

GOSemSim

March 24, 2012

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: 1.9.9
Date: 03-18-2011
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: Clustering
License: GPL Version 2

Author(s)

Guangchuang Yu

Maintainer: Guangchuang Yu <guangchuangyu@gmail.com>

See Also

[GOSet](#), [GeneSet](#) [GeneClusterSet](#)

GOSet-class *Class "GOSet"*

Description

A GOSet contains two sets of GO identifiers.

Slots

GOSet1: containing a vector of GO identifiers.

GOSet2: containing a vector of GO identifiers.

Methods

GO semantic similarity measure :

```
sim signature(object = "GOSet", params="Params")
```

Author(s)

Guangchuang Yu <guangchuangyu@gmail.com>

See Also

[GeneSet Params](#)

Examples

```
## Setting Parameters...
params <- new("Params", ontology="MF", organism="human", method="Wang")
## Setting GOSet...
go1 <- c("GO:0004022", "GO:0004024", "GO:0004023")
go2 <- c("GO:0009055", "GO:0020037")
gos <- new("GOSet", GOSet1=go1, GOSet2=go2)
## Calculating GO Semantic Similarities...
sim(gos, params)
## Setting Combine Method
setCombineMethod(params) <- "rcmax.avg"
## Combining semantic similarity scores of multiple GO terms to one score.
sim(gos, params)
```

GOSet *Methods for GOSet class*

Description

Use GOSet to construct parameter class for calculating Semantic similarity.

Slots

GOSet1: containing a vector of GO Terms.

GOSet2: containing a vector of GO Terms.

Methods

See documentation for [GOSet](#) for examples.

See Also

[GOSet-class](#)

Examples

```
## Construct GOSet class
go1 <- c("GO:0004022", "GO:0004024", "GO:0004023")
go2 <- c("GO:0009055", "GO:0020037")
GOSet <- new("GOSet", GOSet1=go1, GOSet2=go2)
```

GeneClusterSet-class

Class "GeneClusterSet"

Description

A GeneClusterSet contains a list of Gene clusters.

Slots

GeneClusters: containing a list of Gene clusters.

Methods

Gene Clusters semantic similarity measure :

sim signature(object = "GeneClusterSet", params="Params", value = "numeric")

Author(s)

Guangchuang Yu <guangchuangyu@gmail.com>

See Also

[GeneSet Params](#)

Examples

```
## Setting Parameters...
params <- new("Params", ontology="MF", organism="human", method="Wang", combine="rcmax")
## Setting GeneClusterSet...
cluster1 <- c("835", "5261", "241", "994", "514", "517", "533")
cluster2 <- c("578", "582", "583", "400", "409", "411")
cluster3 <- c("307", "308", "317", "321", "506", "540", "378", "388", "396")
clusters <- list(a=cluster1, b=cluster2, c=cluster3)
geneClusters <- new("GeneClusterSet", GeneClusters=clusters)
## Calculating Gene Clusters Semantic Similarities...
# sim(geneClusters, params)
```

GeneSet-class

Class "GeneSet"

Description

A GeneSet contains two sets of Gene identifiers.

Slots

GeneSet1: containing a vector of Gene identifiers.

GeneSet2: containing a vector of Gene identifiers.

Methods

GO semantic similarity measure :

```
sim signature(object = "GeneSet", params="Params", value = "numeric")
```

Author(s)

Guangchuang Yu <guangchuangyu@gmail.com>

See Also

[GeneClusterSet Params](#)

Examples

```
## Setting Parameters...
params <- new("Params", ontology="MF", organism="human", method="Wang", combine="rcmax.av
## Setting GeneSet...
gs1 <- c("835", "5261", "241", "994", "514", "533")
gs2 <- c("578", "582", "400", "409", "411")
gs <- new("GeneSet", GeneSet1=gs1, GeneSet2=gs2)
## Calculating Gene Semantic Similarities...
#sim(gs, params)
```

IC

Information content of GO terms

Description

These datasets are the information contents of GOterms.

 Params-class *Class "Params"*

Description

A `Params` contains parameters for calculating GO semantic similarity among GO Terms or Gene Sets.

Slots

`ontology`: one of "MF", "BP", "CC".

`organism`: one of supported species.

