Package 'mtPRSFIMEL'

October 29, 2025

Type Package

Title Improving multi-trait polygenic risk score prediction using fine-mapping and ensemble learning

Version 0.1.0

Maintainer Song Zhai <zsviolet1993@gmail.com>

Description Construct multi-trait PRS (mtPRS) via performing fine-mapping and ensemble learning. Package is based on "Improving multi-trait polygenic risk score prediction using fine-mapping and ensemble learning" by Zhai, S., Guo, B., and Shen, J., 2025.

License GPL (>= 2)

Depends R (>= 4.0.0)

Imports dplyr (>= 1.0.2), mvtnorm (>= 1.1.1), data.table (>= 1.13.2), Matrix (>= 1.2.18), glmnet (>= 4.0.2), stats (>= 4.0.3), ACAT (>= 0.91)

Suggests knitr, rmarkdown

VignetteBuilder knitr

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

NeedsCompilation yes

Author Song Zhai [aut, cre]

Repository CRAN

R topics documented:

`calculate_PRS` *Calculate individual PRS for each trait based on refined effect size estimates from PolyPred*

Description

Calculate individual PRS for each trait based on refined effect size estimates from PolyPred

Usage

```
calculate_PRS(K, base, pcut, G.target)
```

Arguments

`K`	number of traits
`base`	list of disease GWAS summary statistics of K traits from base cohort
`pcut`	p-value cutoff
`G.target`	genotype in target cohort

Details

calculate_PRS needs disease GWAS summary statistics from K traits, assuming the effect size estimates have been refined by PolyPred after fine-mapping

Value

a matrix of K individual PRSs

Author(s)

Song Zhai

References

Zhai, S., Guo, B., and Shen, J., 2025. Improving multi-trait polygenic risk score prediction using fine-mapping and ensemble learning.

Examples

```
dat <- generate_pgx_data(structure = "CS", sparseness = "no",
rho_DT = c(0.5,0.5,0.5,0.5), rho_T = 0.5, rho_E = 0.5, rho_C = 0.2,
K=4, m=2000, pcausal=0.1, blocksize=100,
gamma=1, samplesize=700, h2_base=0.3, h2_target=0.3)
re <- calculate_PRS(K = 4, base = dat$base, pcut = 1, G.target = dat$target[,-c(1,2)])
head(re)
```

`calculate_weight` *Calculate weights by by performing principal component analysis (PCA) on genetic correlation matrix among traits*

Description

Calculate weights by by performing principal component analysis (PCA) on genetic correlation matrix among traits

Usage

```
calculate_weight(K, corr, varcut)
```

Arguments

`K`	number of traits
`corr`	genetic correlation matrix
`varcut`	variance cutoff

Details

Calculate weights by by performing principal component analysis (PCA) on genetic correlation matrix among traits

Value

a vector of weights

Author(s)

Song Zhai

References

Zhai, S., Guo, B., and Shen, J., 2025. Improving multi-trait polygenic risk score prediction using fine-mapping and ensemble learning.

Examples

```
dat <- generate_pgx_data(structure = "CS", sparseness = "no",
rho_DT = c(0.5,0.5,0.5,0.5), rho_T = 0.5, rho_E = 0.5, rho_C = 0.2,
K=4, m=2000, pcausal=0.1, blocksize=100,
gamma=1, samplesize=700, h2_base=0.3, h2_target=0.3)
re <- calculate_weight(K = 4, corr = dat$corr, varcut = 0.8)
hist(re$mtPRS)
```

`generate_pgx_data` *Simulate PGx GWAS data for mtPRS analysis*

Description

Simulate PGx GWAS data for mtPRS analysis

Usage

```
generate_pgx_data(
  structure = "CS",
  sparseness = "no",
  rho_DT,
  rho_T,
  rho_E,
  rho_C,
  K = 4,
  m = 1000,
  pcausal = 0.1,
  blocksize = 100,
  gamma = 1,
  samplesize = 700,
  h2_base = 0.3,
  h2_target = 0.3
)
```

Arguments

`structure`	genetic correlation structure, either uniformly correlated (i.e., "CS" or "AR1") or clustered correlated ("clustered")
`sparseness`	effect sparsity, either "no" or "half" or "more"
`rho_DT`	effect correlations between traits in base cohort and the phenotype in target cohort
`rho_T`	genetic correlation among traits in base cohort
`rho_E`	correlation between prognostic and predictive effects in target cohort
`rho_C`	between-cluster correlation when the genetic correlation structure is clustered correlated
`K`	number of traits
`m`	number of SNPs
`pcausal`	proportion of causal SNPs
`blocksize`	LD block size (i.e., the number of SNPs in each LD block)
`gamma`	prognostic-to-predictive effect size ratio
`samplesize`	sample size of target cohort
`h2_base`	heritability in base cohort
`h2_target`	heritability in target cohort

Details

Simulate PGx GWAS data for mtPRS analysis

Value

a list of disease GWAS summary statistics in base cohort, individual-level data in validation and target cohorts, genetic correlation matrix among traits, and true effect sizes in base and target cohorts

Author(s)

Song Zhai

References

Zhai, S., Guo, B., and Shen, J., 2025. Improving multi-trait polygenic risk score prediction using fine-mapping and ensemble learning.

