Surveytree
Automatic Generation of Survey Structures for NLP and AI topics

Prawat Trairatvorakul, Alexander Fabbri, Dragomir Radev

Yale
Why is structure important?

Structure increases readability

Consider:
(1) Definition, Task, Evaluation, Results
(2) Results, Task, Definition, Evaluation

Current systems
- Domain-specific templates by clustering (Sauper and Barzilay, 2009)
- HMM of topic ordering (Jha et al., 2015)

Surveys cover many subtopics
Each with their own subtopics
Eg. Automatic Summarization
   -> Evaluation
      -> ROUGE
      -> BLEU
      -> DUC Quality Questions
Approach

Convert document to tree
Nodes are section headings with subheadings as children

Extract features from heading
Adaptation of tf-idf
  - tf: ancestor headings taken into account
  - idf: word frequencies from ACL Anthology

Combine trees into single tree
Nodes merged recursively if similar
Other nodes are simply added
Dataset

Five topics considered
- Linear Algebra
- Statistics and Probability
- Sentiment Analysis
- Automatic Summarization
- Dependency Parsing

4 - 6 input documents from AAN each

Document Frequencies:
Calculated from papers in ACL Anthology
Evaluation & Results

(1) Wikipedia Term Coverage

(2) Human Evaluation of Node Merges

(3) Tree Properties

(4) Inspection of Output
Wikipedia Term Coverage

![Graph showing term coverage for different sections of Wikipedia articles](image)

- **Percentage**
- **Terms in Generated surveytree present in Wikipedia Article**
- **Terms in Generated surveytree present in Generated surveytree**

Sections:
- Automatic Summarization
- Linear Algebra
- Dependency Parsing
- Sentiment Analysis

Values:
- Automatic Summarization: 32.5%
- Linear Algebra: 60.4%
- Dependency Parsing: 32.3%
- Sentiment Analysis: 30.9%
Human Evaluations of Node Merges

Figure 3: Evaluation of Heading Merges for Automatic Summarization
Human Evaluations of Node Merges

Thanks Alex, Michi, and Ryan!

<table>
<thead>
<tr>
<th>Topic</th>
<th>Krippendorff’s $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatic Summarization</td>
<td>0.73</td>
</tr>
<tr>
<td>Linear Algebra</td>
<td>0.54</td>
</tr>
<tr>
<td>Statistics &amp; Probability</td>
<td>0.48</td>
</tr>
<tr>
<td>Dependency Parsing</td>
<td>0.38</td>
</tr>
<tr>
<td>Sentiment Analysis</td>
<td>0.39</td>
</tr>
</tbody>
</table>

Table 2: Inter-Evaluator Agreement

Figure 3: Evaluation of Heading Merges for Automatic Summarization
Human Evaluations of Node Merges

Figure 2: Mean Merging Agreement Score Distributions for each topic
# Tree Properties

<table>
<thead>
<tr>
<th></th>
<th>surveytree</th>
<th>dt 1</th>
<th>dt 2</th>
<th>dt 3</th>
<th>dt 4</th>
<th>dt 5</th>
<th>sum</th>
<th>mean</th>
</tr>
</thead>
<tbody>
<tr>
<td># Nodes</td>
<td>123</td>
<td>90</td>
<td>73</td>
<td>56</td>
<td>51</td>
<td>28</td>
<td>298</td>
<td>59.6</td>
</tr>
<tr>
<td># Nodes in surveytree</td>
<td>123</td>
<td>90</td>
<td>41</td>
<td>49</td>
<td>41</td>
<td>18</td>
<td>239</td>
<td>47.8</td>
</tr>
<tr>
<td>Max Depth</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>N/A</td>
<td>2.2</td>
</tr>
<tr>
<td>Mean Depth</td>
<td>2.8</td>
<td>2.7</td>
<td>2.0</td>
<td>0.8</td>
<td>2.0</td>
<td>2.0</td>
<td>N/A</td>
<td>2.2</td>
</tr>
<tr>
<td>Max # Children</td>
<td>19</td>
<td>11</td>
<td>12</td>
<td>25</td>
<td>8</td>
<td>13</td>
<td>N/A</td>
<td>1.9</td>
</tr>
<tr>
<td>Mean # Children</td>
<td>3.9</td>
<td>3.6</td>
<td>6.5</td>
<td>3.2</td>
<td>5.6</td>
<td>5.4</td>
<td>N/A</td>
<td>4.9</td>
</tr>
</tbody>
</table>

Table 3: Tree Structure Properties for Linear Algebra
Selections from Generated Survey Trees

Figure 4: Selection of nodes from the generated *surveytree* of Sentiment Analysis

Figure 5: Selection of nodes from the generated *surveytree* of Statistics and Probability
Conclusion & Future Work

- This approach has potential
- Heading comparison works fairly well
- Also has a lot of room to improve
- Explore possibility of using vector magnitudes
- We need more reliably parsed input documents