Package 'eaf'

March 5, 2020

Type Package

Title Plots of the Empirical Attainment Function

Version 1.9-1

Description Computation and visualization of the empirical attainment function (EAF) for the analysis of random sets in multi-criterion optimization. M. Lopez-Ibanez, L. Paquete, and T. Stuetzle (2010) <doi:10.1007/978-3-642-02538-9_9>.

Depends R (>= 2.10.0)

Imports modeltools, graphics, grDevices, stats

Suggests testthat

License GPL (>= 2)

BugReports `https://github.com/MLopez-Ibanez/eaf/issues`

URL `http://lopez-ibanez.eu/eaftools`,
`https://github.com/MLopez-Ibanez/eaf`

LazyLoad true

LazyData true

Encoding UTF-8

RoxygenNote 6.1.1

SystemRequirements GNU make

NeedsCompilation yes

Author Manuel López-Ibáñez [aut, cre]
(<https://orcid.org/0000-0001-9974-1295>),
Marco Chiarandini [aut],
Carlos Fonseca [aut],
Luis Paquete [aut],
Thomas Stützle [aut],
Mickaël Binois [ctb]

Maintainer Manuel López-Ibáñez <`manuel.lopez-ibanez@manchester.ac.uk`>

Repository CRAN

Date/Publication 2020-03-05 14:50:06 UTC

R topics documented:

`eaf-package` *Plots of the Empirical Attainment Function*

Description

The empirical attainment function (EAF) describes the probabilistic distribution of the outcomes obtained by a stochastic algorithm in the objective space. This package implements plots of summary attainment surfaces and differences between the first-order EAFs. These plots may be used for exploring the performance of stochastic local search algorithms for biobjective optimization problems and help in identifying certain algorithmic behaviors in a graphical way.

Details

Functions:

`eafdiffplot()`	Empirical attainment function differences
`eafplot()`	Plot the Empirical Attainment Function for two objectives
`read_datasets()`	Read several data.frame sets

Data:

`gcp2x2` Metaheuristics for solving the Graph Vertex Coloring Problem

`HybridGA` Results of Hybrid GA on vanzyl and Richmond water networks

SPEA2minstoptimeRichmond Results of SPEA2 when minimising electrical cost and maximising the minimum idle time of pumps on Richmond water network

Extras are available at `system.file(package="eaf")`:

`extdata`	External data sets (see `read_datasets`)
`scripts/eaf`	EAF command-line program
`scripts/eafplot`	Perl script to generate plots of attainment surfaces
`scripts/eafdiff`	Perl script to generate plots of EAF differences

Author(s)

Maintainer: Manuel López-Ibáñez <`manuel.lopez-ibanez@manchester.ac.uk`>

Contributors: Carlos Fonseca, Luis Paquete, Thomas Stützle, Manuel López-Ibáñez, Marco Chiarandini and Mickaël Binois.

References

V. Grunert da Fonseca, C. M. Fonseca, and A. O. Hall, *Inferential performance assessment of stochastic optimisers and the attainment function*, in Evolutionary Multi-Criterion Optimization. First International Conference, EMO 2001 (E. Zitzler, K. Deb, L. Thiele, C. A. Coello Coello, and D. Corne, eds.), vol. 1993 of Lecture Notes in Computer Science, pp. 213-225, Berlin: Springer, 2001.

V. Grunert da Fonseca and C. M. Fonseca, *The attainment-function approach to stochastic multiobjective optimizer assessment and comparison*. In T. Bartz-Beielstein, M. Chiarandini, L. Paquete, and M. Preuss, editors, Experimental Methods for the Analysis of Optimization Algorithms, pages 103-130, Springer, Berlin, Germany, 2010.

M. López-Ibáñez, L. Paquete, and T. Stützle. *Exploratory Analysis of Stochastic Local Search Algorithms in Biobjective Optimization*. In T. Bartz-Beielstein, M. Chiarandini, L. Paquete, and M. Preuss, editors, Experimental Methods for the Analysis of Optimization Algorithms, pages 209–222. Springer, Berlin, Germany, 2010. doi: 10.1007/978-3-642-02538-9_9

See Also

Useful links:

- `http://lopez-ibanez.eu/eaftools`
- `https://github.com/MLopez-Ibanez/eaf`
- Report bugs at `https://github.com/MLopez-Ibanez/eaf/issues`

Examples

```
data(gcp2x2)
tabucol<-subset(gcp2x2, alg!="TSinN1")
tabucol$alg<-tabucol$alg[drop=TRUE]
eafplot(time+best~run,data=tabucol,subset=tabucol$inst=="DSJC500.5")

eafplot(time+best~run|inst,groups=alg,data=gcp2x2)
eafplot(time+best~run|inst,groups=alg,data=gcp2x2,
percentiles = c(0,50,100), cex = 1.4, lty = c(2,1,2),lwd = c(2,2,2),
       col = c("black","blue","grey50"))
extdata_path <- system.file(package="eaf","extdata")
A1 <- read_datasets(file.path(extdata_path,"ALG_1_dat"))
A2 <- read_datasets(file.path(extdata_path,"ALG_2_dat"))
eafplot(A1, percentiles=c(50))
eafplot(list(A1=A1, A2=A2), percentiles=c(50))
eafdiffplot(A1, A2)
## Save to a PDF file
# dev.copy2pdf(file="eaf.pdf", onefile=TRUE, width=5, height=4)
```

CPFs *Conditional Pareto fronts obtained from Gaussian processes simulations.*

Description

The data has the only goal of providing an example of use of `vorobT()` and `vorobDev()`. It has been obtained by fitting two Gaussian processes on 20 observations of a bi-objective problem, before generating conditional simulation of both GPs at different locations and extracting non-dominated values of coupled simulations.

Usage

```
CPFs
```

Format

A data frame with 2967 observations on the following 3 variables.

`f1` first objective values.

`f2` second objective values.

`set` indices of corresponding conditional Pareto fronts.

