An Improved B-hill Climbing Optimization Technique for Solving the Text Documents Clustering Problem

Laith M Abualigah, Essam Said Hanandeh, Ahamad Tajudin Khader, Mohammed Abdallh Otair, Shishir Kumar Shandilya

Background: Considering the increasing volume of text document information on Internet pages, dealing with such a tremendous amount of knowledge becomes totally complex due to its large size. Text clustering is a common optimization problem used to manage a large amount of text information into a subset of comparable and coherent clusters. Aims: This paper presents a novel local clustering technique, namely, β-hill climbing, to solve the problem of the text document clustering through modeling the β-hill climbing technique for partitioning the similar documents into the same cluster. Methods: The β parameter is the primary innovation in β-hill climbing technique. It has been introduced in order to perform a balance between local and global search. Local search methods are successfully applied to solve the problem of the text document clustering such as; k-medoid and kmean techniques. Results: Experiments were conducted on eight benchmark standard text datasets with different characteristics taken from the Laboratory of Computational Intelligence (LABIC). The results proved that the proposed β-hill climbing achieved better results in comparison with the original hill climbing technique in solving the text clustering problem.