Package 'MMGS'

December 6, 2023

Type Package

Title Reaction Norm and Polygenic Environment Interaction Models for Genomic Selection

Version 1.0.1

Date 2023-12-06

Depends R (>= 4.0.0), Matrix (>= 1.0-6)

Imports rrBLUP, BGLR, glmnet, randomForest, LigntGBM, e1071, stringi

Depends 未找到目录项。

URL <https://github.com/Ryougi-yukiro/MMGS>

BugReports <https://github.com/Ryougi-yukiro/MMGS/discussions>

RoxygenNote 7.1.1

NeedsCompilation no

Author Mingjia Zhu <z980907mj@gmail.com>

Yanjun Zan <>

Maintainer

License GPL (>= 3)

Encoding UTF-8

LazyData true

SystemRequirements

Repository

Date/Publication

Description: Package for Multi-envs Genomic Selection, contains Inbred Lines, CUBIC Lines and Light-cross Lines. For more details of the Norm Reaction Model, you can see the article: <https://doi.org/10.1016/j.molp.2021.03.010>; and for the Polygenic Environment Interaction Model, please see the articles: <https://doi.org/10.1016/j.molp.2022.02.012> and <https://doi.org/10.1016/j.xplc.2022.100473>. Each of these two GS models explains the mechanisms of environmental interactions from a different perspective, so please read them in

detail depending on the type of model you are using. The R package 'MMGS' was developed by Mingjia Zhu <z980907mj@gmail.com> and Yanjun Zan <>. This repository is forked from the original repository <https://github.com/Ryougi-yukiro/MMGS>. If you would like to install the package from GitHub, you can follow this URL.

R topics documented:

Index

MMGS-package *Reaction Norm and Polygenic Environment Interaction Models for Genomic Selection*

Description

This package fits the Reaction Norm and Polygenic Environment Interaction Model paths for Genomic Selection. A variety of predictions can be made from the fitted models. Besides, The algorithm is very user-friendly and provides built-in datasets to help users.

Details

Package: MMGS
Type: Package
Version: 1.0
Date: 2023-12-06

Very simple to use. Accept *trait, pheno* and *env* data for all models, and produces the regularization path as you like, only 5 core functions you need to use:

env_trait_calculate
LbyE_calculate
Exhaustive_search
envMeanPara
MMGP

Author(s)

References

Li X, Guo T, Wang J, et al. An integrated framework reinstating the environmental dimension for GWAS and genomic selection in crops[J]. Molecular Plant, 2021, 14(6): 874-887. *doi.org/10.1016/j.molp.2021.03.010*

Li M, Zhang Y W, Zhang Z C, et al. A compressed variance component mixed model for detecting QTNs and QTN-by-environment and QTN-by-QTN interactions in genome-wide association studies[J]. Molecular Plant, 2022, 15(4): 630-650. *doi.org/10.1016/j.molp.2022.02.012*

Jin M, Liu H, Liu X, et al. Complex genetic architecture underlying the plasticity of maize agronomic traits[J]. Plant Communications, 2023, 4(3). *doi.org/10.1016/j.xplc.2022.100473*

Examples

```
#Load Data
data(trait)
data(geno)
data(env_info)
data("PTT_PTR")
#Data analysis
env_trait<-env_trait_calculate(data=trait,trait="FTgdd",env="env_code")
LbyE<-LbyE_calculate(data=trait,trait="FTgdd",env="env_code",line="line_code")
#corr plot
LbyE_corrplot(LbyE=LbyE,cor_type="heatmap")
Paras <- c('DL', 'GDD', 'PTT', 'PTR', 'PTS')
envMeanPara<-envMeanPara(data=env_trait, env_paras=PTT_PTR, maxR_dap1=18,
            maxR_dap2=43, Paras=Paras)
#Result output
out<-MMGP(pheno=pheno, geno=geno, env=env_info,
      para=envMeanPara, Para_Name="PTT", depend="norm",
      model="rrBLUP", fold=5, reshuffle=5 methods="RM.G")
```

LbyE_caculate *Reorganization of phenotypic data*

Description

Given an adaptation input from origin data. If your data is already collated, you can skip this step

Usage

```
LbyE_calculate(
      object,
      trait=NULL,
      env= NULL
      line= NULL )
```

Arguments

Data	Data frame of phenotypes for the collected population; The first

Trait

Env

Line

Examples

```
data(trait)
LbyE<-LbyE_calculate(data=trait,trait="FTgdd",env="env_code",line="line_code")
```

Index

MMGP *Genomic Prediction with Cross Validation*

Description

The MMGP function carries out cross-validation using genotypic phenotypic data from a reference population across different environments, with options for genotypic matrix processing and genomic breeding value estimation.

