Package: ldatuning (via r-universe)

August 29, 2024

Type Package

Title Tuning of the Latent Dirichlet Allocation Models Parameters

Description This library estimates the best fitting number of topics.

Version 1.0.3

Date 2024-05-31

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

Suggests knitr, rmarkdown, tibble

VignetteBuilder knitr

RoxygenNote 7.3.1

Encoding UTF-8

Repository https://nikita-moor.r-universe.dev

RemoteUrl https://github.com/nikita-moor/ldatuning

RemoteRef HEAD

RemoteSha 661fabb495b69ef9096e433b41bf17f51b6bca4b

Contents

Arun2010 2
CaoJuan2009 2
Deveaud2014 3
FindTopicsNumber 3
FindTopicsNumber_plot 4
Griffiths2004 5
ldatuning 6

Index **7**

`Arun2010` *Arun2010*

Description

Implement scoring algorithm

Usage

```
Arun2010(models, dtm)
```

Arguments

`models`	An object of class "LDA
`dtm`	An object of class "DocumentTermMatrix" with term-frequency weighting or an object coercible to a "simple_triplet_matrix" with integer entries.

Value

A scalar LDA model score

`CaoJuan2009` *CaoJuan2009*

Description

Implement scoring algorithm

Usage

```
CaoJuan2009(models)
```

Arguments

`models`	An object of class "LDA

Value

A scalar LDA model score

Deveaud2014 *Deveaud2014*

Description

Implement scoring algorithm

Usage

```
Deveaud2014(models)
```

Arguments

models	An object of class "LDA

Value

A scalar LDA model score

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 = NA,
  return_models = FALSE,
  verbose = FALSE,
  libpath = NULL
)
```

Arguments

dtm	An object of class "DocumentTermMatrix" with term-frequency weighting or an object coercible to a "simple_triplet_matrix" with integer entries.
topics	Vector 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 "LDAcontrol".
mc.cores	NA, integer or, cluster; the number of CPU cores to process models simultaneously. If an integer, create a cluster on the local machine. If a cluster, use but don't destroy it (allows multiple-node clusters). Defaults to NA, which triggers auto-detection of number of cores on the local machine.
return_models	Whether or not to return the model objects of class "LDA. Defaults to false. Setting to true requires the tibble package.
verbose	If false (default), suppress all warnings and additional information.
libpath	Path to R packages (use only if your R installation can't find 'topicmodels' package, [issue #3](https://github.com/nikita-moor/ldatuning/issues/3). For example: "C:/Program Files/R/R-2.15.2/library" (Windows), "/home/user/R/x86_64-pc-linux-gnu-library/3.2" (Linux)

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)
```

`Griffiths2004` *Griffiths2004*

Description

Implement scoring algorithm. In order to use this algorithm, the LDA model MUST be generated using the keep control parameter >0 (defaults to 50) so that the logLiks vector is retained.

Usage

```
Griffiths2004(models, control)
```

Arguments

models	An object of class "LDA
control	A named list of the control parameters for estimation or an object of class "LDA-control".

Value

A scalar LDA model score

ldatuning *ldatuning: Tuning of the LDA models parameters*

Description

A package for identifying the number of topics in a text corpus by generating LDA models, tuning LDA model parameters, and scoring model results.

Author(s)

Maintainer: Nathan Chaney <nathan@nathanchaney.com> (ORCID) [contributor]

Authors:

- Murzintcev Nikita <nikita@lreis.ac.cn> (ORCID)

See Also

Useful links:

- https://github.com/nikita-moor/ldatuning
- Report bugs at https://github.com/nikita-moor/ldatuning/issues

Index

Arun2010, 2

CaoJuan2009, 2

Deveaud2014, 3
DocumentTermMatrix, *2*, *4*

FindTopicsNumber, 3, *4*
FindTopicsNumber_plot, *4*, 4

Griffiths2004, 5

LDA, *2–5*
LDAcontrol, *4, 5*
ldatuning, 6
ldatuning-package *(*ldatuning*)*, 6

simple_triplet_matrix, *2*, *4*