Package 'moc.gapbk'

March 7, 2019

Type Package

Title Multi-Objective Clustering Algorithm Guided by a-Priori
Biological Knowledge

Version 0.1.0

Date 2019-03-03

Author Jorge Parraga-Alava [aut, cre]
(<https://orcid.org/0000-0001-8558-9122>),
Marcio Dorn [aut],
Mario Inostroza-Ponta [aut]

Maintainer Jorge Parraga-Alava <jorge.parraga@usach.cl>

Description Implements the Multi-Objective Clustering Algorithm Guided by a-
Priori Biological Knowledge (MOC-GaPBK) which was proposed by Parraga-
Alava, J. et. al. (2018) <doi:10.1186/s13040-018-0178-4>.

Depends R (>= 3.2.5)

License GPL-2

Encoding UTF-8

Imports stats, amap, mco, nsga2R, reshape2, miscTools, matrixStats,
fields, Rmisc, plyr, foreach, parallel, doParallel, utils,
doSNOW, doMPI

RoxygenNote 6.1.1

Suggests knitr, rmarkdown

LazyData true

NeedsCompilation no

Repository CRAN

Date/Publication 2019-03-07 17:20:03 UTC

R topics documented:

---

moc.gabk   *Perform the Multi-Objective Clustering Algorithm Guided by a-Priori Biological Knowledge (MOC-GaPBK)*

Description

This function receives two distance matrices and it performs the MOC-GaPBK.

Usage

```
moc.gabk(dmatrix1, dmatrix2, num_k, generation = 50, pop_size = 10,
  rat_cross = 0.8, rat_muta = 0.01, tour_size = 2,
  neighborhood = 0.1, local_search = FALSE, cores = 2)
```

Arguments

dmatrix1	A distance matrix. It should have the same dimensions that dmatrix2. It is mandatory.
dmatrix2	A distance matrix. It should have the same dimensions that dmatrix1. It is mandatory.
num_k	The number k of groups represented by medoids in each individual. It is mandatory.
generation	Number of generations to be performed by MOC-GaPBK. By default 50.
pop_size	Size of population. By default 10.
rat_cross	Probability of crossover. By default 0.80.
rat_muta	Probability of mutation. By default 0.01.
tour_size	Size of tournament. By default 2.
neighborhood	Percentage of neighborhood. A real value between 0 and 1. It is computed as neighborhood*pop_size to determine the size of neighborhood. By default 0.10.
local_search	A boolean value indicating whether the local searches procedures (PR and PLS) are computed. By default *FALSE*.
cores	Number of cores to be used to compute the local searches procedures. By default 2.

Details

MOC-GaPBK is a method proposes by Parraga-Alava, J. et. al. 2018. It carries out the discovery of clusters using NSGA-II algorithm along with Path-Relinking (PR) and Pareto Local Search (PLS) as intensification and diversification strategies, respectively. The algorithm uses as objective functions two versions of the Xie-Beni validity index, i.e., a version for each distance matrix (dmatrix1, dmatrix2). More details about this compute can be found in: <https://doi.org/10.1186/s13040-018-0178-4>. MOC-GaPBK yield a set of the best clustering solutions from a multi-objective point of views.

Value

population	The population of medoids including the objective functions values and order by Pareto ranking and crowding distance values.
matrix.solutions	
	A matrix with results of clustering. Each column represents a clustering solution available in Pareto front.
clustering	A list containing named vectors of integers from 1:k representing the cluster to which each object is assigned.

Author(s)

Jorge Parraga-Alava, Marcio Dorn, Mario Inostroza-Ponta

References

J. Parraga-Alava, M. Dorn, M. Inostroza-Ponta (2018). *A multi-objective gene clustering algorithm guided by apriori biological knowledge with intensification and diversification strategies*. BioData Mining. 11(1) 1-16.

K. Deb, A. Pratap, S. Agarwal, T. Meyarivan (2002). *A fast and elitist multiobjective genetic algorithm: NSGA-II*. IEEE Transactions on Evolutionary Computation, 6(2) 182-197.

F. Glover (1997). *Tabu Search and Adaptive Memory Programming - Advances, Applications and Challenges*. Interfaces in Computer Science and Operations Research: Advances in Metaheuristics, Optimization, and Stochastic Modeling Technologies. 1-75.

J. Dubois-Lacoste, M. Lopez-Ibanez, Stutzle, T. (2015). *Anytime Pareto local search*. European Journal of Operational Research, 243(2) 369-385.

Examples

```
##Generates a data matrix of dimension 50X20

library("amap")
library("moc.gapbk")

x <- matrix(runif(100*20, min = -5, max = 10), nrow=50, ncol=20)

##Compute two distance matrices

dmatrix1<- as.matrix(amap::Dist(x, method = "euclidean"))
dmatrix2<- as.matrix(amap::Dist(x, method = "correlation"))

##Performs MOC-GaPBK with 5 cluster

example<-moc.gabk(dmatrix1, dmatrix2, 5)

example$population
example$matrix.solutions
example$clustering
```

Index