Package ‘GOSemSim’

January 8, 2016

Type Package
Title GO-terms Semantic Similarity Measures
Version 1.28.2
Author Guangchuang Yu <guangchuangyu@gmail.com> with contributions from
Alexey Stukalov.
Maintainer Guangchuang Yu <guangchuangyu@gmail.com>
Description Implemented five methods proposed by Resnik, Schlicker,
Jiang, Lin and Wang respectively for estimating GO semantic
similarities. Support many species, including Anopheles,
Arabidopsis, Bovine, Canine, Chicken, Chimp, Coelicolor, E coli
strain K12 and Sakai, Fly, Gondii, Human, Malaria, Mouse, Pig, Rhesus,
Rat, Worm, Xenopus, Yeast, and Zebrafish.
Depends R (>= 3.1.0)
LinkingTo Rcpp
Imports Rcpp, AnnotationDbi, GO.db
Suggests DOSE, clusterProfiler, org.Hs.eg.db, knitr, BiocStyle,
BiocInstaller
VignetteBuilder knitr
License Artistic-2.0
URL https://github.com/GuangchuangYu/GOSemSim
BugReports https://github.com/GuangchuangYu/GOSemSim/issues
biocViews Annotation, GO, Clustering, Pathways, Network, Software
NeedsCompilation yes

R topics documented:

GOSemSim-package .................................................. 2
clusterSim ........................................................... 3
combineScores ......................................................... 4
geneSim .............................................................. 4
Implementation of semantic similarity measures to estimate the functional similarities among Gene Ontology terms and gene products

### Details

Quantitative measure of functional similarities among gene products is important for post-genomics study, and widely used in gene function prediction, cluster analysis and pathway modeling. This package is designed to estimate the GO terms' and genes' semantic similarities. Implemented five methods proposed by Resnik, Schlicker, Jiang, Lin and Wang respectively. Support many species, including Anopheles, Arabidopsis, Bovine, Canine, Chicken, Chimp, E coli strain K12 and strain Sakai, Fly, Human, Malaria, Mouse, Pig, Rhesus, Rat, Worm, Xenopus, Yeast, Zebrafish.

**Package:** GOSemSim
**Type:** Package
**Version:** 1.24.0
**Date:** 09-11-2012
**biocViews:** GO, Clustering, Pathways, Anopheles_gambiae, Arabidopsis_thaliana, Bos_taurus, Caenorhabditis_elegans, Can
dependencies:
**Imports:** methods, AnnotationDbi, GO.db, org.Hs.eg.db, org.Ag.eg.db, org.At.tair.db, org.Bt.eg.db, org.Ce.eg.db, org.Cf.eg.db, org.Dm.eg.db, org.Re.eg.db, org.Sc.eg.db, org.Sc.scdb
**Suggests:** clusterProfiler
**License:** Artistic-2.0

**Author(s)**

Guangchuang Yu
Maintainer: Guangchuang Yu <guangchuangyu@gmail.com>
References


See Also
gosim mgosim geneSim mgeneSim clusterSim mclusterSim

clusterSim

Semantic Similarity Between Two Gene Clusters

Description

Given two gene clusters, this function calculates semantic similarity between them.

Usage

clusterSim(cluster1, cluster2, ont = "MF", organism = "human", measure = "Wang", drop = "IEA", combine = "BMA")

Arguments

cluster1 A set of gene IDs.
cluster2 Another set of gene IDs.
ont One of "MF", "BP", and "CC" subontologies.
measure One of "Resnik", "Lin", "Rel", "Jiang" and "Wang" methods.
drop A set of evidence codes based on which certain annotations are dropped. Use NULL to keep all GO annotations.
combine One of "max", "average", "rcmax", "BMA" methods, for combining semantic similarity scores of multiple GO terms associated with protein or multiple proteins associated with protein cluster.

