

Package ‘ldatuning’

October 24, 2016

Type Package

Title Tuning of the Latent Dirichlet Allocation Models Parameters

Description For this first version only metrics to estimate the best fitting number of topics are implemented.

Version 0.2.0

Date 2016-10-24

URL <https://github.com/nikita-moor/ldatuning>

BugReports <https://github.com/nikita-moor/ldatuning/issues>

License BSD_2_clause + file LICENSE

LazyData TRUE

Imports parallel, topicmodels, slam, Rmpfr, ggplot2, reshape2, scales, grid

Suggests knitr, rmarkdown

VignetteBuilder knitr

RoxygenNote 5.0.1

NeedsCompilation no

Author Murzintcev Nikita [aut, cre]

Maintainer Murzintcev Nikita <nikita@lreis.ac.cn>

Repository CRAN

Date/Publication 2016-10-24 17:14:24

R topics documented:

FindTopicsNumber	2
FindTopicsNumber_plot	3
ldatuning	3

Index	4
--------------	----------

FindTopicsNumber *FindTopicsNumber*

Description

Calculates different metrics to estimate the most preferable number of topics for LDA model.

Usage

```
FindTopicsNumber(dtm, topics = seq(10, 40, by = 10),
  metrics = "Griffiths2004", method = "Gibbs", control = list(),
  mc.cores = 1L, verbose = FALSE)
```

Arguments

dtm	An object of class " DocumentTermMatrix " with term-frequency weighting or an object coercible to a " simple_triplet_matrix " with integer entries.
topics	Vvector with number of topics to compare different models.
metrics	String or vector of possible metrics: "Griffiths2004", "CaoJuan2009", "Arun2010", "Deveaud2014".
method	The method to be used for fitting; see LDA .
control	A named list of the control parameters for estimation or an object of class " LDA-control ".
mc.cores	Integer; The number of CPU cores to processes models simultaneously.
verbose	If false (default), suppress all warnings and additional information.

Value

Data-frame with one or more metrics. numbers of topics and corresponding values of metric. Can be directly used by [FindTopicsNumber_plot](#) to draw a plot.

Examples

```
## Not run:

library(topicmodels)
data("AssociatedPress", package="topicmodels")
dtm <- AssociatedPress[1:10, ]
FindTopicsNumber(dtm, topics = 2:10, metrics = "Arun2010", mc.cores = 1L)

## End(Not run)
```

FindTopicsNumber_plot *FindTopicsNumber_plot*

Description

Support function to analyze optimal topic number. Use output of the [FindTopicsNumber](#) function.

Usage

```
FindTopicsNumber_plot(values)
```

Arguments

values	Data-frame with first column named 'topics' and other columns are values of metrics.
--------	--

Examples

```
## Not run:  
  
library(topicmodels)  
data("AssociatedPress", package="topicmodels")  
dtm <- AssociatedPress[1:10, ]  
optimal.topics <- FindTopicsNumber(dtm, topics = 2:10,  
  metrics = c("Arun2010", "CaoJuan2009", "Griffiths2004")  
)  
FindTopicsNumber_plot(optimal.topics)  
  
## End(Not run)
```

ldatuning

ldatuning: Tuning of the LDA models parameters

Description

ldatuning: Tuning of the LDA models parameters

Index

DocumentTermMatrix, [2](#)

FindTopicsNumber, [2](#), [3](#)

FindTopicsNumber_plot, [2](#), [3](#)

LDA, [2](#)

LDAcontrol, [2](#)

ldatuning, [3](#)

ldatuning-package (ldatuning), [3](#)

simple_triplet_matrix, [2](#)