Package 'CARMA'

October 21, 2022

Type Package

Title CAusal Robust Mapping Method with Annotations (CARMA)

Version 1.0

Date 2022-01-19

Author Zikun Yang [aut, cre, cph],
Chen Wang [aut, ctb],
Iuliana Ionita-Laza [aut, ctb].

Maintainer Zikun Yang <yangzikun1125@gmail.com>

Description Implementations for the Bayesian fine-mapping method CAusal Robust Mapping method with Annotations (CARMA). CARMA uses common GWAS data and linkage disequilibrium (LD) from external panels to predict causal variants. CARMA allows users to introduce functional annotations to jointly models summary statistics of GWAS and high-dimensional functional annotations. Also, CARMA provides a novel Bayesian hypothesis testing approach to account for discrepancies between summary statistics and LD from external reference panels in order to avoid an increase in false positives.

Depends R (>= 4.0)

License GPL (>= 2)

Encoding UTF-8

Imports methods,
stats,
dplyr,
glmnet,
Matrix,
MASS,
Rcpp (>= 1.0.6)

LinkingTo Rcpp, RcppArmadillo, RcppGSL

RoxygenNote 7.2.1

NeedsCompilation yes

R topics documented:

CARMA 2
CARMA_fixed_sigma 4
CARMA_fixed_sigma_pro 6
CARMA_fixed_sigma_xxx 8

Cauchy_fixed_sigma_marginal 10
Cauchy_marginal 10
compute_summary_statistics 11
hyper_g_fixed_sigma_marginal 12
hyper_g_marginal 12
Normal_fixed_sigma_marginal 12
Normal_marginal 12

Index **13**

CARMA *CAusal Robust Mapping Method with Annotations(CARMA)*

Description

Performs a Bayesian fine-mapping model in order to identify putative causal variants at GWAS loci. This function requires the summary statistics of the SNPs in the testing loci, and the corresponding simple-LD matrices for fine-mapping. Functional annotations can be included as the prior information of the causality of the testing SNPs. The model can be executed chromosome-wise to increase power.

Usage

```
CARMA(
  z.list,
  ld.list,
  w.list = NULL,
  lambda.list,
  tau = NULL,
  effect.size.prior = "Cauchy",
  rho.index = 0.99,
  BF.index = 10,
  Max.Model.Dim = 20000,
  all.iter = 10,
  all.inner.iter = 10,
  label.list = NULL,
  output.labels = ".",
  input.alpha = 0,
  epsilon.threshold = 0.001,
  input.prior.prob = NULL
)
```

Arguments

`z.list`	Input list of the summary statistics of the testing loci, and each element of the list is the summary statistics of each individual locus.
`ld.list`	Input list of the LD correlation matrix of the testing loci, and each element of the list is the LD matrix of each individual locus.
`w.list`	Input list of the functional annotations of the testing loci, and each element of the list is the functional annotation matrix of each individual locus.

`lambda.list` Input list of the hyper-parameter η of the testing loci, and each element of the list is the hyper-parameter of each individual locus.

`effect.size.prior` The prior of the effect size. The choice are 'Cauchy', 'Hyper-g', and 'Normal' priors, where the Cauchy prior is the default prior.

`rho.index` A number between 0 and 1 specifying ρ of the estimated credible sets.

`BF.index` A number greater than 1 to specifying the threshold of the Bayes factor of the estimated credible models.

`Max.Model.Dim` Maximum number of the top candidate models based on the unnormalized posterior probability.

`all.iter` Maximum iterations for EM algorithm to run.

`all.inner.iter` Maximum iterations for Shotgun algorithm to run per iteration within EM algorithm.

`label.list` Input list of the names of the testing loci. Default is NULL.

`output.labels` Output directory where output will be written while CARMA is running. Default is the root directory '.'.

`input.alpha` The elastic net mixing parameter, where $0 \leq \alpha \leq 1$.

`epsilon.threshold` Convergence threshold measured by average of Bayes factors.

Details

The function performs a Bayesian fine-mapping method.

