Package 'seqhandbook'

Type Package

Title Miscellaneous Tools for Sequence Analysis

Version 0.1.1

Depends R (>= 3.5.0), TraMineR

Suggests R.rsp, knitr, rmarkdown, FactoMineR, descriptio,
RColorBrewer, TraMineRextras, WeightedCluster, ade4, cluster,
questionr, rmdformats, dplyr, purrr, ggplot2

VignetteBuilder R.rsp

Author Nicolas Robette

Maintainer Nicolas Robette <nicolas.robette@uvsq.fr>

Description It provides miscellaneous sequence analysis functions for describing episodes in individual sequences, measuring association between domains in multidimensional sequence analysis (see Piccarreta (2017) <doi:10.1177/0049124115591013>), heat maps of sequence data, Globally Interdependent Multidimensional Sequence Analysis (see Robette et al (2015) <doi:10.1177/0081175015570976>), smoothing sequences for index plots (see Piccarreta (2012) <doi:10.1177/0049124112452394>), coding sequences for Qualitative Harmonic Analysis (see Deville (1982)), measuring stress from multidimensional scaling factors (see Piccarreta and Lior (2010) <doi:10.1111/j.1467-985X.2009.00606.x>), symmetrical (or canonical) Partial Least Squares (see Bry (1996)).

License GPL (>= 2)

Encoding UTF-8

LazyData true

URL https://nicolas-robette.github.io/seqhandbook/

NeedsCompilation no

Repository CRAN

R topics documented:

---

assoc.domains	*Association measures between domains in multidimensional sequence analysis*

Description

Computes various measures of association between dimensions of multidimensional sequence data.

Usage

```
assoc.domains(dlist, names, djsa)
```

Arguments

dlist	A list of dissimilarity matrices or dist objects (see [dist](#)), with one element per dimension of the multidimensional sequence data
names	A character vector of the names of the dimensions of the multidimensional sequence data
djsa	A dissimilarity matrix or a dist object (see [dist](#)), corresponding to the distances between the multimdimensional sequences

Author(s)

Nicolas Robette

References

Piccarreta R. (2017). Joint Sequence Analysis: Association and Clustering, *Sociological Methods and Research*, Vol. 46(2), 252-287.

Examples

```
library(TraMineR)
data(biofam)

## Building one channel per type of event (left, children or married)
bf <- as.matrix(biofam[, 10:25])
children <-  bf==4 | bf==5 | bf==6
married <- bf == 2 | bf== 3 | bf==6
left <- bf==1 | bf==3 | bf==5 | bf==6

## Building sequence objects
child.seq <- seqdef(children)
marr.seq <- seqdef(married)
left.seq <- seqdef(left)

## Using Hamming distance
mcdist <- seqdistmc(channels=list(child.seq, marr.seq, left.seq),
  method="HAM")
child.dist <- seqdist(child.seq, method="HAM")
marr.dist <- seqdist(marr.seq, method="HAM")
left.dist <- seqdist(left.seq, method="HAM")

## Association between domains
asso <- assoc.domains(list(child.dist,marr.dist,left.dist), c('child','marr','left'), mcdist)
asso
```

---

seq2qha   *Recoding sequences for qualitative harmonic analysis*

Description

Recodes sequence data into the shape used for qualitative harmonic analysis.

Usage

```
seq2qha(seqdata, periods)
```

Arguments

seqdata   a sequence object (see [seqdef](#) function).

periods   numeric vector of the first positions of the periods used for recoding

Value

A data frame with one column by combination of period and state (i.e. number of columns = number of periods * number of states in the alphabet).

Author(s)

Nicolas Robette

References

Robette N., Thibault N. (2008). Comparing qualitative harmonic analysis and optimal matching. An exploratory study of occupational trajectories, *Population-E*, Vol. 64(3), 533-556. Deville J-C. (1982). Analyse de données chronologiques qualitatives: comment analyser des calendriers ?, *Annales de l'INSEE*, 45, 45-104. Deville J-C., Saporta G. (1980). Analyse harmonique qualitative, in *Data analysis and informatics*, E.Diday (ed.), Amsterdam, North Holland Publishing, 375-389.

Examples

```
data(trajact)
seqact <- seqdef(trajact)
qha <- seq2qha(seqact, periods=c(1,3,7,12,24))
head(qha)
```

---

seqgimsa   *Sample of mothers and daughters employment histories*

Description

A data frame describing mothers employment histories from age 14 to 60 and daughters employment histories from the completion of education to 15 years later. Sequences are sampled (N = 400) from "Biographies et entourage" survey (INED, 2001).

