

Package ‘TFMPvalue’

January 20, 2025

Version 0.0.9

Date 2022-10-19

Title Efficient and Accurate P-Value Computation for Position Weight Matrices

Description In putative Transcription Factor Binding Sites (TFBSs) identification from sequence/alignments, we are interested in the significance of certain match score. TFMPvalue provides the accurate calculation of P-value with score threshold for Position Weight Matrices, or the score with given P-value. It is an interface to code originally made available by Helene Touzet and Jean-Stephane Varre, 2007, Algorithms Mol Biol:2, 15. <[doi:10.1186/1748-7188-2-15](https://doi.org/10.1186/1748-7188-2-15)>.

Author Ge Tan <ge_tan@live.com>

Maintainer Ge Tan <ge_tan@live.com>

Copyright 2007 LIFL-USTL-INRIA

Imports Rcpp(>= 0.11.1)

Depends R (>= 3.0.1)

Suggests testthat

LinkingTo Rcpp

License GPL-2

URL <https://github.com/ge11232002/TFMPvalue>

BugReports <https://github.com/ge11232002/TFMPvalue/issues>

Type Package

NeedsCompilation yes

SystemRequirements C++11

Collate TFMPvalue-sc2pv.R TFMPvalue-pv2sc.R TFMPvalue-lazyScore.R
util.R

Repository CRAN

Date/Publication 2022-10-21 11:55:14 UTC

Contents

| | |
|-----------------------------|----------|
| TFMPvalue-package | 2 |
| TFMLazyScore | 2 |
| TFMpv2sc | 3 |
| TFMsc2pv | 4 |
| Index | 6 |

| | |
|-------------------|--|
| TFMPvalue-package | <i>Efficient and accurate P-value computation for Position Weight Matrices</i> |
|-------------------|--|

Description

This package provides a novel algorithm that solves the P-value calculation problem given the score based on a Position Weight Matrices (PWMs), or the reverse problem: finding the score give the desired P-value. This package is an interface to code originally made available by Helene Touzet and Jean-Stephane Varre, 2007, Algorithms Mol Biol:2, 15.

Details

The original code is taken from <http://bioinfo.lifl.fr/TFM/TFMpv2sc/TFM-Pvalue.tar.gz>, retrived 26/03/2014.

The algorithm is described in Touzet, H., and Varre, J.-S. (2007). Efficient and accurate P-value computation for Position Weight Matrices. Algorithms Mol Biol 2, 15.

Author(s)

Ge Tan

| | |
|--------------|--|
| TFMLazyScore | <i>Compute the score from P-value.</i> |
|--------------|--|

Description

Computes the score threshold associated with P-value p using the algorithm of Beckstette 2006.

Usage

```
TFMLazyScore(mat, pvalue, bg=c(A=0.25, C=0.25, G=0.25, T=0.25),
              type=c("PFM", "PWM"), granularity=1e-5)
```

Arguments

| | |
|-------------|---|
| mat | The input matrix. It can be a Position Frequency Matrix (PFM) or Position Weight Matrix (PWM) in log ratio. The matrix must have row names with "A", "C", "G", "T". |
| pvalue | The required P-value. |
| bg | The background frequency of the sequences. A numeric vector with names "A", "C", "G", "T". |
| type | The type of input matrix. Can be "PFM" or "PWM". |
| granularity | The granularity used in the computation. |

Value

The score is returned based on the matrix, given P-value and granularity.

Author(s)

Ge Tan

Examples

```
## This example is not tested due to running time > 5s
pfm <- matrix(c(3, 5, 4, 2, 7, 0, 3, 4, 9, 1, 1, 3, 3, 6, 4, 1, 11,
                0, 3, 0, 11, 0, 2, 1, 11, 0, 2, 1, 3, 3, 2, 6, 4, 1,
                8, 1, 3, 4, 6, 1, 8, 5, 1, 0, 8, 1, 4, 1, 9, 0, 2, 3,
                9, 5, 0, 0, 11, 0, 3, 0, 2, 7, 0, 5),
              nrow = 4, dimnames = list(c("A", "C", "G", "T")))
)
bg <- c(A=0.25, C=0.25, G=0.25, T=0.25)
pvalue <- 1e-5
type <- "PFM"
granularity <- 1e-5
TFMLazyScore(pfm, pvalue, bg, type, granularity)
```

TFMpv2sc

Compute score from P-value.

