

# Package ‘newsmap’

May 23, 2024

**Type** Package

**Title** Semi-Supervised Model for Geographical Document Classification

**Version** 0.9.0

**Maintainer** Kohei Watanabe <watanabe.kohei@gmail.com>

**Description** Semissupervised model for geographical document classification (Watanabe 2018) <doi:10.1080/21670811.2017.1293487>.

This package currently contains seed dictionaries in English, German, French, Spanish, Italian, Russian, Hebrew, Arabic, Turkish, 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** 7.3.1

**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],  
Barbara Ellynes Zucchi Nobre Silva [aut],  
Lanabi la Lova [aut],  
Lungta Seki [aut]

Repository CRAN

Date/Publication 2024-05-23 02:20:03 UTC

## R topics documented:

accuracy . . . . .	2
afe . . . . .	3
coef.textmodel_newsmap . . . . .	3
data_dictionary_newsmap_ar . . . . .	4
data_dictionary_newsmap_de . . . . .	4
data_dictionary_newsmap_en . . . . .	4
data_dictionary_newsmap_es . . . . .	5
data_dictionary_newsmap_fr . . . . .	5
data_dictionary_newsmap_he . . . . .	5
data_dictionary_newsmap_it . . . . .	6
data_dictionary_newsmap_ja . . . . .	6
data_dictionary_newsmap_pt . . . . .	6
data_dictionary_newsmap_ru . . . . .	7
data_dictionary_newsmap_tr . . . . .	7
data_dictionary_newsmap_zh_cn . . . . .	7
data_dictionary_newsmap_zh_tw . . . . .	8
predict.textmodel_newsmap . . . . .	8
summary.textmodel_newsmap_accuracy . . . . .	9
textmodel_newsmap . . . . .	9

**Index** **11**

---

accuracy	<i>Evaluate classification accuracy in precision and recall</i>
----------	---

---

### 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)
```

---

afe	<i>Compute average feature entropy (AFE)</i>
-----	--

---

**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

---

coef.textmodel_newsmap	<i>Extract coefficients for features</i>
------------------------	--

---

**Description**

Extract coefficients for features

**Usage**

```
## S3 method for class 'textmodel_newsmap'
coef(object, n = 10, select = NULL, ...)

## S3 method for class 'textmodel_newsmap'
coefficients(object, n = 10, select = NULL, ...)
```

**Arguments**

object	a Newsmap model fitted by <code>textmodel_newsmap()</code> .
n	the number of coefficients to extract.
select	returns the coefficients for the selected class; specify by the names of rows in <code>object\$model</code> .
...	not used.

---

data\_dictionary\_newsmap\_ar

*Seed geographical dictionary in Arabic*

---

**Description**

Seed geographical dictionary in Arabic

**Author(s)**

Dai Yamao <daiyamao@scs.kyushu-u.ac.jp>

---

data\_dictionary\_newsmap\_de

*Seed geographical dictionary in German*

---

**Description**

Seed geographical dictionary in German

**Author(s)**

Stefan Müller <mullers@tcd.ie>

---

data\_dictionary\_newsmap\_en

*Seed geographical dictionary in English*

---

**Description**

Seed geographical dictionary in English

**Author(s)**

Kohei Watanabe <watanabe.kohei@gmail.com>

---

data\_dictionary\_newsmap\_es

*Seed geographical dictionary in Spanish*

---

**Description**

Seed geographical dictionary in Spanish

**Author(s)**

Dani Madrid-Morales <dani.madrid@my.cityu.edu.hk>

---

data\_dictionary\_newsmap\_fr

*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>

---

data\_dictionary\_newsmap\_pt

*Seed geographical dictionary in Portuguese*

---

**Description**

Seed geographical dictionary in Portuguese

**Author(s)**

Barbara Ellynes Zucchi Nobre Silva <barbara@zucchi.science>

---

data\_dictionary\_newsmap\_ru

*Seed geographical dictionary in Russian*

---

**Description**

Seed geographical dictionary in Russian

**Author(s)**

Katerina Tertychnaya <katerina.tertychnaya@gmail.com>

Lanabi la Lova <l.lalova@lse.ac.uk>

---

data\_dictionary\_newsmap\_tr

*Seed geographical dictionary in Turkish*

---

**Description**

Seed geographical dictionary in Turkish

**Author(s)**

Lungta Seki <yahoo.co.jp0409@gmail.com>

---

data\_dictionary\_newsmap\_zh\_cn

*Seed geographical dictionary in Chinese (simplified)*

---

**Description**

Seed geographical dictionary in Chinese (simplified)

**Author(s)**

Ke Cheng <kecheng.ac@gmail.com>

---

data\_dictionary\_newsmap\_zh\_tw

*Seed geographical dictionary in Chinese (traditional)*

---

### Description

Seed geographical dictionary in Chinese (traditional)

### Author(s)

Chung-hong Chan <chainsawtiney@gmail.com>

---

predict.textmodel\_newsmap

*Prediction method for textmodel\_newsmap*

---

### Description

Predict document class using trained a Newsmap model

### Usage

