Package 'mbmdr'

August 29, 2013

Type Package

Title Model Based Multifactor Dimensionality Reduction

Version 2.6

Date 2012-07-26

Depends R (>= 2.10), logistf

Author Victor Urrea, Malu Calle, Kristel Van Steen, Nuria Malats

Maintainer Victor Urrea <victor.urrea@uvic.cat>

Description Model Based Multifactor Dimension Reduction proposed by
Calle et al. (2008) as a dimension reduction method for exploring gene-gene interactions.

Acknowledgement This work was supported by the Ministerio de Educacion
y Ciencia (Spain) (Grant: MTM2008-06747-C02-02)

License GPL (>= 2)

LazyLoad yes

Repository CRAN

Date/Publication 2012-07-26 15:09:20

NeedsCompilation no

R topics documented:

mbmdr *Model Based Multifactor Dimensionality Reduction*

Description

mbmdr implements the Model Based Multifactor Dimensionality Reduction (MB-MDR) method proposed by Calle et al.(2008) as a dimension reduction method for exploring gene-gene interactions.

Usage

```
mbmdr(y, data, order, covar=NULL, exclude=NA, risk.threshold=0.1,
      output=NULL, adjust=c("none","covariates","main effects","both"),
      first.model=NULL, list.models=NULL, use.logistf=TRUE,
      printStep1=FALSE, ...)
```

Arguments

y	Vector containing the dependent variable.
data	A data.frame (or object coercible by as.data.frame to a data frame) containing the SNP information with values 0,1,2. For example: 0 = common homozygous genotype, 1 = heterozygous genotype, 2 = variant homozygous genotype.
order	Single integer that specifies the order of interactions to be analyzed. If list.models = NULL (value by default) all possible interactions of the specified order are analyzed.
covar	*(Optional)* A data.frame or object coercible by as.data.frame to a data frame containing the covariates for adjusting regression models. Only used if adjust="covariates" or adjust="both".
exclude	*(Optional)* Value/s of missing data. If missings in data are coded differently than NA it should be specified. For example exclude=c(NA,-1) specifies that both, NA and -1 indicate a missing value.
risk.threshold	Threshold used at the first MB-MDR stage for defining the risk category of a multilocus genotype. It should be a conservative value. The default value is risk.threshold=0.1.
output	*(Optional)* Output file name for storing mbmdr results on file, or NULL (default) for output as R object. If the number of models to be analyzed is too large, it is preferable to store the output in a file. This allows exploring partial results while mbmdr is still running and prevents from loosing all the information if R or mbmdr crashes during the process.
adjust	Type of regressions adjustment. Options are "none", "covariates", "main effects" or "both". By default no adjustment is performed; adjust="none".

`first.model` *(Optional)* Numerical vector of length equal to `order` for specifying the initial interaction model for `mbmdr`; previous models will not be evaluated. This is useful to continue `mbmdr` computation after a stop.
Note that, by default, `mbmdr` explores all possible interactions of a specified order. If there are for example, 50 snps in data and `order=3`, `mbmdr` will start analyzing the model (50,49,48), that means interaction between snps 50, 49 and 48 (column position in data data.frame). The second model that `mbmdr` will analyze is (50,49,47). After model (50,49,1), the next model will be (50,48,47), and the final model will be (3,2,1).
For example, if `mbmdr` stopped after the analysis of model (30,21,14), you can continue the process by specifying `first.model=c(30,21,13)`. Ids of snps must be in descended order. If `first.model=NULL` (by default) all models will be analyzed.

`list.models` *(Optional)* Exhaustive list of models to be analyzed. Only models in list will be analyzed. It can be: a vector of lenght `order` specifying a unique model, a matrix (`n x order`) containing models by rows, or a string for specifying a file with models by rows (all models must be of the same interaction order)
A `NULL` value (by default) indicates that all possible interactions will be analyzed.

`use.logistf` Boolean value indicating whether or not to use the `logistf` package for logistic regressions when separation of data points is observed.
It only has effect if logistic regression (`family=binomial(link = "logit")`) is specified. (See `logistf` help for details). The default value is `TRUE`.

`printStep1` Boolean value. If true, the details of `mbmdr` step 1 are printed for every model. This slows the process, so it is only advisable when the number of models to analyze is small.
By default `printStep1=FALSE`.

`...` For regression arguments: arguments to be passed to `glm` calls.
Mainly to specify the error distribution and link function to be used in the regression models.
For example, use `family=binomial(link=logit)` for specifying logistic regression or `gaussian(link = "identity")` for normal regression.
(See `family` for details of family functions and `glm` for more options of `glm` function).

Details

MB-MDR is a method for identifying multi-locus genotypes that are associated with a phenotype of interest, and allows to adjust for marginal and confounding effects.