Value

The form of the return is a list, for each list:

- pip - The posterior inclusion probability of each individual locus.
- Credible set - The information regarding the credible set given a threshold ρ.
- Credible model - The information regarding the credible model given a threshold of the Bayes factor.

Examples

```
# Example
set.seed(1)
n = 400
p = 500
beta = rep(0,p)
beta[1] = 1
X = matrix(rnorm(n*p),nrow = n,ncol = p)
X = scale(X,center = TRUE,scale = TRUE)
y = drop(X %*% beta + rnorm(n))
SS=compute_summary_statistics(X,y)
z.list<-list()
z.list[[1]]<-(SS$betahat/SS$sebetahat)
ld.list<-list()
ld.list[[1]]<-cov(X)
lambda.list<-list()
lambda.list[[1]]<-1/sqrt(p)
CARMA.result<-CARMA(z.list,ld.list=ld.list,lambda.list = lambda.list,effect.size.prior='Hyper-g')
```

CARMA_fixed_sigma *CARMA (fixed variance)*

Description

Performs a Bayesian fine-mapping model in order to identify putative causal variants at GWAS loci. The model requires the summary statistics of the SNPs in the testing loci, the corresponding LD matrices for fine-mapping, and an estimated variance of traits. Functional annotations can be included as the prior information of the causality of the testing SNPs. The model also provides a procedure of outlier detection, which resolves the discrepancies between the summary statistics and the LD matrix extracted from reference panels. The model can be executed chromosome-wise to increase power.

Usage

```
CARMA_fixed_sigma(
  z.list,
  ld.list,
  w.list = NULL,
  lambda.list = NULL,
  output.labels = ".",
  label.list = NULL,
  effect.size.prior = "Spike-slab",
  rho.index = 0.99,
  BF.index = 10,
  EM.dist = "Logistic",
  outlier.hypo.form = "New.way",
  Max.Model.Dim = 2e+05,
  all.iter = 3,
  all.inner.iter = 10,
  input.alpha = 0,
  epsilon.threshold = 1e-04,
  num.causal = 10,
  y.var = 1,
  tau = 0.04,
  outlier.switch = T,
  outlier.BF.index = 1/3.2,
  prior.prob.computation = "Logistic"
)
```

Arguments

z.list	Input list of the summary statistics of the testing loci, and each element of the list is the summary statistics of each individual locus.
ld.list	Input list of the LD correlation matrix of the testing loci, and each element of the list is the LD matrix of each individual locus.
w.list	Input list of the functional annotations of the testing loci, and each element of the list is the functional annotation matrix of each individual locus.
lambda.list	Input list of the hyper-parameter η of the testing loci, and each element of the list is the hyper-parameter of each individual locus.

`output.labels` Output directory where output will be written while CARMA is running. Default is the OS root directory ".".

`label.list` Input list of the names of the testing loci. Default is NULL.

`effect.size.prior` The prior of the effect size. The choice are 'Cauchy' and 'Spike-slab' priors, where the 'Spike-slab' prior is the default prior.

`rho.index` A number between 0 and 1 specifying ρ of the estimated credible sets.

`BF.index` A number smaller than 1 to specifying the threshold of the Bayes factor of the estimated credible models. The default setting is 3.2.

`Max.Model.Dim` Maximum number of the top candidate models based on the unnormalized posterior probability.

`all.iter` Maximum iterations for EM algorithm to run.

`all.inner.iter` Maximum iterations for Shotgun algorithm to run per iteration within EM algorithm.

`input.alpha` The elastic net mixing parameter, where $0 \leq \alpha \leq 1$.

`epsilon.threshold` Convergence threshold measured by average of Bayes factors.

`num.causal` The maximum number of causal variants assumed per locus, which is 10 causal SNPs per locus by default.

`y.var` The input variance of the summary statistics, the default value is 1 as the summary statistics are standardized.

`outlier.switch` The indicator variable of whether turn on the outlier detection. We suggest that the detection should always turn on if using external LD matrix.

`outlier.threshold` The Bayes threshold of the hypothesis testing of determining outliers, which is 10 by default.

Details

The function performs a Bayesian fine-mapping method.

