MS LDA Package Documentation

August 3, 2021

`lda_original` *Run LDA as in the Blei 2003 paper*

Description

Run LDA as in the Blei 2003 paper

Usage

```
lda_original(docs, K, max_iter = 50, thresh = NULL, seed = NULL)

lda_original_par(
  docs,
  K,
  max_iter = 50,
  thresh = NULL,
  seed = NULL,
  cores = NULL
)

lda_noalpha(
  docs,
  K,
  max_iter = 50,
  thresh = NULL,
  seed = NULL,
  cores = NULL,
  alpha = NULL
)
```

Arguments

`docs`	a list containing all the documents, with the vocabulary encoded e.g. docs[[1]] = c(1, 5, 2) would represent the word indices from a pre-defined vocabulary

K	the number of topics to look for
max_iter	the maximum number of EM iterations to run
thresh	if you want to set a specific threshold for L convergence, if NULL it's defined as 0.01 percent of the previous value
seed	if you want a reproducible result you can set a seed, if NULL the topic distributions are initialised from K random documents,
cores	number of cores to run the E-step in parallel, if NULL all detected cores are used
alpha	if you want to set the exchangeable Dirichlet parameter for theta, if NULL a default value of 1/K is used

Value

A list of all parameters

Functions

- `lda_original_par`: Runs E-step in parallel
- `lda_noalpha`: Alpha is fixed

`lda_reshaped` *Run LDA adapted to use a count matrix*

Description

Run LDA adapted to use a count matrix

Usage

```
lda_reshaped(
  N,
  K,
  max_iter = 50,
  thresh = NULL,
  seed = NULL,
  cores = NULL,
  alpha = NULL
)
```

Arguments

N	matrix of word counts
K	the number of topics to look for
max_iter	the maximum number of EM iterations to run

thresh	if you want to set a specific threshold for L convergence, if NULL it's defined as 0.01 percent of the previous value
seed	if you want a reproducible result you can set a seed, if NULL the topic distributions are initialised from K random documents,
cores	number of cores to run the E-step in parallel, if NULL all detected cores are used
alpha	if you want to set the exchangeable Dirichlet parameter for theta, if NULL a default value of 1/K is used

Value

A list of all parameters

`lda_smoothed` *Run LDA adapted to use a count matrix*

Description

Run LDA adapted to use a count matrix

Usage

```
lda_smoothed(
  N,
  K,
  max_iter = 50,
  thresh = NULL,
  seed = NULL,
  cores = NULL,
  alpha = NULL,
  eta = NULL,
  NMF = FALSE
)
```

Arguments

N	matrix of word counts
K	the number of topics to look for
max_iter	the maximum number of EM iterations to run
thresh	if you want to set a specific threshold for L convergence, if NULL it's defined as 0.01 percent of the previous value
seed	if you want a reproducible result you can set a seed, if NULL the topic distributions are initialised from K random documents,
cores	number of cores to run the E-step in parallel, if NULL all detected cores are used

alpha	if you want to set the exchangeable Dirichlet parameter for theta, if NULL a default value of 1/K is used
eta	if you want to set the exchangeable Dirichlet parameter for beta, if NULL a default value of 1/K is used
NMF	logical indicating if lambda should be initialised using non-negative matrix factorisation, if FALSE it is generated using K random documents

Index