Abstract

Automatic clustering of unstructured documents has become an essentially indispensible task, especially when dealing with the increasing electronic documents. Automatic clustering of documents involves document designation to a sub-group based on its content. K-means is one of the most popular unsupervised clustering algorithms, though the quality of its results relies heavily on the number of clusters chosen and the right selection of the initial cluster centroids. Cuckoo search is one of the most recent soft computing intelligent algorithms that can be chosen as an efficient search method in many optimization problems. In this paper, the original Cuckoo Search algorithm is adapted so that it can be applied efficiently to documents clustering problem. Our proposed modification enable Cuckoo search to use dynamic nests so that different values for the number of clusters can be explored, these nests are initialized with different corresponding forgy selection of initial centroids. During the implementation, these dynamic nests are updated using Lévy flight random walk and evaluated to detect the best nest. The proposed work is implemented and compared to the classical K-means clustering algorithm. The purity measure was used to evaluate the performance. Results show the efficiency of the proposed approach.
Adaptation of Cuckoo search for Documents Clustering

References

Adaptation of Cuckoo search for Documents Clustering


Index Terms

Computer Science

Artificial Intelligence

Keywords

Cuckoo search  Documents clustering  K-means