Package ‘GOSemSim’

October 9, 2013

Type Package

Title GO-terms Semantic Similarity Measures

Version 1.18.0

Author Guangchuang Yu

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, Human, Malaria, Mouse, Pig, Rhesus, Rat, Worm, Xenopus, Yeast, and Zebrafish.

Depends R (>= 2.10), Rcpp

LinkingTo Rcpp

Imports methods, AnnotationDbi, GO.db, org.Hs.eg.db

Suggests DOSE, clusterProfiler, BiocInstaller

biocViews GO, Clustering, Pathways, NetworkAnalysis

Collate 'clusterSim.R' 'combineMethods.R' 'computeIC.R' 'ICMethods.R' 'gene2GO.R' 'geneSim.R' 'goSim.R' 'termSim.R' 'WangMethod.R'

License GPL-2

URL http://bioinformatics.oxfordjournals.org/content/26/7/976.full
Description

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: 1.6.0
Date: 09-11-2012
biocViews: GO, Clustering, Pathways, Anopheles_gambiae, Arabidopsis_thaliana, Bos_taurus, Caenorhabditis_elegans, Carcharias_taurus, Chicken, Chimp, Drosophila_melanogaster, E.coli, Film, Fish, Fruitfly, Human, Mouse, Malaria, Monkey, Pig, Rat, Rhesus, Rodents, Salmon, Scolopendra, Shared, S. cerevisiae, S. coelicolor, S. melanogaster, S. pombe, S. tokodaii, S. tubercularis, S. typhimurium, S. watermolae, Spinal cord, Tetrahymena, Thermoplasma, Tuberculosis, Yeast, Zebrafish
Depends: 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.Sc.sgd.db, org.Sco.eg.db, org.Ss.eg.db, org.Xl.eg.db
Imports: GO.db, org.Hs.eg.db, org.Ag.eg.db, org.At.tair.db, org.Bt.eg.db, org.Ce.eg.db, org.Ss.eg.db, org.Xl.eg.db
Suggests: clusterProfiler
License: GPL Version 2
clusterSim

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

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>cluster1</td>
<td>A set of gene IDs.</td>
</tr>
<tr>
<td>cluster2</td>
<td>Another set of gene IDs.</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>drop</td>
<td>A set of evidence codes based on which certain annotations are dropped. Use NULL to keep all GO annotations.</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
combineScores

Combining similarity matrix to similarity score

Description

Functions for combining similarity matrix to similarity score

Usage

combineScores(SimScores, combine)

Arguments

- SimScores: similarity matrix
- combine: combine method

Value

similarity value

Author(s)

Guangchuang Yu http://ygc.name

References


See Also

gosim mgosim geneSim mgeneSim mclusterSim

Examples

```r
## cluster1 <- c("835", "5261","241", "994")
## clusterSim(cluster1, cluster2, ont="MF", organism="human", measure="Wang")
```
**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**

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

**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**

A 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")
```
goSim  

Semantic Similarity Between Two GO Terms

Description

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

Usage

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

Arguments

- `GOID1`  
  GO ID 1.
- `GOID2`  
  GO ID 2.
- `ont`  
  One of "MF", "BP", and "CC" subontologies.
- `organism`  
- `measure`  
  One of "Resnik", "Lin", "Rel", "Jiang" and "Wang" methods.

Value

- `similarity`

References

[http://bioinformatics.oxfordjournals.org/cgi/content/abstract/26/7/976](http://bioinformatics.oxfordjournals.org/cgi/content/abstract/26/7/976)  
PMID: 20179076

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)
loadICdata

Load IC data

Description
Load Information Content data to DOSEEnv environment

Usage
loadICdata(organism, ont)

Arguments
organism "human"
ont "DO"

Value
NULL

Author(s)
Guangchuang Yu 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
mclusterSim(clusters, ont = "MF", organism = "human",
measure = "Wang", drop = "IEA", combine = "BMA")

Arguments
clusters A list of gene clusters.
ont One of "MF", "BP", and "CC" subontologies.
measure One of "Resnik", "Lin", "Rel", "Jiang" and "Wang" methods.
**mgeneSim**

Pairwise Semantic Similarity for a List of Genes

**Description**

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

**Usage**

```r
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.

Value

similarity matrix

References

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

See Also

mgoSim geneSim mgeneSim clusterSim mclusterSim

Examples

mgeneSim(c("835", "5261", "241"), ont="MF", organism="human", measure="Wang")

mgoSim(GO1, GO2, ont = "MF", organism = "human", measure = "Wang", combine = "BMA")"
termSim

Arguments

GO1          A set of go terms.
GO2          Another set of go terms.
ont          One of "MF", "BP", and "CC" subontologies.
measure      One of "Resnik", "Lin", "Rel", "Jiang" and "Wang" methods.
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 geneSim mgeneSim clusterSim mclusterSim

Examples

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")

termSim

termSim

Description

measuring similarities between two term vectors.

Usage

termSim(t1, t2, method = "Wang", organism = "human", ont)
Arguments

- t1: term vector
- t2: term vector
- method: one of "Wang", "Resnik", "Rel", "Jiang", and "Lin".
- organism: only "human" supported
- ont: ontology

Details

Provide two term vectors, this function will calculate their similarities.

Value

Score matrix

Author(s)

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

---

**Description**

Method Wang for semantic similarity measuring

**Usage**

`wangMethod(ID1, ID2, ont)`

**Arguments**

- ID1: Ontology Term
- ID2: Ontology Term
- ont: Ontology

**Value**

Semantic similarity score

**Author(s)**

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

*Topic **datasets**

IC, 7

*Topic **manip**

clusterSim, 3
geneSim, 5
goSim, 6
mclusterSim, 8
mgeneSim, 9
mgoSim, 10

*Topic **package**

GOSemSim-package, 2

clusterSim, 3, 3, 5, 6, 9–11
combineScores, 4
geneSim, 3, 4, 5, 6, 9–11
GOSemSim (GOSemSim-package), 2
GOSemSim-package, 2
goSim, 3–5, 6, 9–11

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_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
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
loadICdata, 8
mclusterSim, 3–6, 8, 10, 11
mgeneSim, 3–6, 9, 9, 11
mgoSim, 3–6, 9, 10, 10

termSim, 11

wangMethod, 12