

Package ‘distExtraction’

October 19, 2023

Title Functions for efficiently subsetting ``dist" objects

Version 0.0.0.9000

Author Minh Long Nguyen <edelweiss611428@gmail.com>

Maintainer Minh Long Nguyen <edelweiss611428@gmail.com>

Description The package provides functions for efficiently subsetting ``dist" objects commonly used in dissimilarity-based clustering. We cannot use the squared brackets operator to extract units in a ``dist" object as we usually do with regular numeric matrices. A simple method to do this involves `as.dist` and `as.matrix` in base R. However, this method is very slow, especially for large distance matrices. The package is based on C++ functions, allowing us to efficiently extract relevant values directly from the ``dist" object with simple syntax.

License GPL (>= 3)

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

LinkingTo Rcpp

Imports Rcpp

NeedsCompilation yes

R topics documented:

extractDist	1
subDist	2
Index	4

extractDist	<i>Subset a symmetric distance matrix of class "dist"</i>
-------------	---

Description

`extractDist()` allows us to efficiently subset a distance matrix of class "dist" without converting the "dist" object to a numeric matrix.

Usage

```
extractDist(dist, idxRows = NULL, idxCols = NULL)
```

Arguments

<code>dist</code>	A "dist" object, which can be obtained by the "dist" function.
<code>idxRows</code>	An interger vector indicates the indexes of units in the first subset. It cannot be smaller than 1 or larger than the dataset size. By default, <code>idxRows = NULL</code> .
<code>idxCols</code>	An interger vector indicates the indexes of units in the second subset. It cannot be smaller than 1 or larger than the dataset size. By default, <code>idxCols = NULL</code> .

Details

We are interested in extracting pair-wises distances between units in two subsets (indicated by their indexes) from a "dist" object. A simple way to do that is to convert the "dist" object to a numeric matrix using `as.dist()` and extract the relevant values using the squared brackets operator for matrix subsetting. However, it is very slow. `extractDist()` efficiently extract the pair-wise distances from the "dist" object without the need of conversion.

When either `idxRows` or `idxCols` are `NULL`, `extractDist()` extracts column vectors of the "dist" matrix given by the indexes of the not-null vector. Since the distance matrix is symmetric, it does not matter if we extract the row vectors or the column vectors. However, our implementation is more efficient for extracting column vectors from a "dist" object. If `idxRows` and `idxCols` are both `NULL`, the "dist" object is converted to a full numeric distance matrix.

Value

A numeric matrix storing pair-wise distances between the units in each subset.

Author(s)

Minh Long Nguyen <edelweiss611428@gmail.com>

Examples

```
x = rnorm(100)
dx = dist(x)
extractDist(dx, idxRows = 1:10, idxCols = 25:50)
```

subDist

Extract a sub-distance matrix from a "dist" object

Description

`subDist()` allows us to efficiently extract from a "dist" object a sub-distance matrix of class "dist" of a subset of the original dataset.

Usage

```
subDist(dist, idx, diag = F, upper = F)
```

Arguments

dist	A "dist" object, which can be obtained by the "dist" function.
idx	An integer vector indicates the indexes of the extracted units (idx can't be smaller than 1 or larger than the dataset size N).
diag	A boolean value indicates whether or not attr(dist, "Diag") = TRUE (by default, FALSE).
upper	A boolean value indicates whether or not attr(dist, "Upper") = TRUE (by default, FALSE).

Details

We are interested in extracting from a "dist" object a sub-distance matrix of class "dist" of a subset of the original dataset. A simple way to do that is to convert the "dist" object to a numeric matrix using `as.matrix()`, extract the relevant values, and convert the matrix back to a "dist" object using `as.dist()`. However, it is slow. `subDist()` allows us to do that efficiently without the need of conversion.

Value

a sub-distance matrix of class "dist"

Author(s)

Minh Long Nguyen <edelweiss611428@gmail.com>

Examples

```
x = rnorm(1000)
dx = dist(x)
subDist(dx, 1:10)
```

Index

extractDist, [1](#)

subDist, [2](#)