Multi-Document Summarization with Iterative Graph-based Algorithms

Rada Mihalcea and Paul Tarau
Department of Computer Science and Engineering
University of North Texas
Denton, TX, 76203, USA
{rada,tarau}@cs.unt.edu

Keywords: graph algorithms, extractive summarization, automated text analysis

Abstract

In this paper, we show how a meta-summarizer relying on a layered application of graph-based techniques for single-document summarization can be turned into an effective method for multi-document summarization. Through evaluations performed on standard data sets, we show that this method compares favorably with state-of-the-art techniques for multi-document summarization.

1. Introduction

Algorithms for extractive summarization are typically based on methods for sentence extraction, and attempt to identify the set of sentences that are most important for the overall understanding of a given document. In this paper, we present an unsupervised method for extractive summarization relying on iterative graph-based algorithms that exploit the cohesive structure of text. We show how a layered application of this single-document summarization method can result into an efficient multi-document summarization tool.

2. Iterative Graph-based Algorithms for Extractive Summarization

Ranking algorithms, such as Kleinberg's HITS algorithm (Kleinberg 1999) or Google's PageRank (Brin and Page 1998), have been traditionally used in Web-link analysis and social networks. The basic idea implemented by the ranking model is that of "voting" or "recommendation". When one vertex links to another one, it is basically casting a vote for that other vertex. The higher the number of votes that are cast for a vertex, the higher the importance of the vertex.

Let G = (V,E) be a directed graph with the set of vertices V and set of edges E, where E is a subset of V ×V. For a given vertex V_i, let $In(V_i)$ be the set of vertices that point to it (predecessors), and let $Out(V_i)$ be the set of vertices that vertex V_i points to (successors).

PageRank (Brin and Page 1999) is perhaps one of the most popular ranking algorithms, and was designed as a method for Web link analysis.

$$PR(V_i) = (1-d) + d \sum_{V_j \in In(V_i)} \frac{PR(V_j)}{|Out(V_j)|}$$

HITS (Kleinberg 1999) is an iterative algorithm that was designed for ranking Web pages according to their degree of "authority'". For each vertex, HITS produces two sets of scores - an "authority" score, and a "hub" score:

$$HITS_A(V_i) = \sum_{V_j \in In(V_i)} HITS_H(V_j)$$

$$HITS_S(V_i) = \sum_{V_j \in In(V_i)} HITS_A(V_j)$$

For each of these algorithms, starting from arbitrary values assigned to each node in the graph, the computation iterates until convergence below a given threshold is achieved. After running the algorithm, a score is associated with each vertex, which represents the "importance" or "power" of that vertex within the graph.

The ranking algorithm was also adapted to include edge weights, e.g. for PageRank the score is determined using the following formula (a similar change can be applied to the HITS algorithm):

$$PR^W(V_i) = (1-d) + d \sum_{V_j \in In(V_i)} w_{ji} \frac{PR^W(V_j)}{\sum_{V_k \in Out(V_j)} w_{kj}}$$

For the task of single-document extractive summarization, the goal is to rank the sentences in a given text with respect to their importance for the overall understanding of the text. A graph is therefore constructed by adding a vertex for each sentence in the text, and edges between vertices are established using sentence inter-connections, defined using a simple similarity metric based on sentence overlap. The resulting graph is highly connected, with a weight associated with each edge, indicating the strength of the connections between various sentence pairs in the text. The graph can be represented as: (a) simple *undirected* graph; (b) directed weighted graph with the orientation of edges set from a sentence to sentences that follow in the text (*directed forward*); or (c) directed weighted graph with the orientation of edges set from a sentence to previous sentences in the text (*directed backward*). After the ranking algorithm is run on the graph, sentences are sorted in reversed order of their score, and the top ranked sentences are selected for inclusion in the extractive summary

Multi-document summaries for a document cluster are built using a "meta" summarization procedure. First, for each document in the cluster of documents, a single document summary is generated using one of the graph-based ranking algorithms. Next, a "summary of summaries" is produced using the same or a different ranking algorithm.