Value

similarity

References

See Also
gosim mgosim geneSim mgenesim mclustersim

Examples

```r
# cluster1 <- c("835", "5261", "241", "994")
# cluster2 <- c("307", "308", "317", "321", "506", "540", "378", "388", "396")
# clusterSim(cluster1, cluster2, ont="MF", organism="human", measure="Wang")
```

Description

Functions for combining similarity matrix to similarity score

Usage

```r
combineScores <- function(simScores, combine) {
  similarity_score <- 
  return(similarity_score)
}
```

Arguments

- `simScores`: similarity matrix
- `combine`: combine method

Value

similarity value

Author(s)

Guangchuang Yu [http://ygc.name](http://ygc.name)

geneSim

**Semantic Similarity Between two Genes**

Description

Given two genes, this function will calculate the semantic similarity between them, and return their semantic similarity and the corresponding GO terms

Usage

```r
geneSim(gene1, gene2, ont = "MF", organism = "human", measure = "Wang",
        drop = "IEA", combine = "BMA")
```
getDb

Arguments

gene1  Entrez gene id.
gene2  Another entrez gene id.
ont   One of "MF", "BP", and "CC" subontologies.
measure One of "Resnik", "Lin", "Rel", "Jiang" and "Wang" methods.
drop  A set of evidence codes based on which certain annotations are dropped. Use NULL to keep all GO annotations.
combine One of "max", "average", "rcmax", "BMA" methods, for combining semantic similarity scores of multiple GO terms associated with protein or multiple proteins associated with protein cluster.

Value

list of similarity value and corresponding GO.

References


See Also

gosim mgosim mgenesim clustersim mclustersim

Examples

geneSim("241", "251", ont="MF", organism="human", measure="Wang")

getDb

db <- getDb

db

db <- getDb

db

Description

mapping organism name to annotationDb package name

Usage

db <- getDb(organism)

Arguments

organism  one of supported organism
Value
annotationDb name

Author(s)
Yu Guangchuang

getSupported_Org

description
get supported organisms

Usage
getSupported_Org()

Value
supported organisms

Author(s)
Yu Guangchuang

goSim

Semantic Similarity Between Two GO Terms

Description
Given two GO IDs, this function calculates their semantic similarity.

Usage
goSim(GOID1, GOID2, ont = "MF", organism = "human", measure = "Wang")

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOID1</td>
<td>GO ID 1.</td>
</tr>
<tr>
<td>GOID2</td>
<td>GO ID 2.</td>
</tr>
<tr>
<td>ont</td>
<td>One of &quot;MF&quot;, &quot;BP&quot;, and &quot;CC&quot; subontologies.</td>
</tr>
<tr>
<td>measure</td>
<td>One of &quot;Resnik&quot;, &quot;Lin&quot;, &quot;Rel&quot;, &quot;Jiang&quot; and &quot;Wang&quot; methods.</td>
</tr>
</tbody>
</table>
**IC**

Value

similarity

References


See Also

mgoSim geneSim mgeneSim clusterSim mclusterSim

Examples

```r
goSim("GO:0004022", "GO:0005515", ont="MF", measure="Wang")
```

---

**IC**

*Information content of GO terms*

Description

These datasets are the information contents of GO terms.

References


---

**infoContentMethod**

*information content based methods*

Description

Information Content Based Methods for semantic similarity measuring

Usage

```r
infoContentMethod(ID1, ID2, ont = "DO", method, organism = "human")
```
**Arguments**

- **ID1**: Ontology Term
- **ID2**: Ontology Term
- **ont**: Ontology
- **method**: one of "Resnik", "Jiang", "Lin" and "Rel".
- **organism**: one of supported species

**Details**

implemented for methods proposed by Resnik, Jiang, Lin and Schlicker.

**Value**

semantic similarity score

**Author(s)**

Guangchuang Yu [http://ygc.name](http://ygc.name)

---

**Description**

loading GOMap to GOSemSimEnv

**Usage**

`loadGOMap(organism)`

**Arguments**

- **organism**: one of supported organisms

**Value**

`envir`

**Author(s)**

Yu Guangchuang
**loadICdata**

*Load IC data*

**Description**

Load Information Content data to DOSEEnv environment

**Usage**

```r
loadICdata(organismL ont)
```

**Arguments**

- `organism` (character)
- `ont` (character)

**Author(s)**

Guangchuang Yu [http://ygc.name](http://ygc.name)

---

**mclusterSim**

*Pairwise Semantic Similarities for a List of Gene Clusters*

**Description**

Given a list of gene clusters, this function calculates pairwise semantic similarities.

**Usage**

```r
mclusterSim(clustersL ont = "MF", organism = "human", measure = "Wang",
          drop = "IEA", combine = "BMA")
```

**Arguments**

- `clusters` (list)
- `ont` (character)
- `organism` (character)
- `measure` (character)
- `drop` (character)
- `combine` (character)

**Description**

One of "MF", "BP", and "CC" subontologies.


