PAS

An R package for

Genomic Prediction

Ploygenic Analysis System

Authors:

Zhiqiu Hu; Shizhong Xu; Zhiquan Wang; Rongcai Yang

November 20, 2012

Contents

1 PAS *Ploygenic Analysis System*

Description

The PAS package was developed to implement the method and algorithm developed by Zhiqiu Hu, Shizhong Xu, Zhiquan Wang, and Rongcai Yang for genomic value prediction. Although the current version of the package only provided functions for the bin model analysis (Hu et al., 2012), the package will be developed continuously to incorporate new methods of genomic value prediction that will be introduced by the authors in the near future.

Details

Package:	PAS
Type:	Package
Version:	1.1
Data:	2012-10-20
License:	GPL
Depends:	R (2.10 or later); glmnet
URL:	`http://statgen.ualberta.ca/?open=software.html`

Updates

1. A new option (*foldid*) was added into the binmod function to allow users assigning foldid for cross-validations;

2. A new output item (*objoptimalmap.binsnp*) was added bridging the bin map and the SNP map of a binmod object.

Authors

Zhiqiu Hu, Shizhong Xu, Zhiquan Wang, Rongcai Yang
Maintainer: Zhiqiu Hu <zhiqiu.hu@gmail.com>

References

Zhiqiu Hu, Zhiquan Wang, and Shizhong Xu (2012) An infinitesimal model for quantitative trait genomic value prediction. PLoS ONE **7**: e41336.

2 Bin model analysis

The bin model analysis is a novel modeling strategy using high density genotyping data in genomic value prediction. Unlike the other existing strategies, which either requiring variable selection in prior of prediction or directly modeling all markers in a single model, the bin model analysis provided a new way to reduce the dimensions of the genomic data yet without discarding any molecular information.

References

Zhiqiu Hu, Zhiquan Wang, and Shizhong Xu (2012) An infinitesimal model for quantitative trait genomic value prediction. PLoS ONE **7**: e41336.

2.1 beef data *A real data for demonstration*

Description

The data are provide for demonstration purpose only.

Usage

```
data(beef)
```

Value

x genotypic data.

```
int [1:836, 1:300] 0 0 0 -1 -1 -1 -1 0 0 0 ...
```

y phenotypic data.

```
int [1:836, 1] 768 157 508 614 590 777 505 243 509 351 ...
```

map physical map.
'data.frame': 300 obs. of 2 variables:
$chr: num 1 1 1 1 ...
$pos: int 113641 244698 369418 447277 .

Examples

```
#load PAS library
library(PAS)
#load example data
data(beef)
str(x)
str(y)
str(map)
```

2.2 binmod *bin model analysis*

Description

This is the main function for bin model analysis.

Usage

```
binmod(x,y,map,beta0=NA, binsizelist=NA, full.search=FALSE, foldid=NA,...)
```

Arguments

x	input matrix, of dimensions nobs × nvars; each row is a observation vector of an individual and each column is a genotypic indicator vector for a molecular marker.
y	a matrix of response variable (phenotypic observations), of dimensions nobs × 1.
map	A data frame for linkage map or physical map.
beta0	Estimated SNP effects obtained by single marker regression analysis. By default, the glm function in R will be called by the binmod to calculate the estimates of effects.
binsizelist	A list of binsizes to be considered in the analysis. A default list will be generated if the option was ignored or an invalid list has been specified.
full.search	A logic indicator selecting search strategies. If FALSE was assigned, the binmod will complete the running as soon as the optimal binsize was found. Otherwise, analysis will be conducted for all binsizes on the list.

foldid An optional vector of values between 1 and nfold identifying what fold each observation is in. If not supplied, a random vector is generated using nfold=10.

... Other parameters to be pass to glmnet/r or glm/r.

Details

The function invokes binmod analysis for genomic value prediction. The default settings are strongly suggested for new users.

Value

An binmod object, which is a list of:

$ grid :List of 3 *#information of all searched binsizes*

- **$ mselist** :'data.frame': *nbinsizes* of 4 variables *#A list of mean square errors*
  - **$ binsize** :num [...] *#size settings of the bins, eight in bp or cM.*
  - **$ mse** :num [...] *#mean square error*
  - **$ mse_std** :num [...] *#the standard deviation of MSEs*
  - **$ nbin** :num [...] *#number of bins under the binsize setting*
- **$ optbinsize** :num *#optimal binsize*
- **$ optid** :int *#order of the optimal binsize in the grid*

$ optimal :List of 7 *#result obtained under the optimal binsize*

- **$ predict** :'data.frame': of 2 variables
  - **$ y** :num [1:nobs] *#original phenotypic observations*
  - **$ yp_cv** :num [1:nobs] *#predictions by 10-fold cross-validation.*
- **$ beta** :'data.frame': of 6 variables: *#estimated parameters of bins*
  - **$ beta** :num [1:nbins] *#bin effect*
  - **$ SSx** :num [1:nbins] *#sum of square of bin indicator*
  - **$ Se** :num [1:nbins] *#residual error*

