ACM DIGITAL LIBRARY

Daneshyar Network
Amir Kabir University of
Technology

SEARCH

Hybridization of K-Means and Harmony Search Methods for Web Page Clustering

2008 Article

Authors:
Rana Forsati
MohammadReza Meybodi
Mehrdad Mahdavi
AzadehGhari Neiat

Bibliometrics

Tools and Resources

Save to Binder

Export Formats:
BibTeX EndNote ACM Ref

Publisher Site

Share:
|

Tags: algorithms clustering
clustering, data mining,
optimization design
experimentation measurement
performance world wide web

Feedback | Switch to single page view (no tabs)

| Abstract | Authors | References | Cited ... | Index Terms | Publicati... | Reviews | Comments | Table of Contents |

Clustering is currently one of the most crucial techniques for dealing with massive amount of heterogeneous information on the web, which is beyond human being's capacity to digest. Recent studies have shown that the most commonly used partitioning-based clustering algorithm, the K-means algorithm, is more suitable for large datasets. However, the K-means algorithm can generate a local optimal solution. In this paper we present novel harmony search clustering algorithms that deal with documents clustering based on harmony search optimization method. By modeling clustering as an optimization problem, first, we propose a pure harmony search based clustering algorithm that finds near global optimal clusters within a reasonable time. Contrary to the localized searching of the K-means algorithm, the harmony search clustering algorithm performs a globalized search in the entire solution space. Then harmony clustering is integrated with the K-means algorithm in three ways to achieve better clustering. The proposed algorithms improve the K-means algorithm by making it less dependent on the initial parameters such as randomly chosen initial cluster centers, hence more stable. In the experiments we conducted, we applied the proposed algorithms, K-means clustering algorithm on five different document datasets. Experimental results reveal that the proposed algorithms can find better clusters when compared to K-means and the quality of clusters is comparable and converge to the best known optimum faster than it.

Powered by **THE ACM GUIDE TO COMPUTING LITERATURE**