Lecture 22: Summarization

Dan Klein – UC Berkeley
Includes slides from Aria Haghighi, Dan Gillick
Sentence Extraction

\[ D \quad S \]

Selection

- Maximum Marginal Relevance
  [Carbonell and Goldstein, 1998]

Greedy search over sentences

Maximize similarity to the query

\[ \text{MMR} = \max_{S \in \mathcal{S}} \left( \lambda \text{Sim}(S, Q) - \lambda \text{Redundancy}(S, D) \right) \]

Minimize redundancy
Selection

mid-'90s

• Maximum Marginal Relevance
• Graph algorithms [Mihalcea 05++]

present

Selection

mid-'90s

• Maximum Marginal Relevance
• Graph algorithms

present

Nodes are sentences
S1
S2
S3
S4
Selection

- Maximum Marginal Relevance
- Graph algorithms

Nodes are sentences
Edges are similarities

Selection

Stationary distribution represents node centrality
Selection

mid-'90s

- Maximum Marginal Relevance
- Graph algorithms
- Word distribution models

present

Input document distribution

<table>
<thead>
<tr>
<th>w</th>
<th>P_D(w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obama</td>
<td>0.017</td>
</tr>
<tr>
<td>speech</td>
<td>0.024</td>
</tr>
<tr>
<td>health</td>
<td>0.009</td>
</tr>
<tr>
<td>Montana</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Summary distribution

<table>
<thead>
<tr>
<th>w</th>
<th>P_S(w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obama</td>
<td>?</td>
</tr>
<tr>
<td>speech</td>
<td>?</td>
</tr>
<tr>
<td>health</td>
<td>?</td>
</tr>
<tr>
<td>Montana</td>
<td>?</td>
</tr>
</tbody>
</table>

Selection

mid-'90s

- Maximum Marginal Relevance
- Graph algorithms
- Word distribution models

present

SumBasic [Nenkova and Vanderwende, 2005]

Value(w_i) = P_D(w_i)
Value(s_i) = sum of its word values
Choose s_i with largest value
Adjust P_D(w)
Repeat until length constraint
Selection

- Maximum Marginal Relevance
- Graph algorithms
- Word distribution models
- Regression models

<table>
<thead>
<tr>
<th>word values</th>
<th>position</th>
<th>length</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>18</td>
</tr>
</tbody>
</table>

Frequency is just one of many features

Selection

mid-'90s
- Maximum Marginal Relevance
- Graph algorithms
- Word distribution models
- Regression models
- Topic model-based
  [Haghighi and Vanderwende, 2009]
**Summarization Criterion**

\[
P_C(\cdot)
\]

- Barack Obama: 0.15
- Serve America Act: 0.13
- signed: 0.12

\[
P_S(\cdot)
\]

- Barack Obama: 0.18
- Serve America Act: 0.16
- signed: 0.10

[Minimizing the KL divergence]

**Summarization Criterion**

\[
S^* = \min_{S: words(S) \leq L} KL(P_C \parallel P_S)
\]

\[
P_C(\cdot)
\]

- Barack Obama: 0.15
- Serve America Act: 0.13
- signed: 0.12

\[
P_S(\cdot)
\]

- Barack Obama: 0.18
- Serve America Act: 0.16
- signed: 0.10

[Haghighi & Vanderwende, NAACL ’09]
**Raw Count Content Model**

- President Barack Obama received the Serve America Act after congress’ vote.
  The ailing senator was instrumental in its passage.

**Document Structure**

- President Barack Obama received the Serve America Act after congress’ vote...

- The bill is named after Massachusetts Senator Ted Kennedy who was present at its signing. The ailing senator was instrumental...

- The legislation would greatly expand the ranks of Ameri-Corps, which was created by President Bill Clinton in 1993...
Structured Content Models

**General**

Barack Obama: 0.15  
Serve America Act: 0.13  
signed: 0.12

**Ted Kennedy**

Ted Kennedy: 0.18  
introduced: 0.12  
ailing senator: 0.11

**Ameri-Corps**

Ameri-Corps: 0.11  
Bill Clinton: 0.16  
expand: 0.08

**Cost**

cost: 0.11  
republicans: 0.09  
congress: 0.07  
budget: 0.05

[Haghghi & Vanderwende, NAACL ’09 ]

---

Document Structure

**General**

Barack Obama: 0.15  
Serve America Act: 0.13  
signed: 0.12

**Specific**

Ted Kennedy: 0.18  
introduced: 0.12  
ailing senator: 0.11

Ameri-Corps: 0.11  
Bill Clinton: 0.16  
expand: 0.08

Cost: 0.11  
republicans: 0.09  
congress: 0.07  
budget: 0.05
Generative Model

Evaluation: User Study

Pairwise Comparison

<table>
<thead>
<tr>
<th># of User Preferences</th>
<th>PYTHY</th>
<th>H &amp; V 09</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>20</td>
<td>49</td>
</tr>
<tr>
<td>Redundancy</td>
<td>21</td>
<td>48</td>
</tr>
<tr>
<td>Coherence</td>
<td>15</td>
<td>54</td>
</tr>
<tr>
<td>Focus</td>
<td>28</td>
<td>41</td>
</tr>
</tbody>
</table>

Larger is better
Example General Summary

Former House Speaker Newt Gingrich is asking a judge to force his estranged wife to turn over money he says she is hoarding.

On Thursday, accusations of wrongdoing and the mining of dirt in the former U.S. House speaker's divorce case gave way to a secret settlement between Gingrich and his wife of 18 years, Marianne Gingrich.

