

Package ‘nmfem’

April 24, 2018

Type Package

Title NMF-EM Algorithm

Version 1.0.3

Depends R(>= 2.10.0)

Description Provides a version of the Expectation-Maximization algorithm for mix-models, reducing the numbers of parameters to estimate using Non-negative Matrix Factorization methods.
For more explanations, see pre-print of Carel and Alquier (2017) <arXiv:1709.03346>.

License GPL-3

Encoding UTF-8

LazyData true

Imports plyr, dplyr, d3heatmap, mixtools, tidyr

RoxygenNote 6.0.1

Suggests knitr, testthat, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Author Lena Carel [aut, cre]

Maintainer Lena Carel <lena.carel@gmail.com>

Repository CRAN

Date/Publication 2018-04-24 15:00:29 UTC

R topics documented:

display_profile	2
loglik_mult	2
nmfem_mult	3
nmfem_mult_modelselection	5
travelers	5

Index	11
--------------	-----------

display_profile	<i>Display 3D profiles</i>
-----------------	----------------------------

Description

This function display profiles of 3 dimensions (day, hour, number of observations). It has been created to display profiles from the nmfem package data.

Usage

```
display_profile(profile, numclient = FALSE, color = "Blues",
  language = "en", theme = "dark")
```

Arguments

profile	a vector or a matrix row containing the profile to display. The day/hour data are contained in the column names.
numclient	logical. Whether the first value of the row is an identifier.
color	color of the display. Possibilities are the ones provided by colorbrewer2.org .
language	in which language the day/hour names are written. For now, the possibilities are "en" for english and "fr" for french.
theme	A theme to use. The only valid values are "" and "dark".

Value

Creates a 3D-heatmap displayed in the Viewer tab.

Examples

```
display_profile(travelers[sample(nrow(travelers),1), ], numclient = TRUE)
```

loglik_mult	<i>Extract log-likelihood from a mixture of multinomials</i>
-------------	--

Description

Extract log-likelihood from a mixture of multinomials

Usage

```
loglik_mult(X, Theta, Lambda = NULL, p)
```

Arguments

X	a matrix of dimension N (number of observation) x M (number of variables) containing multinomials observations.
Theta	matrix of dimension M x H.
Lambda	matrix of dimension H x K. Can be NULL.
p	vector containing the proportions of each cluster. Must be of dimension K (or H if Lambda is NULL).

Value

The function returns the log-likelihood of the data to the model

Examples

```

travelers <- travelers[ , -1]
M <- ncol(travelers)
K <- 5

Theta0 <- t(dplyr::sample_n(travelers, K))
Theta0 <- Theta0 / matrix(rep(apply(Theta0, 2, sum), M), nrow = M, ncol = K, byrow = TRUE)
travelers <- as.matrix(travelers)
p0 <- rep(1 / K, K)

llh <- loglik_mult(travelers, Theta0, p = p0)
llh

```

nmfem_mult

NMF-EM algorithm for mixture of multinomials

Description

Proceed to an NMF-EM algorithm on mixture of multinomials dataset. In comparison to the classical EM algorithm, the number of parameters to estimate is lower. For more explanation, see pre-print of Carel and Alquier (2017) <arXiv:1709.03346>.

Usage

```
nmfem_mult(X, H, K, path = NULL, eps_init = 0.001, eps_M = 1e-08,
           eps_llh = 1e-05)
```

Arguments

X	a matrix containing multinomials observations of dimension N (number of observation) x M (number of variables).
H	number of words.

K	number of clusters.
path	path to the directory to save the initialization or to load it. NULL by default, won't save or load it.
eps_init	convergence criterion on the initialization. Default value is 1e-3.
eps_M	convergence criterion on the Maximization step. Default value is 1e-8.
eps_llh	convergence criterion on the log-likelihood. Default value is 1e-5.

Value

A list with the elements:

Theta	matrix of dimension $M \times H$. Contains a dictionary of redundant components.
Lambda	matrix of dimension $H \times K$. Contains the expression of the K clusters in the dictionary.
llh	log-likelihood of the model.
p	vector containing the proportions of each cluster.
posterior	matrix containing for each observation the posterior probability to belong to each cluster.

