

Package ‘GOSemSim’

October 17, 2020

Type Package

Title GO-terms Semantic Similarity Measures

Version 2.14.2

Maintainer Guangchuang Yu <guangchuangyu@gmail.com>

Description The semantic comparisons of Gene Ontology (GO) annotations provide quantitative ways to compute similarities between genes and gene groups, and have become important basis for many bioinformatics analysis approaches. GOSemSim is an R package for semantic similarity computation among GO terms, sets of GO terms, gene products and gene clusters. GOSemSim implemented five methods proposed by Resnik, Schlicker, Jiang, Lin and Wang respectively.

Depends R (>= 3.5.0)

LinkingTo Rcpp

Imports AnnotationDbi, GO.db, methods, utils

Suggests AnnotationHub, BiocManager, clusterProfiler, DOSE, knitr, org.Hs.eg.db, prettydoc, testthat

VignetteBuilder knitr

ByteCompile true

License Artistic-2.0

Encoding UTF-8

URL <https://guangchuangyu.github.io/software/GOSemSim>

BugReports <https://github.com/YuLab-SMU/GOSemSim/issues>

biocViews Annotation, GO, Clustering, Pathways, Network, Software

RoxygenNote 7.1.1

git_url <https://git.bioconductor.org/packages/GOSemSim>

git_branch RELEASE_3_11

git_last_commit 15b32bb

git_last_commit_date 2020-09-04

Date/Publication 2020-10-16

Author Guangchuang Yu [aut, cre],
Alexey Stukalov [ctb],
Chuanle Xiao [ctb],
Lluís Revilla Sancho [ctb]

R topics documented:

| | |
|---------------------|----|
| GOSemSim-package | 2 |
| clusterSim | 3 |
| combineScores | 4 |
| geneSim | 4 |
| godata | 5 |
| GOSemSimDATA-class | 6 |
| goSim | 6 |
| go_term_table | 7 |
| infoContentMethod | 7 |
| load_OrgDb | 8 |
| mclusterSim | 8 |
| mgeneSim | 9 |
| mgoSim | 10 |
| termSim | 11 |
| wangMethod_internal | 12 |

| | |
|--------------|-----------|
| Index | 13 |
|--------------|-----------|

| | |
|------------------|---|
| GOSemSim-package | <i>Gene Ontology-based Semantic Similarity Measures</i> |
|------------------|---|

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: 2.0.0
 Date: 09-11-2012
 biocViews: GO, Clustering, Pathways, Anopheles_gambiae, Arabidopsis_thaliana, Bos_taurus, Caenorhabditis_elegans
 Depends:
 Imports: methods, AnnotationDbi, GO.db
 Suggests: clusterProfiler, DOSE
 License: Artistic-2.0

Author(s)

Guangchuang Yu

Maintainer: Guangchuang Yu <guangchuangyu@gmail.com>

References

Yu et al. (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products *Bioinformatics* (Oxford, England), 26:7 976–978, April 2010. ISSN 1367-4803 <http://bioinformatics.oxfordjournals.org/cgi/content/abstract/26/7/976> PMID: 20179076

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,
  semData,
  measure = "Wang",
  drop = "IEA",
  combine = "BMA"
)
```

Arguments

| | |
|----------|--|
| cluster1 | A set of gene IDs. |
| cluster2 | Another set of gene IDs. |
| semData | GOSemSimDATA object |
| 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", "avg", "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

Yu et al. (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products *Bioinformatics* (Oxford, England), 26:7 976–978, April 2010. ISSN 1367-4803 <http://bioinformatics.oxfordjournals.org/cgi/content/abstract/26/7/976> PMID: 20179076

See Also

[goSim](#) [mgoSim](#) [geneSim](#) [mgeneSim](#) [mclusterSim](#)

Examples

```
d <- godata('org.Hs.eg.db', ont="MF", computeIC=FALSE)
cluster1 <- c("835", "5261", "241", "994")
cluster2 <- c("307", "308", "317", "321", "506", "540", "378", "388", "396")
clusterSim(cluster1, cluster2, semData=d, measure="Wang")
```

| | |
|---------------|--|
| combineScores | <i>combining similarity matrix to similarity score</i> |
|---------------|--|

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://guangchuangyu.github.io>

| | |
|---------|--|
| geneSim | <i>Semantic Similarity Between two Genes</i> |
|---------|--|

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, semData, measure = "Wang", drop = "IEA", combine = "BMA")
```