Examples

```
dat <- generate_pgx_data(structure = "CS", sparseness = "no",
rho_DT = c(0.5,0.5,0.5,0.5), rho_T = 0.5, rho_E = 0.5, rho_C = 0.2,
K=4, m=2000, pcausal=0.1, blocksize=100,
gamma=1, samplesize=700, h2_base=0.3, h2_target=0.3)
plot(dat$truesize[,1])
print(dat$corr)
```

---

mtPRS_FIMEL *Construct mtPRS by performing ensemble learning*

Description

Construct mtPRS by performing ensemble learning

Usage

```
mtPRS_FIMEL(
  base,
  target,
  validation,
  corr,
  pcut = 0.05,
  varcut = 0.8,
  K = 4,
  phenotype = "pgx"
)
```

Arguments

base	list of disease GWAS summary statistics of K traits
target	data frame in target cohort, including Y, T (if the target cohort is from a PGx study), and G
validation	data frame from an independent validation cohort, including Y, T (if the validation cohort is from a PGx study), and G
corr	genetic correlation matrix among K traits
pcut	p-value cutoff for C+T method to construct individual stPRS
varcut	variance cutoff to choose top principal components (PCs)
K	number of traits
phenotype	indicator of phenotype in target cohort, either "dis" or "pgx"

Details

Construct mtPRS by performing ensemble learning

Value

a list of fitted model, stPRSs and mtPRSs, and mtPRS-FIMEL

Author(s)

Song Zhai

References

Zhai, S., Guo, B., and Shen, J., 2025. Improving multi-trait polygenic risk score prediction using fine-mapping and ensemble learning.

Examples

```
dat <- generate_pgx_data(structure = "CS", sparseness = "no",
rho_DT = c(0.5,0.5,0.5,0.5), rho_T = 0.5, rho_E = 0.5, rho_C = 0.2,
K=4, m=2000, pcausal=0.1, blocksize=100,
gamma=1, samplesize=700, h2_base=0.3, h2_target=0.3)
re <- mtPRS_FIMEL(dat$base, dat$target, dat$validation, dat$corr, pcut = 1, varcut = 0.8, K = 4, phenotype = "pg
hist(re$slPRS_validation)
```

---

mtPRS_ML *Construct mtPRS by performing elastic net*

Description

Construct mtPRS by performing elastic net

Usage

```
mtPRS_ML(base, target, validation, pcut = 0.05, K = 4, phenotype = "pgx")
```

Arguments

base	list of disease GWAS summary statistics of K traits
target	data frame in target cohort, including Y, T (if the target cohort is from a PGx study), and G
validation	data frame from an independent validation cohort, including Y, T (if the validation cohort is from a PGx study), and G
pcut	p-value cutoff for C+T method to construct individual stPRS
K	number of traits
phenotype	indicator of phenotype in target cohort, either "dis" or "pgx"

Details

Construct mtPRS by performing elastic net

Value

a list of fitted model, individual stPRSs, and mtPRS-ML

Author(s)

Song Zhai

References

Zhai, S., Guo, B., and Shen, J., 2025. Improving multi-trait polygenic risk score prediction using fine-mapping and ensemble learning.

Examples

```
dat <- generate_pgx_data(structure = "CS", sparseness = "no",
rho_DT = c(0.5,0.5,0.5,0.5), rho_T = 0.5, rho_E = 0.5, rho_C = 0.2,
K=4, m=2000, pcausal=0.1, blocksize=100,
gamma=1, samplesize=700, h2_base=0.3, h2_target=0.3)
re <- mtPRS_ML(dat$base, dat$target, dat$validation, pcut = 1, varcut = 0.8, K = 4, phenotype = "pgx")
hist(re$mtPRS_validation)
```

---

mtPRS_PCA *Construct mtPRS by performing PCA on genetic correlation matrix among traits*

Description

Construct mtPRS by performing PCA on genetic correlation matrix among traits

Usage

```
mtPRS_PCA(
  base,
  target,
  validation,
  corr,
  pcut = 0.05,
  varcut = 0.8,
  K = 4,
  phenotype = "pgx"
)
```

Arguments

base	list of disease GWAS summary statistics of K traits
target	data frame in target cohort, including Y, T (if the target cohort is from a PGx study), and G
validation	data frame from an independent validation cohort, including Y, T (if the validation cohort is from a PGx study), and G
corr	genetic correlation matrix among K traits
pcut	p-value cutoff for C+T method to construct individual stPRS
varcut	variance cutoff to choose top principal components (PCs)
K	number of traits
phenotype	indicator of phenotype in target cohort, either "dis" or "pgx"

Details

Construct mtPRS by performing PCA on genetic correlation matrix among traits

Value

a list of weights, individual stPRSs, and mtPRS-PCA

Author(s)

Song Zhai

References

Zhai, S., Guo, B., and Shen, J., 2025. Improving multi-trait polygenic risk score prediction using fine-mapping and ensemble learning.

Examples

```
dat <- generate_pgx_data(structure = "CS", sparseness = "no",
rho_DT = c(0.5,0.5,0.5,0.5), rho_T = 0.5, rho_E = 0.5, rho_C = 0.2,
K=4, m=2000, pcausal=0.1, blocksize=100,
gamma=1, samplesize=700, h2_base=0.3, h2_target=0.3)
re <- mtPRS_PCA(dat$base, dat$target, dat$validation, dat$corr, pcut = 1, varcut = 0.8, K = 4, phenotype = "pgx"
hist(re$mtPRS_validation)
```

Index