Source

M. Binois, D. Ginsbourger and O. Roustant. Quantifying Uncertainty on Pareto Fronts with Gaussian process conditional simulations, *European Journal of Operational Research*, 2015, 243(2), 386-394.

Examples

```
data(CPFs)

res <- vorobT(CPFs, reference = c(2, 200))
eafplot(CPFs[,1:2], sets = CPFs[,3], percentiles = c(0, 20, 40, 60, 80, 100),
       col = gray(seq(0.8, 0.1, length.out = 6)^2), type = "area",
       legend.pos = "bottomleft", extra.points = res$VE, extra.col = "cyan")
```

`eafdiff` *Compute empirical attainment function differences*

Description

Calculate the differences between the empirical attainment functions of two data sets.

Usage

```
eafdiff(x, y, intervals, maximise = c(FALSE, FALSE),
  rectangles = FALSE)
```

Arguments

`x, y`	Data frames corresponding to the input data of left and right sides, respectively. Each data frame has at least three columns, the third one being the set of each point. See also `read_datasets`.
`intervals`	(‘integer(1)‘) The absolute range of the differences $[0, 1]$ is partitioned into the number of intervals provided.
`maximise`	(`logical()` | `logical(1)`) Whether the objectives must be maximised instead of minimised. Either a single logical value that applies to all objectives or a vector of logical values, with one value per objective.
`rectangles`	If TRUE, the output is in the form of rectangles of the same color.

Details

This function calculates the differences between the EAFs of two data sets.

Value

With `rectangle=FALSE`, a `data.frame` containing points where there is a transition in the value of the EAF differences. The last column gives the difference in terms of sets in `x` minus sets in `y` that attain each point (i.e., negative values are differences in favour `y`).

See Also

`read_datasets`, `eafdiffplot`

Examples

```
A1 <- read_datasets(file.path(system.file(package="eaf"), "extdata", "ALG_1_dat"))
A2 <- read_datasets(file.path(system.file(package="eaf"), "extdata", "ALG_2_dat"))
d <- eafdiff(A1, A2)
# This is large
str(d)
```

```
head(d)

d <- eafdiff(A1, A2, rectangles = TRUE)
# This is large
str(d)
head(d)
# FIXME: finish this!
# Small example
X <- read_datasets(text='
1 2
2.5 1

3.5 2

2 3

4 1
')
Y <- read_datasets(text='
2.5 1

3.5 2

2 3

4 1
')
```

`eafdiffplot` *Plot empirical attainment function differences*

Description

Plot the differences between the empirical attainment functions of two data sets as a two-panel plot, where the left side shows the values of the left EAF minus the right EAF and the right side shows the differences in the other direction.

Usage

```
eafdiffplot(data.left, data.right, col = c("#FFFFFF", "#808080",
  "#000000"), intervals = 5, percentiles = c(50), full.eaf = FALSE,
  type = "area", legend.pos = if (full.eaf) "bottomleft" else
  "topright", title.left = deparse(substitute(data.left)),
  title.right = deparse(substitute(data.right)), xlim = NULL,
  ylim = NULL, cex = par("cex"), cex.lab = par("cex.lab"),
  cex.axis = par("cex.axis"), maximise = c(FALSE, FALSE),
  grand.lines = TRUE, sci.notation = FALSE, left.panel.last = NULL,
  right.panel.last = NULL, ...)
```

Arguments

`data.left, data.right`	Data frames corresponding to the input data of left and right sides, respectively. Each data frame has at least three columns, the third one being the set of each point. See also `read_datasets`.
`col`	A character vector of three colors for the magnitude of the differences of 0, 0.5, and 1. Intermediate colors are computed automatically given the value of `intervals`.
`intervals`	An integer or a character vector. The absolute range of the differences [0,1] is partitioned into the number of intervals provided. If an integer is provided, then labels for each interval are computed automatically. If a character vector is provided, its length is taken as the number of intervals.
`percentiles`	The percentiles of the EAF of each side that will be plotted as attainment surfaces. `NA` does not plot any. See `eafplot.default`.
`full.eaf`	Whether to plot the EAF of each side instead of the differences between the EAFs.
`type`	Whether the EAF differences are plotted as points (‘`points`’) or whether to color the areas that have at least a certain value (‘`area`’).
`legend.pos`	The position of the legend. See `legend`. A value of `"none"` hides the legend.
`title.left, title.right`	Title for left and right panels, respectively.
`xlim, ylim, cex, cex.lab, cex.axis`	Graphical parameters, see `plot.default`.
`maximise`	(`logical() | logical(1)`) Whether the objectives must be maximised instead of minimised. Either a single logical value that applies to all objectives or a vector of logical values, with one value per objective.
`grand.lines`	Whether to plot the grand-best and grand-worst attainment surfaces.
`sci.notation`	Generate prettier labels
`left.panel.last, right.panel.last`	An expression to be evaluated after plotting has taken place on each panel (left or right). This can be useful for adding points or text to either panel. Note that this works by lazy evaluation: passing this argument from other `plot` methods may well not work since it may be evaluated too early.
`...`	Other graphical parameters are passed down to `plot.default`.

Details

This function calculates the differences between the EAFs of two data sets, and plots on the left the differences in favour of the left data set, and on the right the differences in favour of the right data set. By default, it also plots the grand best and worst attainment surfaces, that is, the 0% and 100%-attainment surfaces over all data. This two surfaces delimit the area where differences may exist. In addition, it also plots the 50%-attainment surface of each data set.

With `type = "point"`, only the points where there is a change in the value of the EAF difference are plotted. This means that for areas where the EAF differences stays constant, the region will appear in white even if the value of the differences in that region is large. This explains "white holes" surrounded by black points.

With `type = "area"`, the area where the EAF differences has a certain value is plotted. The idea for the algorithm to compute the areas was provided by Carlos M. Fonseca. The implementation uses R polygons, which some PDF viewers may have trouble rendering correctly (See https://cran.r-project.org/doc/FAQ/R-FAQ.html#Why-are-there-unwanted-borders). Plots (should) look correct when printed.