Arguments

| | |
|---------|--|
| gene1 | Entrez gene id. |
| gene2 | Another entrez gene id. |
| semData | GOSemSimDATA object |
| 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", "avg", "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

Yu et al. (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products *Bioinformatics* (Oxford, England), 26:7 976–978, April 2010. ISSN 1367-4803 <http://bioinformatics.oxfordjournals.org/cgi/content/abstract/26/7/976> PMID: 20179076

See Also

[goSim](#) [mgoSim](#) [mgeneSim](#) [clusterSim](#) [mclusterSim](#)

Examples

```
d <- godata('org.Hs.eg.db', ont="MF", computeIC=FALSE)
geneSim("241", "251", semData=d, measure="Wang")
```

godata

godata

Description

prepare GO DATA for measuring semantic similarity

Usage

```
godata(OrgDb = NULL, keytype = "ENTREZID", ont, computeIC = TRUE)
```

Arguments

| | |
|-----------|------------------------------|
| OrgDb | OrgDb object |
| keytype | keytype |
| ont | one of 'BP', 'MF', 'CC' |
| computeIC | logical, whether computer IC |

Value

GOSemSimDATA object

Author(s)

Guangchuang Yu

| | |
|--------------------|---|
| GOSemSimDATA-class | <i>Class "GOSemSimDATA" This class stores IC and gene to go mapping for semantic similarity measurement</i> |
|--------------------|---|

Description

Class "GOSemSimDATA" This class stores IC and gene to go mapping for semantic similarity measurement

Slots

keys gene ID
 ont ontology
 IC IC data
 geneAnno gene to GO mapping
 metadata metadata

| | |
|-------|---|
| goSim | <i>Semantic Similarity Between Two GO Terms</i> |
|-------|---|

Description

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

Usage

```
goSim(GOID1, GOID2, semData, measure = "Wang")
```

Arguments

| | |
|---------|--|
| GOID1 | GO ID 1. |
| GOID2 | GO ID 2. |
| semData | GOSemSimDATA object |
| measure | One of "Resnik", "Lin", "Rel", "Jiang" and "Wang" methods. |

Value

similarity

References

Yu et al. (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products *Bioinformatics* (Oxford, England), 26:7 976–978, April 2010. ISSN 1367-4803 <http://bioinformatics.oxfordjournals.org/cgi/content/abstract/26/7/976> PMID: 20179076

See Also

[mgoSim](#) [geneSim](#) [mgeneSim](#) [clusterSim](#) [mclusterSim](#)

Examples

```
d <- godata('org.Hs.eg.db', ont="MF", computeIC=FALSE)
goSim("GO:0004022", "GO:0005515", semData=d, measure="Wang")
```

go_term_table

Information content of GO terms

Description

These datasets are the information contents of GOterms.

References

Yu et al. (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products *Bioinformatics* (Oxford, England), 26:7 976–978, April 2010. ISSN 1367-4803 <http://bioinformatics.oxfordjournals.org/cgi/content/abstract/26/7/976> PMID: 20179076

infoContentMethod

information content based methods

Description

Information Content Based Methods for semantic similarity measuring

Usage

```
infoContentMethod(ID1, ID2, method, godata)
```

Arguments

| | |
|--------|--|
| ID1 | Ontology Term |
| ID2 | Ontology Term |
| method | one of "Resnik", "Jiang", "Lin" and "Rel". |
| godata | GOSemSimDATA object |

Details

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

Value

semantic similarity score

Author(s)

Guangchuang Yu <https://guangchuangyu.github.io>

| | |
|------------|-------------------|
| load_OrgDb | <i>load_OrgDb</i> |
|------------|-------------------|

Description

load OrgDb

Usage

```
load_OrgDb(OrgDb)
```

Arguments

| | |
|-------|----------------------------|
| OrgDb | OrgDb object or OrgDb name |
|-------|----------------------------|

Value

OrgDb object

Author(s)

Guangchuang Yu

| | |
|-------------|---|
| mclusterSim | <i>Pairwise Semantic Similarities for a List of Gene Clusters</i> |
|-------------|---|

Description

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

Usage

```
mclusterSim(clusters, semData, measure = "Wang", drop = "IEA", combine = "BMA")
```

Arguments

| | |
|----------|--|
| clusters | A list of gene clusters. |
| semData | GOSemSimDATA object |
| 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", "avg", "rcmax", "BMA" methods, for combining semantic similarity scores of multiple GO terms associated with protein or multiple proteins associated with protein cluster. |

