Exploiting Category-Specific Information for Multi-Document Summarization

Jun-Ping Ng Praveen Bysani Ziheng Lin Min-Yen Kan Chew-Lim Tan
Multi-Document Summarization
A Realistic Scenario

2012 US Presidential Election

Hurricane Sandy

iPad Mini Release
Typical Summarizers

2012 US Presidential Election
Typical Summarizers

2012 US Presidential Election

Hurricane Sandy
Typical Summarizers

- 2012 US Presidential Election
- Hurricane Sandy
- iPad Mini Release

Summary
Key Insight

2012 US Presidential Election

Hurricane Sandy

iPad Mini Release

Summarizer

Summary

Summary

Summary
Outline

• Baseline Summarization Pipeline
• Category-Specific Features
• Experiments
Pipeline

Articles → Sentences → Scorer → Ranked Sentences → Summary

Features
Generic Features

- Sentence position
- Sentence length
- Interpolated N-gram Document Frequency
Category-Specific Features

2012 US Presidential Election

Hurricane Sandy

iPad Mini Release
Category-Specific Features

2012 US Presidential Election 2011 UK Election

2012 Indian Assembly Elections

Hurricane Sandy

2012 Mynamar Earthquake

iPad Mini Release

Nokia Lumia 920
Category-Specific Features

**Political News**
- 2012 US Presidential Election
- 2011 UK Election
- 2012 Indian Assembly Elections

**Disasters**
- Hurricane Sandy
- 2012 Myanmar Earthquake

**Technology**
- iPad Mini Release
- Nokia Lumia 920

Tuesday, December 11, 12
Data Set

• Made use of dataset from guided summarization task of TAC 2010/2011

• Sets of articles to be summarized grouped into topics

• Topics are grouped into categories
TAC Categories

- Accidents and Natural Disasters
- Attacks
- Health and Safety
- Endangered Resources
- Investigations and Trials
Justifying Our Intuition

- Words with high log-likelihood ratios are used more frequently inside model summaries

![Bar chart showing density for different TAC categories]
Category Relevance (CRS)

• Importance of a word with respect to a category

\[
CRS_c(s) = \frac{\sum_{w \in s} (\beta \times TLF_c(w) + (1 - \beta) \times DLF_c(w))}{|s|}
\]

Topic Frequency Score

Document Frequency Score
Category KLD (CKLD)

- Measure of divergence of probability distribution of a word in a category, with respect to the whole collection

\[ CKLD_c(s) = \sum_{w \in s} \left( p_c(w) \times \log \frac{p_c(w)}{p_C(w)} \right) \]

- Probability of word in category
- Probability of word in collection
# Top Words

<table>
<thead>
<tr>
<th>CRS</th>
<th>CKLD</th>
</tr>
</thead>
<tbody>
<tr>
<td>official</td>
<td>crane</td>
</tr>
<tr>
<td>people</td>
<td>bridge</td>
</tr>
<tr>
<td>report</td>
<td>construction</td>
</tr>
<tr>
<td>news</td>
<td>java</td>
</tr>
<tr>
<td>accident</td>
<td>people</td>
</tr>
</tbody>
</table>

For articles on “Accidents”
## Summarization Results

- Category-specific features give significant performance enhancement

<table>
<thead>
<tr>
<th>System</th>
<th>ROUGE-2</th>
<th>ROUGE-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic + CRS + CKLD</td>
<td>0.13796</td>
<td>0.16808</td>
</tr>
<tr>
<td>Generic + CRS</td>
<td>0.13702</td>
<td>0.16788</td>
</tr>
<tr>
<td>Generic + CKLD</td>
<td>0.13525</td>
<td>0.16649</td>
</tr>
<tr>
<td>CLASSY</td>
<td>0.12780</td>
<td>0.15812</td>
</tr>
<tr>
<td>POLYCOM</td>
<td>0.12269</td>
<td>0.15974</td>
</tr>
</tbody>
</table>

*ROUGE measures on TAC 2011 testing set*
Less than Perfect Categorization

• In more realistic scenarios, perfect categorization is hard to come by

• Would our proposed CSI still be effective if we don’t have perfect categorization?
## Categorization Experiments

- Perfect clustering information is not essential

<table>
<thead>
<tr>
<th>System</th>
<th>ROUGE-2</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>* Perfect Categories</td>
<td>0.13796</td>
<td>-</td>
</tr>
<tr>
<td>EM Clustering</td>
<td>0.13647</td>
<td>0.154</td>
</tr>
<tr>
<td>X-Means Clustering</td>
<td>0.13546</td>
<td>0.117</td>
</tr>
<tr>
<td>K-Means Clustering</td>
<td>0.13569</td>
<td>0.365</td>
</tr>
</tbody>
</table>

✧ Scores reported for 5 clusters
Analysis

Baseline

− The death toll could rise as thousands are still buried in debris and many are reported missing.

− Therefore, the relevant sectors and personnel should pay attention to disaster prevention.

Generic + CRS + CKLD

+ Chinese authorities did not detect any warning signs ahead of Monday’s earthquake that killed more than 8,600 people.

+ Xinhua said 8,533 people had died in Sichuan alone, citing the local government.
Varying Difficulties

- Some categories are easier to improve on.
- Category-specific features help more with the WHO, WHAT, WHEN, WHERE.
Round Up

• Category-specific information (CSI) can help multi-document summarization
• Effective in improving content selection
• Robust and is useful even when less than ideal categorization results are available
Thank you!

Questions?

You can download the system here:
http://wing.comp.nus.edu.sg/downloads/swing/