Generating a Table of Contents

S.R.K. Branavan, Pawan Deshpande, Regina Barzilay

MIT
Introduction

Current Summarization Methods:

Focus on relatively short articles

   e.g. Newspaper articles ≈ 1,200 words (DUC 2006)

May not be applicable to longer texts

   e.g. Books, lecture transcripts, ...

Our Goal:

Summarize large texts into Tables of Contents.
Automatic Generation of Tables of Contents

Unstructured Document → Structured Document → Table of Contents

Dictionary operations
Direct address table
Computer memory
Worst case running time
Hash table with load factor
Address table
Hash function
Quadratic probing
Double hashing

An actual generated Table of Contents
Our Focus - The Generation Step

Unstructured Document

Structured Document

Table of Contents

Dictionary operations
Direct address table
Computer memory
Worst case running time
Hash table with load factor
Address table
Hash function
Quadratic probing
Double hashing

An actual generated Table of Contents
A Simple Approach
One Title at a Time

Document → Titles

- Hash Tables
- Hash Tables
- Hash Tables
- Collision Resolution by Chaining

Problems:
- Duplication of titles
- Lack of cohesion between titles
A Better Alternative

Our Approach: A Hierarchical Model

- Model relationships between titles
- Integrate a wide variety of local & global constraints
Challenges

Title Generation:
Generated titles should be:
- Representative
- Grammatical
Search space is exponential in title length

Table of Contents Generation:
Tables of Contents have rich internal structure
Search space exponential in number of titles
Model Architecture

Constraints: Word Selection
Grammaticality
Model Architecture

Constraints: Relations among titles

Dictionary operations
  Direct address table
  Computer memory
    Worst case running ...
    Hash table with ...
    Address table
  Hash function
  Quadratic probing
  Double hashing
**Two Linear Models**

Local Model: \((s, t)_i \rightarrow \Phi_{loc}(s, t)\)

Global Model: \((S, T)_j \rightarrow \Phi_{glob}(S, T)\)
Features - Local Model

Selection Features:
- TF*IDF
- Part of speech
- Position of word in section
- Occurrence of word in parent / sibling sections

Contextual Features:
- Bigram + trigram language model scores
  - lexical & part of speech
- Noun phrase relative frequency
Features - Global Model

- Local model rank of title
- Redundancy
  - Title duplication
  - Title similarity
- Parallel construction
  - First word match with parent / sibling titles
  - Last word match with parent / sibling titles
- Local model feature averages
Generating Titles

\[ \Phi_{loc}(s, t) \cdot \alpha_{loc} \]

- Local linear model

\[ \Phi_{loc}(s, t) \]

- Local model feature vector

\[ \alpha_{loc} \]

- Local model parameter vector

All possible titles

<table>
<thead>
<tr>
<th>Candidate Title 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Candidate Title 2</td>
</tr>
<tr>
<td>Candidate Title 3</td>
</tr>
<tr>
<td>Candidate Title 4</td>
</tr>
<tr>
<td>Candidate Title 5</td>
</tr>
<tr>
<td>Candidate Title 6</td>
</tr>
<tr>
<td>Candidate Title 7</td>
</tr>
<tr>
<td>Candidate Title 8</td>
</tr>
</tbody>
</table>

Score

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>Score</td>
</tr>
<tr>
<td>0.154</td>
<td>0.102</td>
</tr>
<tr>
<td>0.099</td>
<td>0.098</td>
</tr>
<tr>
<td>0.081</td>
<td>0.080</td>
</tr>
<tr>
<td>0.077</td>
<td></td>
</tr>
</tbody>
</table>
Generating Titles

Candidate partial titles being extended from 3 to 4 words at iteration 4
Generating Tables of Contents

\[ \Phi_{glo\!b}(s, t) \cdot \alpha_{glo\!b} \] - Global linear model

Table of Contents

Candidate titles from local model

- title 1
- title 2
- title 3
- title 4
Training

Local & Global Models:

The Incremental Perceptron Algorithm
(Collins & Roark, 2004)

\[ \alpha = \alpha + \Phi(s, y) - \Phi(s, z) \]

\( \alpha \) - Parameter Vector
\( \Phi \) - Feature Function
\( s \) - Input (Text Segment)
\( y \) - Target Output (Reference Title)
\( z \) - Actual Output (Generated Title)
Training - Local Model

\[ y \rightarrow Q = \alpha + \Phi(s, y) - \frac{1}{|Q|} \sum_{z \in Q} \Phi(s, z) \]

\[ y \rightarrow Q = \alpha + \Phi(s, y) - \frac{1}{|Q|} \sum_{z \in Q} \Phi(s, z) \]
**Training - Global Model**