Large differences that appear when using `type = "points"` may seem to dissapear when using `type = "area"`. The explanation is the points size is independent of the axes range, therefore, the plotted points may seem to cover a much larger area than the actual number of points. On the other hand, the areas size is plotted with respect to the objective space, without any extra borders. If the range of an area becomes smaller than one-pixel, it won't be visible. As a consequence, zooming in or out certain regions of the plots does not change the apparent size of the points, whereas it affects considerably the apparent size of the areas.

Value

No return value.

See Also

`read_datasets`, `eafplot`

Examples

```
A1 <- read_datasets(file.path(system.file(package="eaf"), "extdata", "ALG_1_dat"))
A2 <- read_datasets(file.path(system.file(package="eaf"), "extdata", "ALG_2_dat"))
# These take time
eafdiffplot(A1, A2, full.eaf = TRUE)
eafdiffplot(A1, A2, type = "area")
eafdiffplot(A1, A2, type = "point", sci.notation = TRUE)

# A more complex example
a1 <- read_datasets(file.path(system.file(package="eaf"), "extdata", "wrots_l100w10_dat"))
a2 <- read_datasets(file.path(system.file(package="eaf"), "extdata", "wrots_l10w100_dat"))
DIFF <- eafdiffplot(a1, a2, col = c("white", "blue", "red"), intervals = 5,
type = "point",
         title.left = expression("W-RoTS, " * lambda * "=" * 100 * ", " * omega * "=" * 10),
            title.right= expression("W-RoTS, " * lambda * "=" * 10 *
", " * omega * "=" * 100),
            right.panel.last={ abline(a = 0, b = 1, col = "red", lty = "dashed")})
DIFF$right[,3] <- -DIFF$right[,3]

## Save the values to a file.
# write.table(rbind(DIFF$left,DIFF$right),
#             file = "wrots_l100w10_dat-wrots_l10w100_dat-diff.txt",
#             quote = FALSE, row.names = FALSE, col.names = FALSE)
```

`eafplot`	*Plot the Empirical Attainment Function for two objectives generic*

Description

Computes and plots the Empirical Attainment Function, either as attainment surfaces for certain percentiles or as points.

Usage

```
eafplot(x, ...)

## S3 method for class 'formula'
eafplot(formula, data, groups = NULL, subset = NULL,
  ...)

## S3 method for class 'list'
eafplot(x, ...)

## Default S3 method:
eafplot(x, sets = NULL, groups = NULL,
  percentiles = c(0, 50, 100), attsurfs = NULL, xlab = NULL,
  ylab = NULL, xlim = NULL, ylim = NULL, log = "",
  type = "point", col = NULL, lty = c("dashed", "solid", "solid",
  "solid", "dashed"), lwd = 1.75, pch = NA, cex.pch = par("cex"),
  las = par("las"), legend.pos = "topright", legend.txt = NULL,
  extra.points = NULL, extra.legend = NULL, extra.pch = 4:25,
  extra.lwd = 0.5, extra.lty = NA, extra.col = "black",
  maximise = c(FALSE, FALSE), xaxis.side = "below",
  yaxis.side = "left", axes = TRUE, sci.notation = FALSE, ...)
```

Arguments

`x`	Either a matrix of data values, or a data frame, or a list of data frames of exactly three columns.
`...`	Other graphical parameters to `plot.default`.
`formula`	A formula of the type: `time + cost ~ run | instance` will draw `time` on the x-axis and `cost` on the y-axis. If `instance` is present the plot is conditional to the instances.
`data`	Dataframe containing the fields mentioned in the formula and in groups.
`groups`	This may be used to plot profiles of different algorithms on the same plot.
`subset`	(‘integer()‘ | ‘NULL‘) A vector indicating which rows of the data should be used. If left to default `NULL` all data in the data frame are used.

`sets`	([numeric]) Vector indicating which set each point belongs to.
`percentiles`	([numeric]) Vector indicating which percentile should be plot. The default is to plot only the median attainment curve.
`attsurfs`	TODO
`xlab, ylab, xlim, ylim, log, col, lty, lwd, pch, cex.pch, las`	Graphical parameters, see `plot.default`.
`type`	(character(1)) string giving the type of plot desired. The following values are possible, '`points`' and '`area`'.
`legend.pos`	the position of the legend, see `legend`. A value of `"none"` hides the legend.
`legend.txt`	a character or expression vector to appear in the legend. If `NULL`, appropriate labels will be generated.
`extra.points`	A list of matrices or data.frames with two-columns. Each element of the list defines a set of points, or lines if one of the columns is `NA`.
`extra.legend`	A character vector providing labels for the groups of points.
`extra.pch, extra.lwd, extra.lty, extra.col`	Control the graphical aspect of the points. See `points` and `lines`.
`maximise`	(`logical()` | `logical(1)`) Whether the objectives must be maximised instead of minimised. Either a single logical value that applies to all objectives or a vector of logical values, with one value per objective.
`xaxis.side`	On which side that xaxis is drawn. Valid values are "below" and "above". See `axis`.
`yaxis.side`	On which side that yaxis is drawn. Valid values are "left" and "right". See `axis`.
`axes`	A logical value indicating whether both axes should be drawn on the plot.
`sci.notation`	Generate prettier labels

Details

This function can be used to plot random sets of points like those obtained by different runs of biobjective stochastic optimization algorithms. An EAF curve represents the boundary separating points that are known to be attainable (that is, dominated in Pareto sense) in at least a fraction (quantile) of the runs from those that are not. The median EAF represents the curve where the fraction of attainable points is 50%. In single objective optimization the function can be used to plot the profile of solution quality over time of a collection of runs of a stochastic optimizer.

Value

No value is returned.

