

Package ‘SimReg’

October 15, 2016

Type Package

Title Similarity Regression

Version 3.0

Date 2016-09-17

Author Daniel Greene

Maintainer Daniel Greene <dg333@cam.ac.uk>

Description Functions for performing Bayesian similarity regression, and evaluating the probability of association between sets of ontological terms and binary response vector. A random model is compared with one in which the log odds of a true response is linked to the semantic similarity between terms and a latent characteristic ontological profile.

License GPL (>= 2)

Imports Rcpp (>= 0.11.1), ontologyIndex (>= 2.0), ontologySimilarity (>= 2.0), ontologyPlot

LinkingTo Rcpp

Depends R (>= 3.0.0)

Suggests knitr

VignetteBuilder knitr

RoxygenNote 5.0.1

NeedsCompilation yes

Repository CRAN

Date/Publication 2016-10-15 18:51:56

R topics documented:

SimReg3-package	2
get_terms	3
get_term_marginals	3
log_BF	4
plot.sim_reg_summary	4

plot_term_marginals	5
print.sim_reg_output	5
print.sim_reg_summary	6
prob_association	6
sim_reg	7
summary.sim_reg_output	8
sum_log_probs	9
term_marginals	9

Index	10
--------------	-----------

SimReg3-package	<i>Similarity Regression Functions</i>
-----------------	--

Description

Functions for performing Bayesian similarity regression, and evaluating the probability of association between sets of ontological terms and binary response vector. A random model is compared with one in which the log odds of a true response is linked to the semantic similarity between terms and a latent characteristic ontological profile.

Details

Package: SimReg3
 Type: Package
 Version: 3.0
 Date: 2016-10-14
 License: GPL (>= 2)

Key functions include `sim_reg`, for similarity regression of binary response variable against an ontologically encoded predictor. An example application would be inferring the probability of association between the presence of a rare genetic variant conditional on an ontologically encoded phenotype.

Author(s)

Daniel Greene <dg333@cam.ac.uk>

Maintainer: Daniel Greene <dg333@cam.ac.uk>

References

D. Greene, NIHR BioResource, S. Richardson, E. Turro, 'Phenotype similarity regression for identifying the genetic determinants of rare diseases', *The American Journal of Human Genetics* 98, 1-10, March 3, 2016.

get_terms	<i>Get full set of terms to use in inference procedure based on similarity function arguments</i>
-----------	---

Description

Get full set of terms to use in inference procedure based on similarity function arguments

Usage

```
get_terms(args)
```

Arguments

args	Named list of named arguments which gets passed to ontological similarity function by <code>sim_reg</code> .
------	--

Value

Character vector of term IDs.

get_term_marginals	<i>Calculate marginal probability of terms inclusion in phi from sim_reg_out object</i>
--------------------	---

Description

Calculate marginal probability of terms inclusion in phi from `sim_reg_out` object

Usage

```
get_term_marginals(sim_reg_out)
```

Arguments

sim_reg_out	Object of class <code>sim_reg_output</code> .
-------------	---

Value

Numeric vector of probabilities, named by term ID.

log_BF	<i>Calculate log Bayes factor for similarity the model, gamma=1 and baseline model, gamma=0.</i>
--------	--

Description

Calculate log Bayes factor for similarity the model, gamma=1 and baseline model, gamma=0.

Usage

```
log_BF(x, ...)
```

Default S3 method:
log_BF(x, ...)

S3 method for class 'sim_reg_output'
log_BF(x, ...)

Arguments

x	list of term sets or sim_reg_output object.
...	If x is a list term sets, other arguments to pass to sim_reg , otherwise this is not used.

Value

Numeric value.

plot.sim_reg_summary	<i>Plot summary of sim_reg_output object</i>
----------------------	--

Description

Plot summary of sim_reg_output object

Usage

```
## S3 method for class 'sim_reg_summary'  
plot(x, ...)
```

S3 method for class 'sim_reg_output'
plot(x, ...)

Arguments

x	Object of class sim_reg_summary.
...	Additional arguments to pass to plot_term_marginals .

plot_term_marginals *Create ontological plot of marginal probabilities of terms*

Description

Create ontological plot of marginal probabilities of terms

Usage

```
plot_term_marginals(ontology, term_marginals, max_terms = 10,  
  min_probability = 0.01, ...)
```

Arguments

ontology	ontology_index object.
term_marginals	Numeric vector of marginal probabilities of inclusion in phi for individual terms, named by the term IDs.
max_terms	Maximum number of terms to include in plot. Note that additional terms may be included when terms have the same marginal probability, and common ancestor terms are included.
min_probability	Threshold probability of inclusion in phi for triggering inclusion in plot.
...	Additional arguments to pass to onto_plot

print.sim_reg_output *Print sim_reg_output object*

Description

Print sim_reg_output object

Usage

```
## S3 method for class 'sim_reg_output'  
print(x, ...)
```

Arguments

x	Object of class sim_reg_output.
...	Non-used arguments.

```
print.sim_reg_summary Print sim_reg_summary object
```

Description

Print sim_reg_summary object

Usage

```
## S3 method for class 'sim_reg_summary'
print(x, ...)
```

Arguments

x	Object of class sim_reg_summary.
...	Non-used arguments.

```
prob_association Calculate probability of association between y and x
```

Description

Calculate probability of association between y and x

Usage

```
prob_association(..., prior = 0.05)
```

Arguments

...	Arguments to pass to log_BF .
prior	Numeric value determining prior probability that gamma=1.

