Package 'HIMA'

April 27, 2023

Type Package

Title High-Dimensional Mediation Analysis

Version 2.2.0

Date 2023-04-27

Description Allows to estimate and test high-dimensional mediation effects based on advanced mediator screening and penalized regression techniques. Methods used in the package refer to Zhang H, Zheng Y, Zhang Z, Gao T, Joyce B, Yoon G, Zhang W, Schwartz J, Just A, Colicino E, Vokonas P, Zhao L, Lv J, Baccarelli A, Hou L, Liu L. Estimating and Testing High-dimensional Mediation Effects in Epigenetic Studies. Bioinformatics. (2016) <doi:10.1093/bioinformatics/btw351>. PMID: 27357171.

License GPL-3

Depends R (>= 3.4.0), ncvreg, glmnet

Imports utils, stats, MASS, survival, HDMT, hommel, iterators, parallel, foreach, doParallel

Collate utils.R hima.R survHIMA.R microHIMA.R hima2.R example1.R example2.R example3.R example4.R onAttach.R HIMA-package.R

Encoding UTF-8

LazyData true

URL https://github.com/YinanZheng/HIMA/

BugReports https://github.com/YinanZheng/HIMA/issues/

RoxygenNote 7.2.1

NeedsCompilation no

Author Yinan Zheng [aut, cre],
Haixiang Zhang [aut],
Lifang Hou [aut],
Lei Liu [aut, cph]

Maintainer Yinan Zheng <y-zheng@northwestern.edu>

Repository CRAN

Date/Publication 2023-04-27 19:40:06 UTC

R topics documented:

`HIMA-package` *High-Dimensional Mediation Analysis for 'Omic' Data*

Description

HIMA is an R package for estimating and testing high-dimensional mediation effects in omic studies. HIMA can perform high-dimensional mediation analysis on a wide range of omic data types as potential mediators, including epigenetics, transcriptomics, proteomics, and metabolomics using function `hima` and microbiome data (function `microHIMA`). HIMA can also handle survival data (function `survHIMA`).

Package:	HIMA
Type:	Package
Version:	2.2.0
Date:	2023-04-27
License:	GPL-3

Details

If package "qvalue" is not found during installation, please first install "qvalue" package # through Bioconductor: https://www.bioconductor.org/packages/release/bioc/html/qvalue.html

Author(s)

Yinan Zheng <y-zheng@northwestern.edu>, Haixiang Zhang <haixiang.zhang@tju.edu.cn>, Lifang Hou <l-hou@northwestern.edu> Lei liu <lei.liu@wustl.edu>

Maintainer: Yinan Zheng <y-zheng@northwestern.edu>

References

1. Zhang H, Zheng Y, Zhang Z, Gao T, Joyce B, Yoon G, Zhang W, Schwartz J, Just A, Colicino E, Vokonas P, Zhao L, Lv J, Baccarelli A, Hou L, Liu L. Estimating and Testing High-dimensional

Mediation Effects in Epigenetic Studies. Bioinformatics. 2016;32(20):3150-4. DOI: 10.1093/bioinformatics/btw351. PubMed PMID: 27357171; PMCID: PMC5048064.

2. Zhang H, Zheng Y, Hou L, Liu L. Mediation Analysis for Survival Data with High-Dimensional Mediators. Bioinformatics. 2021;37(21):3815-21. DOI: 10.1093/bioinformatics/btab564. PubMed PMID: 34343267; PMCID: PMC8570823.

3. Zhang H, Chen J, Feng Y, Wang C, Li H, Liu L. Mediation effect selection in high-dimensional and compositional microbiome data. Stat Med. 2021;40(4):885-96. DOI: 10.1002/sim.8808. PubMed PMID: 33205470; PMCID: PMC7855955.

`Example1` *Example dataset 1 for HIMA: Continuous outcome*

Description

A `list` dataset containing a phenotype dataset and a high-dimension mediator dataset (n=300 participants, p=300 biomarkers). The variables in the phenotype are as follows:

Usage

```
Example1
```

Format

An object of class `list` of length 2.

