# Package 'ldatuning'

April 21, 2020

**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** 1.0.2

**Date** 2020-04-09

**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, tibble

**VignetteBuilder** knitr

**RoxygenNote** 7.1.0

**NeedsCompilation** no

**Author** Murzintcev Nikita [aut], Nathan Chaney [ctb, cre] (<https://orcid.org/0000-0001-8985-2514>)

**Maintainer** Nathan Chaney <nathan@nathanchaney.com>

**Repository** CRAN

**Date/Publication** 2020-04-21 05:20:03 UTC

## R topics documented:

---

`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.

# Index