Ling 573 Summarization Presentation

Thomas Marsh
University of Washington
ling 573, Spring 2015
sugarork@uw.edu

Michael Roylance
University of Washington
ling 573, Spring 2015
roylance@uw.edu

Brandon Gahler
University of Washington
ling 573, Spring 2015
bjg6@uw.edu

Abstract

TODO: fill out

1 Introduction

TODO: fill out

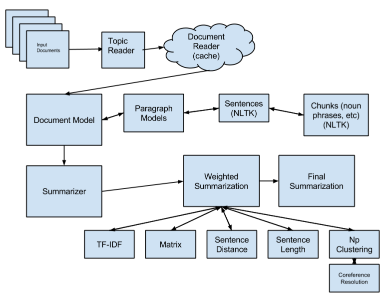
2 System Overview Diagram

TODO: fill out

3 Approach

3.1 System Architecture

Figure 1: System Architecture



3.2 Content Selection

We used a variety of methods initially to help with content selection. First, we the average tf-idf score of every non-stop word for every sentence to help rank the highest sentences.
Next, we implemented a simple graph based metric, in which a dense undirected graph of sentences is constructed with edge weights set to the cosine similarity of the sentences. The most connected sentence is iteratively selected, and the weights of the edges of the previously selected sentences are set to be negative to discourage redundancy.
We also implemented a noun phrase clustering algorithm where we used co-reference resolution to resolve pronouns to their most likely antecedent, and then compared each sentence's noun phrases to each other sentence.

3.3 Information Ordering

Each method had its own ordering criteria.
Tf-idf selection was ordered by the highest sum sentences.
Np-clustering selection was ordered by the number of matches each sentence had to every other sentence.
Simple graph selection was ranked by the most connected sentences.
The highest sentence selection technique was ordered by the longest sentences.
The sentence position technique was ordered by the sentences nearest to the beginning.

3.4 Content Realization

Our realization was simple, the highest ranking sentences were realized into the summary. We used a weighted summarization class to look at every technique, given a defined weight, and selected the top sentences (with a one hundred character limit) to be realized for the summary.

3.5 Best Technique

The best technique we found was tf-idf enhanced with simple graph similarity.

4 Results

Our best score was tf-idf enhanced with matrix similarity:

Rouge Technique	Recall	Precision	F-Score
ROUGE-1	0.54107	0.57388	0.55580
ROUGE-2	0.42822	0.45443	0.43997
ROUGE-3	0.36791	0.39088	0.37819
ROUGE-4	0.31867	0.33882	0.32767

Followed by simple graph similarity by itself:

Rouge Technique	Recall	Precision	F-Score
ROUGE-1	0.55024	0.52418	0.53571
ROUGE-2	0.44809	0.42604	0.43580
ROUGE-3	0.38723	0.36788	0.37643
ROUGE-4	0.33438	0.31742	0.32490

The score for just the np-clustering technique was:

Rouge Technique	Recall	Precision	F-Score
ROUGE-1	0.45691	0.53378	0.49056
ROUGE-2	0.33306	0.39053	0.35813
ROUGE-3	0.28221	0.33196	0.30386
ROUGE-4	0.24758	0.29237	0.26700

The score for just the matrix clustering technique was:

Rouge Technique	Recall	Precision	F-Score
ROUGE-1	0.48228	0.56860	0.52048
ROUGE-2	0.36821	0.43541	0.39787
ROUGE-3	0.31484	0.37348	0.34065
ROUGE-4	0.27465	0.32683	0.29757

4.1 Error Analysis

There are many things to still tweak with both np-clustering and the matrix comparison. For example, we could do better normalization with other co-referents rather than just pronouns. We will be experimenting with these techniques further.

5 Discussion

TODO: fill out

6 Conclusion

Conclusions have been made as can be seen from the following Nenkova, Radev, and Jones (Nenkova et al., 2007) (Jones, 2007) (Radev et al., 2001). We used co-referenced based off ideas from (Cardie and Wagstaff, 1999).

References

C. Cardie and K. Wagstaff. 1999. Noun phrase coreference as clustering. *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, 1001:82–89.

Karen Spärck Jones. 2007. Automatic summarising: The state of the art. *Information Processing Management*, 43(6):1449 – 1481. Text Summarization.

Ani Nenkova, Rebecca Passonneau, and Kathleen McKeown. 2007. The pyramid method: Incorporating human content selection variation in summarization evaluation. *ACM Transactions on Speech and Language Processing (TSLP)*, 4(2):4.

Dragomir R Radev, Sasha Blair-Goldensohn, and Zhu Zhang. 2001. Experiments in single and multi-document summarization using mead. *Ann Arbor*, 1001:48109.