Hybrid Genetic Fuzzy System for Text Summarization

Saziya V Saiyed

As online and offline information is growing day by day, finding related information from large amount of text data is an issue. The aim of text summarization is to create summary from the source text while retaining its information content and overall meaning. It is very difficult for human beings to manually summarize large documents of text. Automatic text summarization is useful to make summary from large document. The purpose of automatic text summarization is to present the source text into a shorter version with semantics. Summarised text reduces the reading time and effort. Automatic text summarization methods can be divided into extractive and abstractive. An extractive summarization method is used for selecting important sentences, paragraphs from the original document. It then concatenates all selected sentences into shorter form. An abstractive summarization is used to understanding the main concepts in a given document and then expresses those concepts in clear natural language.

Extractive text summarization includes pre-processing and processing. Prepossessing involves sentence boundary identification, elimination of common words with no semantic. Processing involves actual summarization process. Features influencing the relevance of sentences are decided and calculated and then weights are assigned to these features using weight learning method. Features used are Content word, Title word, Sentence location, Sentence Length, Proper Noun, Upper-Case, Cue-Phrase, Biased Word. Final score of each sentence is determined using Feature-weight equation. Top ranked sentences are selected for final summary.

Proposed research is based on extractive text summarization which uses genetic algorithm and fuzzy logic to create summary. The genetic approach is helpful to maintain a population of candidate solutions to problem being solved. The number of input features is reduced using the genetic algorithm. This will reduce the number of fuzzy sets to be selected substantially due to optimization process. The genetic algorithm will mutates and alters the candidate solutions to provide a better solution. We can represent candidate solution by a chromosome. Fuzzy fitness function can be used to select some of the features to pass to fuzzy system. The list of features which are obtained as an output of genetic algorithm, are passed to fuzzy system consisting of fuzzy rules and fuzzy sets to generate summary of the text.