Reference Table of Contents

Generated Table of Contents

- Reference Table of Contents is made up of the *reference titles*
- Local model title candidates may *not* include the reference
- Then the global model can never produce the reference

Local model title candidates don't include *reference title*!
Training - Global Model

- Need an alternate if reference title is not present
- Find the title most similar to the reference among the candidates

\[
\hat{y} = \text{argmin}_{z \in Q} L_1(z, y)
\]
## Corpus

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Sections</td>
<td>540</td>
</tr>
<tr>
<td>Number of Trees</td>
<td>39</td>
</tr>
<tr>
<td>Tree Depth</td>
<td>4</td>
</tr>
<tr>
<td>Number of Words</td>
<td>269,650</td>
</tr>
<tr>
<td>Avg. Section Length (words)</td>
<td>609.2</td>
</tr>
<tr>
<td>Avg. Title Length (words)</td>
<td>3.64</td>
</tr>
<tr>
<td>Avg. Title Duplicates</td>
<td>21</td>
</tr>
<tr>
<td>Avg. Branching</td>
<td>3.29</td>
</tr>
</tbody>
</table>

Training: 80% of trees  
Testing: 20% of trees
Experimental Set-up

**Data:**
10 randomization of the 540 segments.
80% training, 20% testing

**Automatic Evaluation:**
Rouge

**Human Evaluation:**
Comparative assessment of title quality
- 3 alternative titles - reference, HD, best baseline
- 6 judges
- Total of 498 titles for 166 unique segments ranked
Comparisons

NP - Noun Phrase with Highest TF*IDF

FG - Noisy-channel Generative (Banko et al. 2000)

HG - Hierarchical Generative (FG + hierarchical model)

FD - Flat Discriminative

HD - Hierarchical Discriminative (FD + hierarchical model)
### Results - Rouge Evaluation

<table>
<thead>
<tr>
<th></th>
<th>Rouge 1</th>
<th>Rouge L</th>
<th>Rouge W</th>
<th>Exact Match</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discriminative</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hierarchical (HD)</td>
<td>0.256</td>
<td>0.249</td>
<td>0.216</td>
<td>13.5</td>
</tr>
<tr>
<td>Flat (FD)</td>
<td>0.241</td>
<td>0.234</td>
<td>0.203</td>
<td>13.1</td>
</tr>
<tr>
<td><strong>Generative</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hierarchical (HG)</td>
<td>0.139</td>
<td>0.133</td>
<td>0.117</td>
<td>5.8</td>
</tr>
<tr>
<td>Flat (FG)</td>
<td>0.094</td>
<td>0.090</td>
<td>0.079</td>
<td>4.1</td>
</tr>
<tr>
<td>Noun Phrase</td>
<td>0.168</td>
<td>0.168</td>
<td>0.157</td>
<td>6.3</td>
</tr>
</tbody>
</table>

Improvement given by **HD** is significant on the Sign test at \( p \leq 0.03 \)
Results - Human Evaluation

Overall pairwise comparisons of the judges' rankings.

<table>
<thead>
<tr>
<th></th>
<th>Better</th>
<th>Worse</th>
<th>Equal</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD vs FD</td>
<td>68</td>
<td>32</td>
<td>49</td>
</tr>
<tr>
<td>Reference vs HD</td>
<td>115</td>
<td>13</td>
<td>22</td>
</tr>
<tr>
<td>Reference vs FD</td>
<td>123</td>
<td>7</td>
<td>20</td>
</tr>
</tbody>
</table>

Improvement given by **HD** is significant on the Sign test at $p \leq 0.0002$
Related Work

Title Generation:
- Dorr et al. (2003)
- Jin & Hauptmann (2001)
- Banko et al. (2000)

Domain Specific Summarization:
- Teufel & Moens (2002)
- Elhadad & McKeown (2001)

Domain Independent Summarization:
- Angheluta et al. (2002)
- Boguraev & Neff (2000)
Conclusions

- Tables of contents can be automatically generated
- Global constraints significantly improve title quality
- Human & automatic evaluations confirm the benefits of joint tree learning

Future Work

Table of contents generation on:
- Automatically hierarchically segmented text
- Lecture transcripts

Code & feature vectors at:
http://people.csail.mit.edu/branavan/code/toc