

Unsupervised Summarization via Cliques Algorithm

Nesreen Alsharman¹ and Inna Pivkina¹

¹Affiliation not available

May 5, 2020

Abstract

The paper introduces new approaches to generate summaries. Summaries are generated by detecting different topics (clusters of sentences) in a document, building summary sentences for each topic, and adding them to the summary. It allows summaries to reflect different ideas from the document. Different clusters in a document are identified by finding cliques in a sentence similarity graph. Selecting representative sentences from the clusters could be done in two ways. The first way is to use Multi-Sentence Compression method for each cluster to generate a compression sentence that reflects the main idea in the cluster (abstractive cliques method). All generated compression sentences are used to build a final abstractive summary. The second way is to extract one sentence from each cluster that is the most similar to sentences in the cluster. This results in an extractive summarization (extractive cliques method). By picking representative sentence for each cluster we reduce potential redundant information in the summary and decrease the chances of missing some important ideas from the text. The approaches are tested on DUC 2004 (Task 3 and 4) data set which contain news documents. ROUGE Perl toolkit is used to compare automatically produced summaries against a set of reference summaries from the data sets. Results show that both approaches perform better than Lex-Rank. Extractive cliques method performs slightly better than abstractive cliques method.

Hosted file

er_UnsupervisedSummarizationviaCliquesAlgorithm_2.pdf available at <https://authorea.com/users/293931/articles/422070-unsupervised-summarization-via-cliques-algorithm>