Details

- Treatment: treated (value = 1) or not treated (value = 0)
- Outcome: outcome of the treatment- a normally distributed continuous variable
- Sex: female (value = 1) or male (value = 0)
- Age: Age of the participant

The datasets are simulated using parameters generated from a real dataset. The code used to generate the data can be found in /inst/script folder of the package.

Value

A list containing two objects: `PhenoData` and `Mediator`

Example2 *Example dataset 2 for HIMA: Binary outcome*

Description

A `list` dataset containing a phenotype dataset and a high-dimension mediator dataset (n=300 participants, p=300 biomarkers). The variables in the phenotype are as follows:

Usage

```
Example2
```

Format

An object of class `list` of length 2.

Details

- Treatment: treated (value = 1) or not treated (value = 0)
- Disease: diseased (value = 1) or healthy (value = 0)
- Sex: female (value = 1) or male (value = 0)
- Age: Age of the participant

The datasets are simulated using parameters generated from a real dataset. The code used to generate the data can be found in /inst/script folder of the package.

Value

A list containing two objects: `PhenoData` and `Mediator`

Example3 *Example dataset 3 for HIMA: Survival data outcome*

Description

A `list` dataset containing a phenotype dataset and a mediator dataset (n=200 participants, p=200 biomarkers). The variables in the phenotype are as follows:

Usage

```
Example3
```

Format

An object of class `list` of length 2.

Details

- Treatment: treated (value = 1) or not treated (value = 0)
- Status: Status indicator: dead (value = 1) or alive (value = 0)
- Time: time to event
- Sex: female (value = 1) or male (value = 0)
- Age: Age of the participant

The datasets are simulated using parameters generated from a real dataset. The code used to generate the data can be found in /inst/script folder of the package.

Value

A list containing two objects: `PhenoData` and `Mediator`

`Example4` *Example dataset 4 for HIMA: Compositional mediator (e.g., microbiome)*

Description

A `list` dataset containing a phenotype dataset and a compositional mediator dataset (n=200 participants, p=50 biomarkers). The variables in the phenotype are as follows:

Usage

```
Example4
```

Format

An object of class `list` of length 2.

Details

- Treatment: treated (value = 1) or not treated (value = 0)
- Outcome: outcome of the treatment- a normally distributed continuous variable
- Sex: female (value = 1) or male (value = 0)
- Age: Age of the participant

The datasets are simulated using parameters generated from a real dataset. The code used to generate the data can be found in /inst/script folder of the package.

Value

A list containing two objects: `PhenoData` and `Mediator`

hima *High-dimensional Mediation Analysis*

Description

hima is used to estimate and test high-dimensional mediation effects.

Usage

```
hima(
  X,
  Y,
  M,
  COV.XM = NULL,
  COV.MY = COV.XM,
  Y.family = c("gaussian", "binomial"),
  M.family = c("gaussian", "negbin"),
  penalty = c("MCP", "SCAD", "lasso"),
  topN = NULL,
  parallel = FALSE,
  ncore = 1,
  scale = TRUE,
  verbose = FALSE,
  ...
)
```

Arguments

X	a vector of exposure. Do not use data.frame or matrix.
Y	a vector of outcome. Can be either continuous or binary (0-1). Do not use data.frame or matrix.
M	a data.frame or matrix of high-dimensional mediators. Rows represent samples, columns represent variables.
COV.XM	a data.frame or matrix of covariates dataset for testing the association M ~ X. Covariates specified here will not participate penalization. Default = NULL. If the covariates contain mixed data types, please make sure all categorical variables are properly formatted as factor type.
COV.MY	a data.frame or matrix of covariates dataset for testing the association Y ~ M. Covariates specified here will not participate penalization. If not specified, the same set of covariates for M ~ X will be applied. Using different sets of covariates is allowed but this needs to be handled carefully.
Y.family	either 'gaussian' (default) or 'binomial', depending on the data type of outcome (Y). See ncvreg
M.family	either 'gaussian' (default) or 'negbin' (i.e., negative binomial), depending on the data type of mediator (M).