Unlike single documents – where sentences with highly similar content are very rarely if at all encountered – it is often the case that clusters of multiple documents, all addressing the same or related topics, would contain very similar or even identical sentences. To avoid such pairs of sentences, which may decrease the readability and the amount of information conveyed by a summary, we introduce a maximum threshold on the sentence similarity measure. Consequently, in the graph construction stage, no link is added between sentences whose similarity exceeds this threshold.

3. Evaluation

Experiments are run using the summarization test collection provided in the framework of the Document Understanding Conference (DUC). In particular, we use the data set of 567 news articles made available during the DUC 2002 evaluations (DUC 2002), and the corresponding 100-word summaries generated for each of the 59 document clusters formed on the same data set. This is the multi-document summarization task undertaken by other systems participating in the DUC 2002 document summarization evaluations.

For evaluation, we use the Rouge evaluation toolkit[2], which is a method based on Ngram statistics, found to be highly correlated with human evaluations (Lin and Hovy 2003). The evaluation is done using the Ngram(1,1) setting of Rouge, found to have the highest correlation with human judgments, at a confidence level of 95%.

We evaluate multi-document summaries generated using combinations of the graph-based ranking algorithms that were found to work best in the single document summarization experiments – PageRank^W and HITS^W_A, on undirected or directed backward graphs. Although the single document summaries used in the "meta" summarization process may conceivably be of any size, in this evaluation their length is limited to 100 words.

As mentioned earlier, different graph algorithms can be used for producing the single document summary and the "meta" summary. Table 1 lists the results for multi-document summarization experiments using various combinations of graph algorithms. For comparison, Table 2 lists the results obtained by the top 5 (out of 9) performing systems in the multi-document summarization task at DUC 2002, and a baseline generated by taking the first sentence in each article.

Single doc	"Meta" summarization algorithm			
summarization	PR^W-U	PR^W-DB	HITS^W_A-U	HITS^W_A-DB
PageRank^W-U	**0.3552**	0.3499	0.3456	0.3465
PageRank^W-DB	0.3502	0.3448	0.3519	0.3439
HITS^W_A-U	0.3368	0.3259	0.3212	0.3423
HITS^W_A-DB	**0.3572**	0.3520	0.3462	0.3473

Table 1: Results for multi-document summarization (U = Undirected; DB = Directed Backward)

Top 5 systems (DUC 2002)					
S26	S19	S29	S25	S20	Baseline
0.3578	0.3447	0.3264	0.3056	0.3047	0.2932

Table 2: Results for top 5 DUC 2002 multi-document summarization systems, and baseline.

For multiple document summarization, the best "meta" summarizer is the PageRank^W algorithm applied on undirected graphs, in combination with a single summarization system using the HITS^W_A ranking algorithm, for a performance similar to the one of the best system in the DUC 2002 multi-document summarization task.

4. Conclusion

The graph-based extractive summarization algorithm succeeds in identifying the most important sentences in a text (or collection of texts) based on information exclusively drawn from the text itself. Unlike other supervised systems, which attempt to learn what makes a good summary by training on collections of summaries built for other articles, the graph-based method is fully unsupervised, and relies only on the given texts to derive an extractive summary.

The results obtained during these experiments prove that graph-based ranking algorithms, previously found successful in Web link analysis and social networks, can be turned into a state-of-the-art tool for extractive summarization when applied to graphs extracted from texts. An important aspect of the graph-based extractive summarization method is that it does not require deep linguistic knowledge, nor domain or language specific annotated corpora, which makes it portable to other domains, genres, or languages.

References

S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. Computer Networks, 30(1-7), 1998.

DUC. Document understanding conference 2002, 2002. http://www-nlpir.nist.gov/projects/duc/.

T. Hirao, Y. Sasaki, H. Isozaki, and E. Maeda. NTT's text summarization system for DUC-2002, DUC, 2002.

J. Kleinberg. Authoritative sources in a hyperlinked environment. Journal of the ACM, 46(5):604-632, 1999.

C. Lin and E. Hovy. Automatic evaluation of summaries using n-gram co-occurrence statistics,HLT-NAACL, Edmonton, 2003.

R. Mihalcea and P. Tarau. TextRank - bringing order into texts. EMNLP 2004, Barcelona, Spain, 2004.