Package 'GOSemSim'

March 26, 2013

Type Package

Title GO-terms Semantic Similarity Measures

Version 1.16.1

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, Ca-
nine, Chicken, Chimp,Coelicolor, E coli strain K12 and Sakai, Fly, Hu-
man, 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

biocViews GO, Clustering, Pathways, NetworkAnalysis

Collate
'clusterSim.R' 'combineMethods.R' 'computeIC.R' 'ICMethods.R''gene2GO.R' 'geneSim.R' 'goSim.R' 'termSim.R'
package.R' 'mgeneSim.R''mgoSim.R' 'mclusterSim.R'

License GPL-2

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

R topics documented:

1

GOSemSim-package   *Gene Ontology-based Sematic Similarity Measures*

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
Depends:	
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
Suggests:	clusterProfiler
License:	GPL Version 2

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, 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.
organism	One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "coeli-color", "ecolik12", "ecsakai", "fly", "human", "malaria", "mouse", "pig", "rat", "rhesus", "worm", "xenopus", "yeast" and "zebrafish".
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 assiciated 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

```
## 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")
```

---

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

---

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.
organism	One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "coeli-color", "ecolik12","ecsakai", "fly", "human", "malaria", "mouse", "pig", "rat","rhesus", "worm", "xenopus", "yeast" and "zebrafish".
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 assiciated 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

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

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   One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "coeli-color", "ecolik12", "ecsakai", "fly", "human", "malaria", "mouse", "pig", "rat","rhesus", "worm", "xenopus", "yeast" and "zebrafish".

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

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

IC	*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, 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

---

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.
organism	One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "coelicolor", "ecolik12", "ecsakai", "fly","human","malaria", "mouse", "pig", "rat", "rhesus", "worm", "xenopus", "yeast" and "zebrafish".
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 assiciated 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

```
## cluster1 <- c("835", "5261","241")
## cluster2 <- c("578","582")
## cluster3 <- c("307", "308", "317")
## clusters <- list(a=cluster1, b=cluster2, c=cluster3)
## mclusterSim(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.
organism	One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "coeli-color", "ecolik12", "ecsakai", "fly", "human", "malaria", "mouse", "pig", "rat", "rhesus", "worm", "xenopus", "yeast" and "zebrafish".
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 assiciated 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

mgeneSim(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

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

Arguments

GO1   A set of go terms.

GO2   Another set of go terms.

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

organism   One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "coeli-color", "ecolik12", "ecsakai", "fly", "human", "malaria", "mouse", "pig", "rat", "rhesus", "worm", "xenopus", "yeast" and "zebrafish".

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

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

```
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

wangMethod *wangMethod*

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

Index