Package ‘newsmap’

May 18, 2021

Type Package
Title Semisupervised Model for Geographical Document Classification
Version 0.7.4
Maintainer Kohei Watanabe <watanabe.kohei@gmail.com>
This package currently contains seed dictionaries in English, German, French, Spanish, Italian, Russian, Hebrew, Arabic Japanese and Chinese (Simplified and Traditional).
License MIT + file LICENSE
URL https://github.com/koheiw/newsmap
BugReports https://github.com/koheiw/newsmap/issues
LazyData TRUE
Encoding UTF-8
Depends R (>= 3.5), methods
Imports utils, Matrix, quanteda (>= 2.1), quanteda.textstats, stringi
Suggests testthat
Language en-GB
RoxygenNote 7.1.1
NeedsCompilation no
Author Kohei Watanabe [aut, cre, cph],
    Stefan Müller [aut],
    Dani Madrid-Morales [aut],
    Katerina Tertychnaya [aut],
    Ke Cheng [aut],
    Chung-hong Chan [aut],
    Claude Grasland [aut],
    Giuseppe Carteny [aut],
    Elad Segev [aut],
    Dai Yamao [aut]
Repository CRAN
Date/Publication 2021-05-18 04:10:02 UTC
R topics documented:

accuracy ........................................................................ 2
afe .............................................................................. 3
data_dictionary_newsmap_ar ..................................... 3
data_dictionary_newsmap_de ..................................... 3
data_dictionary_newsmap_en ..................................... 4
data_dictionary_newsmap_es ..................................... 4
data_dictionary_newsmap_fr ..................................... 4
data_dictionary_newsmap_he ..................................... 5
data_dictionary_newsmap_it ..................................... 5
data_dictionary_newsmap_ja ..................................... 5
data_dictionary_newsmap_ru ..................................... 6
data_dictionary_newsmap_zh_cn ................................ 6
data_dictionary_newsmap_zh_tw ................................ 6
predict.textmodel_newsmap ........................................ 7
print.textmodel_newsmap_summary ............................... 7
summary.textmodel_newsmap_accuracy ....................... 8
textmodel_newsmap ..................................................... 8

Index ............................................................................ 10

-----------------------------------------------

accuracy Evaluate classification accuracy in precision and recall

Description

Evaluate classification accuracy in precision and recall

Usage

accuracy(x, y)

Arguments

x vector of predicted classes
y vector of true classes

Examples

class_pred <- c('US', 'GB', 'US', 'CN', 'JP', 'FR', 'CN') # prediction
class_true <- c('US', 'FR', 'US', 'CN', 'KP', 'EG', 'US') # true class
acc <- accuracy(class_pred, class_true)
print(acc)
summary(acc)
Compute average feature entropy (AFE)

Description
AFE computes randomness of occurrences features in labelled documents.

Usage
afe(x, y, smooth = 1)

Arguments
x a dfm for features
y a dfm for labels
smooth a numeric value for smoothing to include all the features

Seed geographical dictionary in Arabic

Description
Seed geographical dictionary in Arabic

Author(s)
Dai Yamao <daiyamao@scs.kyushu-u.ac.jp>

Seed geographical dictionary in German

Description
Seed geographical dictionary in German

Author(s)
Stefan Müller <mullers@tcd.ie>
Seed geographical dictionary in English

Description
Seed geographical dictionary in English

Author(s)
Kohei Watanabe <watanabe.kohei@gmail.com>

Seed geographical dictionary in Spanish

Description
Seed geographical dictionary in Spanish

Author(s)
Dani Madrid-Morales <dani.madrid@my.cityu.edu.hk>

Seed geographical dictionary in French

Description
Seed geographical dictionary in French

Author(s)
Claude Grasland <claude.grasland@parisgeo.cnrs.fr>
data_dictionary_newsmap_he