`method`: Method for calculating GO semantic similarity, one of "Resnik", "Jiang", "Lin", "Rel", "Wang".

`combine`: Method for combining GO semantic similarity scores, one of "avg", "max", "rcmax", "rcmax.avg"

`dropCodes`: `dropCodes` for mapping Gene to GO Terms.

Methods

Slot access (e.g., `setOntology<-`) and retrieve (e.g., `[]`):

setOntology<- signature(object = "Params")

setOrganism<- signature(object = "Params")

setMethod<- signature(object = "Params")

setCombineMethod<- signature(object = "Params", value = "character")

`[]` signature(x="Params", i="character"): subset the Params by index (i="character")

Loading require data:

loadICdata signature(object = "Params")

loadAnnoPkg signature(object = "Params")

loadGOMap signature(object = "Params")

Useful additional methods include:

computeIC signature(object = "Params"): compute Information Content of GO

Author(s)

Guangchuang Yu <guangchuangyu@gmail.com>

See Also

[GeneClusterSet Params](#)

Examples

```
## Setting Parameters...
params <- new("Params", ontology="MF", organism="human", method="Wang")
```

Params

Methods for Params class

Description

Use Params to construct parameter class for calculating Semantic similarity.

Slots

ontology: An argument determining which ontology were to be measured.

organism: Setting organism for mapping Gene IDs to GO Terms.

method: Method for calculating semantic similarity.

combine: Combine method for combining multiple GO semantic scores to one score.

dropCodes: dropCodes for mapping Gene to GO.

Methods

See documentation for [Params](#) for examples.

See Also

[Params-class](#)

Examples

```
## Construct Params class
params <- new("Params", ontology="MF", organism="human", method="Wang")

## loading IC data
loadICdata(params)

## Setting params
setOntology(params) <- "BP"
setCombineMethod="rcmax"

## Accessing slots.
params["ontology"]
params["organism"]
```

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", dr
```

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", "coelicolor", "ecoli12", "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", "rcmax.avg" methods, for combining semantic similarity scores of multiple GO terms associated with protein or multiple proteins associated with protein cluster.

Value

sim	Semantic Similarity.
-----	----------------------

References

Wang et al.(2007) A new method to measure the semantic similarity of go terms *Bioinformatics* (Oxford, England), 23:0 1274–81, May 2007. ISSN 1460-2059 <http://www.ncbi.nlm.nih.gov/pubmed/17344234> PMID: 17344234

P W Lord et al.(2003) Semantic similarity measures as tools for exploring the gene ontology *PacificSymposiumonBiocomputing* Pacific Symposium on Biocomputing, 2003:601-12, ISSN 1793-5091 <http://www.ncbi.nlm.nih.gov/pubmed/12603061> PMID: 12603061

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")
```

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",
```

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

Value

geneSim	Semantic similarity.
GO1	Corresponding GO terms for gene1.
GO2	Corresponding GO terms for gene2.

References

Wang et al.(2007) A new method to measure the semantic similarity of go terms *Bioinformatics* (Oxford, England), 23:0 1274–81, May 2007. ISSN 1460-2059 <http://www.ncbi.nlm.nih.gov/pubmed/17344234> PMID: 17344234

P W Lord et al.(2003) Semantic similarity measures as tools for exploring the gene ontology *PacificSymposiumonBiocomputing* Pacific Symposium on Biocomputing, 2003:601-12, ISSN 1793-5091 <http://www.ncbi.nlm.nih.gov/pubmed/12603061> PMID: 12603061

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", "coelicolor", "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

goSim	Semantic similarity.
-------	----------------------

References

Wang et al.(2007) A new method to measure the semantic similarity of go terms *Bioinformatics* (Oxford, England), 23:0 1274–81, May 2007. ISSN 1460-2059 <http://www.ncbi.nlm.nih.gov/pubmed/17344234> PMID: 17344234

P W Lord et al.(2003) Semantic similarity measures as tools for exploring the gene ontology *PacificSymposiumonBiocomputing* Pacific Symposium on Biocomputing, 2003:601-12, ISSN 1793-5091 <http://www.ncbi.nlm.nih.gov/pubmed/12603061> PMID: 12603061

See Also

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

Examples