One of "Resnik", "Lin", "Rel", "Jiang" and "Wang" methods.

A set of evidence codes based on which certain annotations are dropped. Use NULL to keep all GO annotations.

One of "max", "average", "rcmax", "BMA" methods, for combining semantic similarity scores of multiple GO terms associated with protein or multiple proteins associated with protein cluster.
mgeneSim

Value
similar similarity matrix

References

See Also
goSim mgoSim geneSim mgeneSim clusterSim

Examples
```r
## cluster1 <- c("835", "5261","241")
## cluster2 <- c("578","582")
## cluster3 <- c("307", "308", "317")
## clusters <- list(a=cluster1, b=cluster2, c=cluster3)
## mclustSim(clusters, ont="MF", organism="human", measure="Wang")
```

---

mgeneSim *Pairwise Semantic Similarity for a List of Genes*

Description
Given a list of genes, this function calculates pairwise semantic similarities.

Usage
```
mgeneSim(genes, ont = "MF", organism = "human", measure = "Wang", drop = "IEA", combine = "BMA", verbose = TRUE)
```

Arguments
genes A list of entrez gene IDs.
ont One of "MF", "BP", and "CC" subontologies.
measure One of "Resnik", "Lin", "Rel", "Jiang" and "Wang" methods.
drop A set of evidence codes based on which certain annotations are dropped. Use NULL to keep all GO annotations.
combine One of "max", "average", "rcmax", "BMA" methods, for combining semantic similarity scores of multiple GO terms associated with protein or multiple proteins associated with protein cluster.
verbose show progress bar or not.
**mgoSim**  

**Value**  

similarity matrix

**References**  

http://bioinformatics.oxfordjournals.org/cgi/content/abstract/26/7/976 PMID: 20179076

**See Also**  

`goSim` `mgoSim` `geneSim` `clusterSim` `mclusterSim`

**Examples**  

```r
mGOsim(c("835", "5261","241"), ont="MF", organism="human", measure="Wang")
```

---

**mgoSim**  

*Semantic Similarity Between two GO terms lists*

**Description**  

Given two GO term sets, this function will calculate the semantic similarity between them, and return their semantic similarity

**Usage**  

```r
mgoSim(G01, G02, ont = "MF", organism = "human", measure = "Wang", combine = "BMA")
```

**Arguments**  

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>G01</td>
<td>A set of go terms.</td>
</tr>
<tr>
<td>G02</td>
<td>Another set of go terms.</td>
</tr>
<tr>
<td>ont</td>
<td>One of &quot;MF&quot;, &quot;BP&quot;, and &quot;CC&quot; subontologies.</td>
</tr>
<tr>
<td>measure</td>
<td>One of &quot;Resnik&quot;, &quot;Lin&quot;, &quot;Rel&quot;, &quot;Jiang&quot; and &quot;Wang&quot; methods.</td>
</tr>
<tr>
<td>combine</td>
<td>One of &quot;max&quot;, &quot;average&quot;, &quot;rcmax&quot;, &quot;BMA&quot; methods, for combining semantic similarity scores of multiple GO terms associated with protein or multiple proteins associated with protein cluster.</td>
</tr>
</tbody>
</table>

**Value**  

similarity
References

See Also
goSim geneSim mgeneSim clusterSim mclusterSim

Examples
```r
go1 <- c("GO:0004022", "GO:0004024", "GO:0004023")
go2 <- c("GO:0009055", "GO:0020037")
mgoSim("GO:0003824", go2, measure="Wang")
mgoSim(go1, go2, ont="MF", organism="human", measure="Wang")
```

Description
measuring similarities between two term vectors.

Usage
termSim(t1, t2, method = c("Wang", "Resnik", "Rel", "Jiang", "Lin"),
organism = "human", ont = "BP")

Arguments
t1 term vector
t2 term vector
method one of "Wang", "Resnik", "Rel", "Jiang", and "Lin".
organism about 20 species supported, please refer to the vignettes
ont ontology

Details
provide two term vectors, this function will calculate their similarities.