penalty	the penalty to be applied to the model. Either 'MCP' (the default), 'SCAD', or 'lasso'. See ncvreg.
topN	an integer specifying the number of top markers from sure independent screening. Default = NULL. If NULL, topN will be either ceiling(n/log(n)) if Y.family = 'gaussian', or ceiling(n/(2*log(n))) if Y.family = 'binomial', where n is the sample size. If the sample size is greater than topN (pre-specified or calculated), all mediators will be included in the test (i.e. low-dimensional scenario).
parallel	logical. Enable parallel computing feature? Default = FALSE.
ncore	number of cores to run parallel computing Valid when parallel == TRUE. By default max number of cores available in the machine will be utilized.
scale	logical. Should the function scale the data? Default = TRUE.
verbose	logical. Should the function be verbose? Default = FALSE.
...	other arguments passed to ncvreg.

Value

A data.frame containing mediation testing results of selected mediators.

- alpha: coefficient estimates of exposure (X) –> mediators (M).
- beta: coefficient estimates of mediators (M) –> outcome (Y) (adjusted for exposure).
- gamma: coefficient estimates of exposure (X) –> outcome (Y) (total effect).
- alpha*beta: mediation effect.
- % total effect: alpha*beta / gamma. Percentage of the mediation effect out of the total effect.
- Bonferroni.p: statistical significance of the mediator (Bonferroni procedure).
- BH.FDR: statistical significance of the mediator (Benjamini-Hochberg procedure).

References

Zhang H, Zheng Y, Zhang Z, Gao T, Joyce B, Yoon G, Zhang W, Schwartz J, Just A, Colicino E, Vokonas P, Zhao L, Lv J, Baccarelli A, Hou L, Liu L. Estimating and Testing High-dimensional Mediation Effects in Epigenetic Studies. Bioinformatics. 2016. DOI: 10.1093/bioinformatics/btw351. PMID: 27357171. PMCID: PMC5048064

Examples

```
## Not run:
# When Y is continuous and normally distributed
# Example 1 (continuous outcome):
data(Example1)
head(Example1$PhenoData)

hima.fit <- hima(X = Example1$PhenoData$Treatment,
                 Y = Example1$PhenoData$Outcome,
                 M = Example1$Mediator,
                 COV.XM = Example1$PhenoData[, c("Sex", "Age")],
                 scale = FALSE,
```

```
                verbose = TRUE)
hima.fit

# When Y is binary (should specify Y.family)
# Example 2 (binary outcome):
data(Example2)
head(Example2$PhenoData)

hima.logistic.fit <- hima(X = Example2$PhenoData$Treatment,
                          Y = Example2$PhenoData$Disease,
                          M = Example2$Mediator,
                          COV.XM = Example2$PhenoData[, c("Sex", "Age")],
                          Y.family = 'binomial',
                          scale = FALSE,
                          verbose = TRUE)
hima.logistic.fit

## End(Not run)
```

`hima2` *Advanced High-dimensional Mediation Analysis*

Description

`hima2` is an upgraded version of hima for estimating and testing high-dimensional mediation effects.

Usage

```
hima2(
  formula,
  data.pheno,
  data.M,
  outcome.family = c("gaussian", "binomial", "survival"),
  mediator.family = c("gaussian", "negbin", "compositional"),
  penalty = c("DBlasso", "MCP", "SCAD", "lasso"),
  topN = NULL,
  scale = TRUE,
  verbose = FALSE
)
```

Arguments

`formula`	an object of class `formula`: a symbolic description of the overall effect model, i.e., `outcome ~ exposure + covariates`, to be fitted. Make sure the "exposure" is the variable of interest, which must be listed as the first variable in the right hand side of the formula. independent variable in the formula. The same covariates will be used in screening and penalized regression.

