

Package ‘GMTM’

September 15, 2026

Title Gaussian Mixture Text Models for Topic Analysis

Version 0.1.0

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

Description Gaussian mixture models and k-means for topic analysis of dense document vectors. The underlying clustering functions rely on the Armadillo library.

License Apache License (>= 2.0)

Encoding UTF-8

Depends R (>= 3.5.0)

Imports quanteda (>= 4.0.0), proxyC, wordvector, Rcpp, RcppArmadillo

Suggests testthat, spelling, rmarkdown, knitr

LinkingTo Rcpp, RcppArmadillo (>= 0.7.600.1.0), quanteda, testthat

Language en-US

Config/roxygen2/version 8.0.0

NeedsCompilation yes

Author Kohei Watanabe [aut, cre, cph] (ORCID:
<<https://orcid.org/0000-0001-6519-5265>>)

Repository CRAN

Date/Publication 2026-09-15 11:20:09 UTC

Contents

terms	2
textmodel_gmm	2
textmodel_kmeans	3
topics	4
Index	5

terms	<i>Extract words for topics from documents</i>
-------	--

Description

Identify distinctive words for each topic by applying TF-IDF weights to the original [quanteda::dfm](#).

Usage

```
terms(x, data, n = 10, ...)
```

Arguments

x	a fitted model.
data	a quanteda::dfm or quanteda::tokens from which words are extracted for each topic.
n	the number of topic words.
...	passed to functions.

Details

To identify distinctive words for topics, original documents must be provided along with a fitted model because the information about individual words are lost in document vectors. The documents in data is grouped by topic and weighted by TF-IDF to select the most distinctive words for each topic. This technique is commonly known as c-TF-IDF.

Value

Returns a character matrix with the most distinctive words for each topic.

textmodel_gmm	<i>Gaussian mixture model for topic analysis</i>
---------------	--

Description

Gaussian mixture model for clustering of document vectors based on the Armadillo library.

Usage

```
textmodel_gmm(
  x,
  k = 10,
  model = NULL,
  ...,
  verbose = quanteda_options("verbose")
)
```

Arguments

x	a <code>wordvector::textmodel_doc2vec</code> or a dense matrix of document vectors in the rows.
k	the number of topics to identify.
model	a fitted model from which initial centroids are extracted.
...	passed to the underlying function.
verbose	print the progress if TRUE.

Details

User can change the number of threads for the parallel computing via `options(GMTM.threads)` or `OMP_THREAD_LIMIT` in the environmental variable.

Value

Returns a fitted `textmodel_gmm` object.

Examples

```
# dummy document vectors with 50 dimensions
mat <- t(replicate(1000, rnorm(50)))
gmm <- textmodel_gmm(mat, k = 10)
table(topics(gmm))
```

textmodel_kmeans	<i>K-means for topic analysis</i>
------------------	-----------------------------------

Description

K-means clustering of document vectors based on the Armadillo library.

Usage

```
textmodel_kmeans(
  x,
  k = 10,
  model = NULL,
  ...,
  verbose = quanteda_options("verbose")
)
```

Arguments

x	a <code>wordvector::textmodel_doc2vec</code> or a dense matrix of document vectors in the rows.
k	the number of topics to identify.
model	a fitted model from which initial centroids are extracted.
...	passed to the underlying function.
verbose	print the progress if TRUE.

Value

Returns a fitted `textmodel_kmeans` object.

Examples

```
# dummy document vectors with 50 dimensions
mat <- t(replicate(1000, rnorm(50)))
km <- textmodel_kmeans(mat, k = 10)
table(topics(km))
```

topics	<i>Extract topics of documents</i>
--------	------------------------------------

Description

Extract topics of documents

Usage

```
topics(x, ...)
```

Arguments

x	a fitted model.
...	not used.

Value

Returns predicted topics as a vector.

Index

quanteda::dfm, [2](#)
quanteda::tokens, [2](#)

terms, [2](#)
textmodel_gmm, [2](#)
textmodel_kmeans, [3](#)
topics, [4](#)

wordvector::textmodel_doc2vec, [3](#), [4](#)