Examples

```
# Example on a data sample
x <- dplyr::sample_n(travelers[,-1],900)
out <- nmfem_mult(x, H = 4, K = 7)
# Display first cluster profile
display_profile(t((out$Theta %*% out$Lambda)[ ,1]))
# Display first word profile
display_profile(t(out$Theta[ ,1]), color = "Greens")

# Example on the complete data - it needs a few minutes to run
## Not run:
nmfem_travelers <- nmfem_mult(travelers[ , -1], H = 5, K = 10)
Theta <- nmfem_travelers$Theta
Lambda <- nmfem_travelers$Lambda

# Display first cluster profile
display_profile(t((Theta %*% Lambda)[ ,1]))

# Display first word profile
display_profile(t(Theta[ ,1]), color = "Greens")
## End(Not run)
```

nmfem_mult_modelselection

Model selection in NMF-EM algorithm for mixture of multinomials

Description

The function proceed to a model selection with NMF-EM algorithm on mixture of multinomials dataset. First, the function plots the log-likelihood in function of K. Second, log-likelihood is plotted in function of H. We recommend the user to choose K and H by slope heuristic method.

Usage

```
nmfem_mult_modelselection(X, maxK = 30, save = FALSE, path = ".")
```

Arguments

X	numeric matrix containing multinomials observations of dimension N (number of observation) x M (number of variables).
maxK	integer. Maximum number of clusters to be tested. By default, function tests from 2 to 30 clusters.
save	logical. Whether the result of each parameter couple (H,K) tested got to be saved.
path	path to save the results if save = TRUE. By default, it is the working directory. Three directories are created to save the results. Directory "Initializations" contains the initialization of the algorithm for each value of K. Matrices are saved in directory "Matrices", and plots in directory "Results".

Examples

```
# Example on the complete data - needs around an hour to run
## Not run:
nmfem_mult_modelselection(travelers[ , -1])

## End(Not run)
```

travelers

Time profiles of 47,000 travelers.

Description

A dataset containing the number of travels made at every moment of the week during one month on a public transportation network for more than 47,000 travelers.

Usage

travelers

Format

A data frame with 47,757 rows and 166 columns.