data.pheno	a data frame containing all the variables listed in the right hand side of the formula. hima2 will scale data.pheno.
data.M	a data.frame or matrix of high-dimensional mediators. Rows represent samples, columns represent variables. hima2 will scale data.M.
outcome.family	either 'gaussian' (default, for normally distributed continuous outcome), 'binomial' (for binay outcome), or 'survival' (for time-to-event outcome), depending on the data type of outcome.
mediator.family	either 'gaussian' (default, for continuous mediators), 'negbin' (i.e., negative binomial, for RNA-seq data as mediators), or 'compositional' (for microbiome data as mediators), depending on the data type of high-dimensional mediators (data.M).
penalty	the penalty to be applied to the model. Either 'DBlasso' (De-biased LASSO, default), 'MCP', 'SCAD', or 'lasso'.
topN	an integer specifying the number of top markers from sure independent screening. Default = NULL. If NULL, topN will be either ceiling(n/log(n)) if outcome.family = 'gaussian', or ceiling(n/(2*log(n))) if outcome.family = 'binomial', where n is the sample size. If the sample size is greater than topN (pre-specified or calculated), all mediators will be included in the test (i.e. low-dimensional scenario).
scale	logical. Should the function scale the data? Default = TRUE.
verbose	logical. Should the function be verbose and shows the progression? Default = FALSE.

Value

A data.frame containing mediation testing results of selected mediators.

References

Zhang H, Zheng Y, Zhang Z, Gao T, Joyce B, Yoon G, Zhang W, Schwartz J, Just A, Colicino E, Vokonas P, Zhao L, Lv J, Baccarelli A, Hou L, Liu L. Estimating and Testing High-dimensional Mediation Effects in Epigenetic Studies. Bioinformatics. 2016. DOI: 10.1093/bioinformatics/btw351. PMID: 27357171. PMCID: PMC5048064

Perera C, Zhang H, Zheng Y, Hou L, Qu A, Zheng C, Xie K, Liu L. HIMA2: high-dimensional mediation analysis and its application in epigenome-wide DNA methylation data. BMC Bioinformatics. 2022. DOI: 10.1186/s12859-022-04748-1. PMID: 35879655. PMCID: PMC9310002

Zhang H, Zheng Y, Hou L, Zheng C, Liu L. Mediation Analysis for Survival Data with High-Dimensional Mediators. Bioinformatics. 2021. DOI: 10.1093/bioinformatics/btab564. PMID: 34343267. PMCID: PMC8570823

Zhang H, Chen J, Feng Y, Wang C, Li H, Liu L. Mediation effect selection in high-dimensional and compositional microbiome data. Stat Med. 2021. DOI: 10.1002/sim.8808. PMID: 33205470; PMCID: PMC7855955.

Zhang H, Chen J, Li Z, Liu L. Testing for mediation effect with application to human microbiome data. Stat Biosci. 2021. DOI: 10.1007/s12561-019-09253-3. PMID: 34093887; PMCID: PMC8177450.

Examples

```
## Not run:
# Example 1 (continous outcome):
data(Example1)
head(Example1$PhenoData)

e1 <- hima2(Outcome ~ Treatment + Sex + Age,
      data.pheno = Example1$PhenoData,
      data.M = Example1$Mediator,
      outcome.family = "gaussian",
      mediator.family = "gaussian",
      penalty = "MCP",
      scale = FALSE)
e1
attributes(e1)$variable.labels

# Example 2 (binary outcome):
data(Example2)
head(Example2$PhenoData)

e2 <- hima2(Disease ~ Treatment + Sex + Age,
      data.pheno = Example2$PhenoData,
      data.M = Example2$Mediator,
      outcome.family = "binomial",
      mediator.family = "gaussian",
      penalty = "MCP",
      scale = FALSE)
e2
attributes(e2)$variable.labels

# Example 3 (time-to-event outcome):
data(Example3)
head(Example3$PhenoData)

e3 <- hima2(Surv(Status, Time) ~ Treatment + Sex + Age,
      data.pheno = Example3$PhenoData,
      data.M = Example3$Mediator,
      outcome.family = "survival",
      mediator.family = "gaussian",
      penalty = "DBlasso",
      scale = FALSE)
e3
attributes(e3)$variable.labels

# Example 4 (compositional data as mediator, e.g., microbiome):
data(Example4)
head(Example4$PhenoData)

e4 <- hima2(Outcome ~ Treatment + Sex + Age,
      data.pheno = Example4$PhenoData,
      data.M = Example4$Mediator,
      outcome.family = "gaussian",
```

```
      mediator.family = "compositional",
      penalty = "DBlasso",
      scale = FALSE)
e4
attributes(e4)$variable.labels

## End(Not run)
```

`microHIMA` *High-dimensional mediation analysis for compositional microbiome data*

Description

`microHIMA` is used to estimate and test high-dimensional mediation effects for compositional microbiome data.

