

Package: gsdmm (via r-universe)

September 20, 2024

Type Package

Title GSDMM Short Text Clustering via Dirichlet Mixture Models

Version 0.0.6

Maintainer Till Tietz <ttietz2014@gmail.com>

Description This package implements a Dirichlet Mixture Model and accompanying Gibbs sampler for short text clustering proposed by Yin and Wang 2014.

License MIT + file LICENSE

Imports fastmatch, Matrix, Rcpp, vctrs

Suggests testthat (>= 3.0.0)

LinkingTo Rcpp, RcppArmadillo, RcppThread

Config/testthat/edition 3

Encoding UTF-8

RoxygenNote 7.3.2

Repository <https://paithiov909.r-universe.dev>

RemoteUrl <https://github.com/paithiov909/gsdmm>

RemoteRef HEAD

RemoteSha 589020057575447572ff983ba31668344df82c33

Contents

gsdmm-package	2
gsdmm	2
Index	4

 gsdmm-package

gsdmm: GSDMM Short Text Clustering via Dirichlet Mixture Models

Description

This package implements a Dirichlet Mixture Model and accompanying Gibbs sampler for short text clustering proposed by Yin and Wang 2014.

Author(s)

Maintainer: Till Tietz <ttietz2014@gmail.com>

Other contributors:

- Akiru Kato [contributor]

 gsdmm

Fit Dirichlet Mixture Model

Description

Fit Dirichlet Mixture Model

Usage

```
gsdmm(
  texts,
  n_iter = 30L,
  n_clust = 8L,
  alpha = 0.1,
  beta = 0.1,
  progress = TRUE
)
```

Arguments

texts	a list of character vectors
n_iter	integer number of iterations to run gibbs sampler for
n_clust	integer upper bound on number of clusters. The returned number of cluster will be smaller or equal to n_clust
alpha	double governing the probability of assigning a text to a currently empty cluster (larger alpha means higher probability).
beta	double governing the tradeoff between cluster size and fit. Smaller betas make clustering more sensitive to congruence between cluster-word and document-word distributions while larger betas make it more sensitive to cluster size.
progress	logical indicating whether to print progress bar.

gsdmm

3

Value

a list that contains an integer vector of clusters and a word-cluster matrix.

Index

[gsdmm](#), [2](#)

[gsdmm-package](#), [2](#)