ID_card smart card encrypted identifier

Friday00 number of travels made on Fridays between 12:00 a.m and 12:59 a.m

Friday01 number of travels made on Fridays between 1:00 a.m and 1:59 a.m

Friday02 number of travels made on Fridays between 2:00 a.m and 2:59 a.m

Friday03 number of travels made on Fridays between 3:00 a.m and 3:59 a.m

Friday04 number of travels made on Fridays between 4:00 a.m and 4:59 a.m

Friday05 number of travels made on Fridays between 5:00 a.m and 5:59 a.m

Friday06 number of travels made on Fridays between 6:00 a.m and 6:59 a.m

Friday07 number of travels made on Fridays between 7:00 a.m and 7:59 a.m

Friday08 number of travels made on Fridays between 8:00 a.m and 8:59 a.m

Friday09 number of travels made on Fridays between 9:00 a.m and 9:59 a.m

Friday10 number of travels made on Fridays between 10:00 a.m and 10:59 a.m

Friday11 number of travels made on Fridays between 11:00 a.m and 11:59 a.m

Friday12 number of travels made on Fridays between 12:00 p.m and 12:59 p.m

Friday13 number of travels made on Fridays between 1:00 p.m and 1:59 p.m

Friday14 number of travels made on Fridays between 2:00 p.m and 2:59 p.m

Friday15 number of travels made on Fridays between 3:00 p.m and 3:59 p.m

Friday16 number of travels made on Fridays between 4:00 p.m and 4:59 p.m

Friday17 number of travels made on Fridays between 5:00 p.m and 5:59 p.m

Friday18 number of travels made on Fridays between 6:00 p.m and 6:59 p.m

Friday19 number of travels made on Fridays between 7:00 p.m and 7:59 p.m

Friday20 number of travels made on Fridays between 8:00 p.m and 8:59 p.m

Friday21 number of travels made on Fridays between 9:00 p.m and 9:59 p.m

Friday22 number of travels made on Fridays between 10:00 p.m and 10:59 p.m

Friday23 number of travels made on Fridays between 11:00 p.m and 11:59 p.m

Monday00 number of travels made on Mondays between 12:00 a.m and 12:59 a.m

Monday01 number of travels made on Mondays between 1:00 a.m and 1:59 a.m

Monday03 number of travels made on Mondays between 3:00 a.m and 3:59 a.m

Monday04 number of travels made on Mondays between 4:00 a.m and 4:59 a.m

Monday05 number of travels made on Mondays between 5:00 a.m and 5:59 a.m

Monday06 number of travels made on Mondays between 6:00 a.m and 6:59 a.m

Monday07 number of travels made on Mondays between 7:00 a.m and 7:59 a.m

- Monday08** number of travels made on Mondays between 8:00 a.m and 8:59 a.m
- Monday09** number of travels made on Mondays between 9:00 a.m and 9:59 a.m
- Monday10** number of travels made on Mondays between 10:00 a.m and 10:59 a.m
- Monday11** number of travels made on Mondays between 11:00 a.m and 11:59 a.m
- Monday12** number of travels made on Mondays between 12:00 p.m and 12:59 p.m
- Monday13** number of travels made on Mondays between 1:00 p.m and 1:59 p.m
- Monday14** number of travels made on Mondays between 2:00 p.m and 2:59 p.m
- Monday15** number of travels made on Mondays between 3:00 p.m and 3:59 p.m
- Monday16** number of travels made on Mondays between 4:00 p.m and 4:59 p.m
- Monday17** number of travels made on Mondays between 5:00 p.m and 5:59 p.m
- Monday18** number of travels made on Mondays between 6:00 p.m and 6:59 p.m
- Monday19** number of travels made on Mondays between 7:00 p.m and 7:59 p.m
- Monday20** number of travels made on Mondays between 8:00 p.m and 8:59 p.m
- Monday21** number of travels made on Mondays between 9:00 p.m and 9:59 p.m
- Monday22** number of travels made on Mondays between 10:00 p.m and 10:59 p.m
- Monday23** number of travels made on Mondays between 11:00 p.m and 11:59 p.m
- Saturday00** number of travels made on Saturdays between 12:00 a.m and 12:59 a.m
- Saturday01** number of travels made on Saturdays between 1:00 a.m and 1:59 a.m
- Saturday02** number of travels made on Saturdays between 2:00 a.m and 2:59 a.m
- Saturday03** number of travels made on Saturdays between 3:00 a.m and 3:59 a.m
- Saturday04** number of travels made on Saturdays between 4:00 a.m and 4:59 a.m
- Saturday05** number of travels made on Saturdays between 5:00 a.m and 5:59 a.m
- Saturday06** number of travels made on Saturdays between 6:00 a.m and 6:59 a.m
- Saturday07** number of travels made on Saturdays between 7:00 a.m and 7:59 a.m
- Saturday08** number of travels made on Saturdays between 8:00 a.m and 8:59 a.m
- Saturday09** number of travels made on Saturdays between 9:00 a.m and 9:59 a.m
- Saturday10** number of travels made on Saturdays between 10:00 a.m and 10:59 a.m
- Saturday11** number of travels made on Saturdays between 11:00 a.m and 11:59 a.m
- Saturday12** number of travels made on Saturdays between 12:00 p.m and 12:59 p.m
- Saturday13** number of travels made on Saturdays between 1:00 p.m and 1:59 p.m
- Saturday14** number of travels made on Saturdays between 2:00 p.m and 2:59 p.m
- Saturday15** number of travels made on Saturdays between 3:00 p.m and 3:59 p.m
- Saturday16** number of travels made on Saturdays between 4:00 p.m and 4:59 p.m
- Saturday17** number of travels made on Saturdays between 5:00 p.m and 5:59 p.m
- Saturday18** number of travels made on Saturdays between 6:00 p.m and 6:59 p.m
- Saturday19** number of travels made on Saturdays between 7:00 p.m and 7:59 p.m
- Saturday20** number of travels made on Saturdays between 8:00 p.m and 8:59 p.m

