Solving text clustering problem using a memetic differential evolution algorithm

Hossam M. J. Mustafa , Masri Ayob , Dheeb Albashish and Sawsan Abu-Taleb

The text clustering is considered as one of the most effective text document analysis methods, which is applied to cluster documents as a consequence of the expanded big data and online information. Based on the review of the related work of the text clustering algorithms, these algorithms achieved reasonable clustering results for some datasets, while they failed on a wide variety of benchmark datasets. Furthermore, the performance of these algorithms was not robust due to the inefficient balance between the exploitation and exploration capabilities of the clustering algorithm. Accordingly, this research proposes a Memetic Differential Evolution algorithm (MDETC) to solve the text clustering problem, which aims to address the effect of the hybridization between the differential evolution (DE) mutation strategy with the memetic algorithm (MA). This hybridization intends to enhance the quality of text clustering and improve the exploitation and exploration capabilities of the algorithm. Our experimental results based on six standard text clustering benchmark datasets (i.e. the Laboratory of Computational Intelligence (LABIC)) have shown that the MDETC algorithm outperformed other compared clustering algorithms based on AUC metric, F-measure, and the statistical analysis. Furthermore, the MDETC is compared with the state of art text clustering algorithms and obtained almost the best results for the standard benchmark datasets.

Hossam M. J. Mustafa , Masri Ayob , Dheeb Albashish and Sawsan Abu-Taleb, (2020), Solving text clustering problem using a memetic differential evolution algorithm, PLoS One.