Query Specific Summarization

Information Retrieval Track
Group 1

Ashish Gupta, 3rd year UG, IIT Kanpur
Ankit Kumar, 3rd year UG, IIT Kanpur
Kamal Sahni, 3rd year UG, IIT Kanpur
Tarun Kr. Baranwal, 3rd year UG, IIT Kanpur
• Research Problem
• Motivation
• Earlier Works
• What people do and what we are doing
• Extracting Keywords from **semantic networks**
• Re-ranking of existing ranked sentences
• Evaluation and our Results.

**Outline**
• To summarize a single text document in accordance with the **query** specified by the user

• What are the important features of a text summarization system that extracts the words related to query from original documents?
• By just looking at the summary of a document, a user will be able to decide whether the document is of interest to him/her without looking at the whole document.

• Although a number of tools like MS AutoSum, Summarist etc. that are available to facilitate the text summarization process automatically, but the summarized text output is still imprecise or inaccurate.

**Motivation**
<table>
<thead>
<tr>
<th>Author/ Year</th>
<th>Techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luhn, 1958</td>
<td>Word Frequency, Statistical Approach</td>
</tr>
<tr>
<td>Baxendale, 1958</td>
<td>Text Positions</td>
</tr>
<tr>
<td>Edmunson, 1969</td>
<td>Cue Words and Heading</td>
</tr>
<tr>
<td>Miller, 1995</td>
<td>WordNet Lexical Terms</td>
</tr>
<tr>
<td>Lin and Hovy, 1997</td>
<td>Sentence Position</td>
</tr>
<tr>
<td>Marcu, 1998</td>
<td>Rhetorical Structural Theory</td>
</tr>
<tr>
<td>Daume &amp; Marcu, 2002-04</td>
<td>Log Probability &amp; Rhetorical Structural Theory</td>
</tr>
<tr>
<td>Kaustubh Patil, 2007</td>
<td>Graph Theory &amp; Node centrality</td>
</tr>
<tr>
<td>Bawakid, 2008</td>
<td>Semantic similarity between user query &amp; sentences</td>
</tr>
<tr>
<td>Liu, 2009</td>
<td>Correlation Matrix between user queries and sentence</td>
</tr>
</tbody>
</table>

Earlier Work
- **Features Used:**
  - Cue words, Heading words, Sentence Location, TF-IDF significance, Named Entities etc.

- **Sentence weighting to rank sentence:**
  \[ S_i = w_1 F_1 + w_2 F_2 + w_3 F_3 + \ldots + w_n F_n \]

- **Drawback:**
  - There may be sentences which are not statistically expected but more query oriented.

---

**What people do**
Our approach

- Sentence weighting to rank sentence:
  \[ S_i = (w_1 F_1 + w_2 F_2 + w_3 F_3 + \ldots + w_n F_n) + w_s S_{fi} \]
  
  - \( S_{fi} \) are few extra query based keywords that we are introducing in existing model.

- These weights should be calculated simultaneously with regression models that we couldn’t have done in given time line.

- Solution: Re-ranking of existing ranked sentences.
Query: **Efforts** made toward **peace** between **India** and **Pakistan** over **Kashmir** conflict.

- **Peace**
- **India**
- **Conflict**

- Negotiation
- Settlement
- Agreement

- West Bengal
- Northern
- Country

- Resolve
- Political
- War

**Bag of words**

**Extra query based keywords**
Methodology
• Initially **ranked** sentences are taken from an **existing** summarization tool.

• Re-ranking of sentences done as

\[
S_i = \text{initial score} + \text{extra score}.
\]

• Re-ranked sentences were input back into the existing summarization tool to generate summary.

**What is Re-ranking?**
• Number of **word overlap** between sentence and semantic bag of words is found: \( S_f \)

• **Method-1**:  
  All words obtained from semantic network assigned **equal weighting**  
  \[ S_i = \text{(initial score)} + W_s \times (\text{no of words overlap}) \]

• **Method-2**:  
  Summary generated with **Jaccard Indexing**:
  \[ S_i = \text{(initial score)} + W_s \times (\text{no of words overlap})/\text{(no of words in sentence)} \]

**Extra Score ?**
Evaluation with ROUGE Scores

- Measures similarity between our generated summary and gold set summary

- Gold set Summary is available from TAC 2009 dataset

- Evaluation: comparison of Rouge-N Scores of existing query based summarization baseline model of (IIIT-H) and our generated summary
### Rouge Scores w/o Jaccard Indexing