Seed geographical dictionary in Hebrew

Description
Seed geographical dictionary in Hebrew

Author(s)
Elad Segev <eladseg@gmail.com>

data_dictionary_newsmap_it

Seed geographical dictionary in Italian

Description
Seed geographical dictionary in Italian

Author(s)
Giuseppe Carteny <giuseppe.carteny@unimi.it>

data_dictionary_newsmap_ja

Seed geographical dictionary in Japanese

Description
Seed geographical dictionary in Japanese

Author(s)
Kohei Watanabe <watanabe.kohei@gmail.com>
<table>
<thead>
<tr>
<th>Code</th>
<th>Name</th>
<th>Description</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>data_dictionary_newmap_ru</td>
<td>Seed geographical dictionary in Russian</td>
<td>Seed geographical dictionary in Russian</td>
<td>Katerina Tertytchnaya <a href="mailto:katerina.tertytchnaya@gmail.com">katerina.tertytchnaya@gmail.com</a></td>
</tr>
<tr>
<td>data_dictionary_newmap_zh_cn</td>
<td>Seed geographical dictionary in Chinese (simplified)</td>
<td>Seed geographical dictionary in Chinese (simplified)</td>
<td>Ke Cheng <a href="mailto:kecheng.ac@gmail.com">kecheng.ac@gmail.com</a></td>
</tr>
<tr>
<td>data_dictionary_newmap_zh_tw</td>
<td>Seed geographical dictionary in Chinese (traditional)</td>
<td>Seed geographical dictionary in Chinese (traditional)</td>
<td>Chung-hong Chan <a href="mailto:chainsawtiney@gmail.com">chainsawtiney@gmail.com</a></td>
</tr>
</tbody>
</table>
predict.textmodel_newsmap

Prediction method for textmodel_newsmap

Description

Predict document class using trained a Newsmap model

Usage

```r
## S3 method for class 'textmodel_newsmap'
predict(
  object,
  newdata = NULL,
  confidence.fit = FALSE,
  rank = 1L,
  type = c("top", "all"),
  ...
)
```

Arguments

- `object`: a fitted Newsmap textmodel
- `newdata`: dfm on which prediction should be made
- `confidence.fit`: if TRUE, likelihood ratio score will be returned
- `rank`: rank of class to be predicted. Only used when `type = "top"`. 
- `type`: if top, returns the most likely class specified by `rank`; otherwise return a matrix of likelihood ratio scores for all possible classes
- `...`: not used.

print.textmodel_newsmap_summary

Print method for a fitted Newsmap model

Description

Print method for a fitted Newsmap model

Usage

```r
## S3 method for class 'textmodel_newsmap_summary'
print(x, ...)
```
Arguments

x  a fitted Newsmap textmodel
...

summary.textmodel_newsmap_accuracy

*Calculate micro and macro average measures of accuracy*

Description

This function calculates micro-average precision (p) and recall (r) and macro-average precision (P) and recall (R) based on a confusion matrix from accuracy().

Usage

```r
## S3 method for class 'textmodel_newsmap_accuracy'
summary(object, ...)
```

Arguments

object  output of accuracy()
...

Semi-supervised Bayesian multinomial model for geographical document classification

Description

Train a Newsmap model to predict geographical focus of documents using a pre-defined seed dictionary. Currently seed dictionaries are available in English (en), German (de), Spanish (es), Japanese (ja), Russian (ru) and Chinese (zh).

Usage

```r
textmodel_newmap(x, y, smooth = 1, verbose = quanteda_options("verbose"))
```

Arguments

x  dfm from which features will be extracted
y  dfm in which features will be class labels
smooth  smoothing parameter for word frequency
verbose  if TRUE, show progress of training
References


Examples

```
require(quanteda)
text_en <- c(text1 = "This is an article about Ireland.",
            text2 = "The South Korean prime minister was re-elected.")

toks_en <- tokens(text_en)
label_toks_en <- tokens_lookup(toks_en, data_dictionary_newsmap_en, levels = 3)
label_dfm_en <- dfm(label_toks_en)

feat_dfm_en <- dfm(toks_en, tolower = FALSE)

model_en <- textmodel_newsmap(feat_dfm_en, label_dfm_en)
predict(model_en)
```
Index

* data
  data_dictionary_newsmap_ar, 3
  data_dictionary_newsmap_de, 3
  data_dictionary_newsmap_en, 4
  data_dictionary_newsmap_es, 4
  data_dictionary_newsmap_fr, 4
  data_dictionary_newsmap_he, 5
  data_dictionary_newsmap_it, 5
  data_dictionary_newsmap_ja, 5
  data_dictionary_newsmap_ru, 6
  data_dictionary_newsmap_zh_cn, 6
  data_dictionary_newsmap_zh_tw, 6

  accuracy, 2
  afe, 3

  data_dictionary_newsmap_ar, 3
  data_dictionary_newsmap_de, 3
  data_dictionary_newsmap_en, 4
  data_dictionary_newsmap_es, 4
  data_dictionary_newsmap_fr, 4
  data_dictionary_newsmap_he, 5
  data_dictionary_newsmap_it, 5
  data_dictionary_newsmap_ja, 5
  data_dictionary_newsmap_ru, 6
  data_dictionary_newsmap_zh_cn, 6
  data_dictionary_newsmap_zh_tw, 6

  predict.textmodel_newsmap, 7
  print.textmodel_newsmap_summary, 7

  summary.textmodel_newsmap_accuracy, 8

  textmodel_newsmap, 8