Value

Numeric value.

sim_reg

*Similarity regression***Description**

Performs Bayesian ‘similarity regression’ on given logical response vector y against list of ontological term sets x . It returns an object of class `sim_reg_output`. Of particular interest are the probability of an association, which can be calculated with [prob_association](#), and the characteristic ontological profile ϕ , which can be visualised using the functions [plot_term_marginals](#), and [term_marginals](#)). The results can be summarised with `summary`.

Usage

```
sim_reg(ontology, x, y, information_content = get_term_info_content(ontology,
  x), sim_params = list(ontology = ontology, information_content =
  information_content), using_terms = get_terms(sim_params),
  term_weights = rep(0, length(using_terms)),
  prior = discrete_gamma(using_terms), min_BF = -Inf, max_select = 2000L,
  max_phi_count = 200L, two_way = TRUE, selection_fn = fg_step,
  lik_method = NULL, lik_method_args = list(), gamma0_ml = bg_rate,
  min_ratio = 1e-04, ...)
```

Arguments

<code>ontology</code>	ontology_index object.
<code>x</code>	list of character vectors of ontological terms.
<code>y</code>	logical response vector.
<code>information_content</code>	Numeric vector of information contents of terms named by term ID. Defaults to information content based on frequencies of annotation in x .
<code>sim_params</code>	List of arguments to pass to <code>get_asym_sim_grid</code> .
<code>using_terms</code>	Character vector of term IDs giving the complete set of terms to include in the the ϕ parameter space.
<code>term_weights</code>	Numeric vector of prior weights for individual terms.
<code>prior</code>	Function for computing the unweighted prior probability of a ϕ value.
<code>min_BF</code>	Bayes factor threshold below which to terminate computation, enabling faster execution time at the expense of accuracy and precision.
<code>max_select</code>	Upper bound for number of ϕ values to sample.
<code>max_phi_count</code>	Upper bound for number of ϕ values to include in final likelihood sum.
<code>two_way</code>	Boolean value determining whether to calculate semantic similarity ‘in both directions’ (i.e. compute s_x and s_ϕ or just s_ϕ).
<code>selection_fn</code>	Function for selecting values of ϕ with high posterior mass.
<code>lik_method</code>	Function for calculating marginal likelihood conditional on values of ϕ .

lik_method_args	List of additional arguments to pass to lik_method.
gamma0_ml	Function for computing marginal likelihood of data under baseline model gamma=0.
min_ratio	Lower bound on ratio below which to discard phi values.
...	Additional arguments to pass to selection_fn.

Examples

```
## Not run:
set.seed(0)
data(hpo)
disease_terms <- c("HP:0005537", "HP:0000729", "HP:0001873")
all_terms <- get_ancestors(hpo,
c(disease_terms, sample(hpo$id, size=50)))
y <- c(rep(FALSE, 96), rep(TRUE, 3))
x <- lapply(y, function(.y) minimal_set(
hpo, if (!.y) sample(all_terms, size=3) else
c(sample(all_terms, size=1), disease_terms[runif(n=3) < 0.8])))
sim_reg_out <- sim_reg(ontology=hpo, x=x, y=y)

## End(Not run)
```

```
summary.sim_reg_output
```

Get summary of sim_reg_output object

Description

Get summary of sim_reg_output object

Usage

```
## S3 method for class 'sim_reg_output'
summary(object, prior = 0.05, ...)
```

Arguments

object	Object of class sim_reg_output.
prior	Prior probability of association.
...	Non-used arguments.

sum_log_probs	<i>Calculate sum of log probabilities on log scale without over/under-flow</i>
---------------	--

Description

Calculate sum of log probabilities on log scale without over/under-flow

Usage

```
sum_log_probs(log_probs)
```

Arguments

log_probs Numeric vector of probabilities on log scale.

Value

Numeric value on log scale.

term_marginals	<i>Calculate marginal probability of terms inclusion in phi</i>
----------------	---

Description

Calculate marginal probability of terms inclusion in phi

Usage

```
term_marginals(...)
```

Arguments

... Arguments to pass to [sim_reg](#).

Value

Numeric vector of probabilities, named by term ID.

Index

*Topic **HPO , MCMC, ontology**

SimReg3-package, 2

get_term_marginals, 3

get_terms, 3

log_BF, 4, 6

plot.sim_reg_output

(plot.sim_reg_summary), 4

plot.sim_reg_summary, 4

plot_term_marginals, 4, 5, 7

print.sim_reg_output, 5

print.sim_reg_summary, 6

prob_association, 6, 7

sim_reg, 4, 7, 9

SimReg3 (SimReg3-package), 2

SimReg3-package, 2

sum_log_probs, 9

summary.sim_reg_output, 8

term_marginals, 7, 9