$ Sb :num [1:nbins] *#estimating error of bin effect*

$ Wald :num [1:nbins] *#Wald-test statistics*

$ LOD :num [1:nbins] *#LOD-test statistics*

$ xbin :num [1:nobs, 1:nbins] *#indicator matrix of the bins under the optimal binsize*

$ map :'data.frame': of 5 variables: *#bin map*

$ chr :num [1:nbins] *#chromosome id*

$ pos :num [1:nbins] *#bin position*

$ pos_id :num [1:nbins] *#mean of the orders of markers in the bin*

$ start_id :num [1:nbins] *#the order the first maker in a bin*

$ end_id :num [1:nbins] *#the order the last maker in a bin*

$ binsize :num *#optimal binsize*

$ cv :num *#cross-validation results*

$ binsize :num *#binsize*

$ nbin :num *#number of bins under the binsize setting*

$ mse :num *#mean squared error obtained from cross-validation*

$ r :num *#Pearson's correlation coefficient obtained from cross-validation*

$ map.binsnp :'data.frame': *nvar* of 5 variables

$ chr :int [1:nvar] *#chromosome id*

$ pos :int [1:nvar] *#marker position*

$ pos_id :int [1:nvar] *#marker order*

$ snp.effect :num [1:nvar] *#marker effect*

$ snp.weight :num [1:nvar] *#weight calculated from marker effect*

$ bin.id :int [1:nvar] *#bin order*

$ bin.effect :num [1:nvar] *#bin effect estimation*

$ snp :List of 3 *#SNP information*

$ map :'data.frame': nvar of 3 variables *#linkage map or physical map*

$ chr :num [1:nvar] *#chromosome id*

$ pos :int [1:nvar] *#marker position*

$ pos_id :int [1:nvar] *#marker order*

$ effect :'data.frame': obs. of 6 variables: *#single marker analysis result*

$ beta :num [1:nvar] *#SNP effect*

$ SSx :num [1:nvar] *#sum of square of genotypic indicator*

$ Se :num [1:nvar] *#residual variance*

$ Sb :num [1:nvar] *#estimating error of marker effect*

$ Wald :num [1:nvar] *#Wald-test statistics*

$ LOD :num [1:nvar] *#LOD test statistics*

$ mapinfo :'data.frame': nchr of 7 variables *#chromosome information*

$ chr :num [1:nchr] *#chromosome id*

$ start :int [1:nchr] *#the position of the first marker on the chromosome*

$ end :int [1:nchr] *#the position of the last marker on the chromosome*

$ length :int [1:nchr] *#length of the chromosome*

$ nmark :int [1:nchr] *#number of markers on the chromosome*

$ aver :num [1:nchr] *#average interval of the chromosome*

$ min.interval :int [1:nchr] *#the smallest interval size on the chromosome*

$ cvfit :List of 10 *#A cv.glmnet project. See manual of glmnet for details.*

Examples

```
#load PAS library
library(PAS)
#load the demo data
data(beef)
#perform binmod analysis under the default settings.
binmod.result=binmod(x, y, map)
#plot binmod result
plot(binmod.result)
```

```
str(binmod.result)
#Output the predicted phenotypic values that obtained
#by 10-fold cross validation.
predict(binmod.result)
#predict the phenotypic values for new individuals
x1=x[sort(sample(1:NROW(x), 20)), ]
bin.pred.x1=predict(binmod.result, newx=x1)
str(bin.pred.x1)
```

2.3 binmod.plot *plotting binmod result*

Description

Generate figures using an object created by the binmod function.

Usage

```
plot(obj, file=NULL, width=7, height=5, getdata=FALSE,...)
```

Arguments

obj	An object generated by the binnod function.
file	The prefix of the figure files to be saved.
width	width of the figures (inch).
height	height of the figures (inch).
getdata	A logic indicator. The default value is FALSE, which mean not to return the data for plotting.

Examples

```
#load PAS library
library(PAS)
#load the demo data
data(beef)
#conduct bin model analysis and plotting the result.
plot(binmod(x, y, map))
```

2.4 binmod.predict *genomic value prediction using a binmod object*

Description

Extract predicted genomic breeding values from the 10-fold cross-validation result that has been saved in a binmod object, or predict the breeding values for a new sample.

Usage

```
predict(object, newx=NULL, ...)
```

Arguments

object An object generated by the binnod function.

newx The numeric genotype indicator matrix of a new sample, which need to be coded in the same way as the genotypicdata generating the binmod object.

... Further parameters may also be supplied as arguments.

Examples

```
#load PAS library
library(PAS)
#load the demo data
data(beef)
#conduct bin model analysis.
binmod.result=binmod(x, y, map)
#generate a new sample by sampling 20 individuals from the demo data
```

```
x0=x[sample[1:NROW(x), 20],]
#predict the genomic values of the new sample.
predict(binmod.result, newx=x0)
```

2.5 binmod.print *display a binmod object*

Description

Show a terse summarize for a binmod object.

Usage

```
x or print(x)
```

Argument

x	An binmod object.

Examples

```
#load PAS library
library(PAS)
#load the demo data
data(beef)
#conduct bin model analysis.
binmod.result=binmod(x, y, map)
print(binmod.result)
#show the structure of a binmod object
str(obj)
```