The exploration of interacions is performed in three steps:

Step1
Each genotype is tested for association with the response and is classified as high risk, low risk or not significant, and all genotypes of the same class are merged. The threshold for considering significant evidence is the value specified in `risk.threshold` (by default `risk.threshold=0.1`).
If `printStep1=TRUE`, the MBMDR function prints this classification.

Step2
For each risk categories, high and low, a new association test is performed. The result provides a Wald statistic for the high and for the low categories.

Step3
The significance is explored through a permutation test on the maximum Wald statistics.

Value

`mbmdr` returns an object of class `mbmdr` with the following attributes:

`call`	The matched call.
`y`	The outcome used.
`data`	The SNPs data used.
`covar`	The covariate data used.
`result`	Dataframe with those interactions that have at least a significant genotype. For each interaction (rows), the following information is returned:

`SNP1...SNPx`	Names of snps in interaction.
`NH`	Number of significant High risk genotypes in the interaction.
`betaH`	Regresion coeficient in step2 for High risk exposition.
`WH`	Wald statistic for High risk category.
`PH`	P-value of the Wald test for the High risk category.
`NL`	Number of significant Low risk genotypes in the interaction.
`betaL`	Regresion coeficient in step2 for Low risk exposition.
`WL`	Wald statistic for Low risk category.
`PL`	P-value of the Wald test for the Low risk category.
`MIN.P`	Minimun p-value (`min(PH,PL)`) for the interaction model.

If `printStep1` argument is set to `TRUE`, the result of the first step in `mbmdr` is printed for each genotype with the following information:

`...`	Genotype.
`cases`	(only for case/control outcome) Number of cases with the specific genotype.
`controls`	(only for case/control outcome) Number of controls with the specific genotype.
`beta`	Regression coefficient for this genotype.
`p.value`	Wald test p-value for this genotype.
`category`	Predicted risk category for this genotype.

Author(s)

Victor Urrea, Malu Calle, Kristel Van Steen, Nuria Malats

References

Calle M.L., Urrea V., Vellalta G., Malats N., Steen K.V. (2008) *Improving strategies for detecting genetic patterns of disease susceptibility in association studies.* Statistics in Medicine 27, 6532-6546.

Examples

```
#---  Dicotomous outcome  -------------

# load example data
data(simSNP)

# MB-MDR analysis of all possible 2nd order interactions (it may take some time)
# The order of the interactions to be explored is specified by order=2
# fit <- mbmdr(y=simSNP$Y,data=simSNP[,3:12],order=2,family=binomial(link=logit))
# print(fit)

# MB-MDR analysis of the epistatic effect of SNP1 and SNP2 (Model 2 1)
# The specific model to be analyzed is specified by list.models=c(2,1)
fit <- mbmdr(y=simSNP$Y,data=simSNP[,3:12],order=2,list.models=c(2,1),
             family=binomial(link=logit),printStep1=TRUE)
print(fit)


# MB-MDR analysis of the epistatic effect of SNP1 and SNP2, adjusted for variable X
# The specific model to be analyzed is specified by list.models=c(2,1)
# The adjustment statement is specified by adjust="covariates"
fit <- mbmdr(y=simSNP$Y,data=simSNP[,3:12],order=2,list.models=c(2,1),
             covar=simSNP$X,adjust="covariates",family=binomial(link=logit))
print(fit)


# MB-MDR analysis of all 2nd order interactions restricted to a subset of snps
# (all possible 2nd order interactions among SNP1, SNP2 and SNP3 are explored,
# it may take some time)
# SNP1, SNP2 and SNP3 are placed in columns 3 to 5 of simSNP. This is specified
# by data=simSNP[,3:5]
# fit <- mbmdr(y=simSNP$Y,data=simSNP[,3:5],order=2,family=binomial(link=logit),
#             printStep1=TRUE)
# print(fit)


# MB-MDR analysis of all possible 3rd order interactions (it may take some time)
# The order of the interactions to be explored is specified by order=3
# fit <- mbmdr(y=simSNP$Y,data=simSNP[,3:12],order=3,family=binomial(link=logit))
# print(fit)


# MB-MDR analysis of the 3rd order epistatic effect of SNP1, SNP2 and SNP3
# The specific model to be analyzed is specified by list.models=c(3,2,1)
# fit <- mbmdr(y=simSNP$Y,data=simSNP[,3:12],order=3,list.models=c(3,2,1),
```

```
#              family=binomial(link=logit),printStep1=TRUE)
# print(fit)


#---  Continuous outcome  --------------
# load example data
data(simSNPcont)

# MB-MDR analysis of all possible 2nd order interactions (it may take some time)
# The order of the interactions to be explored is specified by order=2
# fit <- mbmdr(y=simSNPcont$Y,data=simSNPcont[,2:11],order=2)
# print(fit)

# MB-MDR analysis of the epistatic effect of SNP1 and SNP2 (Model 2 1)
# The specific model to be analyzed is specified by list.models=c(2,1)
fit <- mbmdr(y=simSNPcont$Y,data=simSNPcont[,2:11],order=2,
           list.models=c(2,1),printStep1=TRUE)
print(fit)
```

`mbmdr.PermTest` *Permutation Test for exploring significance of MB-MDR result.*

Description

Performs a permutation test for specified interaction models from `mbmdr` object.