Value

The form of the return is a list, for each list:

- pip - The posterior inclusion probability of each individual locus.
- Credibleset - The information regarding the credible set given a threshold ρ.
- Credible model - The information regarding the credible model given a threshold of the Bayes factor.
- Outliers - The information regarding the detected outliers.

Examples

```
# Example
set.seed(1)
n = 400
p = 500
beta = rep(0,p)
beta[1] = 1
```

```
X = matrix(rnorm(n*p),nrow = n,ncol = p)
X = scale(X,center = TRUE,scale = TRUE)
y = drop(X %*% beta + rnorm(n))
SS=compute_summary_statistics(X,y)
z.list<-list()
z.list[[1]]<-(SS$betahat/SS$sebetahat)
ld.list<-list()
ld.list[[1]]<-cov(X)
lambda.list<-list()
lambda.list[[1]]<-1/sqrt(p)
CARMA.result<-CARMA_fixed_sigma(z.list,ld.list=ld.list,
lambda.list = lambda.list,effect.size.prior='Hyper-g')
```

`CARMA_fixed_sigma_pro` *CARMA (fixed variance)*

Description

Performs a Bayesian fine-mapping model in order to identify putative causal variants at GWAS loci. The model requires the summary statistics of the SNPs in the testing loci, the corresponding LD matrices for fine-mapping, and an estimated variance of traits. Functional annotations can be included as the prior information of the causality of the testing SNPs. The model also provides a procedure of outlier detection, which resolves the discrepancies between the summary statistics and the LD matrix extracted from reference panels. The model can be executed chromosome-wise to increase power.

Usage

```
CARMA_fixed_sigma_pro(
  z.list,
  ld.list,
  w.list = NULL,
  lambda.list = NULL,
  output.labels = ".",
  label.list = NULL,
  effect.size.prior = "Cauchy",
  rho.index = 0.99,
  BF.index = 10,
  Max.Model.Dim = 2e+05,
  all.iter = 10,
  all.inner.iter = 10,
  input.alpha = 0,
  epsilon.threshold = 1e-04,
  num.causal = 10,
  y.var = 1,
  outlier.switch = T,
  outlier.BF.index = 10
)
```

Arguments

`z.list`	Input list of the summary statistics of the testing loci, and each element of the list is the summary statistics of each individual locus.

`ld.list` Input list of the LD correlation matrix of the testing loci, and each element of the list is the LD matrix of each individual locus.

`w.list` Input list of the functional annotations of the testing loci, and each element of the list is the functional annotation matrix of each individual locus.

`lambda.list` Input list of the hyper-parameter η of the testing loci, and each element of the list is the hyper-parameter of each individual locus.

`output.labels` Output directory where output will be written while CARMA is running. Default is the OS root directory ".".

`label.list` Input list of the names of the testing loci. Default is NULL.

`effect.size.prior` The prior of the effect size. The choice are 'Cauchy', 'Hyper-g', and 'Normal' priors, where the Cauchy prior is the default prior.

`rho.index` A number between 0 and 1 specifying ρ of the estimated credible sets.

`BF.index` A number greater than 1 to specifying the threshold of the Bayes factor of the estimated credible models.

`Max.Model.Dim` Maximum number of the top candidate models based on the unnormalized posterior probability.

`all.iter` Maximum iterations for EM algorithm to run.

`all.inner.iter` Maximum iterations for Shotgun algorithm to run per iteration within EM algorithm.

`input.alpha` The elastic net mixing parameter, where $0 \leq \alpha \leq 1$.

`epsilon.threshold` Convergence threshold measured by average of Bayes factors.

`num.causal` The maximum number of causal variants assumed per locus, which is 10 causal SNPs per locus by default.

`y.var` The input variance of the summary statistics, the default value is 1 as the summary statistics are standardized.

`outlier.switch` The indicator variable of whether turn on the outlier detection. We suggest that the detection should always turn on if using external LD matrix.

`outlier.threshold` The Bayes threshold of the hypothesis testing of determining outliers, which is 10 by default.

Details

The function performs a Bayesian fine-mapping method.