Methods (by class)

- `formula`: Formula interface
- `list`: List interface for lists of data.frames or matrices
- `default`: Main function

See Also

read_datasets eafdiffplot

Examples

```
data(gcp2x2)
tabucol <- subset(gcp2x2, alg != "TSinN1")
tabucol$alg <- tabucol$alg[drop=TRUE]
eafplot(time + best ~ run, data = tabucol, subset = tabucol$inst=="DSJC500.5")

## Not run:  # These take time
eafplot(time + best ~ run | inst, groups=alg, data=gcp2x2)
eafplot(time + best ~ run | inst, groups=alg, data=gcp2x2,
percentiles=c(0,50,100), cex = 1.4, lty = c(2,1,2), lwd = c(2,2,2),
     col = c("black","blue","grey50"))

A1 <- read_datasets(file.path(system.file(package = "eaf"), "extdata", "ALG_1_dat"))
A2 <- read_datasets(file.path(system.file(package = "eaf"), "extdata", "ALG_2_dat"))
eafplot(A1, percentiles = 50, sci.notation = TRUE)
eafplot(list(A1 = A1, A2 = A2), percentiles = 50)

## Save as a PDF file.
# dev.copy2pdf(file = "eaf.pdf", onefile = TRUE, width = 5, height = 4)

## End(Not run)

## Using extra.points
## Not run:
data(HybridGA)
data(SPEA2relativeVanzyl)
eafplot(SPEA2relativeVanzyl, percentiles = c(25, 50, 75),
        xlab = expression(C[E]), ylab = "Total switches", xlim = c(320, 400),
        extra.points = HybridGA$vanzyl, extra.legend = "Hybrid GA")

data(SPEA2relativeRichmond)
eafplot (SPEA2relativeRichmond, percentiles = c(25, 50, 75),
         xlab = expression(C[E]), ylab = "Total switches",
         xlim = c(90, 140), ylim = c(0, 25),
         extra.points = HybridGA$richmond, extra.lty = "dashed",
         extra.legend = "Hybrid GA")

eafplot (SPEA2relativeRichmond, percentiles = c(25, 50, 75),
         xlab = expression(C[E]), ylab = "Total switches",
         xlim = c(90, 140), ylim = c(0, 25), type = "area",
         extra.points = HybridGA$richmond, extra.lty = "dashed",
         extra.legend = "Hybrid GA", legend.pos = "bottomright")

data(SPEA2minstoptimeRichmond)
SPEA2minstoptimeRichmond[,2] <- SPEA2minstoptimeRichmond[,2] / 60
eafplot (SPEA2minstoptimeRichmond, xlab = expression(C[E]),
         ylab = "Minimum idle time (minutes)", maximise = c(FALSE, TRUE),
         las = 1, log = "y", main = "SPEA2 (Richmond)",
          legend.pos = "bottomright")

## End(Not run)
```

---

eafs *Exact computation of the EAF*

Description

This function computes the EAF given a set of points and a vector `set` that indicates to which set each point belongs.

Usage

```
eafs(points, sets, groups = NULL, percentiles = NULL)
```

Arguments

`points`	Either a matrix or a data frame of numerical values, where each row gives the coordinates of a point.
`sets`	A vector indicating which set each point belongs to.
`groups`	Indicates that the EAF must be computed separately for data belonging to different groups.
`percentiles`	The percentiles of the EAF of each side that will be plotted as attainment surfaces. `NA` does not plot any. See `eafplot.default`.

Value

A data frame (`data.frame`) containing the exact representation of EAF. The last column gives the percentile that corresponds to each point. If groups is not `NULL`, then an additional column indicates to which group the point belongs.

Note

There are several examples of data sets in `file.path(system.file(package="eaf"),"extdata")`.

Author(s)

Manuel López-Ibáñez

See Also

`read_datasets`

Examples

```
eaf.path <- system.file(package="eaf")

x <- read_datasets(file.path(eaf.path, "extdata","example1_dat"))
# Compute full EAF
str(eafs(x[,1:2], x[,3]))

# Compute only best, median and worst
str(eafs(x[,1:2], x[,3], percentiles = c(0, 50, 100)))

x <- read_datasets(file.path(eaf.path,"extdata", "spherical-250-10-3d.txt"))
y <- read_datasets(file.path(eaf.path,"extdata", "uniform-250-10-3d.txt"))
x <- data.frame(x, groups = "spherical")
x <- rbind(x, data.frame(y, groups = "uniform"))
# Compute only median separately for each group
z <- eafs(x[,1:3], sets = x[,4], groups = x[,5], percentiles = 50)
str(z)
# library(plotly)
# plot_ly(z, x = ~X1, y = ~X2, z = ~X3, color = ~groups,
#         colors = c('#BF382A', '#0C4B8E')) %>% add_markers()
```

`epsilon_additive` *Epsilon metric*

Description

Computes the epsilon metric, either additive or multiplicative.

Usage

```
epsilon_additive(data, reference, maximise = FALSE)

epsilon_mult(data, reference, maximise = FALSE)
```

Arguments

`data`	(`matrix` | `data.frame`) Matrix or data frame of numerical values, where each row gives the coordinates of a point.
`reference`	(`matrix` | `data.frame`) Reference set as a matrix or data.frame of numerical values.
`maximise`	(`logical()` | `logical(1)`) Whether the objectives must be maximised instead of minimised. Either a single logical value that applies to all objectives or a vector of logical values, with one value per objective.

Details

Given objective vectors a and b, epsilon(a,b) is computed in the case of minimization as a/b for the multiplicative variant (respectively, a - b for the additive variant), whereas in the case of maximization it is computed as b/a for the multiplicative variant (respectively, b - a for the additive variant). This allows computing a single value for mixed optimization problems, where some objectives are to be maximized while others are to be minimized. Moreover, a lower value corresponds to a better approximation set, independently of the type of problem (minimization, maximization or mixed). However, the meaning of the value is different for each objective type. For example, imagine that f1 is to be minimized and f2 is to be maximized, and the multiplicative epsilon computed here for epsilon(A,B) = 3. This means that A needs to be multiplied by 1/3 for all f1 values and by 3 for all f2 values in order to weakly dominate B.