Usage

```
microHIMA(X, Y, OTU, COV = NULL, FDPcut = 0.05, scale = TRUE)
```

Arguments

`X`	a vector of exposure.
`Y`	a vector of outcome.
`OTU`	a `data.frame` or `matrix` of high-dimensional compositional OTUs (mediators). Rows represent samples, columns represent variables.
`COV`	a `data.frame` or `matrix` of adjusting covariates. Rows represent samples, columns represent microbiome variables. Can be `NULL`.
`FDPcut`	FDP (false discovery proportions) cutoff applied to define and select significant mediators. Default = `0.05`.
`scale`	logical. Should the function scale the data? Default = `TRUE`.

Value

A data.frame containing mediation testing results of selected mediators (FDP < `FDPcut`).

- ID: index of selected significant mediator.
- alpha: coefficient estimates of exposure (X) –> mediators (M).
- alpha_se: standard error for alpha.
- beta: coefficient estimates of mediators (M) –> outcome (Y) (adjusted for exposure).
- beta_se: standard error for beta
- p_FDP: false discovery proportions of selected significant mediator.

References

Zhang H, Chen J, Feng Y, Wang C, Li H, Liu L. Mediation effect selection in high-dimensional and compositional microbiome data. Stat Med. 2021. DOI: 10.1002/sim.8808. PMID: 33205470; PMCID: PMC7855955.

Zhang H, Chen J, Li Z, Liu L. Testing for mediation effect with application to human microbiome data. Stat Biosci. 2021. DOI: 10.1007/s12561-019-09253-3. PMID: 34093887; PMCID: PMC8177450.

Examples

```
## Not run:
data(Example4)
head(Example4$PhenoData)

microHIMA.fit <- microHIMA(X = Example4$PhenoData$Treatment,
                           Y = Example4$PhenoData$Outcome,
                           OTU = Example4$Mediator,
                           COV = Example4$PhenoData[, c("Sex", "Age")],
                           scale = FALSE)
microHIMA.fit

## End(Not run)
```

---

survHIMA *High-dimensional mediation analysis for survival data*

Description

survHIMA is used to estimate and test high-dimensional mediation effects for survival data.

Usage

```
survHIMA(X, Z, M, OT, status, FDRcut = 0.05, scale = TRUE, verbose = FALSE)
```

Arguments

X	a vector of exposure.
Z	a matrix of adjusting covariates. Rows represent samples, columns represent variables. Can be NULL.
M	a data.frame or matrix of high-dimensional mediators. Rows represent samples, columns represent mediator variables.
OT	a vector of observed failure times.
status	a vector of censoring indicator (status = 1: uncensored; status = 0: censored)
FDRcut	FDR cutoff applied to define and select significant mediators. Default = 0.05.
scale	logical. Should the function scale the data? Default = TRUE.
verbose	logical. Should the function be verbose? Default = FALSE.

Value

A data.frame containing mediation testing results of selected mediators (FDR <FDPcut).

- ID: index of selected significant mediator.
- alpha: coefficient estimates of exposure (X) –> mediators (M).
- alpha_se: standard error for alpha.
- beta: coefficient estimates of mediators (M) –> outcome (Y) (adjusted for exposure).
- beta_se: standard error for beta
- p.joint: joint p-value of selected significant mediator.

References

Zhang H, Zheng Y, Hou L, Zheng C, Liu L. Mediation Analysis for Survival Data with High-Dimensional Mediators. Bioinformatics. 2021. DOI: 10.1093/bioinformatics/btab564. PMID: 34343267. PMCID: PMC8570823

Examples

```
## Not run:
data(Example3)
head(Example3$PhenoData)

survHIMA.fit <- survHIMA(X = Example3$PhenoData$Treatment,
                Z = Example3$PhenoData[, c("Sex", "Age")],
                M = Example3$Mediator,
                OT = Example3$PhenoData$Time,
                status = Example3$PhenoData$Status,
                FDRcut = 0.05,
                scale = FALSE,
                verbose = TRUE)
survHIMA.fit

## End(Not run)
```

Index