Text summarization features selection method using pseudo genetic-based model

Abstract :

The features are considered the cornerstone of text summarization. The most important issue is what feature to be considered in a text summarization process. Including all the features in the summarization process may not be considered as an optimal solution. Therefore, other methods need to be deployed. In this paper, random five features used and investigated using a (pseudo) Genetic concept as an optimized trainable features selection mechanism. The Document Understanding Conference (DUC2002) used to train our proposed model; hence the objective of this paper is to learn the weight (importance) of each used feature. For each input document using the genetic concept, the size of the generation is defined and the chromosome dimension (genes) is equal to number of features used5. Each gene is represents a feature and in binary format. A chromosome with high fitness value is selected to be enrolled in the final round. The average of each gene is computed for all best chromosomes and considered the weight of that feature. Our experimental result shows that our proposed model is able performing features selection process.