This also means that the computation of the multiplicative version for negative values doesn't make sense.

Value

A single numerical value.

Author(s)

Manuel López-Ibáñez

References

E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. Grunert da Fonseca. Performance Assessment of Multiobjective Optimizers: an Analysis and Review. IEEE Transactions on Evolutionary Computation, 7(2):117-132, 2003.

Examples

```
extdata_path <- system.file(package="eaf","extdata")
path.A1 <- file.path(extdata_path, "ALG_1_dat")
path.A2 <- file.path(extdata_path, "ALG_2_dat")
A1 <- read_datasets(path.A1)[,1:2]
A2 <- read_datasets(path.A2)[,1:2]
ref <- filter_dominated(rbind(A1, A2))
epsilon_additive(A1, ref)
epsilon_additive(A2, ref)

# Multiplicative version of epsilon metric
ref <- filter_dominated(rbind(A1, A2))
epsilon_mult(A1, ref)
epsilon_mult(A2, ref)
```

gcp2x2 *Metaheuristics for solving the Graph Vertex Coloring Problem*

Description

Two metaheuristic algorithms, TabuCol (Hertz et al., 1987) and simulated annealing (Johnson et al., 1991), to find a good approximation of the chromatic number of two random graphs. The data here has the only goal of providing an example of use of eafplot for comparing algorithm performance with respect to both time and quality when modelled as two objectives in trade off.

Usage

```
gcp2x2
```

Format

A data frame with 3133 observations on the following 6 variables.

`alg` a factor with levels `SAKempeFI` and `TSinN1`

`inst` a factor with levels `DSJC500.5` and `DSJC500.9`. Instances are taken from the DIMACS repository.

`run` a numeric vector indicating the run to which the observation belong.

`best` a numeric vector indicating the best solution in number of colors found in the corresponding run up to that time.

`time` a numeric vector indicating the time since the beginning of the run for each observation. A rescaling is applied.

`titer` a numeric vector indicating iteration number corresponding to the observations.

Details

Each algorithm was run 10 times per graph registering the time and iteration number at which a new best solution was found. A time limit corresponding to 500*10^5 total iterations of TabuCol was imposed. The time was then normalized on a scale from 0 to 1 to make it instance independent.

Source

M. Chiarandini (2005). Stochastic local search methods for highly constrained combinatorial optimisation problems. Ph.D. thesis, Computer Science Department, Darmstadt University of Technology, Darmstadt, Germany. page 138.

References

A. Hertz and D. de Werra. Using Tabu Search Techniques for Graph Coloring. Computing, 1987, 39(4), 345-351.

D.S. Johnson, C.R. Aragon, L.A. McGeoch and C. Schevon. Optimization by Simulated Annealing: An Experimental Evaluation; Part II, Graph Coloring and Number Partitioning. Operations Research, 1991, 39(3), 378-406

Examples

```
data(gcp2x2)
```

---

hv_contributions — *Hypervolume contribution of a set of points*

Description

Computes the hypervolume contribution of each point given a set of points with respect to a given reference point assuming minimization of all objectives. Dominated points have zero contribution. Duplicated points have zero contribution even if not dominated, because removing one of them does not change the hypervolume dominated by the remaining set.

Usage

```
hv_contributions(data, reference, maximise = FALSE)
```

Arguments

`data`	`(matrix | data.frame)` Matrix or data frame of numerical values, where each row gives the coordinates of a point.
`reference`	`(numeric())` Reference point as a vector of numerical values.
`maximise`	`(logical()	logical(1))` Whether the objectives must be maximised instead of minimised. Either a single logical value that applies to all objectives or a vector of logical values, with one value per objective.

Value

([numeric]) A numerical vector

Author(s)

Manuel López-Ibáñez

References

C. M. Fonseca, L. Paquete, and M. López-Ibáñez. An improved dimension-sweep algorithm for the hypervolume indicator. In IEEE Congress on Evolutionary Computation, pages 1157-1163, Vancouver, Canada, July 2006.

Nicola Beume, Carlos M. Fonseca, Manuel López-Ibáñez, Luís Paquete, and J. Vahrenhold. On the complexity of computing the hypervolume indicator. IEEE Transactions on Evolutionary Computation, 13(5):1075-1082, 2009.

See Also

`hypervolume`

Examples

```
data(SPEA2minstoptimeRichmond)
# The second objective must be maximized
# We calculate the hypervolume contribution of each point of the union of all sets.
hv_contributions(SPEA2minstoptimeRichmond[, 1:2], reference = c(250, 0),
            maximise = c(FALSE, TRUE))

# Duplicated points show zero contribution above, even if not
# dominated. However, filter_dominated removes all duplicates except
# one. Hence, there are more points below with nonzero contribution.
hv_contributions(filter_dominated(SPEA2minstoptimeRichmond[, 1:2], maximise = c(FALSE, TRUE)),
                 reference = c(250, 0), maximise = c(FALSE, TRUE))
```

`HybridGA` *Results of Hybrid GA on vanzyl and Richmond water networks*

Description

The data has the only goal of providing an example of use of eafplot.

Usage

```
HybridGA
```

Format

A list with two data frames, each of them with three columns, as produced by `read_datasets()`.

`$vanzyl` data frame of results on vanzyl network

`$richmond` data frame of results on Richmond network. The second column is filled with `NA`

Source

Manuel López-Ibáñez. **Operational Optimisation of Water Distribution Networks**. PhD thesis, School of Engineering and the Built Environment, Edinburgh Napier University, UK, 2009.

Examples

```
data(HybridGA)
print(HybridGA$vanzyl)
print(HybridGA$richmond)
```

`hypervolume` *Hypervolume metric*

Description

Computes the hypervolume metric with respect to a given reference point assuming minimization of all objectives.

