Package 'GRSxE'

January 20, 2025

Type Package

Title Testing Gene-Environment Interactions Through Genetic Risk
Scores

Version 1.0.1

Description Statistical testing procedures for detecting
GxE (gene-environment) interactions. The main focus lies on
GRSxE interaction tests that aim at detecting GxE interactions
through GRS (genetic risk scores). Moreover, a novel testing
procedure based on bagging and OOB (out-of-bag) predictions is
implemented for incorporating all available observations at
both GRS construction and GxE testing (Lau et al., 2023,
<doi:10.1038/s41598-023-28172-4>).

License MIT + file LICENSE

Encoding UTF-8

Imports glmnet, ranger, stats, utils

RoxygenNote 7.1.2

NeedsCompilation no

Author Michael Lau [aut, cre] (<https://orcid.org/0000-0002-5327-8351>)

Maintainer Michael Lau <michael.lau@hhu.de>

Repository CRAN

Date/Publication 2023-10-30 14:00:05 UTC

Contents

Description

Fitting and evaluating GRS (genetic risk scores) for testing the presence of GxE (gene-environment) interactions.

Usage

```
GRSxE(
  X,
  y,
  E,
  C = NULL,
  test.type = "bagging",
  B = 500,
  replace = TRUE,
  subsample = ifelse(replace, 1, 0.632),
  test.ind = sample(nrow(X), floor(nrow(X)/2)),
  grs.type = "rf",
  grs.args = list()
)
```

Arguments

X	Matrix or data frame of genetic variables such as SNPs usually coded as 0-1-2.
y	Numeric vector of the outcome/phenotype. Binary outcomes such as a disease status should be coded as 0-1 (control-case).
E	Numeric vector of the environmental exposure.
C	Optional data frame containing potentially confounding variables to be adjusted for.
test.type	Testing type. The standard setting is "bagging", which employs its OOB (out-of-bag) prediction mechanism such that the full data can be used for both training the GRS and testing the GxE interaction. Alternatively, this can be set to "holdout", which requires splitting the available data into a training data set and test data set. For that, test.ind needs to be set to the data indices used for testing.
B	The number of bagging iterations if test.type = "bagging" is used. Also used as the number of trees grown in the random forest if grs.type = "rf" is set.
replace	Should sampling with or without replacement be performed? Only used if test.type = "bagging" is set.
subsample	Subsample fraction if test.type = "bagging" is used.

test.ind	Vector of indices in the supplied data for testing the GxE interaction. Only used if test.type = "holdout" is set. The standard setting corresponds to a random 50:50 training-test split.
grs.type	Type of GRS to be constructed. Either "rf" for a random forest or "elnet" for an elastic net.
grs.args	Optional list of arguments passed to the GRS fitting procedure.

Details

The GRS is usually constructed through random forests for taking gene-gene interactions into account and using its OOB (out-of-bag) prediction mechanism. Alternatively, a classical GRS construction approach can be employed by fitting an elastic net. Bagging can also be applied to fit multiple elastic net models to also be able to perform OOB predictions.

The advantage of OOB predictions is that they allow the GRS model to be constructed on the full available data, while performing unbiased predictions also on the full available data. Thus, both the GRS construction and the GxE interaction testing can utilize all observations.

If desired, sampling can be performed without replacement in contrast to the classical bagging approach that utilizes bootstrap sampling.

Potentially confounding variables can also be supplied that will then be adjusted for in the GxE interaction testing.

This function uses a GLM (generalized linear model) for modelling the marginal genetic effect, marginal environmental effect, the GRSxE interaction effect, and potential confounding effects. The fitted GLM is returned, which can be, e.g., inspected via summary(...) to retrieve the Wald test p-values for the individual terms. The p-value corresponding to the G:E term is the p-value for testing the presence of a GRSxE interaction.

Value

An object of class glm is returned, in which G:E describes the GRSxE term.

References

- Lau, M., Kress, S., Schikowski, T. & Schwender, H. (2023). Efficient gene–environment interaction testing through bootstrap aggregating. Scientific Reports 13:937. doi: 10.1038/s41598023281724

- Lau, M., Wigmann C., Kress S., Schikowski, T. & Schwender, H. (2022). Evaluation of tree-based statistical learning methods for constructing genetic risk scores. BMC Bioinformatics 23:97. doi: 10.1186/s1285902204634w

- Breiman, L. (1996). Bagging predictors. Machine Learning 24:123–140. doi: 10.1007/BF00058655

- Breiman, L. (2001). Random Forests. Machine Learning 45:5–32. doi: 10.1023/A:1010933404324

- Friedman J., Hastie T. & Tibshirani R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software 33(1):1–22. doi: 10.18637/jss.v033.i01

Examples

```
# Generate toy data
set.seed(101299)
maf <- 0.25
n.snps <- 10
N <- 500
X <- matrix(sample(0:2, n.snps * N, replace = TRUE,
                    prob = c((1-maf)^2, 1-(1-maf)^2-maf^2, maf^2)),
            ncol = n.snps)
colnames(X) <- paste("SNP", 1:n.snps, sep="")
E <- rnorm(N, 20, 10)
E[E < 0] <- 0

# Generate outcome with a GxE interaction
y.GxE <- -0.75 + log(2) * (X[,"SNP1"] != 0) +
  log(4) * E/20 * (X[,"SNP2"] != 0 & X[,"SNP3"] == 0) +
  rnorm(N, 0, 2)
# Test for GxE interaction (Wald test for G:E)
summary(GRSxE(X, y.GxE, E))

# Generate outcome without a GxE interaction
y.no.GxE <- -0.75 + log(2) * (X[,"SNP1"] != 0) +
  log(4) * E/20 + log(4) * (X[,"SNP2"] != 0 & X[,"SNP3"] == 0) +
  rnorm(N, 0, 2)
# Test for GxE interaction (Wald test for G:E)
summary(GRSxE(X, y.no.GxE, E))
```

GxE   *Testing individual gene-environment interactions*

Description

Function for testing univariate GxE interactions, e.g., using single SNPs or a GRS.

Usage

```
GxE(G, y, E, C = NULL)
```

Arguments

G	Numeric vector of a genetic variable such as a GRS (genetic risk score) or a SNP coded as 0-1-2.
y	Numeric vector of the outcome/phenotype. Binary outcomes such as a disease status should be coded as 0-1 (control-case).
E	Numeric vector of the environmental exposure.
C	Optional data frame containing potentially confounding variables to be adjusted for.

Details

This function uses a GLM (generalized linear model) for modelling the marginal genetic effect, marginal environmental effect, the GxE interaction effect, and potential confounding effects. The fitted GLM is returned, which can be, e.g., inspected via summary(...) to retrieve the Wald test p-values for the individual terms. The p-value corresponding to the G:E term is the p-value for testing the presence of a GxE interaction.

Value

An object of class glm is returned, in which G:E describes the GxE term.

References

- Lau, M., Kress, S., Schikowski, T. & Schwender, H. (2023). Efficient gene–environment interaction testing through bootstrap aggregating. Scientific Reports 13:937. doi: 10.1038/s41598023281724

Index