Usage

```
data("seqgimsa")
```

Format

A data frame with 400 observations and 62 numeric variables. The first 15 variables (prefixed 'f') describe the daughters employment status a given year : 1 = education, 2 = inactivity, 3 = part-time job, 4 = full-time job. The following 47 variables (prefixed 'm') describe the mothers employment status at a given age : 1 = self-employment, 3 = higher level or intermediate occupation, 5 = lower level occupation, 8 = inactivity, 9 = education.

Examples

```
data(seqgimsa)
str(seqgimsa)
```

---

seqi1epi   *At least one episode in the states*

Description

Returns whether each sequence comprises at least one episode in the states.

Usage

```
seqi1epi(seqdata)
```

Arguments

seqdata   a sequence object (see [seqdef](#) function).

Author(s)

Nicolas Robette

References

Gabadinho, A., G. Ritschard, N. S. Müller and M. Studer (2011). Analyzing and Visualizing State Sequences in R with TraMineR. *Journal of Statistical Software* **40**(4), 1-37.

See Also

[seqistatd](#), [seqinepi](#), [seqifpos](#)

Examples

```
data(trajact)
seqact <- seqdef(trajact)
stat <- seqi1epi(seqact)
head(stat)
```

---

seqifpos   *First position in each state*

Description

Returns the first position in each state.

Usage

```
seqifpos(seqdata)
```

Arguments

seqdata   a sequence object (see [seqdef](#) function).

Author(s)

Nicolas Robette

References

Gabadinho, A., G. Ritschard, N. S. Müller and M. Studer (2011). Analyzing and Visualizing State Sequences in R with TraMineR. *Journal of Statistical Software* **40**(4), 1-37.

See Also

[seqistatd](#), [seqi1epi](#), [seqinepi](#)

Examples

```
data(trajact)
seqact <- seqdef(trajact)
stat <- seqifpos(seqact)
head(stat)
```

---

seqinepi   *Number of episodes in each state*

Description

Returns the number of episodes in the states.

Usage

```
seqinepi(seqdata)
```

Arguments

seqdata   a sequence object (see [seqdef](#) function).

Author(s)

Nicolas Robette

References

Gabadinho, A., G. Ritschard, N. S. Müller and M. Studer (2011). Analyzing and Visualizing State Sequences in R with TraMineR. *Journal of Statistical Software* **40**(4), 1-37.

See Also

[seqistatd](), [seqi1epi](), [seqifpos]()

Examples

```
data(trajact)
seqact <- seqdef(trajact)
stat <- seqinepi(seqact)
head(stat)
```

---

seqmds.stress	*Stress measure of multidimensional scaling factors*

Description

Computes stress measure of multidimensional scaling data for different number of dimensions of the represented space

Usage

```
seqmds.stress(seqdist, mds)
```

Arguments

seqdist	a dissimilarity matrix or a dist object (see [dist]())
mds	a matrix with coordinates in the represented space (dimension 1 in column 1, dimension 2 in column 2, etc.)

Value

A numerical vector of stress values.

Author(s)

Nicolas Robette

References

Piccarreta R., Lior O. (2010). Exploring sequences: a graphical tool based on multi-dimensional scaling, *Journal of the Royal Statistical Society (Series A)*, Vol. 173(1), 165-184.

Examples

```
data(trajact)
seqact <- seqdef(trajact)
dissim <- seqdist(seqact, method="HAM")
mds <- cmdscale(dissim, k=20, eig=TRUE)
stress <- seqmds.stress(dissim, mds)
plot(stress, type='l', xlab='number of dimensions', ylab='stress')
```

---

seqmsa   *Sample of marital, parental and residential sequences*

Description

A data frame describing respectively the matrimonial, parental and residential status from age 14 to age 35. It's sampled (N=500) from "Biographies et entourage" survey (INED, 2001).

Usage

```
data("seqmsa")
```

Format

A data frame with 500 observations and 66 variables. The first 22 variables (prefixed 'log') describe the residential status at a given age : 0 = not independent, 1 = independent. The next 22 variables (prefixed 'mat') describe the matrimonial status at a given age : 1 = never been in a relationship, 2 = cohabiting union, 3 = married, 4 = separated. The last 22 variables (prefixed 'nenf') describe the parental status at a given age : 0 = no child, 1 = one child, 2 = two children, 3 = three children or more.