Usage

```
hypervolume(data, reference, maximise = FALSE)
```

Arguments

`data`	`(matrix | data.frame)` Matrix or data frame of numerical values, where each row gives the coordinates of a point.
`reference`	`(numeric())` Reference point as a vector of numerical values.
`maximise`	`(logical()	logical(1))` Whether the objectives must be maximised instead of minimised. Either a single logical value that applies to all objectives or a vector of logical values, with one value per objective.

Details

The algorithm has $O(n^{d-2} \log n)$ time and linear space complexity in the worst-case, but experimental results show that the pruning techniques used may reduce the time complexity even further.

Value

A single numerical value.

Author(s)

Manuel López-Ibáñez

References

C. M. Fonseca, L. Paquete, and M. López-Ibáñez. An improved dimension-sweep algorithm for the hypervolume indicator. In IEEE Congress on Evolutionary Computation, pages 1157-1163, Vancouver, Canada, July 2006.

Nicola Beume, Carlos M. Fonseca, Manuel López-Ibáñez, Luís Paquete, and J. Vahrenhold. On the complexity of computing the hypervolume indicator. IEEE Transactions on Evolutionary Computation, 13(5):1075-1082, 2009.

Examples

```
data(SPEA2minstoptimeRichmond)
# The second objective must be maximized
# We calculate the hypervolume of the union of all sets.
hypervolume(SPEA2minstoptimeRichmond[, 1:2], reference = c(250, 0),
            maximise = c(FALSE, TRUE))
```

---

igd *Inverted Generational Distance (IGD and IGD+)*

Description

Computes the inverted generational distance (IGD and IGD+)

Usage

```
igd(data, reference, maximise = FALSE)

igd_plus(data, reference, maximise = FALSE)
```

Arguments

`data`	(`matrix` | `data.frame`) Matrix or data frame of numerical values, where each row gives the coordinates of a point.
`reference`	(`matrix` | `data.frame`) Reference set as a matrix or data.frame of numerical values.
`maximise`	(`logical()` | `logical(1)`) Whether the objectives must be maximised instead of minimised. Either a single logical value that applies to all objectives or a vector of logical values, with one value per objective.

Details

The generational distance (GD) of a set A is defined as the distance between each point $a \in A$ and the closest point r in a reference set R, averaged over the size of A. Formally,

$$GD(A, R) = \frac{1}{|A|} \left(\sum_{a \in A} \min_{r \in R} d(a, r)^p \right)^{\frac{1}{p}}$$

where:

$$d(a, r) = \sqrt{\sum_{k=1}^{M} (a_k - r_k)^2}$$

The inverted generational distance (IGD) is calculated as $IGD(A,R) = GD(R,A)$ with $p = 1$.

GD was first proposed by Van Veldhuizen and Lamont (1997)in [1] with p=2. IGD seems to have been mentioned first by Coello Coello & Reyes-Sierra (2004), however, some people also used the name D-metric for the same thing with p=1 and later papers have often used IGD/GD with p=1.

The modified inverted generational distanced (IGD+) was proposed by Ishibuchi et all. (2015) to ensure that IGD+ is weakly Pareto compliant, similarly to unary epsilon. It averages over $|R|$ within the exponent $1/p$ and modifies the distance measure as:

$$d^+(r,a) = \sqrt{\sum_{k=1}^{M}(\max\{r_k - a_k, 0\})^2}$$

See Bezerra et al. (2017) for a comparison.

Value

A single numerical value.

Author(s)

Manuel López-Ibáñez

References

D. A. Van Veldhuizen and G. B. Lamont. Evolutionary Computation and Convergence to a Pareto Front. In J. R. Koza, editor, Late Breaking Papers at the Genetic Programming 1998 Conference, pages 221–228, Stanford University, California, July 1998. Stanford University Bookstore. Keywords: generational distance.

Coello Coello, C.A., Reyes-Sierra, M.: A study of the parallelization of a coevolutionary multi-objective evolutionary algorithm. In: Monroy, R., et al. (eds.) Proceedings of MICAI, LNAI, vol. 2972, pp. 688–697. Springer, Heidelberg, Germany (2004).

Q. Zhang and H. Li. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. IEEE Transactions on Evolutionary Computation, 11(6):712–731, 2007. doi:10.1109/TEVC.2007.892759.

Schutze, O., Esquivel, X., Lara, A., Coello Coello, C.A.: Using the averaged Hausdorff distance as a performance measure in evolutionary multiobjective optimization. IEEE Trans. Evol. Comput. 16(4), 504–522 (2012)

H. Ishibuchi, H. Masuda, Y. Tanigaki, and Y. Nojima. Modified Distance Calculation in Generational Distance and Inverted Generational Distance. In A. Gaspar-Cunha, C. H. Antunes, and C. A. Coello Coello, editors, Evolutionary Multi-criterion Optimization, EMO 2015 Part I, volume 9018 of Lecture Notes in Computer Science, pages 110–125. Springer, Heidelberg, Germany, 2015.

Leonardo C. T. Bezerra, Manuel López-Ibáñez, and Thomas Stützle. An empirical assessment of the properties of inverted generational distance indicators on multi- and many-objective optimization. In H. Trautmann, G. Rudolph, K. Klamroth, O. Schütze, M. M. Wiecek, Y. Jin, and C. Grimme, editors, Evolutionary Multi-criterion Optimization, EMO 2017, Lecture Notes in Computer Science, pages 31–45. Springer, 2017.

Examples

```
extdata_path <- system.file(package="eaf","extdata")
path.A1 <- file.path(extdata_path, "ALG_1_dat")
path.A2 <- file.path(extdata_path, "ALG_2_dat")
A1 <- read_datasets(path.A1)[,1:2]
A2 <- read_datasets(path.A2)[,1:2]
ref <- filter_dominated(rbind(A1, A2))
igd(A1, ref)
igd(A2, ref)

# IGD+ (Pareto compliant)
ref <- filter_dominated(rbind(A1, A2))
igd_plus(A1, ref)
igd_plus(A2, ref)
```

`is_nondominated` *Identify, remove and rank dominated points according to Pareto optimality*

Description

Identify nondominated points with `is_nondominated` and remove dominated ones with `filter_dominated`.