Value

similarity matrix

References

Yu et al. (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products *Bioinformatics* (Oxford, England), 26:7 976–978, April 2010. ISSN 1367-4803 <http://bioinformatics.oxfordjournals.org/cgi/content/abstract/26/7/976> PMID: 20179076

See Also

[goSim](#) [mgoSim](#) [geneSim](#) [mgeneSim](#) [clusterSim](#)

Examples

```
d <- godata('org.Hs.eg.db', ont="MF", computeIC=FALSE)
cluster1 <- c("835", "5261", "241")
cluster2 <- c("578", "582")
cluster3 <- c("307", "308", "317")
clusters <- list(a=cluster1, b=cluster2, c=cluster3)
mclusterSim(clusters, semData=d, 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,
  semData,
  measure = "Wang",
  drop = "IEA",
  combine = "BMA",
  verbose = TRUE
)
```

Arguments

| | |
|---------|--|
| genes | A list of entrez gene IDs. |
| semData | GOSemSimDATA object |
| 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", "avg", "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

Yu et al. (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products *Bioinformatics* (Oxford, England), 26:7 976–978, April 2010. ISSN 1367-4803 <http://bioinformatics.oxfordjournals.org/cgi/content/abstract/26/7/976> PMID: 20179076

See Also

[goSim](#) [mgoSim](#) [geneSim](#) [clusterSim](#) [mclusterSim](#)

Examples

```
d <- godata('org.Hs.eg.db', ont="MF", computeIC=FALSE)
mgeneSim(c("835", "5261", "241"), semData=d, 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

```
mgoSim(GO1, GO2, semData, measure = "Wang", combine = "BMA")
```

Arguments

| | |
|---------|--|
| GO1 | A set of go terms. |
| GO2 | Another set of go terms. |
| semData | GOSemSimDATA object |
| measure | One of "Resnik", "Lin", "Rel", "Jiang" and "Wang" methods. |
| combine | One of "max", "avg", "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

Yu et al. (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products *Bioinformatics* (Oxford, England), 26:7 976–978, April 2010. ISSN 1367-4803 <http://bioinformatics.oxfordjournals.org/cgi/content/abstract/26/7/976> PMID: 20179076

See Also

[goSim](#) [geneSim](#) [mgeneSim](#) [clusterSim](#) [mclusterSim](#)

Examples

```
d <- godata('org.Hs.eg.db', ont="MF", computeIC=FALSE)
go1 <- c("GO:0004022", "GO:0004024", "GO:0004023")
go2 <- c("GO:0009055", "GO:0020037")
mgoSim("GO:0003824", go2, semData=d, measure="Wang")
mgoSim(go1, go2, semData=d, measure="Wang")
```

termSim

termSim

Description

measuring similarities between two term vectors.

Usage

```
termSim(t1, t2, semData, method = c("Wang", "Resnik", "Rel", "Jiang", "Lin"))
```

Arguments

| | |
|---------|---|
| t1 | term vector |
| t2 | term vector |
| semData | GOSemSimDATA object |
| method | one of "Wang", "Resnik", "Rel", "Jiang", and "Lin". |

Details

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

Value

score matrix

Author(s)

Guangchuang Yu <http://guangchuangyu.github.io>

wangMethod_internal *wangMethod*

Description

Method Wang for semantic similarity measuring

Usage

```
wangMethod_internal(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

- * **classes**
 - GOSemSimDATA-class, 6
- * **datasets**
 - go_term_table, 7
- * **manip**
 - clusterSim, 3
 - geneSim, 4
 - goSim, 6
 - mclusterSim, 8
 - mgeneSim, 9
 - mgoSim, 10
- * **package**
 - GOSemSim-package, 2

clusterSim, 3, 3, 5, 7, 9–11

combineScores, 4

geneSim, 3, 4, 4, 7, 9–11

GO (go_term_table), 7

go_term_table, 7

godata, 5

GOSemSim (GOSemSim-package), 2

GOSemSim-package, 2

GOSemSimDATA-class, 6

goSim, 3–5, 6, 9–11

gotbl (go_term_table), 7

infoContentMethod, 7

load_OrgDb, 8

mclusterSim, 3–5, 7, 8, 10, 11

mgeneSim, 3–5, 7, 9, 9, 11

mgoSim, 3–5, 7, 9, 10, 10

show, GOSemSimDATA-method
(GOSemSimDATA-class), 6

termSim, 11

wangMethod_internal, 12