Package ‘meanr’

July 19, 2019

Type Package

Title Sentiment Analysis Scorer

Version 0.1-2

Description Sentiment analysis is a popular technique in text mining that attempts to determine the emotional state of some text. We provide a new implementation of a common method for computing sentiment, whereby words are scored as positive or negative according to a dictionary lookup. Then the sum of those scores is returned for the document. We use the 'Hu' and 'Liu' sentiment dictionary ('Hu' and 'Liu', 2004) <doi:10.1145/1014052.1014073> for determining sentiment. The scoring function is 'vectorized' by document, and scores for multiple documents are computed in parallel via 'OpenMP'.

License BSD 2-clause License + file LICENSE

Depends R (>= 3.0.0)

LazyData yes

LazyLoad yes

NeedsCompilation yes

ByteCompile yes

Maintainer Drew Schmidt <`wrathematics@gmail.com`>

URL `https://github.com/wrathematics/meanr`

BugReports `https://github.com/wrathematics/meanr/issues`

RoxygenNote 6.1.1

Author Drew Schmidt [aut, cre]

Repository CRAN

Date/Publication 2019-07-19 19:40:02 UTC

R topics documented:

meanr-package *meanr: Sentiment Analysis Scorer*

Description

Sentiment analysis is a popular technique in text mining that attempts to determine the emotional state of some text. We provide a new implementation of a common method for computing sentiment, whereby words are scored as positive or negative according to a dictionary lookup. Then the sum of those scores is returned for the document. We use the Hu and Liu sentiment dictionary (Hu and Liu, 2004) <doi:10.1145/1014052.1014073> for determining sentiment. The scoring function is vectorized by document, and scores for multiple documents are computed in parallel via OpenMP.

Author(s)

Drew Schmidt `<wrathematics AT gmail.com>`

meanr.nthreads *meanr.nthreads*

Description

Returns the number of cores + hyperthreads on the system. The function respects the environment variable `OMP_NUM_THREADS`.

Usage

```
meanr.nthreads()
```

Value

The number of cores + hyperthreads on the system (an integer).

score *score*

Description

Computes the sentiment score, the sum of the total number of positive and negative scored words. The function is vectorized so that it will return one row per string. The scoring function ignores (upper/lower) case and punctuation.

Usage

```
score(s, nthreads = meanr.nthreads())
```

Arguments

`s` A string or vector of strings.

`nthreads` Number of threads to use. By default it will use the total number of cores + hyperthreads.

Details

The scoring function uses OpenMP to process text in parallel.

The function uses the Hu and Liu sentiment dictionary (same as everybody else) available here: https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html

Value

A dataframe, consisting of columns "positive", "negative", "score", and "wc". With the exception of "score", these are counts; that is, "positive" is the number of positive sentiment words, "negative" is the number of negative sentiment words, and "wc" is the wordcount (total number of words).

References

Hu, M., & Liu, B. (2004). Mining opinion features in customer reviews. National Conference on Artificial Intelligence.

Examples

```
library(meanr)
s1 = "Abundance abundant accessable."
s2 = "Banana apple orange."
s3 = "Abnormal abolish abominable."
s = c(s1, s2, s3)

# as separate 'documents'
score(s, nthreads=1)

# as one document
score(paste0(s, collapse=" "), nthreads=1)
```