`pareto_rank` ranks points according to Pareto-optimality, which is also called nondominated sorting (Deb et al., 2002).

Usage

```
is_nondominated(data, maximise = FALSE, keep_weakly = FALSE)

filter_dominated(data, maximise = FALSE, keep_weakly = FALSE)

pareto_rank(data, maximise = FALSE)
```

Arguments

`data`	(`matrix` | `data.frame`) Matrix or data frame of numerical values, where each row gives the coordinates of a point.
`maximise`	(`logical()` | `logical(1)`) Whether the objectives must be maximised instead of minimised. Either a single logical value that applies to all objectives or a vector of logical values, with one value per objective.
`keep_weakly`	If `FALSE`, return `FALSE` for any duplicates of nondominated points.

Details

pareto_rank is meant to be used like rank(), but it assigns ranks according to Pareto dominance. Duplicated points are kept on the same front. When ncol(data) == 2, the code uses the $O(n \log n)$ algorithm by Jensen (2003).

Value

is_nondominated returns a logical vector of the same length as the number of rows of data, where TRUE means that the point is not dominated by any other point.

filter_dominated returns a matrix or data.frame with only mutually nondominated points.

pareto_rank returns an integer vector of the same length as the number of rows of data, where each value gives the rank of each point.

Author(s)

Manuel López-Ibáñez

References

Deb, K., S. Agrawal, A. Pratap, and T. Meyarivan. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. IEEE Transactions on Evolutionary Computation, 6(2): 182-197, 2002.

M. T. Jensen. Reducing the run-time complexity of multiobjective EAs: The NSGA-II and other algorithms. IEEE Transactions on Evolutionary Computation, 7(5):503–515, 2003.

Examples

```
path_A1 <- file.path(system.file(package="eaf"),"extdata","ALG_1_dat")
set <- read_datasets(path_A1)[,1:2]

is_nondom <- is_nondominated(set)
cat("There are ", sum(is_nondom), " nondominated points\n")

plot(set, col = "blue", type = "p", pch = 20)
ndset <- filter_dominated(set)
points(ndset[order(ndset[,1]),], col = "red", pch = 21)

ranks <- pareto_rank(set)
colors <- colorRampPalette(c("red","yellow","springgreen","royalblue"))(max(ranks))
plot(set, col = colors[ranks], type = "p", pch = 20)
```

normalise *Normalise points*

Description

Normalise points per coordinate to a range, e.g., `c(1,2)`, where the minimum value will correspond to 1 and the maximum to 2. If bounds are given, they are used for the normalisation.

Usage

```
normalise(data, to.range = c(1, 2), lower = NA, upper = NA,
  maximise = FALSE)
```

Arguments

`data`	(`matrix` | `data.frame`) Matrix or data frame of numerical values, where each row gives the coordinates of a point.
`to.range`	Normalise values to this range. If the objective is maximised, it is normalised to `c(to.range[1],to.range[0])` instead.
`lower, upper`	Bounds on the values. If NA, the maximum and minimum values of each coordinate are used.
`maximise`	(`logical()` | `logical(1)`) Whether the objectives must be maximised instead of minimised. Either a single logical value that applies to all objectives or a vector of logical values, with one value per objective.

Value

A numerical matrix

Author(s)

Manuel López-Ibáñez

Examples

```
data(SPEA2minstoptimeRichmond)
# The second objective must be maximized
head(SPEA2minstoptimeRichmond[, 1:2])

head(normalise(SPEA2minstoptimeRichmond[, 1:2], maximise = c(FALSE, TRUE)))

head(normalise(SPEA2minstoptimeRichmond[, 1:2], to.range = c(0,1), maximise = c(FALSE, TRUE)))
```

read_datasets *Read several data sets*

Description

Reads a text file in table format and creates a matrix from it. The file may contain several sets, separated by empty lines. Lines starting by `'#'` are considered comments and treated as empty lines. The function adds an additional column `set` to indicate to which set each row belongs.

Usage

```
read_datasets(file, col_names, text)

read.data.sets(file, col.names)
```

Arguments

`file`	Filename that contains the data. Each row of the table appears as one line of the file. If it does not contain an *absolute* path, the file name is *relative* to the current working directory, `getwd()`. Tilde-expansion is performed where supported.
`col_names, col.names`	Vector of optional names for the variables. The default is to use '`"V"`' followed by the column number.
`text`	('character()') If `file` is not supplied and this is, then data are read from the value of `text` via a text connection. Notice that a literal string can be used to include (small) data sets within R code.

Value

('matrix()') containing a representation of the data in the file. An extra column `set` is added to indicate to which set each row belongs.

Warning

A known limitation is that the input file must use newline characters native to the host system, otherwise they will be, possibly silently, misinterpreted. In GNU/Linux the program `dos2unix` may be used to fix newline characters.

Note

There are several examples of data sets in `file.path(system.file(package="eaf"),"extdata")`.

`read.data.sets` is a deprecated alias. It will be removed in the next major release.

Author(s)

Manuel López-Ibáñez

See Also

read.table, eafplot, eafdiffplot

Examples

```
A1 <- read_datasets(file.path(system.file(package="eaf"),"extdata","ALG_1_dat"))
str(A1)
A2 <- read_datasets(file.path(system.file(package="eaf"),"extdata","ALG_2_dat"))
str(A2)

read_datasets(text="1 2\n3 4\n\n5 6\n7 8\n", col_names=c("obj1", "obj2"))
```

SPEA2minstoptimeRichmond *Results of SPEA2 when minimising electrical cost and maximising the minimum idle time of pumps on Richmond water network.*

Description

The data has the only goal of providing an example of use of eafplot.

Usage

```
SPEA2minstoptimeRichmond
```

Format

A data frame as produced by `read_datasets()`. The second column measures time in seconds and corresponds to a maximisation problem.

