A parallel hybrid krill herd algorithm for feature selection

Laith Abualigah, Bisan Alsalibi, Mohammad Shehab, Mohammad Alshinwan, Ahmad M Khasawneh, Hamzeh Alabool

In this paper, a novel feature selection method is introduced to tackle the problem of high-dimensional features in the text clustering application. Text clustering is a prevailing direction in big text mining; in this manner, documents are grouped into cohesive groups by using neatly selected informative features. Swarm-based optimization techniques have been widely used to select the relevant text features and shown promising results on multi-sized datasets. The performance of traditional optimization algorithms tends to fail miserably when using large-scale datasets. A novel parallel membrane-inspired framework is proposed to enhance the performance of the krill herd algorithm combined with the swap mutation strategy (MHKHA). In which the krill herd algorithm is hybridized the swap mutation strategy and incorporated within the parallel membrane framework. Finally, the k-means technique is employed based on the results of feature selection-based Krill Herd Algorithm to cluster the documents. Seven benchmark datasets of various characterizations are used. The results revealed that the proposed MHKHA produced superior results compared to other optimization methods. This paper presents an alternative method for the text mining community through cohesive and informative features.