

MS LDA Package Documentation

November 16, 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 = 1e-04, seed = NULL)
```

```
lda_original_par(  
  docs,  
  K,  
  max_iter = 50,  
  thresh = 1e-04,  
  seed = NULL,  
  cores = NULL  
)
```

```
lda_noalpha(  
  docs,  
  K,  
  max_iter = 50,  
  thresh = 1e-04,  
  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 | threshold for L convergence, $(L_i - L_{i-1})/L_i < \text{thresh}$ |
| seed | set a seed for the random documents to initialise beta |
| 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 | <i>Run LDA adapted to use a count matrix</i> |
|--------------|--|

Description

Run LDA adapted to use a count matrix

Usage

```
lda_reshaped(
  N,
  K,
  max_iter = 50,
  thresh = 1e-04,
  seed = NULL,
  cores = NULL,
  alpha = NULL
)

lda_reshaped_noalpha(
  N,
  K,
  max_iter = 50,
  thresh = 1e-04,
  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 | threshold for L convergence, $(L_i - L_{i-1})/L_i < \text{thresh}$ |
| seed | set a seed for the random documents to initialise beta |
| 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

A list of all parameters

Functions

- lda_resaped_noalpha: Alpha is fixed

| | |
|--------------|--|
| lda_smoothed | <i>Run LDA adapted to use a count matrix</i> |
|--------------|--|

Description

Run LDA adapted to use a count matrix

Usage

```
lda_smoothed(
  N,
  K,
  max_iter = 50,
  thresh = 1e-04,
  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 | threshold for L convergence, $(L_i - L_{i-1})/L_i < \text{thresh}$ |
| seed | set a seed for the random documents to initialise beta |
| 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 | 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 |

| | |
|-----|-------------------------------------|
| nmf | <i>Run NMF using a count matrix</i> |
|-----|-------------------------------------|

Description

Run NMF using a count matrix

Usage

```
nmf(counts, K, max_iter = 50, thresh = 1e-04, seed = NULL)
```

Arguments

| | |
|----------|--|
| counts | matrix of word counts |
| K | internal dimension of matrix factors |
| max_iter | the maximum number of iterations to run |
| thresh | threshold for L convergence, $(L_i - L_{i-1})/L_i < \text{thresh}$ |
| seed | for the random initialisation of factors W and H |

Value

A list of all parameters

Index

`lda_noalpha (lda_original)`, [1](#)
`lda_original`, [1](#)
`lda_original_par (lda_original)`, [1](#)
`lda_resaped`, [2](#)
`lda_resaped_noalpha (lda_resaped)`, [2](#)
`lda_smoothed`, [3](#)

`nmf`, [4](#)