Examples

```
data(seqmsa)
str(seqmsa)
```

---

seqsmooth   *Smoothing sequence data*

Description

Smoothing of sequence data, using for each sequence the medoid of the sequences in its neighborhood. The results can be used to get a smoothed index plot.

Usage

```
seqsmooth(seqdata, diss, k=20, r=NULL)
```

Arguments

seqdata	a sequence object (see [seqdef](#) function).
diss	a dissimilarity matrix, giving the pairwise distances between sequences.
k	size of the neighborhood. Default is 20.
r	radius of the neighborhood. If NULL (default), the radius is not used for smoothing.

Value

A list with the following elements:

seqdata   a sequence object (see [seqdef](#) function)

R2   pseudo-R2 measure of the goodness of fit of the smoothing

S2   stress measure of the goodness of fit of the smoothing

Author(s)

Nicolas Robette

References

Piccarreta R. (2012). Graphical and Smoothing Techniques for Sequence Analysis, *Sociological Methods and Research*, Vol. 41(2), 362-380.

Examples

```
data(trajact)
seqact <- seqdef(trajact)
dissim <- seqdist(seqact, method="LCS")
mds <- cmdscale(dissim, k=1)
smoothed <- seqsmooth(seqact, dissim, k=30)$seqdata
seqIplot(smoothed, sortv=mds, xtlab=14:50, with.legend=FALSE, yaxis=FALSE, ylab=NA)
```

---

seq_heatmap   *Index plot of sequences ordered according to a dendrogram*

Description

Index plot of state sequences. Sequences are ordered according to the specified dendrogram. The dendrogram is also plotted on the side of the index plot.

Usage

```
seq_heatmap(seq, tree, with.missing = FALSE, ...)
```

Arguments

seq   a state sequence object created with the [seqdef](#) function

tree   a dendrogram of the sequences (an object of class [hclust](#), [dendrogram](#) or [agnes](#))

with.missing is there a 'missing value' state in the sequences?

...   additional parameters sent to [heatmap](#)

Source

http://joseph.larmarange.net/?Representer-un-tapis-de-sequences

See Also

seqIplot

Examples

```
if (require(TraMineR)) {
  data(mvad)
  mvad.seq <- seqdef(mvad[,17:86])
  mvad.lcs <- seqdist(mvad.seq, method = "LCS")
  mvad.hc <- hclust(as.dist(mvad.lcs), method = "ward.D2")
  seq_heatmap(mvad.seq, mvad.hc)
}
```

---

socdem   *Sample of sociodemographic variables*

Description

A data frame with sociodemographic variables for a sample of 500 interviewees from "Biographies et entourage" survey (INED, 2001).

Usage

```
data("socdem")
```

Format

A data frame with 500 observations on the following 9 variables.

annais year of birth (numeric)

nbenf number of children (factor)

nbunion number of relationships (factor)

mereactive whether mother was active or not (factor)

sexe gender (factor)

PCS occupational category (factor)

PCSpere occupational category of the father (factor)

diplome degree (factor)

nationalite nationality (factor)

Examples

```
data(socdem)
str(socdem)
```

---

symPLS   *Symmetric (or canonical) PLS*

Description

Computes symmetric (or canonical) PLS for two groups of continuous variables

Usage

```
symPLS(a,b)
```

Arguments

a   data frame of the first group of continuous variables
b   data frame of the second group of continuous variables

Author(s)

Nicolas Robette, Xavier Bry

References

Bry X. (1996). Analyses Factorielles Multiples. Paris, Economica Poche. de Jong S., Wise B.M. and Ricker N.L. (2001). Canonical Partial Least Squares and Continuum Power Regression. *Journal of Chemometrics*, Vol. 15, 85–100.

---

trajact   *Sample of employment histories*

Description

A data frame describing the employment status from age 14 to age 50. It's a sample of 500 interviewees from "Biographies et entourage" survey (INED, 2001).

Usage

```
data("trajact")
```

Format

A data frame with 500 observations and 37 variables. Each variable is numeric and describes the employment status at a given age : 1 = education, 2 = full-time job, 3 = part-time job, 4 = small jobs, 5 = inactivity, 6 = military service.

Examples

```
data(trajact)
str(trajact)
```

Index