```
## S3 method for class 'textmodel_newsmap'
predict(
  object,
  newdata = NULL,
  confidence = FALSE,
  rank = 1L,
  type = c("top", "all"),
  rescale = FALSE,
  min_conf = -Inf,
  min_n = 0L,
  ...
)
```

### Arguments

object	a fitted Newsmap textmodel.
newdata	dfm on which prediction should be made.
confidence	if TRUE, it returns likelihood ratio score.
rank	rank of the 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.



rescale	if TRUE, likelihood ratio scores are normalized using <code>scale()</code> . This affects both types of results.
min_conf	return NA when confidence is lower than this value.
min_n	set the minimum number of polarity words in documents.
...	not used.

---

```
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

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

### Arguments

object	output of <code>accuracy()</code>
...	not used.

---

```
textmodel_newsmap
```

*Semi-supervised Bayesian multinomial model for geographical document classification*

---

### Description

Train a Newsmap model to predict geographical focus of documents with labels given by a dictionary.

### Usage

```
textmodel_newsmap(
  x,
  y,
  label = c("all", "max"),
  smooth = 1,
  drop_label = TRUE,
  verbose = quanteda_options("verbose"),
  entropy = c("none", "global", "local", "average"),
  ...
)
```

**Arguments**

x	a dfm or fcm created by <code>quanteda::dfm()</code>
y	a dfm or a sparse matrix that record class membership of the documents. It can be created applying <code>quanteda::dfm_lookup()</code> to x.
label	if "max", uses only labels for the maximum value in each row of y.
smooth	a value added to the frequency of words to smooth likelihood ratios.
drop_label	if TRUE, drops empty columns of y and ignore their labels.
verbose	if TRUE, shows progress of training.
entropy	[experimental] the scheme to compute the entropy to regularize likelihood ratios. The entropy of features are computed over labels if <code>global</code> or over documents with the same labels if <code>local</code> . Local entropy is averaged if <code>average</code> . See the details.
...	additional arguments passed to internal functions.

**Details**

Newsmap learns association between words and classes as likelihood ratios based on the features in x and the labels in y. The large likelihood ratios tend to concentrate to a small number of features but the entropy of their frequencies over labels or documents helps to disperse the distribution.

**References**

Kohei Watanabe. 2018. "[Newsmap: semi-supervised approach to geographical news classification.](#)" *Digital Journalism* 6(3): 294-309.

**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, 4
- data\_dictionary\_newsmap\_de, 4
- data\_dictionary\_newsmap\_en, 4
- data\_dictionary\_newsmap\_es, 5
- data\_dictionary\_newsmap\_fr, 5
- data\_dictionary\_newsmap\_he, 5
- data\_dictionary\_newsmap\_it, 6
- data\_dictionary\_newsmap\_ja, 6
- data\_dictionary\_newsmap\_pt, 6
- data\_dictionary\_newsmap\_ru, 7
- data\_dictionary\_newsmap\_tr, 7
- data\_dictionary\_newsmap\_zh\_cn, 7
- data\_dictionary\_newsmap\_zh\_tw, 8

accuracy, 2

afe, 3

coef.textmodel\_newsmap, 3

coefficients.textmodel\_newsmap  
(coef.textmodel\_newsmap), 3

- data\_dictionary\_newsmap\_ar, 4
- data\_dictionary\_newsmap\_de, 4
- data\_dictionary\_newsmap\_en, 4
- data\_dictionary\_newsmap\_es, 5
- data\_dictionary\_newsmap\_fr, 5
- data\_dictionary\_newsmap\_he, 5
- data\_dictionary\_newsmap\_it, 6
- data\_dictionary\_newsmap\_ja, 6
- data\_dictionary\_newsmap\_pt, 6
- data\_dictionary\_newsmap\_ru, 7
- data\_dictionary\_newsmap\_tr, 7
- data\_dictionary\_newsmap\_zh\_cn, 7
- data\_dictionary\_newsmap\_zh\_tw, 8

predict.textmodel\_newsmap, 8

quanteda::dfm(), 10

quanteda::dfm\_lookup(), 10

scale(), 9

summary.textmodel\_newsmap\_accuracy, 9

textmodel\_newsmap, 9

textmodel\_newsmap(), 3