Saturday21 number of travels made on Saturdays between 9:00 p.m and 9:59 p.m
Saturday22 number of travels made on Saturdays between 10:00 p.m and 10:59 p.m
Saturday23 number of travels made on Saturdays between 11:00 p.m and 11:59 p.m
Sunday00 number of travels made on Sundays between 12:00 a.m and 12:59 a.m
Sunday01 number of travels made on Sundays between 1:00 a.m and 1:59 a.m
Sunday02 number of travels made on Sundays between 2:00 a.m and 2:59 a.m
Sunday03 number of travels made on Sundays between 3:00 a.m and 3:59 a.m
Sunday04 number of travels made on Sundays between 4:00 a.m and 4:59 a.m
Sunday05 number of travels made on Sundays between 5:00 a.m and 5:59 a.m
Sunday06 number of travels made on Sundays between 6:00 a.m and 6:59 a.m
Sunday07 number of travels made on Sundays between 7:00 a.m and 7:59 a.m
Sunday08 number of travels made on Sundays between 8:00 a.m and 8:59 a.m
Sunday09 number of travels made on Sundays between 9:00 a.m and 9:59 a.m
Sunday10 number of travels made on Sundays between 10:00 a.m and 10:59 a.m
Sunday11 number of travels made on Sundays between 11:00 a.m and 11:59 a.m
Sunday12 number of travels made on Sundays between 12:00 p.m and 12:59 p.m
Sunday13 number of travels made on Sundays between 1:00 p.m and 1:59 p.m
Sunday14 number of travels made on Sundays between 2:00 p.m and 2:59 p.m
Sunday15 number of travels made on Sundays between 3:00 p.m and 3:59 p.m
Sunday16 number of travels made on Sundays between 4:00 p.m and 4:59 p.m
Sunday17 number of travels made on Sundays between 5:00 p.m and 5:59 p.m
Sunday18 number of travels made on Sundays between 6:00 p.m and 6:59 p.m
Sunday19 number of travels made on Sundays between 7:00 p.m and 7:59 p.m
Sunday20 number of travels made on Sundays between 8:00 p.m and 8:59 p.m
Sunday21 number of travels made on Sundays between 9:00 p.m and 9:59 p.m
Sunday22 number of travels made on Sundays between 10:00 p.m and 10:59 p.m
Sunday23 number of travels made on Sundays between 11:00 p.m and 11:59 p.m
Thursday00 number of travels made on Thursdays between 12:00 a.m and 12:59 a.m
Thursday01 number of travels made on Thursdays between 1:00 a.m and 1:59 a.m
Thursday02 number of travels made on Thursdays between 2:00 a.m and 2:59 a.m
Thursday03 number of travels made on Thursdays between 3:00 a.m and 3:59 a.m
Thursday04 number of travels made on Thursdays between 4:00 a.m and 4:59 a.m
Thursday05 number of travels made on Thursdays between 5:00 a.m and 5:59 a.m
Thursday06 number of travels made on Thursdays between 6:00 a.m and 6:59 a.m
Thursday07 number of travels made on Thursdays between 7:00 a.m and 7:59 a.m
Thursday08 number of travels made on Thursdays between 8:00 a.m and 8:59 a.m
Thursday09 number of travels made on Thursdays between 9:00 a.m and 9:59 a.m

