

Package ‘moc.gapbk’

May 14, 2026

Type Package

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

Version 0.2.1

Description Implements the Multi-Objective Clustering Algorithm Guided by a-Priori Biological Knowledge (‘MOC-GaPBK’) proposed by Parraga-Alava and others (2018) <[doi:10.1186/s13040-018-0178-4](https://doi.org/10.1186/s13040-018-0178-4)>. The algorithm performs gene clustering using ‘NSGA-II’ as the underlying multi-objective evolutionary engine, together with Path-Relinking and Pareto Local Search as intensification and diversification strategies. Two versions of the Xie-Beni validity index are used as objective functions, one per distance matrix, so that prior biological knowledge can be incorporated through the second matrix.

License GPL-2

Encoding UTF-8

Language en-US

Depends R (>= 3.5.0)

Imports stats, utils, nsga2R, foreach, parallel, doParallel

Suggests amap, testthat (>= 3.0.0), knitr, rmarkdown

URL <https://github.com/jorgeklz/package-moc.gapbk>

BugReports <https://github.com/jorgeklz/package-moc.gapbk/issues>

VignetteBuilder knitr

Config/testthat/edition 3

Config/roxygen2/version 8.0.0

NeedsCompilation no

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

Maintainer Jorge Parraga-Alava <jorge.parraga@utm.edu.ec>

Repository CRAN

Date/Publication 2026-05-14 14:10:10 UTC

Contents

moc.gapbk	2
Index	5

moc.gapbk	<i>Multi-Objective Clustering Guided by a-Priori Biological Knowledge (MOC-GaPBK)</i>
-----------	---

Description

Performs the MOC-GaPBK algorithm proposed by Parraga-Alava and others (2018). It receives two distance matrices and returns a set of non-dominated clustering solutions.

Usage

```

moc.gapbk(
  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
)

moc.gabk(...)

```

Arguments

dmatrix1	A square distance matrix. Must have the same dimensions as dmatrix2.
dmatrix2	A square distance matrix. Must have the same dimensions as dmatrix1. Typically encodes a-priori biological knowledge.
num_k	The number k of clusters represented by medoids in each individual. Must be greater than 1.
generation	Number of generations to be performed. Default 50.
pop_size	Size of the population. Default 10.

<code>rat_cross</code>	Probability of crossover. Default 0.80.
<code>rat_muta</code>	Probability of mutation. Default 0.01.
<code>tour_size</code>	Size of the tournament for parent selection. Default 2.
<code>neighborhood</code>	Percentage of neighborhood used by Pareto Local Search. A real value between 0 and 1. The neighborhood size is computed as <code>neighborhood * num_objects</code> . Default 0.10.
<code>local_search</code>	Logical. If TRUE, Path-Relinking (PR) and Pareto Local Search (PLS) are applied as intensification and diversification strategies. Default FALSE.
<code>cores</code>	Number of cores used by Path-Relinking. Default 2.
<code>...</code>	Arguments passed to <code>moc.gapbk</code> .

Details

MOC-GaPBK couples NSGA-II with Path-Relinking and Pareto Local Search. Two versions of the Xie-Beni validity index are used as objectives, one per distance matrix.

`moc.gabk` (note the single p) is a deprecated alias kept for backward compatibility with versions 0.1.x. New code should call `moc.gapbk` directly.

Value

A named list with three elements:

`population` A data frame containing the final population of medoids together with the values of the two objective functions, the Pareto ranking and the crowding distance, ordered accordingly.

`matrix.solutions` A data frame whose columns are clustering solutions on the Pareto front. Each row corresponds to an object and each cell to its assigned cluster.

`clustering` A list of named integer vectors. Element `i` is the partition produced by the `i`-th solution on the Pareto front.

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. doi:10.1186/s1304001801784.

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*. 1-75.

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

Examples

```
set.seed(1)
x <- matrix(stats::runif(50 * 20, min = -5, max = 10),
            nrow = 50, ncol = 20)

# Two distance matrices from base R; in real applications dmatrix2
# typically encodes a-priori biological knowledge (e.g. GO semantic
# similarity). See vignette("moc-gapbk-intro") for examples using
# amap::Dist() with correlation-based distances.
dmatrix1 <- as.matrix(stats::dist(x, method = "euclidean"))
dmatrix2 <- as.matrix(stats::dist(x, method = "manhattan"))

res <- moc.gapbk(dmatrix1, dmatrix2, num_k = 3,
                generation = 5, pop_size = 6)

head(res$matrix.solutions)
```

Index

`moc.gabk` (`moc.gapbk`), [2](#)

`moc.gapbk`, [2](#), [3](#)