Value
score matrix

Author(s)
Guangchuang Yu http://ygc.name
wangMethod

Description
Method Wang for semantic similarity measuring

Usage
wangMethod(ID1, ID2, ont = "BP")

Arguments
ID1 Ontology Term
ID2 Ontology Term
ont Ontology

Value
semantic similarity score

Author(s)
Guangchuang Yu http://ygc.name
Index

∗Topic datasets
IC, 7

∗Topic manip
  clustersim, 3
genesim, 4
goSim, 6
mclustersim, 9
mgenesim, 10
mgosim, 11

∗Topic package
  GOSemSim-package, 2

clustersim, 3, 3, 5, 7, 10–12
combineScores, 4
genSim, 3, 4, 4, 7, 10–12
getDb, 5
getSupported_Org, 6
GOSemSim/GOSemSim-package, 2
GOSemSim-package, 2
goSim, 3–5, 6, 10–12

IC, 7

Info_Contents_anopheles_BP (IC), 7
Info_Contents_anopheles_CC (IC), 7
Info_Contents_anopheles_MF (IC), 7
Info_Contents_arabidopsis_BP (IC), 7
Info_Contents_arabidopsis_CC (IC), 7
Info_Contents_arabidopsis_MF (IC), 7
Info_Contents_bovine_BP (IC), 7
Info_Contents_bovine_CC (IC), 7
Info_Contents_bovine_MF (IC), 7
Info_Contents_canine_BP (IC), 7
Info_Contents_canine_CC (IC), 7
Info_Contents_canine_MF (IC), 7
Info_Contents_chicken_BP (IC), 7
Info_Contents_chicken_CC (IC), 7
Info_Contents_chicken_MF (IC), 7
Info_Contents_chimp_BP (IC), 7
Info_Contents_chimp_CC (IC), 7
Info_Contents_chimp_MF (IC), 7
Info_Contents_coelicolor_BP (IC), 7
Info_Contents_coelicolor_CC (IC), 7
Info_Contents_coelicolor_MF (IC), 7
Info_Contents_ecolik12_BP (IC), 7
Info_Contents_ecolik12_CC (IC), 7
Info_Contents_ecolik12_MF (IC), 7
Info_Contents ecsakai_BP (IC), 7
Info_Contents ecsakai_CC (IC), 7
Info_Contents ecsakai_MF (IC), 7
Info_Contents_fly_BP (IC), 7
Info_Contents_fly_CC (IC), 7
Info_Contents_fly_MF (IC), 7
Info_Contents_gondii_BP (IC), 7
Info_Contents_gondii_CC (IC), 7
Info_Contents_gondii_MF (IC), 7
Info_Contents_human_BP (IC), 7
Info_Contents_human_CC (IC), 7
Info_Contents_human_MF (IC), 7
Info_Contents_malaria_BP (IC), 7
Info_Contents_malaria_CC (IC), 7
Info_Contents_malaria_MF (IC), 7
Info_Contents_mouse_BP (IC), 7
Info_Contents_mouse_CC (IC), 7
Info_Contents_mouse_MF (IC), 7
Info_Contents_pig_BP (IC), 7
Info_Contents_pig_CC (IC), 7
Info_Contents_pig_MF (IC), 7
Info_Contents_rat_BP (IC), 7
Info_Contents_rat_CC (IC), 7
Info_Contents_rat_MF (IC), 7
Info_Contents_rhesus_BP (IC), 7
Info_Contents_rhesus_CC (IC), 7
Info_Contents_rhesus_MF (IC), 7
Info_Contents_worm_BP (IC), 7
Info_Contents_worm_CC (IC), 7
Info_Contents_worm_MF (IC), 7
Info_Contents_xenopus_BP (IC), 7
Info_Contents_xenopus_CC (IC), 7

14
Info_Contents_xenopus_MF (IC), 7
Info_Contents_yeast_BP (IC), 7
Info_Contents_yeast_CC (IC), 7
Info_Contents_yeast_MF (IC), 7
Info_Contents_zebrafish_BP (IC), 7
Info_Contents_zebrafish_CC (IC), 7
Info_Contents_zebrafish_MF (IC), 7
infoContentMethod, 7
loadGOMap, 8
loadICdata, 9

mclusterSim, 3–5, 7, 9, 11, 12
mgeneSim, 3–5, 7, 10, 10, 12
mgoSim, 3–5, 7, 10, 11, 11

termSim, 12

wangMethod, 13