Usage

Arguments

Geno Matrix (n x m) of genotypes for the training population: n lines with m markers.Genotypes should be coded -1, 0, 1. Missing data are not allowed, please imputed before input**.**

Pheno Vector (n x 1) of "phenotypes", i.e. observations or pre-processed, corrected values. This vector suggested have no missing values, otherwise missing values (NA) will be omitted in both pheno and geno. In a first step, MMGP checks whether rownames(geno) match with names(pheno). If not the case, the common elements (intersect) are selected in both geno and pheno for further analyses. However, if you have missing value in pheno data, it would continued, except some methods.

Env

Para

Para_Name

Depends

Models The options for genomic breeding value prediction methods. The available options are:

- rrBLUP:
- GBLUP: performs G-BLUP using a marker-based relationship matrix, implemented through BGLR R-library. Equivalent to ridge regression (RR-BLUP) of marker effects.
- RR: ridge regression, using package glmnet. In theory, strictly equivalent to gblup.

- LASSO: Least Absolute Shrinkage and Selection Operator is another penalized regression methods which yield more shrinked estimates than RR.Run by glmnet library.
- EN: Elastic Net (Zou and Hastie, 2005), which is a weighted combination of RR and LASSO, using glmnet library.

Several Bayesian methods, using the BGLR library:

- BRR: Bayesian ridge regression: same as rr-blup, but bayesian resolution.Induces homogeneous shrinkage of all markers effects towards zero with Gaussian distribution (de los Campos et al, 2013)
- BL: Bayesian LASSO: uses an exponential prior on marker variances priors, leading to double exponential distribution of marker effects (Park & Casella 2008)
- BA: Bayes A uses a scaled-t prior distribution of marker effects. (Meuwissen et al 2001).
- BB: Bayes B, uses a mixture of distribution with a point mass at zero and with a slab of non-zero marker effects with a scaled-t distribution (Habier et al 2011).
- BC: Bayes C same as Bayes B with a slab with Gaussian distribution. A more detailed description of follow method can be found in Perez & de los Campos 2014 (http://genomics.cimmyt.org/BGLR-extdoc.pdf).
- RKHS: reproductive kernel Hilbert space and multiple kernel MRKHS, using BGLR (Gianola and van Kaam 2008). Based on genetic distance and a kernel function to regulate the distribution of marker effects. This methods is claimed to be effective for detecting non additive effects.

Other Options contained Machine Learning as follow:

- SVM: support vector machine, run by e1071 library. For details, see Chang, Chih-Chung and Lin, Chih-Jen: LIBSVM: a library for Support Vector Machines http://www.csie.ntu.edu.tw/~cjlin/libsvm
- RF: Random forest regression, using the randomForest library (Breiman, 2001, Breiman and Cutler 2013). This method uses regression models on tree nodes which are rooted in bootstrapping data. Supposed to be able to capture interactions between markers.
- LightGBM: LightGBM is a gradient boosting framework that uses tree-based learning algorithms. It is designed to be distributed and efficient with the following advantages: faster training speed and higher efficiency, lower memory usage and support of parallel, distributed, and GPU learning. First capture for genomic breeding on this article published on Genomic Biology (Jun Yan ,2021). A more detailed descriptions are in < https://lightgbm.readthedocs.io/en/latest/index.html >, designed by Microsoft.

Fold Number of folds for the cross-validation. Smallest value recommended is Fold=2.

Reshuffle Number of independent replicates for the cross-validation. Smallest value recommended is

Reshuffle= 5.

Methods

ENalpha

SVM_cost

Gamma

GBM_params

GBM_rounds

Value

The class bwgs.cv returns a list containing:

- Summary: Matrix of dimension m x 4. Columns are:
  -Obs: i.e. pheno vector
  -Pre: the nx1 vector of GEBVs
  -Env: the nx1 vector of all environments
  -Color: the nx1 vector of environments colors used for plotting
- bv_table**:** Matrix of dimension m x n. m Columns are gpreSD (Standart deviation of estimated GEBV) among each environment. n Rows are reshuffles.
- R2_table**:** A vector of cross R2, its length equal to reshuffles

Example

```
#Load Data
data(trait)
data(geno)
data(env_info)
data("PTT_PTR")
#Data analysis
env_trait<-env_trait_calculate(data=trait,trait="FTgdd",env="env_code")
LbyE<-LbyE_calculate(data=trait,trait="FTgdd",env="env_code",line="line_code")
#corr plot
LbyE_corrplot(LbyE=LbyE,cor_type="heatmap")
Paras <- c('DL', 'GDD', 'PTT', 'PTR', 'PTS')
envMeanPara<-envMeanPara(data=env_trait, env_paras=PTT_PTR, maxR_dap1=18,
        maxR_dap2=43, Paras=Paras)
#Result output
out<-MMGP(pheno=pheno, geno=geno, env=env_info,
    para=envMeanPara, Para_Name="PTT", depend="norm",
    model="rrBLUP", fold=5, reshuffle=5 methods="RM.G")
```