Package 'LBLGXE'

July 20, 2015

Type Package

Title Bayesian Lasso for detecting Rare (or Common) Haplotype Association and their interactions with Environmental Covariates

Version 1.2

Date 2015-07-09

Author Shuang Xia, Yuan Zhang, Swati Biswas, and Shili Lin

Maintainer Yuan Zhang <yxz112020@utdallas.edu>

Description This function takes a dataset of haplotypes and environmental covariates in which rows for indi- viduals of uncertain phase have been augmented by ``pseudo-individuals'' who carry the possible multilocus genotypes consistent with the single-locus phenotypes. Bayesian lasso is used to find the posterior distributions of logistic regression coefficients, which are then used to calculate Bayes Factor to test for association with haplotypes, environmental covariates and interactions.

License GPL-3

LazyLoad yes

Archs x64

R topics documented:

LBLGXE-package 1
LBL 2

Index **6**

`LBLGXE-package` *Bayesian Lasso for detecting Rare (or Common) Haplotype Association and their interactions with Environmental Covariates*

Description

The main function of this package is LBL. For details, see ?LBL.

Details

Package:	LBLGXE
Type:	Package
Version:	1.2
Date:	2015-07-09
License:	GPL-3
LazyLoad:	yes

Currently available functions: LBL. Type ?LBL for more details.

Author(s)

Shuang Xia, Yuan Zhang, Swati Biswas, and Shili Lin

Maintainer: Yuan Zhang <yxz112020@utdallas.edu>

References

Zhang, Y. and Biswas, S (2015). An Improved Version of Logistic Bayesian LASSO for Detecting Rare Haplotype-Environment Interactions With Application to Lung Cancer, Cancer Informatics, 14(S2): 11-16.

Biswas S, Xia S and Lin S (2014). Detecting Rare Haplotype-Environment Interaction with Logistic Bayesian LASSO. Genetic Epidemiology, 38: 31-41.

Biswas S and Lin S (2012). Logistic Bayesian LASSO for Identifying Association with Rare Haplotypes and Application to Age-related Macular Degeneration. Biometrics, 68(2): 587-97.

Burkett K, Graham J and McNeney B (2006). hapassoc: Software for Likelihood Inference of Trait Associations with SNP Haplotypes and Other Attributes. Journal of Statistical Software, 16(2): 1-19.

See Also

<hapassoc> <pre.hapassoc>

Examples

```
#see ?LBL
```

LBL *Bayesian Lasso for detecting Rare (or Common) Haplotype Association and their interactions with Environmental Covariates*

Description

Bayesian LASSO is used to find the posterior distributions of logistic regression coefficients, which are then used to calculate Bayes Factor and credible sets to test for association with haplotypes, environmental covariates, and interactions. This function first calls pre.hapassoc function from the hapassoc package, and some of the options such as "dat", "numSNPs", "maxMissingGenos" and "allelic" are used by pre.hapassoc. It takes as an argument a dataframe with non-SNP and SNP data. The rows of the input data frame should correspond to subjects. Missing single-locus genotypes, up to a maximum of maxMissingGenos (see below), are allowed, but subjects with missing data in more than maxMissingGenos, or with missing non-SNP data, are removed.

Usage

```
LBL(dat, numSNPs, maxMissingGenos = 1, allelic = TRUE, haplo.baseline = "missing", cov.baseline
```

