

Package ‘ClickClustCont’

August 19, 2019

Type Package

Title Mixtures of Continuous Time Markov Models

Version 0.1.7

Date 2019-08-18

Author Michael P.B. Gallagher, Paul D. McNicholas

Maintainer Michael P.B. Gallagher <gallaump@mcmaster.ca>

Description Provides an expectation maximization (EM) algorithm to fit a mixture of continuous time Markov models for use with clickstream or other sequence type data. Gallagher, M.P.B and McNicholas, P.D. (2018) <arXiv:1802.04849>.

License GPL (>= 2)

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

Imports gtools

NeedsCompilation no

Repository CRAN

Date/Publication 2019-08-19 07:50:05 UTC

R topics documented:

ClickClust_EM	2
mMSNBC	3
SimData	3

Index	5
--------------	----------

ClickClust_EM

*EM Algorithm for Continuous Time Markov Models***Description**

This function fits the continuous time first-order Markov model for a specified set of groups and returns the model chosen by the BIC. This is an implementation of the methodology developed in Gallagher and McNicholas (2019).

Usage

```
ClickClust_EM(x, t, J, G, itemEM = 5, starts = 100, maxit = 5000,
  tol = 0.001, Contin = TRUE, Verbose = TRUE, seed = 1,
  known = NULL, crit = "BIC", returnall = FALSE)
```

Arguments

x	A list of states
t	A list of times spent in each state
J	The total number of states
G	A vector containing the number of groups to test
itemEM	The number of emEM iterations for initialization (defaults to 5)
starts	The number of random starting values for the emEM algorithm (defaults to 100)
maxit	The maximum number of iterations after initialization (defaults to 5000)
tol	The tolerance for convergence (defaults to 0.001)
Contin	Fit the continuous time model (defaults to TRUE). If FALSE, fit the discrete model.
Verbose	Display Messages (defaults to TRUE)
seed	Sets the seed for the emEM algorithm (defaults to 1)
known	A vector of labels for semi-supervised classification. 0 indicates unknown observations. The known labels are denoted by their group number (1,2,3, etc.).
crit	The model selection criterion to use ("BIC" or "ICL"). Defaults to "BIC".
returnall	If true, returns the results for all groups considered. Defaults to FALSE.

Value

Returns a list with parameter and classification estimates for the best model chosen by the selection criterion.

References

Michael P.B. Gallagher and Paul D. McNicholas (2019). Clustering and semi-supervised classification for clickstream data via mixture models. arXiv preprint arXiv:1802.04849v2.

Examples

```
library(gtools)
data(SimData)
x<-SimData[[1]]
t<-SimData[[2]]
Click_2G<-ClickClust_EM(x=x, t=t, J=5, G=2, starts=10)
```

mMSNBC

Revised MSNBC Data

Description

This is a revised version of the MSNBC323 dataset in the R package ClickClust (Melnykov, 2016). This dataset contains the clickstreams without within-state transitions x and simulated time points t . See Gallagher and McNicholas (2019) for further details.

Usage

```
data(mMSNBC)
```

Format

An object of class `list` of length 2.

References

Michael P.B. Gallagher and Paul D. McNicholas (2019). Clustering and semi-supervised classification for clickstream data via mixture models. arXiv preprint arXiv:1802.04849v2.

Volodymyr Melnykov (2016). ClickClust: An R Package for Model-Based Clustering of Categorical Sequences. *Journal of Statistical Software* **74**(9), 1-34.

SimData

Simulated Data

Description

This is a simulated dataset with two groups. It is in the form of a list with the first element being the list of states and the second element being the list of time stamps. This is an example of the simulated data used in Simulation 1B in Gallagher and McNicholas (2019).

Usage

```
data(SimData)
```

Format

An object of class `list` of length 2.

References

Michael P.B. Gallagher and Paul D. McNicholas (2019). Clustering and semi-supervised classification for clickstream data via mixture models. arXiv preprint arXiv:1802.04849v2.

Index

*Topic **data**

mMSNBC, 3

SimData, 3

ClickClust_EM, 2

mMSNBC, 3

SimData, 3