Usage

```
mbmdr.PermTest(x, n, model = NULL, sig.level=1)
```

Arguments

`x`	An mbmdr object returned by `mbmdr` function.
`n`	Number of permutations.
`model`	Vector specifiying an interaction model or matrix with rows referring to interaction models to be submitted for permutation testing. If `model=NULL` (default), permutation testing is performed on the best model derived from the `x` object.
`sig.level`	Significance level for the confidence intervals of the permutation based p-values, using a normal approximation (Nettleton, 2000). If `sig.level=1` (as default), confidence intervals are not computed. When the permutation p-value is too small with respect to `n`, the normal aproximation is not appropiated and the C.I. are not provided (Nettleton, 2000).

Details

A permutation testing is performed for each specified model by permuting the outcome variable and calling the mbmdr function. The call to the mbmdr function is made by recovering the call from mbmdr object and replacing the outcome (by a permuted outcome vector) and SNP data (by the subset of the specified model). All other arguments in the initial call to mbmdr are transfered to mbmdr.PermTest.

Value

An object is returned of a new class, mbmdr.PermTest, with following attributes:

n	The number of permutations.
mbmdr	The mbmdr object used.
PermTest	A data.frame with results of the permutation tests. The following information is returned:

SNP1...SNPx	Names of snps for each specified interaction models.
NH	Number of significant High risk genotypes in the interaction.
betaH	Regresion coeficient in step2 for High risk exposition.
WH	Wald statistic for High risk category.
NL	Number of significant Low risk genotypes in the interaction.
betaL	Regresion coeficient in step2 for Low risk exposition.
WL	Wald statistic for Low risk category.
Wmax	Maximun Wald statistic for the interaction model.
Perm.P	Permutation p-value for the interaction model.
IC.lower	Lower limit of the confidence interval for permutation p-value.
IC.upper	Upper limit of the confidence interval for permutation p-value.

References

Nettleton D., Doerge R.W. (2000) *Accounting for Variability in the Use of Permutation Testing to Detect Quantitative Trait Loci.* Biometrics, Vol. 56, No. 1, pp. 52-58.

See Also

mbmdr

Examples

```
data(simSNP)
fit <- mbmdr(y=simSNP$Y,data=simSNP[,3:12],order=2,list.models=c(2,1),
           family=binomial(link=logit))

# Single model permutation test
mbmdr.PermTest(fit,100)

# Next steps takes some time
```

```
# Permutation test for all models with MIN.P <= 0.05
# order <- 2
# models <- subset(fit$result, MIN.P <= 0.05, select = 1:order)
# mbmdr.PermTest(fit,100,models)

# Permutation test and confidence interval for all models with MIN.P <= 0.05
# mbmdr.PermTest(fit,100,models,sig.level=0.05)
```

`simSNP` *Simulated data of SNPs*

Description

simSNP is a data.frame containing a bivariate outcome as in a case-control study, a continous covariate and 10 simulated SNPs. SNPs 1 and 2 have an interaction effect on the outcome. Penetrances are specified as in model 1 of Ritchie et al.(2003)

simSNPcont is a data.frame containing a continuous normally distributed outcome and 10 simulated SNPs, where SNPs 1 and 2 have an interaction effect on the outcome.

Usage

```
data(simSNP)
data(simSNPcont)
```

Format

simSNP is a data frame with 400 observations on the following 12 variables.

`Y`	Outcome: 1=case, 0=control
`X`	Age
`snp.1`	SNP genotype data 1
`snp.2`	SNP genotype data 2
...	.....
`snp.10`	SNP genotype data 10

SNP values are coded as: 0=common homozygous, 1=heterozygous, 2=variant homozygous.

simSNPcont has a continous outcome and the same SNPs.

References

Ritchie M.D., Hahn L.W., Moore J.H. (2003) *Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and*

genetic heterogeneity. Genetic Epidemiology 24, 150-157

`summary.mbmdr` *Summary Method for Model Based Multifactor Dimensionality Reduction*

Description

A summary method for mbmdr.

Usage

```
 ## S3 method for class 'mbmdr'
summary(object,sig.level=0.05,...)
```

Arguments

`object`	An mbmdr object returned from `mbmdr` function.
`sig.level`	Threshold value for significant p-values.
`...`	Further arguments passed to or from other methods.

Value

It returns a data.frame with those models with a MIN.P value smaller than `sig.level`.

See Also

`mbmdr`

Index