```
goSim("GO:0043121", "GO:0019838", measure="Wang")
goSim("GO:0043121", "GO:0019838", ont="MF", organism="human", measure="Wang")
```

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")
```

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

Value

simmat	A Matrix of Pairwise Semantic Similarities
--------	--

References

Wang et al.(2007) A new method to measure the semantic similarity of go terms *Bioinformatics* (Oxford, England), 23:0 1274–81, May 2007. ISSN 1460-2059 <http://www.ncbi.nlm.nih.gov/pubmed/17344234> PMID: 17344234

P W Lord et al.(2003) Semantic similarity measures as tools for exploring the gene ontology *PacificSymposiumonBiocomputing* Pacific Symposium on Biocomputing, 2003:601-12, ISSN 1793-5091 <http://www.ncbi.nlm.nih.gov/pubmed/12603061> PMID: 12603061

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", comb
```

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

Value

simMatrix	A Matrix of Pairwise Semantic Similarities
-----------	--

References

Wang et al.(2007) A new method to measure the semantic similarity of go terms *Bioinformatics* (Oxford, England), 23:0 1274–81, May 2007. ISSN 1460-2059 <http://www.ncbi.nlm.nih.gov/pubmed/17344234> PMID: 17344234

P W Lord et al.(2003) Semantic similarity measures as tools for exploring the gene ontology *PacificSymposiumonBiocomputing* Pacific Symposium on Biocomputing, 2003:601-12, ISSN 1793-5091 <http://www.ncbi.nlm.nih.gov/pubmed/12603061> PMID: 12603061

See Also

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

Examples

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

mgoSim	<i>Semantic Similarity Between two GO terms lists</i>
--------	---

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="rcmax.avg")
```

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", "coelior", "ecoli12", "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", "rcmax.avg" methods, for combining semantic similarity scores of multiple GO terms associated with protein or multiple proteins associated with protein cluster.

Value

mgoSim Semantic similarity.

References

Wang et al.(2007) A new method to measure the semantic similarity of go terms *Bioinformatics* (Oxford, England), 23:0 1274–81, May 2007. ISSN 1460-2059 <http://www.ncbi.nlm.nih.gov/pubmed/17344234> PMID: 17344234

P W Lord et al.(2003) Semantic similarity measures as tools for exploring the gene ontology *PacificSymposiumonBiocomputing* Pacific Symposium on Biocomputing, 2003:601-12, ISSN 1793-5091 <http://www.ncbi.nlm.nih.gov/pubmed/12603061> PMID: 12603061

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")
```

sim-methods

Methods for calculating semantic similarity

Description

This generic and methods calculating semantic similarities among GO terms, gene sets and gene clusters.

Methods

Defined methods include:

These methods calculating semantic similarities about `GOSet`, `GeneSet`, and `GeneClusterSet`.

See Also