Thursday10 number of travels made on Thursdays between 10:00 a.m and 10:59 a.m
Thursday11 number of travels made on Thursdays between 11:00 a.m and 11:59 a.m
Thursday12 number of travels made on Thursdays between 12:00 p.m and 12:59 p.m
Thursday13 number of travels made on Thursdays between 1:00 p.m and 1:59 p.m
Thursday14 number of travels made on Thursdays between 2:00 p.m and 2:59 p.m
Thursday15 number of travels made on Thursdays between 3:00 p.m and 3:59 p.m
Thursday16 number of travels made on Thursdays between 4:00 p.m and 4:59 p.m
Thursday17 number of travels made on Thursdays between 5:00 p.m and 5:59 p.m
Thursday18 number of travels made on Thursdays between 6:00 p.m and 6:59 p.m
Thursday19 number of travels made on Thursdays between 7:00 p.m and 7:59 p.m
Thursday20 number of travels made on Thursdays between 8:00 p.m and 8:59 p.m
Thursday21 number of travels made on Thursdays between 9:00 p.m and 9:59 p.m
Thursday22 number of travels made on Thursdays between 10:00 p.m and 10:59 p.m
Thursday23 number of travels made on Thursdays between 11:00 p.m and 11:59 p.m
Tuesday00 number of travels made on Tuesdays between 12:00 a.m and 12:59 a.m
Tuesday01 number of travels made on Tuesdays between 1:00 a.m and 1:59 a.m
Tuesday03 number of travels made on Tuesdays between 3:00 a.m and 3:59 a.m
Tuesday04 number of travels made on Tuesdays between 4:00 a.m and 4:59 a.m
Tuesday05 number of travels made on Tuesdays between 5:00 a.m and 5:59 a.m
Tuesday06 number of travels made on Tuesdays between 6:00 a.m and 6:59 a.m
Tuesday07 number of travels made on Tuesdays between 7:00 a.m and 7:59 a.m
Tuesday08 number of travels made on Tuesdays between 8:00 a.m and 8:59 a.m
Tuesday09 number of travels made on Tuesdays between 9:00 a.m and 9:59 a.m
Tuesday10 number of travels made on Tuesdays between 10:00 a.m and 10:59 a.m
Tuesday11 number of travels made on Tuesdays between 11:00 a.m and 11:59 a.m
Tuesday12 number of travels made on Tuesdays between 12:00 p.m and 12:59 p.m
Tuesday13 number of travels made on Tuesdays between 1:00 p.m and 1:59 p.m
Tuesday14 number of travels made on Tuesdays between 2:00 p.m and 2:59 p.m
Tuesday15 number of travels made on Tuesdays between 3:00 p.m and 3:59 p.m
Tuesday16 number of travels made on Tuesdays between 4:00 p.m and 4:59 p.m
Tuesday17 number of travels made on Tuesdays between 5:00 p.m and 5:59 p.m
Tuesday18 number of travels made on Tuesdays between 6:00 p.m and 6:59 p.m
Tuesday19 number of travels made on Tuesdays between 7:00 p.m and 7:59 p.m
Tuesday20 number of travels made on Tuesdays between 8:00 p.m and 8:59 p.m
Tuesday21 number of travels made on Tuesdays between 9:00 p.m and 9:59 p.m
Tuesday22 number of travels made on Tuesdays between 10:00 p.m and 10:59 p.m
Tuesday23 number of travels made on Tuesdays between 11:00 p.m and 11:59 p.m

Wednesday00 number of travels made on Wednesdays between 12:00 a.m and 12:59 a.m
Wednesday01 number of travels made on Wednesdays between 1:00 a.m and 1:59 a.m
Wednesday03 number of travels made on Wednesdays between 3:00 a.m and 3:59 a.m
Wednesday04 number of travels made on Wednesdays between 4:00 a.m and 4:59 a.m
Wednesday05 number of travels made on Wednesdays between 5:00 a.m and 5:59 a.m
Wednesday06 number of travels made on Wednesdays between 6:00 a.m and 6:59 a.m
Wednesday07 number of travels made on Wednesdays between 7:00 a.m and 7:59 a.m
Wednesday08 number of travels made on Wednesdays between 8:00 a.m and 8:59 a.m
Wednesday09 number of travels made on Wednesdays between 9:00 a.m and 9:59 a.m
Wednesday10 number of travels made on Wednesdays between 10:00 a.m and 10:59 a.m
Wednesday11 number of travels made on Wednesdays between 11:00 a.m and 11:59 a.m
Wednesday12 number of travels made on Wednesdays between 12:00 p.m and 12:59 p.m
Wednesday13 number of travels made on Wednesdays between 1:00 p.m and 1:59 p.m
Wednesday14 number of travels made on Wednesdays between 2:00 p.m and 2:59 p.m
Wednesday15 number of travels made on Wednesdays between 3:00 p.m and 3:59 p.m
Wednesday16 number of travels made on Wednesdays between 4:00 p.m and 4:59 p.m
Wednesday17 number of travels made on Wednesdays between 5:00 p.m and 5:59 p.m
Wednesday18 number of travels made on Wednesdays between 6:00 p.m and 6:59 p.m
Wednesday19 number of travels made on Wednesdays between 7:00 p.m and 7:59 p.m
Wednesday20 number of travels made on Wednesdays between 8:00 p.m and 8:59 p.m
Wednesday21 number of travels made on Wednesdays between 9:00 p.m and 9:59 p.m
Wednesday22 number of travels made on Wednesdays between 10:00 p.m and 10:59 p.m
Wednesday23 number of travels made on Wednesdays between 11:00 p.m and 11:59 p.m

Source

These data are aggregated from Transdev datasets.

Index

*Topic **datasets**

travelers, [5](#)

display_profile, [2](#)

loglik_mult, [2](#)

nmfem_mult, [3](#)

nmfem_mult_modelselection, [5](#)

travelers, [5](#)