Value

The form of the return is a list, for each list:

- pip - The posterior inclusion probability of each individual locus.
- Credibleset - The information regarding the credible set given a threshold ρ.
- Credible model - The information regarding the credible model given a threshold of the Bayes factor.
- Outliers - The information regarding the detected outliers.

Examples

```
# Example
set.seed(1)
n = 400
p = 500
beta = rep(0,p)
beta[1] = 1
X = matrix(rnorm(n*p),nrow = n,ncol = p)
X = scale(X,center = TRUE,scale = TRUE)
y = drop(X %*% beta + rnorm(n))
SS=compute_summary_statistics(X,y)
z.list<-list()
z.list[[1]]<-(SS$betahat/SS$sebetahat)
ld.list<-list()
ld.list[[1]]<-cov(X)
lambda.list<-list()
lambda.list[[1]]<-1/sqrt(p)
CARMA.result<-CARMA_fixed_sigma(z.list,ld.list=ld.list,
lambda.list = lambda.list,effect.size.prior='Hyper-g')
```

CARMA_fixed_sigma_xxx *CARMA (fixed variance)*

Description

Performs a Bayesian fine-mapping model in order to identify putative causal variants at GWAS loci. The model requires the summary statistics of the SNPs in the testing loci, the corresponding LD matrices for fine-mapping, and an estimated variance of traits. Functional annotations can be included as the prior information of the causality of the testing SNPs. The model also provides a procedure of outlier detection, which resolves the discrepancies between the summary statistics and the LD matrix extracted from reference panels. The model can be executed chromosome-wise to increase power.

Usage

```
CARMA_fixed_sigma_xxx(
  z.list,
  ld.list,
  w.list = NULL,
  lambda.list = NULL,
  output.labels = ".",
  label.list = NULL,
  effect.size.prior = "Cauchy",
  rho.index = 0.99,
  BF.index = 10,
  Max.Model.Dim = 2e+05,
  all.iter = 10,
  all.inner.iter = 10,
  input.alpha = 0,
  epsilon.threshold = 1e-04,
  num.causal = 10,
  y.var = 1,
```

```
  tau = 0.05,
  outlier.switch = T,
  outlier.BF.index = 0.1
)
```

Arguments

z.list	Input list of the summary statistics of the testing loci, and each element of the list is the summary statistics of each individual locus.
ld.list	Input list of the LD correlation matrix of the testing loci, and each element of the list is the LD matrix of each individual locus.
w.list	Input list of the functional annotations of the testing loci, and each element of the list is the functional annotation matrix of each individual locus.
lambda.list	Input list of the hyper-parameter η of the testing loci, and each element of the list is the hyper-parameter of each individual locus.
output.labels	Output directory where output will be written while CARMA is running. Default is the OS root directory ".".
label.list	Input list of the names of the testing loci. Default is NULL.
effect.size.prior	The prior of the effect size. The choice are 'Cauchy', 'Hyper-g', and 'Normal' priors, where the Cauchy prior is the default prior.
rho.index	A number between 0 and 1 specifying ρ of the estimated credible sets.
BF.index	A number greater than 1 to specifying the threshold of the Bayes factor of the estimated credible models.
Max.Model.Dim	Maximum number of the top candidate models based on the unnormalized posterior probability.
all.iter	Maximum iterations for EM algorithm to run.
all.inner.iter	Maximum iterations for Shotgun algorithm to run per iteration within EM algorithm.
input.alpha	The elastic net mixing parameter, where $0 \leq \alpha \leq 1$.
epsilon.threshold	Convergence threshold measured by average of Bayes factors.
num.causal	The maximum number of causal variants assumed per locus, which is 10 causal SNPs per locus by default.
y.var	The input variance of the summary statistics, the default value is 1 as the summary statistics are standardized.
outlier.switch	The indicator variable of whether turn on the outlier detection. We suggest that the detection should always turn on if using external LD matrix.
outlier.threshold	The Bayes threshold of the hypothesis testing of determining outliers, which is 10 by default.

Details

The function performs a Bayesian fine-mapping method.

