

Package ‘textmatch’

March 25, 2019

Title Toolkit for Matching Textual Data and Evaluating Textual Similarity

Version 0.0.0.9000

Description What the package does (one paragraph).

Depends R (>= 3.5.2)

License What license is it under?

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1.9000

Imports dplyr,
data.table,
quanteda

Suggests knitr,
rmarkdown

VignetteBuilder knitr

R topics documented:

get_pair_distances	1
get_similarity_scores	2
textmatch	2

Index	4
--------------	----------

get_pair_distances	<i>Similarity and distance computation between documents or features</i>
--------------------	--

Description

These functions compute distance matrices from a text representation where each row is a document and each column is a feature to measure distance over based on treatment indicator Z

Usage

```
get_pair_distances(dat, Z, include = c("cosine", "jaccard", "euclidean",  
  "mahalanobis", "propensity"), exclude = NULL, docnames = NULL,  
  verbose = FALSE)
```

Arguments

Z	A logical or binary vector indicating treatment and control for each unit in the study. TRUE or 1 represents a treatment unit, FALSE or 0 represents a control unit.
docnames	A vector of document names equal in length to the number of documents
x	a valid quanteda dfm object

Value

A matrix showing pairwise distances for all potential matches of treatment and control units under various distance metrics

`get_similarity_scores` *This function calculates an input character vector's similarity matrix according to the measures contained in the predictive model.*

Description

This function calculates an input character vector's similarity matrix according to the measures contained in the predictive model.

Usage

```
get_similarity_scores(x)
```

Arguments

x	A character vector where each element is a document
---	---

Value

A data frame of rows ($n * n-1$) and columns 16; each column is one of the constituent similarity measures

`textmatch` *This function runs the main ML model as specified in Mozer et al. (2018)*

Description

This function runs the main ML model as specified in Mozer et al. (2018)

Usage

```
textmatch(x, outcome = "matrix")
```

Arguments

x	A character vector where each element is a document
---	---

Value

An n by n matrix where n is the length of parameter x. Each entry is a standardized similarity score.

Examples

```
textmatch(c("I am a dog", "I am a cat", "The rain in Spain falls mainly on the plain."),  
outcome = "matrix")
```

Index

`get_pair_distances`, [1](#)
`get_similarity_scores`, [2](#)
`textmatch`, [2](#)