Source

Manuel López-Ibáñez. Operational Optimisation of Water Distribution Networks. PhD thesis, School of Engineering and the Built Environment, Edinburgh Napier University, UK, 2009.

Examples

```
data(HybridGA)
data(SPEA2minstoptimeRichmond)
SPEA2minstoptimeRichmond[,2] <- SPEA2minstoptimeRichmond[,2] / 60
eafplot (SPEA2minstoptimeRichmond, xlab = expression(C[E]),
         ylab = "Minimum idle time (minutes)", maximise = c(FALSE, TRUE),
         las = 1, log = "y", legend.pos = "bottomright")
```

`SPEA2relativeRichmond` *Results of SPEA2 with relative time-controlled triggers on Richmond water network.*

Description

The data has the only goal of providing an example of use of eafplot.

Usage

```
SPEA2relativeRichmond
```

Format

A data frame as produced by `read_datasets()`.

Source

Manuel López-Ibáñez. Operational Optimisation of Water Distribution Networks. PhD thesis, School of Engineering and the Built Environment, Edinburgh Napier University, UK, 2009.

Examples

```
data(HybridGA)
data(SPEA2relativeRichmond)
eafplot (SPEA2relativeRichmond, percentiles = c(25, 50, 75),
        xlab = expression(C[E]), ylab = "Total switches",
        xlim = c(90, 140), ylim = c(0, 25),
        extra.points = HybridGA$richmond, extra.lty = "dashed",
        extra.legend = "Hybrid GA")
```

`SPEA2relativeVanzyl` *Results of SPEA2 with relative time-controlled triggers on Vanzyl's water network.*

Description

The data has the only goal of providing an example of use of eafplot.

Usage

```
SPEA2relativeVanzyl
```

Format

A data frame as produced by `read_datasets()`.

Source

Manuel López-Ibáñez. Operational Optimisation of Water Distribution Networks. PhD thesis, School of Engineering and the Built Environment, Edinburgh Napier University, UK, 2009.

Examples

```
data(HybridGA)
data(SPEA2relativeVanzyl)
eafplot(SPEA2relativeVanzyl, percentiles = c(25, 50, 75),
        xlab = expression(C[E]), ylab = "Total switches", xlim = c(320, 400),
        extra.points = HybridGA$vanzyl, extra.legend = "Hybrid GA")
```

`vorobT` *Vorob'ev computations*

Description

Compute Vorob'ev threshold, expectation and deviation. Also, displaying the symmetric deviation function is possible. The symmetric deviation function is the probability for a given target in the objective space to belong to the symmetric difference between the Vorob'ev expectation and a realization of the (random) attained set.

Usage

```
vorobT(x, reference)

vorobDev(x, VE, reference)

symDifPlot(x, VE, threshold, nlevels = 11, ve.col = "blue",
  xlim = NULL, ylim = NULL, legend.pos = "topright",
  main = "Symmetric deviation function", col.fun = function(n)
  gray(seq(0, 0.9, length.out = n)^2))
```

Arguments

`x`	Either a matrix of data values, or a data frame, or a list of data frames of exactly three columns. The third column gives the set (run, sample, ...) identifier.
`reference`	(`numeric()`) Reference point as a vector of numerical values.
`VE, threshold`	Vorob'ev expectation and threshold, e.g., as returned by `vorobT()`.
`nlevels`	number of levels in which is divided the range of the symmetric deviation.
`ve.col`	plotting parameters for the Vorob'ev expectation.
`xlim, ylim, main`	Graphical parameters, see `plot.default()`.
`legend.pos`	the position of the legend, see `legend()`. A value of `"none"` hides the legend.
`col.fun`	function that creates a vector of `n` colors, see `heat.colors()`.

Value

`vorobT` returns a list with elements `threshold`, `VE`, and `avg_hyp` (average hypervolume)

`vorobDev` returns the Vorob'ev deviation.

Author(s)

Mickael Binois

References

M. Binois, D. Ginsbourger and O. Roustant (2015), Quantifying Uncertainty on Pareto Fronts with Gaussian process conditional simulations, European Journal of Operational Research, 243(2), 386-394.

C. Chevalier (2013), Fast uncertainty reduction strategies relying on Gaussian process models, University of Bern, PhD thesis.

I. Molchanov (2005), Theory of random sets, Springer.

Examples

```
data(CPFs)
res <- vorobT(CPFs, reference = c(2, 200))
print(res$threshold)

## Display Vorob'ev expectation and attainment function
# First style
eafplot(CPFs[,1:2], sets = CPFs[,3], percentiles = c(0, 25, 50, 75, 100, res$threshold),
        main = substitute(paste("Empirical attainment function, ",beta,"* = ", a, "%"),
                          list(a = formatC(res$threshold, digits = 2, format = "f"))))

# Second style
eafplot(CPFs[,1:2], sets = CPFs[,3], percentiles = c(0, 20, 40, 60, 80, 100),
        col = gray(seq(0.8, 0.1, length.out = 6)^0.5), type = "area",
        legend.pos = "bottomleft", extra.points = res$VE, extra.col = "cyan",
        extra.legend = "VE", extra.lty = "solid", extra.pch = NA, extra.lwd = 2,
        main = substitute(paste("Empirical attainment function, ",beta,"* = ", a, "%"),
                          list(a = formatC(res$threshold, digits = 2, format = "f"))))

# Now print Vorob'ev deviation
VD <- vorobDev(CPFs, res$VE, reference = c(2, 200))
print(VD)
# Now display the symmetric deviation function.
symDifPlot(CPFs, res$VE, res$threshold, nlevels = 11)
# Levels are adjusted automatically if too large.
symDifPlot(CPFs, res$VE, res$threshold, nlevels = 200, legend.pos = "none")

# Use a different palette.
symDifPlot(CPFs, res$VE, res$threshold, nlevels = 11, col.fun = heat.colors)
```

Index