Description

Computes the score threshold associated with a P-value.

Usage

```
TFMpv2sc(mat, pvalue, bg=c(A=0.25, C=0.25, G=0.25, T=0.25),
         type=c("PFM", "PWM"))
```

Arguments

| | |
|--------|---|
| mat | The input matrix. It can be a Position Frequency Matrix (PFM) or Position Weight Matrix (PWM) in log ratio. The matrix must have row names with "A", "C", "G", "T". |
| pvalue | The required P-value. |
| bg | The background frequency of the sequences. A numeric vector with names "A", "C", "G", "T". |
| type | The type of input matrix. Can be "PFM" or "PWM". |

Value

The score is returned based on the matrix, given P-value.

Author(s)

Ge Tan

References

Touzet, H., and Varre, J.-S. (2007). Efficient and accurate P-value computation for Position Weight Matrices. *Algorithms Mol Biol* 2, 15.

Examples

```

pfm <- matrix(c(3, 5, 4, 2, 7, 0, 3, 4, 9, 1, 1, 3, 3, 6, 4, 1, 11,
               0, 3, 0, 11, 0, 2, 1, 11, 0, 2, 1, 3, 3, 2, 6, 4, 1,
               8, 1, 3, 4, 6, 1, 8, 5, 1, 0, 8, 1, 4, 1, 9, 0, 2, 3,
               9, 5, 0, 0, 11, 0, 3, 0, 2, 7, 0, 5),
              nrow = 4, dimnames = list(c("A", "C", "G", "T")))
)
bg <- c(A=0.25, C=0.25, G=0.25, T=0.25)
pvalue <- 1e-5
type <- "PFM"
score <- TFMpv2sc(pfm, pvalue, bg, type)

```

TFMsc2pv

Compute P-value from score.

Description

Computes the P-value associated with a score threshold.

Usage

```
TFMsc2pv(mat, score, bg=c(A=0.25, C=0.25, G=0.25, T=0.25),
         type=c("PFM", "PWM"))
```

Arguments

| | |
|-------|---|
| mat | The input matrix. It can be a Position Frequency Matrix (PFM) or Position Weight Matrix (PWM) in log ratio. The matrix must have row names with "A", "C", "G", "T". |
| score | The required score. |
| bg | The background frequency of the sequences. A numeric vector with names "A", "C", "G", "T". |
| type | The type of input matrix. Can be "PFM" or "PWM". |

Value

The P-value is returned based on the matrix, given the desired score.

Author(s)

Ge Tan

References

Touzet, H., and Varre, J.-S. (2007). Efficient and accurate P-value computation for Position Weight Matrices. *Algorithms Mol Biol* 2, 15.

Examples

```
pfm <- matrix(c(3, 5, 4, 2, 7, 0, 3, 4, 9, 1, 1, 3, 3, 6, 4, 1, 11,
               0, 3, 0, 11, 0, 2, 1, 11, 0, 2, 1, 3, 3, 2, 6, 4, 1,
               8, 1, 3, 4, 6, 1, 8, 5, 1, 0, 8, 1, 4, 1, 9, 0, 2, 3,
               9, 5, 0, 0, 11, 0, 3, 0, 2, 7, 0, 5),
              nrow = 4, dimnames = list(c("A", "C", "G", "T")))
bg <- c(A=0.25, C=0.25, G=0.25, T=0.25)
score <- 8.77
type <- "PFM"
pvalue <- TFMsc2pv(pfm, score, bg, type)
```

Index

* **package**

TFMPvalue-package, [2](#)

TFMLazyScore, [2](#)

TFMpv2sc, [3](#)

TFMPvalue-package, [2](#)

TFMsc2pv, [4](#)