

Multi-Doc Summarization COVID-19

Sharon Wang

11/12/2020

Overview

1. Extract sentences that contain the theme.
2. Remove redundant sentences.
3. Perform clustering on current (unpartitioned) summary.
4. Include previous and/or subsequent sentences from the original article that may not contain a time or location, but have the same theme. Will help make the summary more coherent and complete.

Partitioning via Clustering

1. If there is one cluster, no partition is needed. Else, determine what kind of partition that recreates the clustered groups.
2. Then, perform clustering on each partitioned group until clustering makes no effect on the summary, .

Clustering

- When clustering, weigh different words in the summary differently
- Weigh CatE keywords the heaviest
- Use AutoPhrase to do frequent-pattern phrase mining on other words and weigh them as well
- This method of clustering should yield better results, as we are only focusing on grouping the sentences by important words now