Arguments

dat	the non-SNP and SNP data as a data frame. For the non-SNP data, the first column is the affection status, others (optional) are environmental covariates. The SNP data should comprise the last 2*numSNPs columns (allelic format) or last numSNPs columns (genotypic format). Missing allelic data should be coded as NA or "" and missing genotypic data should be coded as, e.g., "A" if one allele is missing and "" if both alleles are missing. Covariates should to be coded as numerical variables, e.g., 0, 1, etc.
numSNPs	number of SNPs per haplotype.
maxMissingGenos	maximum number of single-locus genotypes with missing data to allow for each subject. (Subjects with more missing data, or with missing non-SNP data are removed.) The default is 1.
allelic	TRUE if single-locus SNP genotypes are in allelic format and FALSE if in genotypic format; default is TRUE.
haplo.baseline	Haplotype to be used for baseline coding; default is the most frequent haplotype according to the initial haplotype frequency estimates returned by pre.hapassoc.
cov.baseline	Needed only if the non-SNP data contains environmental covariates. Indicates covariates to be used for baseline coding; default is a vector of zeros whose length is the number of covariates.
interaction	Needed only if the non-SNP data contains environmental covariates. Indicates whether or not to put gene-environment interactions in the model; default is TRUE.
interaction.model	Needed only if the interaction option is set to be TRUE. Indicates whether G-E independence is assumed or not for fitting haplotype-environment interactions. "i" represents G-E independent model, "d" represents G-E dependent model, and "u" represents uncertainty about G-E independence, i.e., allows possibility of both models. The default is "i".
names.dep	Needed only if the interaction.model option is set to be "d" or "u". It indicates the environmental covariates that may cause G-E dependence. The default is a vector consisting of all covariates.
a	first hyperparameter of the prior for regression coefficients, beta. The prior variance of beta is 2/lambda^2 and lambda has Gamma(a,b) prior. The Gamma parameters a and b are such that the mean and variance of the Gamma distribution are a/b and a/b^2. The default is 20.
b	b parameter of the Gamma(a,b) distribution described above; default is 20.
start.beta	starting value of all regression coefficients, beta; default is 0.01.
lambda	starting value of the lambda parameter described above; default is 1.
gamma	starting value of the gamma parameters (slopes), which are used to model G-E dependence through a multinomial logistic regression model; default is 0.01.
D	starting value of the D parameter, which is the within-population inbreeding coefficient; default is 0.

e	A (small) number epsilon in the null hypothesis of no association, H0: |beta| <= epsilon. Changing e from default of 0.1 may need choosing a different threshold for Bayes Factor (one of the outputs) to infer association. The default is 0.1.
seed	the seed to be used for the MCMC in Bayesian Lasso; default is a random seed. If exactly same results need to be reproduced, seed should be fixed to the same number.
burn.in	Burn-in period of the MCMC sampling scheme; default is 20000.
num.it	Total number of MCMC iterations including burn-in; default is 50000 when there are no covariates or the interaction.model option = "i". For interaction.model = "d" and "u", the default values are 70000 and 100000, respectively.

Value

BF	A vector of Bayes Factors for all regression coefficients. If BF exceeds a certain threshold (e.g., 2 or 3) association may be concluded.
OR	A vector of estimated odds ratios of the corresponding haplotype against the reference haplotype (haplo.baseline). This is the exponential of the posterior means of the regression coefficients.
CI.OR	95% credible sets for the ORs. If CI.OR excludes 1, association may be concluded.
freq	A vector of posterior means of the haplotype frequencies.
CI.freq	95% credible sets for each haplotype frequency.
percentage.indep	Available only if the interaction.model option is set to be "u". Percentage of iterations in which independent model is choosen.
percentage.dep	Available only if the interaction.model option is set to be "u". Percentage of iterations in which dependent model is chosen.
CI.gamma	Available only if the interaction.model option is set to be "d" or "u". 95% credible sets for the gamma parameters as described above.
CI.lambda	95% credible sets for the lambda parameter as described above.
CI.D	95% credible sets for D as described above.

Author(s)

Shuang Xia, Yuan Zhang, Swati Biswas, Shili Lin

References

Zhang, Y. and Biswas, S (2015). An Improved Version of Logistic Bayesian LASSO for Detecting Rare Haplotype-Environment Interactions With Application to Lung Cancer, Cancer Informatics, 14(S2): 11-16.

Biswas S, Xia S and Lin S (2014). Detecting Rare Haplotype-Environment Interaction with Logistic Bayesian LASSO. Genetic Epidemiology, 38: 31-41.

Biswas S, Lin S (2012). Logistic Bayesian LASSO for Identifying Association with Rare Haplotypes and Application to Age-related Macular Degeneration. Biometrics, 68(2): 587-97.

Burkett K, Graham J and McNeney B (2006). hapassoc: Software for Likelihood Inference of Trait Associations with SNP Haplotypes and Other Attributes. Journal of Statistical Software, 16(2): 1-19.

See Also

'pre.hapassoc', 'hapassoc', 'rGLM'

Examples

```
# Load an example data
data(LBL.ex)

#Install hapassoc and dummies package
library(hapassoc)
library(dummies)

## Run LBL to make inference on association of haplotypes with the disease. Note the default setting for bu

out.LBL<-LBL(LBL.ex, numSNPs=5, burn.in=100, num.it=1000)
```

Index

∗Topic **package**
    LBLGXE-package, 1
<hapassoc>, *2*
<pre.hapassoc>, *2*

LBL, 2
LBLGXE *(*LBLGXE-package*)*, 1
LBLGXE-package, 1