysis) topic mode
• \( pw\theta_{\downarrow j} = \frac{\sum_{d \in C^\uparrow c(w,d)} p(z_{\downarrow d,w,j})}{\sum_{w' \in V^\uparrow c} \sum_{d' \in C^\uparrow c(w',d')} p(z_{\downarrow d',w',j})} \)
• \( \max_{\neg j} pd\big\{\theta_{\downarrow j}\big\} = \max_{\neg j} \sum_{w \in V^\uparrow c} p(w, d_{\downarrow i}) pw\theta_{\downarrow j} \)

(2) Semi-supervised PLSA (Lu & Zhai, 2008)
• Adding prior: \( pw_{\uparrow r l i} = \frac{c(w, r_{\downarrow l i})}{\sum_{w' \in V^\uparrow c} c(w', r_{\downarrow l i})} \)
• \( pw\theta_{\downarrow j} = \frac{\sum_{d \in C^\uparrow c(w,d)} p(z_{\downarrow d,w,j})}{\sum_{w' \in V^\uparrow c} \sum_{d' \in C^\uparrow c(w',d')} p(z_{\downarrow d',w',j})} \)
Semi-supervised PLSA for Clustering

(1) Basic PLSA (Probabilistic Latent Semantic Analysis) topic mode

\[ p(w \theta | j) = \sum_{d \in C \uparrow \downarrow c(w, d)} p(z \downarrow d, w, j) / \sum_{w' \in V \uparrow \downarrow c(w')} \sum_{d' \in C \uparrow \downarrow c(w', d')} p(z \downarrow d', w', j) \]

• \[ \max_{j} p(d \theta | j) = \max_{j} \sum_{w \in V \uparrow \downarrow c(w, d \downarrow i)} p(w \theta | j) \]

(2) Semi-supervised PLSA (Lu & Zhai, 2008)

• Adding prior: \[ p(w | r \downarrow l) = c(w, r \downarrow l) / \sum_{w' \in V \uparrow \downarrow c(w')} \]

\[ p(w \theta | j) = \sum_{d \in C \uparrow \downarrow c(w, d)} p(z \downarrow d, w, j) + \sigma_{j} p, w, r \downarrow j / \sum_{w' \in V \uparrow \downarrow c(w')} \sum_{d' \in C \uparrow \downarrow c(w', d')} p(z \downarrow d', w', j) + \sigma_{j} \]
Compute Prior Probability of Expert Keywords

• 15 pros and 13 cons about “gay marriage”
• Convert to 8 aspects of argument
  • Add keywords (probability) for each aspect, namely $p_{wrij}$ as prior
  • Or calculate $p_{wrij}$ from the document
<table>
<thead>
<tr>
<th>Prior Category</th>
<th>Positive Cluster</th>
<th>Negative Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children &amp; Adoption</td>
<td>women, church, mean, move, hypocrisi, contemporari, keen, examin, time, issu...</td>
<td>state, vote, voter, believ, express, defin, meanwhil, bauer, gari, guess, children...</td>
</tr>
<tr>
<td>Economic Problem</td>
<td>cuz, say, get, peopl, hate, marri, defend, stop, destroy, real...</td>
<td>fight, law, thing, issu, time, real, need, corrupt, horror, write</td>
</tr>
<tr>
<td>Religion</td>
<td>believ, rule, ban, women, church, go, reason, equal, kill, agre...</td>
<td>state, god, thank, church, misrepresent, caricatur, everybodi, fool, goon, liber</td>
</tr>
<tr>
<td>Procreation</td>
<td>talk, thing, posit, breed, allow, base, also, disallow, let, us...</td>
<td>anyon, bibl, speak, freedom, let, want, christian, much, rememb, marri</td>
</tr>
<tr>
<td>None</td>
<td>even, realli, know, debat, problem, human, andddd, argument, valid, believ...</td>
<td>want, cuz, that, vote, happen, go, call, im, gon, na</td>
</tr>
</tbody>
</table>
Outline

• Current Approaches on Opinion Summarization
• Expert-Guided Contrastive Opinion Summarization (ECOS)
• Use ECOS on Contrastive Opinion Summarization (COS) task
  • Semi-supervised PLSA model for argument clustering
  • Contrastive Similarity
  • Experiment
• Conclusion
Contrastive Similarity

• Content similarity between positive and negative tweets
  • Similarity with all words
  • Similarity with content words only (remove negation, adjectives, etc.)