Gingrich filed for divorce July 29 amid allegations he is having an affair with 33-year-old congressional aide Callista Bisek.

Example Topical Summary

Gingrich Bio / Post-Speaker Life

Gingrich is best known leading the Republican Party's takeover of the House in 1994. During that so-called Republican Revolution, Gingrich emphasized that "family values" should be a core pillar in American society.

Since resigning as speaker and from the congressional seat he held for 20 years, Gingrich has been making a living giving speeches, sitting on corporate boards, consulting and appearing as a political analyst on Fox News.

U.S. Rep. J.D. Hayworth (R-Ariz.) argued that Gingrich's new job as a political commentator for Fox News makes it inappropriate to include him in political gatherings. "Time marches on. He's gone on to other pursuits," Hayworth said.
Selection

- Maximum Marginal Relevance
- Graph algorithms
- Word distribution models
- Regression models
- Topic models
- Globally optimal search
  [McDonald, 2007]

Optimal search using MMR

Integer Linear Program

Maximize:\n\[ \sum_i R_e(i) s_i - \sum_{ij} R_{eij} s_{ij} \]

Subject to:\n\[ \sum_j s_{ij} \leq L \]
\[ s_{ij} \leq s_i \leq s_j \quad \forall i, j \]
\[ s_i + s_j - s_{ij} \leq 1 \quad \forall i, j \]
\[ s_i \in \{0, 1\} \quad \forall i \]
\[ s_{ij} \in \{0, 1\} \quad \forall i, j \]

Selection

[Gillick and Favre, 2008]

<table>
<thead>
<tr>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
</tr>
</thead>
<tbody>
<tr>
<td>The health care bill is a major test for the Obama administration.</td>
<td>Universal health care is a divisive issue.</td>
<td>President Obama remained calm.</td>
<td>Obama addressed the House on Tuesday.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>concept</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Universal health care is a divisive issue. Obama addressed the House on Tuesday.

President Obama remained calm. The health care bill is a major test for the Obama administration.

<table>
<thead>
<tr>
<th>Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Gillick and Favre, 2008]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>concept</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>obama</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Gillick and Favre, 2008]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>concept</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>obama</td>
<td>3</td>
</tr>
<tr>
<td>health</td>
<td>2</td>
</tr>
</tbody>
</table>
Universal health care is a divisive issue.

The health care bill is a major test for the Obama administration.

President Obama remained calm.

Obama addressed the House on Tuesday.

<table>
<thead>
<tr>
<th>Concept</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>obama</td>
<td>3</td>
</tr>
<tr>
<td>health</td>
<td>2</td>
</tr>
<tr>
<td>house</td>
<td>1</td>
</tr>
</tbody>
</table>

Length limit: 18 words

<table>
<thead>
<tr>
<th>Summary</th>
<th>Length</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>{s_1, s_3}</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>{s_2, s_3, s_4}</td>
<td>17</td>
<td>6</td>
</tr>
</tbody>
</table>
Selection

Integer Linear Program for the maximum coverage model

[Gillick, Riedhammer, Favre, Hakkani-Tur, 2008]

Maximize: \( \sum_i w_i c_i \)  \hspace{1cm} \text{total concept value}

Subject to: \( \sum_j l_j s_j \leq L \)  \hspace{1cm} \text{summary length limit}

\( s_j \text{Occ}_{ij} \leq c_i \), \( \forall i, j \)

\( \sum_j s_j \text{Occ}_{ij} \geq c_i \), \( \forall i \)

\( c_i \in \{0, 1\} \)  \( \forall i \)

\( s_j \in \{0, 1\} \)  \( \forall j \)

This ILP is tractable for reasonable problems
How to include sentence position?

First sentences are unique

Selection

Only allow first sentences in the summary
Up-weight concepts appearing in first sentences
Identify more sentences that look like first sentences

surprisingly strong baseline
included in TAC 2009 system
first sentence classifier is not reliable enough yet

How to include sentence position?
Selection

Some interesting work on sentence ordering
[Barzilay et. al., 1997; 2002]

But choosing independent sentences is easier
  • First sentences usually stand alone well
  • Sentences without unresolved pronouns
  • Classifier trained on OntoNotes: <10% error rate

Baseline ordering module (chronological) is not obviously worse than anything fancier

---

Results [G & F, 2009]

- Overall Quality
  - Rating scale: 1-10
  - Humans in [8.3, 9.3]
- Pyramid
  - Rating scale: 0-1
  - Humans in [0.62, 0.77]
- Linguistic Quality
  - Rating scale: 1-10
  - Humans in [8.5, 9.3]
- ROUGE-2
  - Rating scale: 0-1
  - Humans in [0.11, 0.15]

- 52 submissions
- 27 teams
- 44 topics
- 10 input docs
- 100 word summaries
  
  Gillick & Favre
Error Breakdown?

[Gillick and Favre, 2008]

Sentence extraction is limiting

... and boring!

But abstractive summaries are much harder to generate...

in 25 words?
Huckleberry Finn
By Mark Twain
Ultra-Condensed by David J. Parker

Huckleberry Finn
(Goes rafting. Goes home.)

THE END

Don Quixote
By Cervantes
Ultra-Condensed by Scott Kinides

Don Quixote
Chovvy demands I destroy that evil thing.

Sancho Panza
No, master. It is something ordinary and harmless.

Don Quixote
(falls down)

THE END

http://www.rinkworks.com/bookaminute/