Package: iGSEA (via r-universe)

October 18, 2024

Type Package

Title Integrative Gene Set Enrichment Analysis Approaches

Version 1.2

Date 2017-05-07

Author Wentao Lu, Xinlei Wang, Xiaowei Zhan

Maintainer Wentao Lu <wlu1026@yahoo.com>

Description To integrate multiple GSEA studies, we propose a hybrid strategy, iGSEA-AT, for choosing random effects (RE) versus fixed effect (FE) models, with an attempt to achieve the potential maximum statistical efficiency as well as stability in performance in various practical situations. In addition to iGSEA-AT, this package also provides options to perform integrative GSEA with testing based on a FE model (iGSEA-FE) and testing based on a RE model (iGSEA-RE). The approaches account for different set sizes when testing a database of gene sets. The function is easy to use, and the three approaches can be applied to both binary and continuous phenotypes.

License GPL-2

NeedsCompilation no

Date/Publication 2017-05-12 23:12:53 UTC

Repository https://luwt.r-universe.dev

RemoteUrl https://github.com/cran/iGSEA

RemoteRef HEAD

RemoteSha 5a4150b8208c3bfaf5e131c358196a4185683aea

Contents

iGSEA-package *Integrative GSEA Approaches*

Description

This package provides three approaches, testing based on an fixed-effect (FE) model (iGSEA-FE), testing based on a random-effect (RE) model (iGSEA-RE), and adaptive testing (iGSEA-AT) to integrate multiple gene set enrichment studies. These approaches can be applied to both binary and continuous phenotypes. The output of the function will be the Q-values of gene sets to control the false discovery rate (FDR). We recommend iGSEA-AT due to its stability in performance in various practical situations.

Usage

```
igsea.test(gel, pheno, ssize, gind, gsind, B = 500,
     vtype = "binary", method = "AT", alpha1 = 0.0253, pihat = 1)
```

Arguments

`gel` a numeric matrix of gene expression levels merged from component studies. Rows represent genes and columns represent samples.

`pheno` a numeric vector of phenotypes merged from component studies.

`ssize` a numeric vector indicating the number of samples in each study.

`gind` a matrix indicating if genes are included in studies. Use 1 as included and 0 as not. Rows represent genes and columns represent studies.

`gsind` a matrix indicating if genes belongs to gene sets. Use 1 as in a gene set and 0 as not. Rows represent genes and columns represent gene sets.

`B` an integer indicating the times of shuffling gene labels in order to compute the permuted enrichment scores. It is 500 by default.

`vtype` a character string specifying the type of phenotypes. It can only be "binary" or "continuous" at this moment.

`method` a character string specifying the approach you want to use, must be one of "AT" (default), "FE", or "RE".

`alpha1` a number indicating the first-stage significance level for iGSEA-AT. It is 0.0253 by default.

`pihat` a number indicating a rough estimate of the proportion of non-enriched sets. It is 1 by default.

Details

Package:	iGSEA
Type:	Package
Version:	1.2

Date:	2017-05-07
License:	GPL-2

The core function in this package is `igsea.test`, which checks whether your input is correct at the very beginning. Specifically, the length of the vector pheno, the number of columns of the matrix `gel`, and the sum of the vector `ssize` should be equal, indicating the total number of samples. The number of rows of the matrix `gel`, the matrix `gind`, and the matrix `gsind` should be equal, indicating the total number genes involved in studies. The number of columns of the matrix `gind` and the length of the vector `ssize` should be equal, indicating the number of studies.

Please note genes with the same IDs should mean the same genes in reality. In the matrix gind, a gene should be involved in at least one study.

The approaches account for different set sizes when testing multiple gene sets.

The parameter alpha1 should be greater than 0 and smaller than the overall significance level alpha. The choice of alpha1 could be chosen based on some exploratory data analysis or the prior knowledge about the between-study heterogeneity. It slightly influences the performance of iGSEA-AT. The default value is chosen to be 1-sqrt(1-alpha). When alpha=0.05, the default value is 0.0253.

Value

the Q-values of gene sets after false discovery rate (FDR) procedure

Author(s)

Wentao Lu, Xinlei Wang, Xiaowei Zhan

Maintainer: Wentao Lu <wlu1026@yahoo.com>

References

Lu, Wentao (2016), Meta-analysis approaches to combine multiple gene set enrichment studies. Unpublished doctoral dissertation, Southern Methodist University.

Examples

```
#Set seed to make sure the permutaiton test gives the same results
set.seed(1234)

#In the following binray example, there are 200 genes in total.
#Genes 1-40 are set to be up-regulated genes as their gene expression levels are
#associated with phenotypes.
#The remaining 160 genes are equally expressed genes.
#Gene set 1, which contains 40% up-regulated genes, is enriched.
#Gene set 2, which contains 20% up-regulated genes, is not enriched.
#As there are no RE genes, FE and AT should perform well.
G <- matrix(rnorm(200 * 60), c(200, 60))              #200 genes and 60 samples in total
P <- c(rep(c(rep(1, 5), rep(0, 5)), 2), rep(c(rep(1, 10), rep(0, 10)), 2))     #phenotypes
G[1:40, c(1:5, 11:15, 21:30, 41:50)] <- G[1:40, c(1:5, 11:15, 21:30, 41:50)] + 0.45
S <- c(10, 10, 20, 20)                              #the number of samples in each study
I <- matrix(rep(1, 200*4), 200)                     #all genes are included in 4 studies
```

```
GS <- matrix(0, 200, 2)
GS[c(1:20, 151:180), 1] <- 1                                       #gene set 1 is enriched
GS[c(31, 80), 2] <- 1                                          #gene set 2 is non-enriched
igsea.test(G, P, S, I, GS)   #the output vector consists of two Q-values for the gene sets

#A similar normal example is also provided below:
set.seed(1234)
G <- matrix(rnorm(200 * 60), c(200, 60))                #200 genes and 60 samples in total
P <- rnorm(60)                                                                 #phenotypes
S <- c(10, 10, 20, 20)                                #the number of samples in each study
rho_raw <- matrix(0, 200, 4)
for (i in 1:40) rho_raw[i, ] <- rnorm(4, mean = 0.3, sd = 0.1)
beta <- matrix(0, 200, 60)
for (i in 1:200) beta[i, ] <- beta[i, ] + c(rep(rho_raw[i, 1], 10), rep(rho_raw[i, 2], 10),
rep(rho_raw[i, 3], 20), rep(rho_raw[i, 4], 20))
for (i in 1:200) {
  for (j in 1:60){
    G[i, j] <- rnorm(1, mean = beta[i, j] * P[j], sd = sqrt(1 - beta[i, j] ^ 2))
  }
}
I <- matrix(rep(1, 200*4), 200)                       #all genes are included in 4 studies
GS <- matrix(0, 200, 2)
GS[c(1:20, 151:180), 1] <- 1                                       #gene set 1 is enriched
GS[c(31, 80), 2] <- 1                                          #gene set 2 is non-enriched
igsea.test(G, P, S, I, GS, vtype = ”continuous”)
```

Index