```
signature(object = "GOSet", params="Params"),signature(object = "GeneSet", params="")
GOSet-class GeneSet-class GeneClusterSet-class
```

Index

*Topic classes

GeneClusterSet-class, 3
GeneSet-class, 4
GOSet-class, 2
Params-class, 5

*Topic datasets

IC, 4

*Topic manip

clusterSim, 6
geneSim, 7
goSim, 8
mclusterSim, 9
mgeneSim, 10
mgoSim, 11

*Topic methods

GOSet, 2
Params, 6
sim-methods, 12

*Topic package

GOSemSim-package, 1
[, Params, character-method
(Params-class), 5

clusterSim, 6, 8–12
computeIC (Params-class), 5
computeIC, Params-method
(Params-class), 5

GeneClusterSet, 1, 4, 5
GeneClusterSet-class, 12
GeneClusterSet-class, 3
GeneSet, 1–3
GeneSet-class, 12
GeneSet-class, 4
geneSim, 7, 7, 9–12
GOSemSim (GOSemSim-package), 1
GOSemSim-package, 1
GOSet, 1, 2, 3
GOSet-class, 3, 12
GOSet-class, 2
GOSet-methods (GOSet), 2
goSim, 7, 8, 8, 10–12

IC, 4

Info_Contents_anopheles_BP (IC), 4
Info_Contents_anopheles_CC (IC), 4
Info_Contents_anopheles_MF (IC), 4
Info_Contents_arabidopsis_BP
(IC), 4
Info_Contents_arabidopsis_CC
(IC), 4
Info_Contents_arabidopsis_MF
(IC), 4
Info_Contents_bovine_BP (IC), 4
Info_Contents_bovine_CC (IC), 4
Info_Contents_bovine_MF (IC), 4
Info_Contents_canine_BP (IC), 4
Info_Contents_canine_CC (IC), 4
Info_Contents_canine_MF (IC), 4
Info_Contents_chicken_BP (IC), 4
Info_Contents_chicken_CC (IC), 4
Info_Contents_chicken_MF (IC), 4
Info_Contents_chimp_BP (IC), 4
Info_Contents_chimp_CC (IC), 4
Info_Contents_chimp_MF (IC), 4
Info_Contents_coelicolor_BP (IC),
4
Info_Contents_coelicolor_CC (IC),
4
Info_Contents_coelicolor_MF (IC),
4
Info_Contents_ecolik12_BP (IC), 4
Info_Contents_ecolik12_CC (IC), 4
Info_Contents_ecolik12_MF (IC), 4
Info_Contents_ecsakai_BP (IC), 4
Info_Contents_ecsakai_CC (IC), 4
Info_Contents_ecsakai_MF (IC), 4
Info_Contents_fly_BP (IC), 4
Info_Contents_fly_CC (IC), 4
Info_Contents_fly_MF (IC), 4
Info_Contents_human_BP (IC), 4
Info_Contents_human_CC (IC), 4
Info_Contents_human_MF (IC), 4
Info_Contents_malaria_BP (IC), 4
Info_Contents_malaria_CC (IC), 4
Info_Contents_malaria_MF (IC), 4
Info_Contents_mouse_BP (IC), 4

- Info_Contents_mouse_CC (IC), 4
- Info_Contents_mouse_MF (IC), 4
- Info_Contents_pig_BP (IC), 4
- Info_Contents_pig_CC (IC), 4
- Info_Contents_pig_MF (IC), 4
- Info_Contents_rat_BP (IC), 4
- Info_Contents_rat_CC (IC), 4
- Info_Contents_rat_MF (IC), 4
- Info_Contents_rhesus_BP (IC), 4
- Info_Contents_rhesus_CC (IC), 4
- Info_Contents_rhesus_MF (IC), 4
- Info_Contents_worm_BP (IC), 4
- Info_Contents_worm_CC (IC), 4
- Info_Contents_worm_MF (IC), 4
- Info_Contents_xenopus_BP (IC), 4
- Info_Contents_xenopus_CC (IC), 4
- Info_Contents_xenopus_MF (IC), 4
- Info_Contents_yeast_BP (IC), 4
- Info_Contents_yeast_CC (IC), 4
- Info_Contents_yeast_MF (IC), 4
- Info_Contents_zebrafish_BP (IC), 4
- Info_Contents_zebrafish_CC (IC), 4
- Info_Contents_zebrafish_MF (IC), 4

- loadAnnoPkg (Params-class), 5
- loadAnnoPkg, Params-method
(Params-class), 5
- loadGOMap (Params-class), 5
- loadGOMap, Params-method
(Params-class), 5
- loadICdata (Params-class), 5
- loadICdata, Params-method
(Params-class), 5

- mclusterSim, 7-9, 9, 11, 12
- mgeneSim, 7-10, 10, 12
- mgoSim, 7-11, 11

- Params, 2-6, 6
- Params-class, 6
- Params-class, 5
- Params-methods (Params), 6

- setCombineMethod<-
(Params-class), 5
- setCombineMethod<-, Params-method
(Params-class), 5
- setMethod<- (Params-class), 5
- setMethod<-, Params-method
(Params-class), 5
- setOntology<- (Params-class), 5
- setOntology<-, Params-method
(Params-class), 5

- setOrganism<- (Params-class), 5
- setOrganism<-, Params-method
(Params-class), 5
- sim (sim-methods), 12
- sim, GeneClusterSet-method
(sim-methods), 12
- sim, GeneSet-method (sim-methods),
12
- sim, GOSet-method (sim-methods), 12
- sim-methods, 12