Package ‘GOSemSim’

Type Package
Title GO-terms Semantic Similarity Measures
Version 2.0.4
Author Guangchuang Yu <guangchuangyu@gmail.com> with contributions from
Alexey Stukalov and Chuanle Xiao.
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 became 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 (>= 3.3.0)
LinkingTo Rcpp
Imports AnnotationDbi, GO.db, methods, utils
Suggests AnnotationHub, BiocInstaller, BiocStyle, clusterProfiler,
DOSE, knitr, org.Hs.eg.db, testthat
VignetteBuilder knitr
ByteCompile true
License Artistic-2.0
URL https://guangchuangyu.github.io/GOSemSim
BugReports https://github.com/GuangchuangYu/GOSemSim/issues
BiocViews Annotation, GO, Clustering, Pathways, Network, Software
RoxygenNote 5.0.1
NeedsCompilation yes

R topics documented:

GOSemSim-package ................................................................. 2
clusterSim ................................................................. 3
combineScores ............................................................... 4
geneSim ................................................................. 4
godata ................................................................. 5
GOSemSim-package

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, ... Plasmodium_falciparum, Rattus_norvegicus, Saccharomyces_cerevisiae, Streptomyces_coelicolor, Sus-scrofa, Xenopus_laevis
Depends: GO.db
Imports: methods, AnnotationDbi
Suggests: clusterProfiler, DOSE
License: Artistic-2.0

Author(s)

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

References

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", "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 clusterSim mclusterSim

Examples

d <- godata('org.Hs.eg.db', ont="MF", computeIC=FALSE)
cluster1 <- c("835", "5261", "241", "994")
clusterSim(cluster1, cluster2, semData=d, measure="Wang")
## combineScores

### Description

Functions for combining similarity matrix to similarity score

### Usage

```r
combineScores(SimScores, combine)
```

### Arguments

- **SimScores**: similarity matrix
- **combine**: combine method

### Value

similarity value

### Author(s)

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

## geneSim

### 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, 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", "average", "rcmax", "BMA" methods, for combining semantic similarity scores of multiple GO terms associated with protein or multiple proteins associated with protein cluster.
**godata**

**Value**

list of similarity value and corresponding GO.

**References**


**See Also**

`goSim` `mgoSim` `mgeneSim` `clusterSim` `mclusterSim`

**Examples**

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

**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 compute IC

**Value**

GOSemSimDATA object

**Author(s)**

Guangchuang Yu
goSim

GOSemSimDATA-class

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

Semantic Similarity Between Two GO Terms

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


See Also

mgoSim geneSim mgeneSim clusterSim mclusterSim
Examples

```r
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 GO terms.

**References**


**infoContentMethod**

**information content based methods**

**Description**

Information Content Based Methods for semantic similarity measuring

**Usage**

```r
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 http://ygc.name
load_OrgDb

Description
load OrgDb

Usage
load_OrgDb(OrgDb)

Arguments
OrgDb

Value
OrgDb object

Author(s)
Guangchuang Yu

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, 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", "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 matrix
mgeneSim

References


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", "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
mgoSim

References


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

Value

  similarity

References


See Also
goSim geneSim mgeneSim clusterSim mclusterSim
**termSim**

**Examples**

```r
d <- godata('org.Hs egret.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")
```

---

**Description**

measuring similarities between two term vectors.

**Usage**

```r
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://ygc.name](http://ygc.name)
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

*Topic **classes**
  GOSemSimDATA-class, 6

*Topic **datasets**
  go_term_table, 7

*Topic **manip**
  clusterSim, 3
geneSim, 4
goSim, 6
mclusterSim, 8
mgeneSim, 9
mgoSim, 10

*Topic **package**
  GOSemSim-package, 2

clusterSim, 3, 3, 5, 6, 9, 10
combineScores, 4
geneSim, 3, 4, 6, 9, 10
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, 10
gotbl (go_term_table), 7

infoContentMethod, 7

load_OrgDb, 8

mclusterSim, 3, 5, 6, 8, 10
mgeneSim, 3, 5, 6, 9, 9, 10
mgoSim, 3, 5, 6, 9, 10, 10

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

termSim, 11

wangMethod_internal, 12