• Different content similarity measure
  • Cosine similarity
  • Semantic similarity
Outline

• Current Approaches on Opinion Summarization
• Expert-Guided Contrastive Opinion Summarization (ECOS)
• Use ECOS on Contrastive Opinion Summarization (COS) task
  • Semi-supervised PLSA model for argument clustering
  • Contrastive Similarity
  • Experiment
• Conclusion
Experiment: Data

• Data
  • Expert opinions
    • 15 pros and 13 cons for topic “gay marriage” from procon.org
    • Expert opinions are converted to 8 aspects/themes
  • Ordinary opinions
    • A subset of tweets collected from 2014-11-19 to 2014-12-05 with keyword “gay marriage” and its synonyms (7624 tweets, one day).
    • 633 left after cleaning and filtering

• Annotation
  • 633 tweets labeled as positive, negative, neutral or none of the above
  • 50 positive and 50 negative tweets are randomly selected for aspect/theme annotation
## Experiment: Annotation Results

<table>
<thead>
<tr>
<th>Aspect ID</th>
<th>Aspect Category</th>
<th>No. in Positive Tweets</th>
<th>No. in Negative Tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Children &amp; Adoption</td>
<td>None</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>Economic Problem</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>3</td>
<td>Civil Rights</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Discrimination</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>Tradition &amp; Definition</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>Psychological Problem</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>7</td>
<td>Religion</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>Procreation</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>9</td>
<td>Emotions</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>Liberalism</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>Priority</td>
<td>None</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>
Experiment Results

• Precision: percentage of the k (k=10) pairs that are correct.
• Coverage: percentage of argument aspects covered in the k pairs.
• Baseline
  • only top contrastive similarity
  • free cluster information (set prior as 0) + top contrastive similarity
• ECOS: expert guided cluster information (top contrastive similarity) + top contrastive similarity

<table>
<thead>
<tr>
<th></th>
<th>Word All</th>
<th>Content Word Only</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td>Coverage</td>
</tr>
<tr>
<td>Baseline 1</td>
<td>0.200</td>
<td>0.167</td>
</tr>
<tr>
<td>Baseline 2</td>
<td>0.200</td>
<td>0.333</td>
</tr>
<tr>
<td>ECOS</td>
<td>0.600</td>
<td>0.667</td>
</tr>
</tbody>
</table>
Extra Benefits of ECOS

• Opinion integration
  • Contrastive summarization of ordinary opinion together with the expert opinion (enhance + complement)

• Interactive Analysis
  • Adjust keywords for your own needs to guide the process of argument clustering
Limitations & Future Work

• Limitations
  • Experiment data is relatively small
  • Accuracy of sentiment classification
  • Naïve way of structuring controversial arguments as positive (pros) vs. negative (cons)

• Future work
  • Improve experiments: compare to other COS approaches
  • Generalize the model: try other semi-supervised models (e.g. LDA)
  • Advanced structures for controversial arguments (argumentation mining)
Outline

• Current Approaches on Opinion Summarization
• Expert-Guided Contrastive Opinion Summarization (ECOS)
• Use ECOS on Contrastive Opinion Summarization (COS) task
  • Semi-supervised PLSA model for argument clustering
  • Contrastive Similarity
  • Experiment

• Conclusion
Conclusion

- ECOS model shows promising results in **integrating** expert and ordinary opinions (both positive and negative) in a unified way.
- ECOS provides a new **semi-supervised** approach for the COS task.
Thank you for listening!
Questions?
Reference