<table>
<thead>
<tr>
<th></th>
<th>ROUGE-1</th>
<th></th>
<th>ROUGE-2</th>
<th></th>
<th>ROUGE-SU4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>P</td>
<td>F</td>
<td>R</td>
<td>P</td>
</tr>
<tr>
<td><strong>Baseline Summaries</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>0.3573</td>
<td>0.3551</td>
<td>0.3560</td>
<td>0.0827</td>
<td>0.0820</td>
</tr>
<tr>
<td>P</td>
<td>0.1236</td>
<td>0.1229</td>
<td>0.1232</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Weight=0.02</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>0.3557</td>
<td>0.3523</td>
<td>0.3539</td>
<td>0.0875</td>
<td>0.0870</td>
</tr>
<tr>
<td>P</td>
<td>0.1238</td>
<td>0.1229</td>
<td>0.1233</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Weight=0.06</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>0.3278</td>
<td>0.3305</td>
<td>0.3290</td>
<td>0.0670</td>
<td>0.0678</td>
</tr>
<tr>
<td>P</td>
<td>0.0603</td>
<td>0.1068</td>
<td>0.1062</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R : Recall  \quad P : Precision  \quad F : F- measure
<table>
<thead>
<tr>
<th>ROUGE-1</th>
<th>ROUGE-2</th>
<th>ROUGE-SU4</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>P</td>
<td>F</td>
</tr>
<tr>
<td>Baseline Summary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3573</td>
<td>0.3551</td>
<td>0.3560</td>
</tr>
<tr>
<td>Weight=0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3577</td>
<td>0.3554</td>
<td>0.3564</td>
</tr>
<tr>
<td>Weight=1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3493</td>
<td>0.3467</td>
<td>0.3479</td>
</tr>
</tbody>
</table>

R : Recall  
P : Precision  
F : F- measure

Rouge Scores with Jaccard Indexing
• These ROUGE scores do not appear to improve the existing system.

• Possible reasons
  • The Gold summary set for this dataset is 100 words only and is also an abstractive one
  • The Gold summary is diverse while our summary is more query focused. Hence, the low ROUGE scores.

• New evaluation technique required.

Conclusion with Rouge Scores
• Manually ranked sentences in binary (i.e. sentence can be relevant or irrelevant) are generated.

• Generated with compress ratio of 5%, 10%, and 20%

• This is used as reference summary sentences and we evaluated how close are our ranked sentences to this reference summary sentences.

• A correlation Score is calculated:
  o If ( score > 0 ) : Our summary is better
  o If ( score < 0 ) : Baseline summary is better
<table>
<thead>
<tr>
<th>Document Id</th>
<th>Compress Ratio = 5%</th>
<th>Compress Ratio = 10%</th>
<th>Compress Ratio = 20%</th>
</tr>
</thead>
<tbody>
<tr>
<td>D0902A</td>
<td>21.1875</td>
<td>14.45161</td>
<td>6.966101</td>
</tr>
<tr>
<td>D0905A</td>
<td>2.84210</td>
<td>2.19512</td>
<td>-0.05405</td>
</tr>
<tr>
<td>D0901A</td>
<td>-1.23076</td>
<td>7.95</td>
<td>13.4705</td>
</tr>
<tr>
<td>D0910B</td>
<td>5.25</td>
<td>4.3333</td>
<td>3.5789</td>
</tr>
<tr>
<td>D0908B</td>
<td>-2.4285</td>
<td>-4.2142</td>
<td>5.2413</td>
</tr>
<tr>
<td>Average</td>
<td>5.124468</td>
<td>6.628846</td>
<td>5.8405502</td>
</tr>
</tbody>
</table>

Correlation score
Future work

- This removes noise introduced in the system by semantic networks
- Expected to improve Re ranking of sentences.

**Future work**
Thanks!!!

Acknowledgement

- Dr. Carolyn Rose
- Dr. Vasudeva Varma
- Elijah Mayfield

Special thanks to..

- Rohit Bharadawaj, MS, IIIT-H
- Sudheer kovelamudi, MS, IIIT-H

Q/A
• Assigning different weights depending upon relation of word in semantic with query word.

\[ S_i = (\text{initial weight}) + \text{Wr}_1 \ast (\text{no}\_\text{of}\_\text{Sfi}\_r1) + \text{Wr}_2 \ast (\text{no}\_\text{of}\_\text{Sfi}\_r1) + \ldots + \text{Wr}_n \ast (\text{no}\_\text{of}\_\text{Sfi}\_r1). \]
• Parsing query sentences, **stemming** and deleting stop words from query sentence to generate additional keywords

• Retrieving **Semantic** network with **nodes** and **relations** for keywords obtained above using online tools viz. Microsoft Research: MNEX

**Extracting Keywords**
• Semantic Networks:
  • Using Microsoft Research: MNEX, the online MindNet explorer:  
    
  • Bag of Keywords:
    
    With step-1\bag_of_words_kashmir.txt
With different weighting...

With weights, \( W_s = 0.1 \), assigned to each word of our bag of words are quite higher, So summary will be **biased** to our bag of words

- This partial generated summary seems more query oriented.

With weights, \( W_s = 0.02 \), assigned to each word of our bag of word, we got summary i.e. more biased toward initial summarization.

- Very much similar to initial one that we got.
General Features used to weight sentences

- Cue words
- Heading words
- Sentence Location
- Sentence Length
- Presence of uppercase words
- TF-IDF significance of sentence
- Named Entities in sentence
- Dates in sentence
- Quotation marks in sentence
- Pronouns in sentence
- Numbers in sentence