Value

The form of the return is a list, for each list:

- pip - The posterior inclusion probability of each individual locus.
- Credibleset - The information regarding the credible set given a threshold ρ.
- Credible model - The information regarding the credible model given a threshold of the Bayes factor.
- Outliers - The information regarding the detected outliers.

Examples

```
# Example
set.seed(1)
n = 400
p = 500
beta = rep(0,p)
beta[1] = 1
X = matrix(rnorm(n*p),nrow = n,ncol = p)
X = scale(X,center = TRUE,scale = TRUE)
y = drop(X %*% beta + rnorm(n))
SS=compute_summary_statistics(X,y)
z.list<-list()
z.list[[1]]<-(SS$betahat/SS$sebetahat)
ld.list<-list()
ld.list[[1]]<-cov(X)
lambda.list<-list()
lambda.list[[1]]<-1/sqrt(p)
CARMA.result<-CARMA_fixed_sigma(z.list,ld.list=ld.list,
lambda.list = lambda.list,effect.size.prior='Hyper-g')
```

Cauchy_fixed_sigma_marginal *Marginal likelihood of Cauchy prior when varinace is fixed*

Description

Marginal likelihood of Cauchy prior when varinace is fixed

Marginal likelihood of Cauchy prior when varinace is fixed

Cauchy_marginal *Marginal likelihood of Cauchy prior*

Description

Marginal likelihood of Cauchy prior

compute_summary_statistics
Perform Univariate Linear Regression Separately for Columns of X

Description

This is a function provided in the package of "susieR", Wang et al (2020) <doi:10.1101/501114>, for performing the univariate linear regression y ~ x separately for each column x of X to generate summary statistics. Each regression is implemented using .lm.fit(). The estimated effect size and stardard error for each variable are outputted.

Usage

```
compute_summary_statistics(X, y, Z = NULL, center = TRUE,
                                   scale = TRUE, return_residuals = FALSE)
```

Arguments

X	n by p matrix of regressors.
y	n-vector of response variables.
Z	Optional n by k matrix of covariates to be included in all regresions. If Z is not NULL, the linear effects of covariates are removed from y first, and the resulting residuals are used in place of y.
center	If center = TRUE, center X, y and Z.
scale	If scale = TRUE, scale X, y and Z.
return_residuals	Whether or not to output the residuals if Z is not NULL.

Details

A list with two vectors containing the least-squares estimates of the coefficients (betahat) and their standard errors (sebetahat). Optionally, and only when a matrix of covariates Z is provided, a third vector residuals containing the residuals is returned.

Examples

```
# Example
set.seed(1)
n = 400
p = 500
beta = rep(0,p)
beta[1] = 1
X = matrix(rnorm(n*p),nrow = n,ncol = p)
X = scale(X,center = TRUE,scale = TRUE)
y = drop(X %*% beta + rnorm(n))
SS=compute_summary_statistics(X,y)
```

`hyper_g_fixed_sigma_marginal` *Marginal likelihood of hyper-g prior when varinace is fixed*

Description

Marginal likelihood of hyper-g prior when varinace is fixed

`hyper_g_marginal` *Marginal likelihood of hyper-g prior*

Description

Marginal likelihood of hyper-g prior

`Normal_fixed_sigma_marginal` *Marginal likelihood of Normal prior when varinace is fixed*

Description

Marginal likelihood of Normal prior when varinace is fixed

Marginal likelihood of Normal prior when varinace is fixed

Marginal likelihood of Normal prior when varinace is fixed

Marginal likelihood of Normal prior when varinace is fixed

`Normal_marginal` *Marginal likelihood of Normal prior*

Description

Marginal likelihood of Normal prior

Index

CARMA, 2
CARMA_fixed_sigma, 4
CARMA_fixed_sigma_pro, 6
CARMA_fixed_sigma_xxx, 8
Cauchy_fixed_sigma_marginal, 10
Cauchy_marginal, 10
compute_summary_statistics, 11

hyper_g_fixed_sigma_marginal, 12
hyper_g_marginal, 12

Normal_fixed_sigma_marginal, 12
Normal_marginal, 12