Surveytree: Automatic Generation of Survey Structures for NLP and AI topics

Prawat Trairatvorakul,[1] Alexander Fabbri,[1] and Dragomir Radev PhD[1]

[1]Department of Computer Science, Yale University

LILY Lab

Introduction

Surveys are helpful in quick knowledge acquisition in a given topic. However, in rapidly advancing and specialized fields like Natural Language Processing (NLP) and Artificial Intelligence (AI), surveys quickly become outdated or simply do not exist. Currently, automatically generated surveys in these two domains are not yet reliable enough for general use. One aspect of automatically generated surveys that has room to improve on is its structure. Current systems that attempt this typically disregard the hierarchical structure within documents and instead focus on a one dimensional ordering of extracted text.

In this paper, we propose a method where the hierarchical structures of input documents are parsed into trees which are then combined to generate a hierarchical structure for a survey. Initial tests of the system yield promising results that show both potential of using hierarchical structure in survey generation and that certain aspects of the system still have room to improve.

Materials and Methods

Five topics were selected from the AAN TutorialBank corpus. For each topic, three to six documents were selected as inputs to the system.

The hierarchical structure (headings and subheadings) were extracted into a tree and each node was converted into a vector representation using modified tf-idf. Due to the short length of headings, the tf portion also takes ancestor nodes of a given heading into account. The idf portion was trained on papers from the ACL Anthology in order to include relevant terms.

Each document tree was then combined to form a final tree that forms the structure of a survey. Cosine similarity was used to decide node merges.

The final generated structure can be displayed using a pre-order depth-first traversal of the generated tree

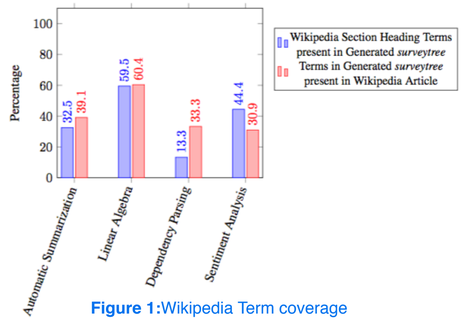

Figure 1: Wikipedia Term coverage

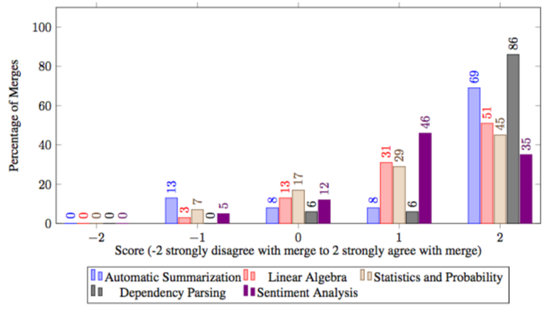

Figure 2: Mean Merging Agreement Score Distributions for each topic

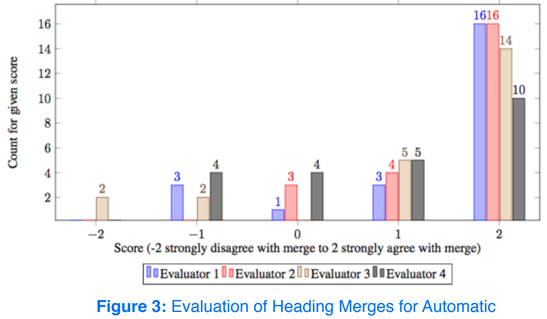

Figure 3: Evaluation of Heading Merges for Automatic Summarization by Different Evaluators

	surveytree	*dt* 1	*dt* 2	*dt* 3	*dt* 4	*dt* 5	sum	mean
# Nodes	123	90	73	56	51	28	298	59.6
# Nodes in *surveytree*	123	90	41	49	41	18	239	47.8
Max Depth	4	3	2	2	2	2	N/A	2.2
Mean Depth	2.8	2.7	2.0	0.8	2.0	2.0	N/A	1.9
Max # Children	19	11	12	25	8	13	N/A	13.8
Mean # Children	3.9	3.6	6.5	3.2	5.6	5.4	N/A	4.9

Table 1: Tree Structure Properties for Linear Algebra

```
(1) Sentiment Analysis and Opinion Mining
├─ (1) Sentiment Analysis:  A Fascinating Problem
│  └─ ...
├─ (2) The Problem of Sentiment Analysis
│  └─ ...
├─ (1) Document Sentiment Classification
│  └─ ...
├─ (2) Sentence Subjectivity and Sentiment Classification
│  ├─ (3) Subjectivity Classification
│  │  └─ (3) Classification Based on Supervised Learning
│  ├─ (1) Dealing with Conditional Sentences
│  ├─ (1) Dealing with Sarcastic Sentences
│  └─ ...
├─ ...
└─ (1) Concluding Remarks
```

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

```
(1) Introduction to Bayesian Statistics
├─ ...
├─ (2) Hypothesis Testing and Model Selection
│  ├─ (1) An Example Hypothesis Test
│  ├─ (1) The "Testing" Prior
│  ├─ (1) Some Terminology
│  ├─ (1) Hypothesis Testing and the Marginal Likelihood
│  ├─ (3) The hypothesis-testing framework
│  └─ ...
├─ ...
├─ (3) Probability
│  ├─ (1) Random Variables
│  │  └─ (3) Discrete random variables
│  ├─ (5) Continuous Random variables
│  └─ ...
└─ ...
```

Figure 5: Selection of nodes from the generated surveytree of Statistics and Probability

Results

Four tests were used to evaluate the output:

(1) Wikipedia Term Coverage
Tests coverage of ideas by checking if terms from headings of Wikipedia article in generated tree, and checks irrelevant information by comparison of terms in survey tree not in body of article. Results shown in fig. 1

(2) Human Evaluation of Node Merges (fig. 2 & fig. 3)
Four evaluators were asked to rate their agreement with node merges. Fig. 2 for comparison between topics, fig. 3 to show inter-evaluator agreement.

(3) Comparison of Generated Tree Properties
Properties of generated tree compared to properties of input trees. Example for Linear Algebra shown in table 1.

(4) Inspection of Output
Outputs were manually inspected. Interesting results that show sections of the tree with headings from different documents shown in fig. 4 and fig. 5.

Conclusion

The initial results are promising and suggest that the general hierarchical approach to generating the structure of surveys may be effective. The term similarity portion appears to work well, while other components will benefit from further testing and development. To aid this, a larger corpus of well-parsed documents in different topics will be helpful. Overall, the current implementation is able to produce a reasonable result with certain sections of the generated tree that show potential for future development.

Acknowledgement

Thank you to Alex Fabbri and Professor Dragomir Radev for advising me. I would also like to thank Michihiro Yasunaga and Ryan Lim for helping to evaluate my results.