Automatic Text Summarization with Multi-objective Optimization (ATS-MOO)

Chihoon Jung · Rituparna Datta ·
Aviv Segev

Abstract Multi-document summarization aims to solve information overload problem by providing condensed summary from a given set of documents on a topic. In this work, we propose a novel multi-objective optimization approach for generic multi-document summarization. The proposed method is based on two conflicting objective functions, namely, coverage and diversity. The generated summary aims to maximize the content coverage of the original documents, while maximizing the diversity between sentences within the summary. We apply k-means clustering to the original document set and use cluster centroids to extract topics for the coverage computation. We sum the cosine similarity of each pair of sentences in the summary to estimate diversity. The similarity is estimated based on the Vector Space Model (VSM) representation of the text snippets with TF-IDF (term frequency - inverse document frequency) term weighting scheme. The advantage of our method comes from using a global optimization technique, so that the problem is considered in a global perspective as compared to a greedy approach that may get stuck on the local optimum. The novelty of the proposed method is in the ability to generate multiple Pareto optimal solutions of candidate summaries on which we can develop a parameter optimization technique to get the most effective summary compared to the human generated summary. We evaluate the proposed method with the Document Understanding Conference (DUC) open benchmark dataset widely used in automatic text summarization evaluation. The results are compared with the human generated gold standard summary using the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) metric. The preliminary experiment on DUC2002 dataset shows that the proposed method successfully generates effective summaries and shows better performance compared to the existing methods.

Keywords Text Summarization · Optimization · Clustering · Text Similarity

Chihoon Jung, Rituparna Datta
Graduate School of Knowledge Service Engineering,
KAIST, 291 Daehak-ro, Yuseong-gu, 305-701, South Korea
E-mail: {chihoon.jung,rdatta}@kaist.ac.kr

Aviv Segev
Graduate School of Knowledge Service Engineering and School of Computing,
KAIST, 291 Daehak-ro, Yuseong-gu, 305-701, South Korea
E-mail: aviv@kaist.edu